Vikas Anand Saharan *Editor*

Computer Aided Pharmaceutics and Drug Delivery

An Application Guide for Students and Researchers of Pharmaceutical Sciences

Springer

Computer Aided Pharmaceutics and Drug Delivery

Vikas Anand Saharan
Editor

Computer Aided Pharmaceutics and Drug Delivery

An Application Guide for Students and Researchers of Pharmaceutical Sciences

Springer

Editor
Vikas Anand Saharan
Department of Pharmaceutics, School of
Pharmaceutical Sciences and Technology
Sardar Bhagwan Singh University
Dehradun, Uttarakhand, India

ISBN 978-981-16-5179-3 ISBN 978-981-16-5180-9 (eBook)
https://doi.org/10.1007/978-981-16-5180-9

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Dedicated to
my Mother Mrs. Chanderpati
&
my Father Mr. Dayanand Saharan

Foreword

I remember starting to use a computer in the early 1990s. PCs were chunky and very costly and it was a matter of pride to have a PC on the desk at that time. The initial scientific uses were centered on the preparation of scientific reports and publications. Computational chemists and scientists working on molecule modeling had sophisticated workstations which catered to the requirements of in silico drug discovery. Computers on the desk of formulation scientists reached in 1992. Later in the 2000s, it almost became a routine to start the work of the day by checking and responding to emails. The internet fired the popularity of PC as scientists and they started communicating through emails which were easier, cheaper, and faster than any other mode. Further, gradually with better functionality of search engines, the task of retrieving information became easier. Journals started appearing online and this transition gradually reduced the visits of researchers and students to libraries. Hand searching of journals gradually shifted to proprietary database searching (SciFinder, International Pharmaceutical Abstracts, etc.) and online searching on platforms like PubMed. The internet was being considered as a host of libraries with subscription services from publishers for either individual journal or online libraries of publishers, like ScienceDirect, Ingenta, etc. My first software purchase in the area of pharmaceutics was Design Expert in 2002.

Life is very different now; I am dependent on various apps on mobile phone, tablet, and PCs.

I know Prof. Vikas Anand since his master's program at NIPER, India. During his studies, I observed him developing a keen interest in computers. He used to spend most of his time either in the Library or Computer Centre of NIPER. In those days, he was pretty good at scholarly information retrieval, e.g., patent searching and data analysis on computers relevant to pharmaceutical sciences. Since then, he continued to develop his interest. I am happy to see that his and his team's efforts have resulted in a great compilation of the most interesting topics on computer applications in pharmaceutics.

There are not many books written on computer applications in the area of pharmaceutics or focusing on in silico developments in drug delivery. The title *Computer Aided Pharmaceutics and Drug Delivery* fascinated me. I am sure that everyone in the area of Pharmaceutics would love to read it and recommend it. A lot has been written on Computer-Aided Drug Design (CADD), but this is a unique

attempt at combining computers with pharmaceutics. I would like to see more volumes in future with the rising role of molecular pharmaceutics, DoE, prediction of physicochemical properties and ADMET predictions, robotics and automation, and in silico analysis of pharmaceutical formulations. The idea of authors to conceive the term "Computer-Aided Pharmaceutics" will flourish with advancements in AI, data science, and quantum computing.

I would like to recommend this book to all graduate and postgraduate students of pharmacy as the chapters are good at conceptual understanding and updated with recent knowledge in the field. Scientists working in industries and other research institutions will also find it attractive with applications of commercial software/expert systems/AI packages through the most recent publications.

I wish to congratulate the team of authors and Springer publisher for excellent work and contribution to the pharmaceutical sciences. It is especially pleasing to note that Vikas's contributions in the field, keep it up.

University of South Australia
Adelaide, SA, Australia
12 May 2021

Sanjay Garg

Preface

I got an opportunity to learn computer applications for the very first time in my school days, in the 1980s, when I had learnt a little bit of programming in DOS and the thing I remember most is playing Pac-Man when our teacher was not around us. I had no course on computers in our Bachelor of Pharmacy program. In 1999–2001, I, as a student of Master's Program in Pharmaceutics, had a full course on computers at NIPER (Mohali). We, the students of the postgraduate program, had received long hand practical trainings from teachers of not only the computer department but also from other departments like pharmaceutics, pharmaceutical analysis, pharmaceutical chemistry, and pharmacology. We were taught and trained for various applications of computers like chemical structure drawing, molecular modeling, information retrieval, reference managing, applications of Excel in data handling and visualization, communication tools, and many others. At that time, we were curious to find whether all that training will ever be of any use in formulation development. But now after about 20 years of experience in the field, I can say with confidence that our dependency on computers has increased tremendously for designing quality in formulations, in silico ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) predictions, and automation of manufacturing, laboratory, and analytical procedures.

Advancements in Computer-Aided Drug Design (CADD) and the availability of three-dimensional structures of polymers, receptors, and other macromolecules have allowed formulation scientists to explore in silico drug-polymer interactions for drug targeting, controlling drug release, and other applications. The rise of animal and human ethical concerns has also propagated in silico characterization of ADMET eliminating/reducing the use of animals/humans. Computer simulations have provided new avenues to discover pharmacokinetics/pharmacodynamics of new molecules on whole organisms, isolated tissues, cells, proteins, and genes. Data collection and maintenance is now very easy through electronic notebooks and numerous data storage options like cloud storage, thanks to the ultra-high internet speed, search engines, and databases which have eased the process of scholarly literature retrieval. Artificial intelligence, robotics, big data (omics, chemical and clinical), IoT, and health devices/apps are accelerating innovative computer applications in drug formulation design, manufacturing, and delivery of drugs. Computational approaches, statistical and molecular modeling have radically

transformed dosage form development and drug delivery programs of pharmaceutical industries through experimental design and quality predictions via Artificial Intelligence (AI) based software. The concept of QbD (Quality by Design) has emerged as an interesting application of computers for formulation development. DoE (Design of Experiments) and optimization of process/formulation have been made a lot easier with the help of software and expert systems.

This book *Computer Aided Pharmaceutics and Drug Delivery* (CAPDD) explores the role of computers, software, databases, artificial neural networks, expert systems, and information technologies in the fields of pharmaceutical formulation development, manufacturing, drug delivery, biopharmaceutical and pharmacokinetic characterization, and pharmaceutical analysis. This book aims to fulfill the curricular requirements of graduate and postgraduate students of pharmacy. Formulation scientists and research scholars find few resources on computer applications outside the primary literature. Scientists at midcareer found it most difficult to deal with the issue of scarcity of books on this topic. This book also caters to the requirements of scientists and research scholars to present most updated computer applications in a very lucid manner to make the learning curve somewhat shallow. Abundant applications of computers in formulation designing and characterization are provided as examples, case studies, and illustrations. Short reviews of software, databases, and expert systems have been added to culminate the interest of readers and inspire them to apply these tools for developing novel formulations for effective drug delivery.

I do hope that the vision of this book will be realized with the rise in computer-based formulation development, characterization, and manufacturing in pharmaceutical industries to speed up the delivery of drug products to patient's bedside. I sincerely believe that you will enjoy reading and learning from history the present and future applications of computers as I have enjoyed while editing this book.

Dehradun, Uttarakhand, India Vikas Anand Saharan

Acknowledgments

First and foremost, I would like to thank God, the almighty, for all the graces, blessings, health, knowledge, and opportunity to me so that I can accomplish this book project in time.

I am extremely grateful to the management and administration of Sardar Bhagwan Singh University, especially Dr. Gaurav Deep Singh (Hon'able Chancellor), Sh. S. P. Singh (Chairman, Gaurav Bharti Shiksha Sansthan), Prof. Veerma Ram (Hon'able Vice Chancellor), and Sh. KBS Rawat (The Registrar) for all-round support during my entire tenure at the University. I am highly privileged to have some of my great teachers, like Dr. Luvkush, Prof. Veerma Ram and Ms. Urmi Chaurasia, Prof. Sanjay Garg, and Prof. P. K. Choudhury, as a constant and continuous source of inspiration and motivation.

I appreciate the support received from Prof. R. J. Krupadam (CSIR-NEERI), Prof. Hemant Jadhav (BITS, Pilani), Prof. Sokindra Kumar (Subharthi University), Prof. Alka Chaudhary (SGRR University), Prof. Ganesh Bhatt (Shivalik College of Pharmacy), Prof. Mamta F. Singh (Sardar Bhagwan Singh University), Dr. Seema Sadrai (Tehran University of Medical Sciences), Dr. Anand Sharma (NIPER Mohali), Dr. Hitesh Kulhari (Central University of Gujarat), Dr. Mahesh K. Kataria (Seth GL Bihani SD College of Technical Education), Dr. Bhavna (DIT University), Dr. Ujwal Kolhe (NIPER, Reabareli), Dr. Supriyo Saha (Sardar Bhagwan Singh University), and Mr. Prashant Beniwal (Seth GL Bihani SD College of Technical Education) for reviewing book chapters and providing their critical comments for improving the quality of book chapters.

I would like to thank all the chapter authors for their hard work, valuable contributions, and their patience in various phases of publication of this book. I am extremely thankful to Dr. Gaurav Singh (Editor, Springer Nature), Mr. Ejaz Ahmad (Project Coordinator, Springer Nature), Ms. Raman Shukla (Associate Editor, Springer Nature), and Ms. ArulVani Parttibane (Project Manager, Straive) for their continuous support in providing necessary guidance throughout this project. I thank Springer for providing a platform from where this book can reach to its readers.

Without the support of my mother, father, and wife, it was really impossible to complete this arduous task. They have sacrificed a lot of my personal time while I was working on this book. I also missed my sons, Inesh and Krit, when I was working on this book and they were not beside me.

18th May 2022

Vikas Anand Saharan
Sardar Bhagwan Singh University
Dehradun, Uttarakhand, India

Contents

About the Editor

Vikas Anand Saharan had obtained his B.Pharm from SBSPGI Dehradun, master's from NIPER Mohali, and PhD from Mohanlal Sukhadia University Udaipur. He has 20 years of teaching and research experience in pharmacy, published 60+ papers in journals, presented 75+ papers in conferences, and supervised 39+ M.Pharm and 2 PhD thesis. He has authored 3 books and 19 book chapters. He has received an award for GATE guidance and appreciation awards for teaching and research. He is a reviewer and editorial board member of various national and international journals. He is a recipient of grant-in-aid under the AICTE-MODROBS project. Presently he is also entrusted with the responsibilities of chief program coordinator for PhD and nodal officer for AISHE at Sardar Bhagwan Singh University.

History and Present Scenario of Computers in Pharmaceutical Research and Development

1

Vikas Anand Saharan, Surojit Banerjee, Swati Penuli, and Swati Dobhal

Contents

Abstract

Twentieth and early twenty-first centuries have observed a lot of fascinating developments in pharmaceutical research and development with the rising role of computers. Computers have transformed drug discovery form hit-and-trial approach to rational drug design. Initial quantitative structure-activity relationship (QSAR) studies led to the foundation of computer-aided drug design (CADD), which subsequently evolved to structure-based drug design (SBDD), ligand-based drug design (LBDD), and fragment-based drug design (FBDD). The successes of computational chemistry and CADD have brought many interesting

V. A. Saharan (✉) · S. Banerjee · S. Penuli
Department of Pharmaceutics, School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Dehradun, Uttarakhand, India

S. Dobhal
Department of Pharmacology, School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Dehradun, Uttarakhand, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_1

drug molecules from bench to the patient's bedside. Drug discovery, being a multidisciplinary field, has been benefited with the advancements from the field of not only computers but also from associated technologies like softwares, Internet, big data, omics, internet of things, and artificial intelligence (AI). This chapter is the author's attempt to explore history from 1960 to the present time for finding key developments that have played a larger role in evolving drug discovery and development.

Keywords

Computational chemistry · CADD · Artificial intelligence · Drug discovery · Hardware · Software · Database · Program · Molecular modeling · SAR

Chapter Objectives

Upon reading of this chapter, it is expected that reader will be able to understand:

- History of use of computers in pharmaceutical companies
- The history of advancements in computers, computer languages, and softwares for drug discovery and development
- Key developments in the fields of computational chemistry, computer-aided drug design (CADD), and artificial intelligence (AI)
- The advancements of drug discovery programs *vis a vis* to advancements in science with the rising role of computers and softwares
- Some significant discoveries and developments, which have a long-term impact on pharmaceutical industries, either in research and development or in manufacturing of drug products
- Marketed drugs whose discovery was assisted by computers and softwares

1.1 Introduction

Development of the human race has brought about revolutionary changes in several aspects of life, be it lifestyle or technology. Computers, in terms of technology, have been a boon for the society owing to the ease of some high-end tasks it provided. Charles Babbage conceived the idea of computer with his "Analytical Engine" and "Difference Engine" designs in the nineteenth century and passed on plans for its completion [1]. Babbage could not construct any of his designs but later the designs of Babbage laid foundations of modern computing. Alan M. Turing conceived the idea of a universal machine (Universal Turing Machine) in 1936 [2]. Universal Turing machine laid foundations for the successful construction of initial computers like Turing's Automatic Computing Engine (ACE), and John von Neumann's Electronic Numerical Integrator and Computer (ENIC) and Electronic Discrete Variable Computer (EDVAC) [3]. Universal Automatic Computer (UNIVAC) was the first commercial computer developed by J. Presper Eckert and John Mauchly and delivered by Remington Rand, an office equipment company in 1951 [4]. IBM701, of pioneer IBM700 series, was the first commercially available scientific computer,

which came in the 1950s [5]. IBM 704, IBM 709, IBM 650, and IBM 610 were some computers which were purchased for their use in finance and business operations of industries or business firms [6]. In science, initially, use of computers was restricted to solving mathematical problems but later the use of computers grew in the fields of chemistry and biology albeit through more involvement of mathematics in science. With the development of dispersive computer technologies and widespread use of softwares and the Internet, drug discovery in pharmaceutical industries has gradually become a multidisciplinary field.

Programming languages are the languages used by the programmers to communicate with the computers, that is, it is the language of the computers, for example, DOS, C, C++, JAVA, SQL, etc. Programming languages allow humans to give instructions to a computer in a language the computer understands. These programming languages form the framework for the development of the various softwares and applications that are run on computers [7]. Software is an array of instructions or programs that are defined to perform a specific function. ChemDraw, AutoDock, DesignExpert, and GraphPad Prism are some softwares used in pharmaceuticals and other research fields [8]. Hardware are the physical parts of the computers like the central processing unit (CPU), monitor/screen, mouse, keyboard, graphics card, etc. that are involved in storage and execution of the software. The programming language, software and hardware together, end up in the efficient working of the computer system.

The initial computer language used was the Regional Assembly Language that required a great deal of efforts by programmers. With the development of high-level programming languages, several new languages came into existence that eased the work of the programmers and resulted in development of various application softwares that can be employed for scientific purposes in pharmaceutical industries and other institutions. The efforts and exertion of performing data analysis, statistical computations, and data compilation was eased as all of this is now just a click away with the help of softwares like SPSS, XLSTAT, Stata, etc. The experiments that were hard to perform or used animal models can now be easily performed using simulation softwares in computers, for example, Ex-Pharm. The softwares are also being employed in the process of drug discovery like docking models of drugs, that is, if the structure of protein to which the drug will bind is available from a database, the structure of potential biologically active molecule can be easily elucidated using software, for example, AutoDock, etc. [9]. Software packages have been advanced with computing speed of computers and the use of integrated AI technologies for accurate predictions of biological activity, physicochemical property, toxicity, and pharmacokinetics.

This chapter is an attempt of authors to look back into the history to find some remarkable developments of hardware, programming languages, and softwares for their specific applications in the field of drug discovery and development. Emphasis has been laid down to track key developments of computational chemistry, computer-aided drug discovery (CADD), and AI. In addition to this, significant developments of early twenty-first century are also presented.

1.2 The Birth of Computational Chemistry: The 1960s

The 1960s initiated the use of computers in academics by quantum chemists, theoretical chemists, X-Ray crystallographers, and later by pharmaceutical companies for drug discovery [10]. Computers available to chemists were of very large size and called mainframes (central systems). IBM 7094 was a scientific mainframe computer, which was used by crystallographers and theoretical chemists [6, 11]. IBM 1620 was in use for statistical and analytical chemistry applications. Softwares were distributed on magnetic tape or computer punch cards (IBM Cards). Programs were written in FORmula TRANslation (FORTRAN), which allowed easy transition of mathematics formula into codes; FORTRAN IV being a mid-1960s version [12]. Continuous System Modeling Program III (CSMP III) was introduced by IBM in 1967 for scientists and engineers who were not computer experts [13]. This generalized simulation program was used by a number of authors for publishing papers.

The idea of having a mouse on the computer system was first conceived by Douglas Engelbart; he reportedly conceived the mouse during a conference lecture in 1960 [14]. Not only the mouse, Doug Engelbart also demonstrated a computer system and hyperlinks, video conferencing, teleconferencing, word processing, and many more. A video titled "Mother of All Demos" is available on YouTube [15]. The RKS 100-86 "Rollkugel-Steuerung" was the first rolling ball mouse, designed by Rainer Mallebrein of Telefunken (Germany), and was shipped with the Telefunken TR86 computer in 1968 [16].

FORmula MAnipulation Compiler (FORMAC), as an extension of FORTRAN IV and developed by Jean E. Sammet, was the first widely available programming language of 1960s for manipulating nonnumeric algebraic expressions and computations [17]. Simula, developed by Ole-Johan Dahland Kristen Nygaard at Norwegian Computing Center, Oslo [18], was the programming language for UNIVersal Automatic Computer 1 (UNIVAC 1107) computer since January 1965. UNIVAC machines were the first general-purpose electronic digital computer design. StriNg Oriented and symBOlic Language (SNOBOL), developed at AT&T Bell Laboratories by David J. Farber, Ralph E. Griswold, and Ivan P. Polonsky in the 1960s, was created as a tool to work with the symbolic manipulation of polynomials [19]. SNOBOL was later written in assembly language for the IBM 7090 machines. In the 1960s, statistical software package SPSS (by Statistical Package for the Social Sciences and later by IBM in 1964) was a leading statistical software package [20], which is still in extensive use for optimization and other statistical calculations.

The year 1962 marked the conceptualization of The Quantum Chemistry Program Exchange (QCPE) as an international repository of softwares at Gordon Conference on Theoretical Chemists [21]. It was a brainchild of Prof. Harrison Shull and colleagues at Indiana University, Bloomington. The idea was appreciated for the benefits of sharing softwares coupled with other benefits like the benefits of having an intermediary between the users and owners/code writers, preventing "reinvention of the wheel" and publishing codes of programs as intellectual writing. With the initial support received from Indiana University and Air Force Office of Aerospace

Research (ARAC), Prof. Shull and colleagues started QCPE in 1963. QCPE initially offered programs for matrix multiplication, matrix diagonalization, and determination of integrals over elliptical orbitals. The programs were written in FORTRAN II and FORTRAN Assembly Program (FAP). Softwares depositions increased with time, though it was a little slow initially. QCPE was a milestone in the growth of computational chemistry as there were no software companies in the 1960s and even in the 1970s. Initially, chemical structure calculations and ab initio programs (like molecular orbital calculations) were shared on QCPE platform, which further evolved to molecular mechanics programs. QCPE Newsletter (later renamed as QCPE Bulletin) was a forum of information sharing among quantum/theoretical chemists for new programs or news in the field. Some of the industrial chemists were also members of QCPE.

In 1965, Oak Ridge Thermal Ellipsoid Program (ORTEP), written in FORTRAN by Carroll K. Johnson of the Oak Ridge National Laboratory (ORNL), was released for preparing illustrations of structures for publication and conference presentations [22]. ORTEP was a noninteractive molecular graphics program which was used to draw ball and stick diagrams. ORTEP was widely used and a favorite program of crystallographers. Hand-held models used were space-filling Corey-Pauling-Koltun (CPK) models with color-coded elements (carbon, hydrogen, nitrogen, and oxygen as black, white, blue, and red, respectively) and models with ball and joints held with metal or plastic rods. Hartree-Fock method, complete-neglect-of-differential-overlap/second parameterization (CNDO/2), Huckel theory, and Pariser-Parr-Pople theory were good at predicting molecular geometries of planar molecules but the application on molecules of pharmaceutical interest was limited due to their nonplanar structure [6, 23]. Lilly, Schering-Plough, Upjohn, and Dow Chemical were some pharmaceutical companies which started using computers in their drug discovery programs for finding correlation of structure with biological activity.

Some pioneer works, which laid the foundation of computational chemistry, were from Corwin Hansch and Toshio Fujita [24] and Spencer M. Free and James W. Wilson [25]. Corwin Hansch and Toshio Fujita established quantitative structure-activity relationship (QSAR) modeling with the discovery of consistent relationship between logP (a physicochemical property) and molecules' in vivo biological activity. Corwin Hansch is regarded as the father of QSAR. Spencer M. Free and James W. Wilson used a series of chemical analogs to fit mathematical equations (models) to anticipate complex facets of toxicity. Later in 1968, Yvonne Martin began working with Corwin Hansch on QSAR. These landmark discoveries marked the birth of computational chemistry as an offshoot of physical chemistry. Significant developments of 1960s are summarized in Table 1.1.

Table 1.1 Some significant developments of 1960s

Interesting development	Application	Developer/team/ organization	Year	Ref.
FORTRAN IV	First programming language mainly used for numeric and scientific computations	IBM	1962 (FORTRAN IV)	[12]
Simula	Simulation programming language	Kristen Nygaard at Norwegian Computing Center, Oslo	1962	[18]
SNOBOL	Programming language—string-oriented and symbolic language	David J. Farber, Ralph E. Griswold, and Ivan P. Polonsky, AT&T Bell Laboratories	1962	[19]
Quantum Chemistry Program Exchange (QCPE)	Repository of programs and softwares for sharing codes	Prof. Shull and colleagues from Air Force Office of Aerospace Research (ARAC), Indiana University	1963	[21]
FORMAC	The first computer algebra system	Jean E. Simmet from IBM	1964	[17]
QSAR Modelling	Method for determination of biological activity by correlating it with chemical structure of compound	Corwin Hansch and Toshio Fujita	1964	[24]
Mathematical equations	Developed methods to mathematically report the toxicity profile of chemicals	Spencer M. Free and James W. Wilson	1964	[25]
BMDP	Software used for Statistical data analysis	Wilfrid Dixon from University of California	1965	[26]
SAMCEF	Software used for finite-element analysis	SAMTECH	1965	[27]
ORTEP	Software used for crystal structure illustrations	Oak Ridge National Laboratory (ORNL)	1965	[28]
CSMP III	Software used for Solving and modeling of differential equations	IBM	1967	[29]
SPSS	Software used for statistical data analysis	University of Stanford (Norman H. Nie, C. Hadlai (Tex) Hull, and Dale H. Bent)	1968	[20]
Mouse	For pointing and for improving speed and accuracy of computers	Douglas Engelbart	1968	[14]
RKS 100-86 (Rollkugel-Steuerung)	First rolling ball mouse	Rainer Mallebrein of Telefunken, Germany	1968	[16]

1.3 Evolving Collaboration of Computational Chemists with Other Chemists: The 1970s

Early-to-mid 1970s marked a mini revolution in computer machines. However, some companies even launched their mainframes. BS2000 was launched in 1975 as the mainframe computer by Siemens [30]. Small computers, like Datacraft 6024, with low-volume card readers, printers, and disk systems were also available [12]. Digital Equipment Collaboration introduced a minicomputer VAX11/780. VAX was much cheaper (300 k US $) than the multimillion dollar mainframes of the 1960s. Widely used models of 1960s belonged to IBM 360 and IBM 370 series. Statistics programs, like MINITAB, were used on DEC machines (Discwriter II) connected to IBM computers [6]. Machines based on card punches were still in use. Pharmaceutical companies started giving word processing (called Wang machines) to secretaries. At that time, very few scientists were keying in the data by themselves, rather it was done by either data entry technicians, secretaries, or the computational chemists. In 1972, floppy disk 8 inch, with the storage capacity of 3000 punched cards, was brought to the market by IBM in IBM 370 machines, which later became almost essential in word processing computer systems and minicomputers [31]. The year 1977 marked the entry of Apple II Person Computer (PC) with 5¼ inch floppy drives. The addition of floppy discs to PCs allowed ordinary people to store operating systems and softwares on their PC.

FORTRAN IV was still extensively used for writing programs. Dennis M. Ritchie at Bell Labs (late AT&T) developed the C language in the early 1970s by Dennis M. Ritchie for Unix [32]. Around 1976, Cleve Moler started working on developing an educational software package for matrix calculations; MATrix LABoratory (MATLAB) first appeared in 1970s [33]. Structured Query Language (SQL) was launched by IBM in 1972 as a domain-specific language for data analyst, database administrators, pharmacoeconomists, and for the development of various healthcare applications [34].

QCPE expanded with more memberships. Prof. John A Pople's (Carnegie-Mellon University), future Noble Laureate, started direct distribution of ab initio code with name Gaussian 70 (first version) to numerous laboratories in around 1970 [12, 35]. Pople submitted Gaussian 76 and Gaussian 80 to QCPE. Later John Pople commercialized the Gaussian program in 1987 and withdrew all previous versions from QCPE [21]. Since then, Gaussian was developed and licensed by Gaussian, Inc. The current 16C.01 version of this program was released in 2019 [36]. The 1960s ab initio calculations and programs were not beneficial to drug molecules. However, for the first time, computational efforts were conceived for biology and pharmacology with coining of terms like quantum pharmacology [37] and quantum biology [38].

Prof. N. L. Allinger's MMI/MMPI program was developed for molecular mechanics and made available on QCPE in 1976 for purchase at a nominal cost [21]. This program was faster than quantum mechanics and used for generating organic chemical structures with prediction accuracy of 0.01 °A. Computational chemists from pharmaceutical companies started applying their academically

acquired knowledge of QSAR, force field methods, and statistics for transforming quantum chemistry to versatile computational chemistry.

Cambridge Structural Database (CSD) and Protein Data Bank (PDB), launched in the 1970s and before, were of great use to pharmaceutical companies as they provided 3D molecular structures of compounds [39, 40]. CSD was a repository for small organic and organometallic crystals structures. SAS developed by North Carolina State University [41] and Minitab developed by Minitab LLC [42] were the statistical softwares launched in this decade. GAMESS, written in FORTRAN and C, program for computational chemistry was launched at Iowa State University in 1977 [43] for computational chemists. GROMOS developed by University of Groningen in 1978 was the software used to study field force [44]. Syntex developed a software called XTL in 1974 for the purpose of crystallographic study of molecules [45].

Sidney Fernbach and Abraham Haskell Taub mentioned the term "computational chemistry" probably for the first time in 1970 in a book "*Computers and Their Role in the Physical Sciences*" where they state "It seems, therefore, that 'computational chemistry' can finally be more and more of a reality." [46]. With the help of physical chemistry, computational chemists were able to solve spectroscopy and molecular conformation and such studies brought various correlations between experimental and calculated properties. However, the era lacked collaboration among computational chemists and medicinal chemists. Medicinal chemists, due to their higher hierarchy in pharmaceutical companies, were unwilling to work on the ideas of synthesizing chemical structures (designs), which were the outcome of computers. Designing chemical structures was considered a much easy task rather than performing its synthesis. Synthesizing so many compounds in laboratories was not a feasible task for experimental (medicinal) chemists. Pharmaceutical companies tried to overcome this collaboration gap by organizing workshops and conferences but medicinal chemists were very slow to accept computer-generated structures. Some new pharmaceutical companies such as Merck and SmithKline and French started using computers while other companies like Lilly focused more on strengthening their foothold in the area of computation chemistry [47]. In 1975, The QSAR Gordon conference was held for the first time and this seminal event brought a widespread acceptance of QSAR as a field [48]. In 1978, Y Yvonne Martin published a book "*Quantitative Drug Design: A Critical Introduction*" [49]. Many other books and papers were published on QSAR studies. The National Cancer Institute of the United States financially supported many research projects in this direction which led to the outcome in the form of papers on anticancer agents published by theoretical chemists [6].

Significant advancements were observed for structure-based drug design (SBDD) that could not yield fruitful results as only a few drug targets with their 3D structure were available. Dihydrofolate reductase (DHRF) structure came in 1970s and this protein was used by many research teams for their drug designing efforts to explore better inhibitors than trimethoprim or methotrexate [50, 51]. Some significant developments of 1970s are summarized in Table 1.2.

Table 1.2 Some significant developments of 1970s

Interesting development	Application	Developer/team/ organization	Year	Ref.
Gaussian	Software used for computational chemistry	Prof. John A Pople's from Carnegie Mellon University	1970	[35]
ARPANET's Netword e-mail System; First electronic mail program	Wide-area packet-switching network	Ray Tomilson at Advanced Research Projects Agency (ARPA)	1971	[52]
Protein Data Bank (PDB)	Database of protein structures	Walter Hamilton at Brookhaven National Laboratory	1971	[39]
UNIX	Computer operating system	Ken Thompson, Dennis Ritchie, Brian Kernighan, Douglas McIlroy, and Joe Ossanna at Bell Labs	1971	[53]
Floppy disk	Used for data storage	IBM	1972	[31]
SQL	Structured Query Language used in data analysis	IBM	1972	[54]
Minitab	Statistical software	Minitab LLC	1972	[42]
XTL	Software used for crystallographic study	Syntex	1974	[45]
Version 6 Unix	First version of Unix operating system	AT&T Bell Laboratories	1975	[55]
CLU	Programming language	Massachusetts Institute of Technology	1975	[56]
BS2000	Mainframe computer operating system	Siemens AG and later by Fujitsu Technology Solutions	1975	[30, 57]
MAtrix LABoratory (MALTAB)	Educational Software package for calculations related to matrices	Cleve Moler	1976	[33]
MMI/MMPI	Software launched on QCEP platform for making chemical structures with increased accuracy	N L Allinger from University of Georgia	1976	[21]
SAS	Software for statistical data analysis	North Carolina State University and SAS Institute	1976	[33]
VAX11/780	Minicomputer with low cost and extendable storage	Digital Equipment Corporation (DEC)	1977	[12]
Apple II	Personal computers with floppy drive so that the user can store softwares	Developed by Apple Inc.	1977	[58]

(continued)

Table 1.2 (continued)

Interesting development	Application	Developer/team/ organization	Year	Ref.
General Atomic and Molecular Electronic Structure System (GAMESS)	Software for computational chemistry	Iowa state University	1977	[43]
Quantitative Drug Design: A Critical Introduction	A book on drug design	Yvonne Connolly Martin	1978	[49]
GROningen MOlecular Simulation (GROMOS)	Software used for field force for molecular dynamics simulation	University of Groningen	1978	[44]
Word star	Word processing application for microcomputers	Rob Barnaby	1978	[59]
VisiCalc	Software used as data entry	Bob Frankston	1979	[60]

1.4 Multifaceted Acceptance of Computational Chemistry and the Birth of CADD: The 1980s

Pharmaceutical industries were in a naïve euphoria phase in the early 1980s [48]. In the timeframe of the late 1980s and early 1990s, there was a considerable rise in interest in CADD in academia (Fig. 1.1). At that peak of hype, a common saying was "we can design drugs atom-by-atom." CADD was then emerging as a distinct discipline out of computational chemistry. *Fortune* magazine proclaimed it as the "next Industrial Revolution" in 1981. Computational chemists restricted themselves to focus on explaining chemical phenomena by providing atomic-/molecular-level explanations. While the discipline of CADD involves predictions about molecules to be synthesized with desired biological properties, CADD emergence led to the subsequent developments like virtual library design, virtual screening, and de novo design, though by the then computational chemists. Prediction of physicochemical properties and biological activity started with the advancements of AI. However, AI was not new as one can even track it back to the 1960s, in its theoretical form at least, with the work of Ivakhnenko and Lapa [61]. AI can even trace its roots even further back to a workshop that was run at Dartmouth College in 1956.

Parallel and vector computing architectures emerged in the early 1980s for the fast spread of minicomputers [12]. Noncentralized control on computing resources was observed in this era with VAX 11/780 computer (Digital Equipment Collaboration), which was of departmental size. Many pharmaceutical companies such as Lilly acquired VAX machines for research purposes [6]. IBM PC with DOS was also

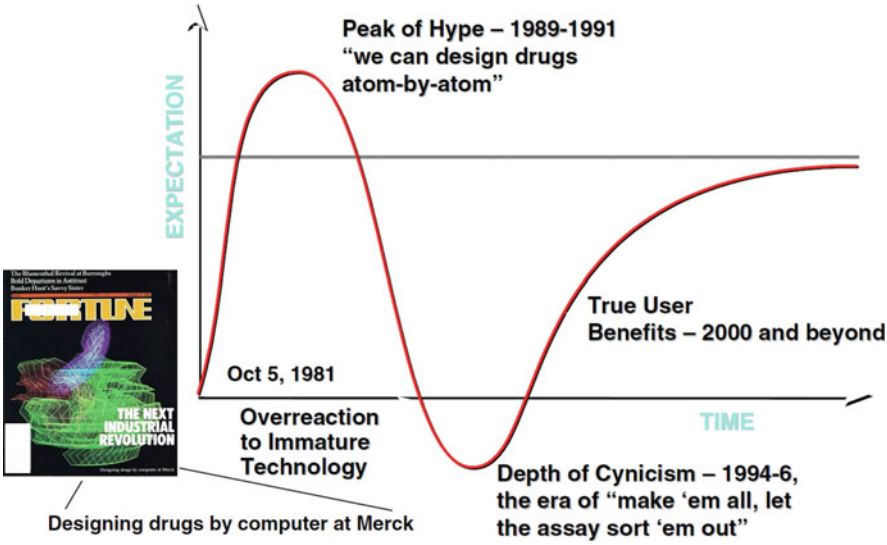


Fig. 1.1 An illustration on rise and fall of CADD from true user expectations. (Copyright with Springer Nature) Reproduced from [48]

evident at some of the companies. The Apple Macintosh later came in 1984. All these machines provided a good working environment for word processing, graph plotting, and running computational programs. Apple's Mac was considered more interactive and user friendly. IBM and Microsoft launched personal computers in the1980s [62]. By the end of 1980s, these minicomputers became as powerful as mainframes of 1970s. However, these PCs were not powered enough to perform molecular modeling calculations. All these developments in low-cost PC segment resulted in a large spread of PCs among users almost everywhere. IBM sold its first PC in 1981 with two floppy drives [31]. IBM also introduced high-density 3½ inch floppy drives in 1984 with storage capacity of 1.2 megabytes, which remained in the market till the 1990s.

FORTRAN remained on lead for Unix-based workstations and minicomputers in the 1980s [63]. Visual basic gave a very convenient environment for designing Windows interfaces. The use of C language expanded further with the UNIX-based workstations in the 1980s [32]. C++, a successor of C, was developed by Bjarne Stroustrup at Bell Labs and originated in the mid of 1980s. C++ was launched as the most common programming language for machine-independent applications [32]. R is a platform-independent language for statistical computing and graphics [64]. John Chambers and colleagues at Bell Laboratories (later AT&T, now Lucent Technologies) conceived the idea of R in mid-1970s and its first version appeared in public in 1980s. MATLAB evolved from a programming language to a commercial software package MATLAB by Mathworks [65]. The use of this educational software package was very popular among students of varied backgrounds like science, engineering, and economics.

The 1980s was the era of resurgence with newer developments in the fields of molecular mechanics, QSAR, molecular graphics, molecular simulations, and quantum chemistry. QCPE Newsletter was changed to QCPE Bulletin to make its content citable in the scientific literature. *Theoretica Chimica Acta, International Journal of Quantum Chemistry, Computers and Chemistry,* and *Journal of Computational Chemistry* were some journals which were publishing papers of quantum and theoretical chemists [21]. *The Journal of Computational Chemistry* was established in 1980 by John Wiley & Sons with the efforts of Prof. Allinger, who was the founding editor of this journal [66]. This journal was instrumental in bringing up the name of computational chemistry to the field, which was previously called with numerous names like calculational chemistry, theoretical chemistry, chemical modeling, etc. Since then, this journal has been published with the same name by John Wiley and Sons.

In this era, several softwares were developed for statistical data analysis, analyzing NMR and FTIR spectrum and graphics editing. ChemDraw, developed by David A. Evans and Stewart Rubenstein, was launched for Mac machines in 1986 by Cambridge Scientific Computing (later renamed as CambridgeSoft) [67]. ChemDraw allowed chemists to draw two-dimensional chemical structures on computer and exported it to other documents. ChemDraw was user-friendly software and the learning part of it was not more than an hour. Computer graphics development of this decade allowed colored 3D structures to be displayed on computer screens. ChemDraw was commercialized by Cambridge Soft, and later, ChemOffice suite (comprising Chem3D and ChemFinder) came to PerkinElmer in 2011 [67, 68]. This software is still one of the widely used software for drawing chemical structure, conversion of name from structure, NMR and FTIR data analysis, mass spectrum simulation, etc. Currently its 17.1 version is available, which was released in 2018.

The first electronic mail program was invented by Ray Tomilson 1971 as Arpanet System [52]. It was not until the 1980s that the email programs started connecting people across the globe with developments in the Internet and internet service providers. The developments on electronic mail was another significant landmark, which allowed sharing of documents among scientists and also minimized the use of cables for connecting computers and storage on floppy disks.

Most of the pharmaceutical companies were embracing a new culture of computational chemistry in drug discovery programs. Softwares by Molecular Design Ltd. (MDL), Tripos (SYBYL), and Chemical Design Ltd. (CHEMGRAF) were popular among pharmaceutical companies [47]. Companies like Lilly and Merck were aggressively expanding their computational expert teams. Abott, Lederle, Rohm and Haas, SmithKline Beecham, Searle, and Upjohn were hiring full-time computational chemists. IBM and VAX machines were being used for molecular modeling. Universities started offering postdoctoral programs on molecular modeling for organic chemists to cater the rising demand. More and more students started pursuing Ph.D. in quantum chemistry for the sake of job prospects.

Gaussian Inc. and Tripos Associates were dominant players of molecular modeling softwares in the 1980s [47]. MDL softwares were preferred for chemical

structure measurements. MACCS and REACCS (MDL) databases of chemical structures and chemical reactions were extensively used in pharmaceutical companies [69]. In 1981, MOPAC, a semi-empirical molecular orbital program, was developed by Dr. James P. Steward for automatic optimization of molecular geometry [70].

Some softwares developed for molecular modeling, QSAR/QSPR prediction, toxicity prediction, and statistics in 1980s are still widely used. Assisted Model Building with Energy Refinement (*AMBER*) was developed by Peter Kollman and his team at University of California, San Francisco, in 1981, for UNIX systems, for molecular dynamics and free energy calculations [71]. Over the time and with the support of 40+ collaborators and contributors, AMBER has now transformed to a package of biomolecular simulation package [72]. AutoDock was developed by The Scripps Research Institute, California, in 1989, for protein-ligand docking [73]. From the last five decades, AutoDock remained a widely used and most cited molecular modeling simulation package. CLOGP developed by Prof. Al Leo at Pomona College was commercialized by Daylight Chemical Information Systems [74]. CLOG was designed for predicting lipophilicity of organic molecules. *Toxicity* Prediction by Komputer Assisted Technology (TOPKAT), a correlative SAR system developed by Kurt Enslein, from Health Designs (Accelrys later acquired by Biovia), was introduced for predicting the toxicity of molecules from their chemical structures [75–77]. Ligand-based molecular modeling software Macro Model was developed and distributed free by Fariborz Mohamadi and colleagues in 1986 for VAX machines [78]. Macro Model was later acquired by Schordinger and developed further for various other platforms/machines in 2000. Stata, a small regression package with some data management features, was developed by StataCorp in 1985 for PCs with DOS operating system [79]. Later in 1988 and 1882, its Unix and Macintosh versions came. Now Stata 17 package fulfills all the data science needs and comes for almost all platforms/operating systems. Design-Expert software, developed by Stat-Ease, Inc., in 1988, is a commercial statistical data analysis software for applying design of experiments (DoE) in optimizing formulations or processes [80]. Since then, Design-Expert remained a widely used software among industry, academics, and other institutions.

IBM PC, developed in 1981 by Philip Donald Estridge, was a PC with multiuser support and multiwindow processing [81]. Xerox 8010 star system PC was developed in 1981 by Xerox corporation [82]. Word star was a word processor application developed by Rob Barnaby and used for microcomputers [59]. Visicalc was developed in 1979 by Bob Frankston and this was a first spreadsheet program for data entry [60]. Supercalc, developed by Sorcim in 1980, was a spreadsheet developed for PCs [83]. All these developments supported word and spreadsheet processing in PCs, which are now routine work for almost all computer users.

The 1985s Nobel Prize in Chemistry was awarded to Herbert Hauptman and Jerome Karle, both mathematicians, for developing direct methods for the determination of crystal structures [84].

Koga used classic QSAR (hansch type) for the discovery of norfloxacin as an analog of nalidixic acid [85, 86]. This discovery led to the first molecule based on

SBDD to reach the market and also attracted scientists across the globe to molecular dynamics, X-ray crystallography, and NMR spectroscopy. Some significant software and hardware based developments of 1980s are highlighted in Table 1.3.

1.5 CADD Regaining Momentum and Entry of ML: The 1990s

Low-priced Windows-based personal computers from Microsoft were started getting favor at pharmaceutical companies. In 1990s, Windows NT was introduced as a robust operating system that made connecting PCs easy in parallel [62]. PCs in pharmaceutical companies were networked and put under centralized control. Computers started reaching on tables of scientists. However, Apple's Mac and Cary's supercomputers were still very popular for molecular research and SGI Unix-based workstations (Silicon Graphics) dominated for molecular modeling [62]. However, these changes were in favor of PCs and motivated the development of sophisticated computational chemistry programs/applications for either PCs or multiple platforms or platform-independent versions. ChemDraw and ISIS/Draw were made available for PCs to draw molecules and perform single energy minimization. Linux, an open-source operating system, developed by Linus Torvalds at Helsinki University of Technology (Finland) was distributed by Free Software Foundation [100]. Linux became very popular in the late 1990s as an inexpensive substitute to Windows in PCs. PCs' prices further slashed with the wide use of open-source Linux operating systems.

The IBM's 3½ inch floppy drives were more durable than previous versions of floppy disk [101]. The disk was made of plastic with a sliding metal shutter. CDs drives and flash drives with more storage capacity and durability eventually ran floppy drives out of the data storage business.

New languages like Java, Python, and R originated in the 1990s. The idea of Java was conceived by James Gosling, Patrick Naughton, Chris Warth, Ed Frank, and Mike Sheridan at Sun Microsystems in 1991 [102]. They wanted to develop a platform-independent language that can run on any type of computer and thereby could be more useful with the infiltration of World Wide Web (the Internet). Initially, the language was named Oak in 1991 but later its name was changed to Java, in 1995. Many Java characters relate to characters of C and C++. Python was conceived by Guido van Rossum and first implemented in December 1989 at Centrum Wiskunde and Informatica, the Netherlands [99]. Python is a general-purpose, high-level programming language that emphasizes code readability and its syntax allows programmers to express concepts in fewer lines of code. Robert and Ross established R as an open-source project in 1995. The development of R started in the early 1990s as a project of Robert Gentleman and Ross Ihaka, both faculty members at the University of Auckland, and later established as open-source project in 1995 [103]. R came into existence for statistical computing and data analysis. R commands can be utilized in making graphs and inference reports [104]. Also released in 1991, Visual Basic.NET is a programming language employed in the development of some pharmaceutical programs, for example, software for acid-base

Table 1.3 Some significant developments of 1980s

Interesting development	Application	Developer/team/ organization	Year	Ref.
SYBYL	Software for molecular modeling, homolog recognition, structure, and function prediction	Tripos (now Certara from 2008)	Early 1980s	[87, 88]
Super Calc	Software used as advanced calculator app, which is used available on smartphones and tablets	Sorcim	1980	[83]
VAX-11/780; VAX 8000 series	Minicomputers, 8000 series, offered higher performance than that of VAX-11/780	Digital Equipment Corporation	1977–1989	[6, 89, 90]
MATLAB	Software used for matrix manipulations and plotting of functions and data and implementation of algorithms	MathWorks	1980	[65]
IBM PC	Personal computer with multiuser support and multiwindow processing	Philip Donald Estridge at IBM	1981	[91]
Xerox 8010 star system	Personal computer	Xerox Corporation	1981	[82]
Assisted Model Building with Energy Refinement (AMBER)	Software to apply force fields for molecular dynamics of biomolecules	Peter Kollman and his team at University of California	1981	[71, 92, 93]
CLOGP	Software used for predicting lipophilicity	Prof. Al Leo at Pomona College	1982	[74]
Chemistry at Harvard Macromolecular Mechanics (CHARMM)	Software for molecular mechanics and molecular dynamics simulation program	Prof. Martin Karplus at Harvard University	1983	[94]
Macro Model	Program for molecular modeling of organic compounds and biopolymers; force fields plus energy minimizing algorithms	Fariborz Mohamadi and colleagues, Columbia University, New York; Schrödinger, LLC (2000 Onwards)	1986	[78, 95]
Stata	Statistical program	StataCorp	1985	[79]
Microsoft Paint	Microsoft paint is a software used as painting and graphics editing	Microsoft	1985	[96]

(continued)

Table 1.3 (continued)

Interesting development	Application	Developer/team/ organization	Year	Ref.
ChemDraw	Software for chemical structure drawing, representation, NMR, and mass spectrum simulation	David A. Evans and Stewart Rubenstein, and later by CambridgeSoft, and recently by PerkinElmer from 2011	1985	[67, 68]
X-PLOR	X-PLOR is a software used for protein crystallography	Dr. Axel T. Brunger	1987	[97]
TOPKAT	Correlative SAR system for predicting preclinical toxicity	Kurt Enslein, from Health Designs (Accelry), and later acquired by Biovia	Late 1980s	[75–77, 98]
Design-Expert	Software used in statistical data analysis and design of experiments	Stat-Ease, Inc.	1988	[80]
AutoDock	Molecular modeling software	Scripps Research	1989	[73]
Python	Programming language	Guido van Rossum	1989	[99]

balance disorder [105]. JavaScript and PHP developed in the year 1995 are the programming languages that find use in the development of some applications used in healthcare [34].

Programs like Insight/Discover (Biosym), SYBYL (Tripos), and Quanta/CHARMM (Polygen/Molecular Simulations Inc./Accelrys) were in demand around the globe for molecular modeling and simulations [6, 87, 88]. With the wide use of commercial softwares, in-house computational chemists were able to focus more on specific applications of these softwares as per the requirement of the industry rather than laborious program/software development tasks. This trend had significantly reduced the membership of QCPE, and the depositions under QCPE also decreased to only 1000 [21].

In the 1990s, several popular softwares for computational chemistry, molecular modeling, and pharmacokinetic modeling were introduced. Chemistry at Harvard Macromolecular Mechanics (CHARMM) was developed by Martin Karplus of Accelrys (Now Biovia) in 1991 for molecular dynamics and force simulations [94]. It was widely used for free energy perturbation, methods for quantum mechanics, etc. Integrated Scientific Information System (ISIS)/Draw was a 2D chemical structure drawing software, developed and commercially distributed by MDL from 1991 [106]. *GastroPlus*® is a physiologically based pharmacokinetic simulation program, developed by Simulations Plus in 1998, for the prediction of pharmacokinetic parameters, drug-drug interactions, in vitro in vivo correlation (IVIVC), conduct of in silico trials, and biowaiver estimations [107]. Now this software comes in

ten different modules, and users may purchase one or more module as per their requirements [108].

Public databases containing records of molecules of potential interest to pharmaceutical companies started appearing in the late 1990s. Erlangen/Bethesda Data and Online Service was started as a collaborative effort of researchers at the Computer Chemistry Center (CCC) at the University of Erlangen-Nuremberg, Germany, and National Cancer Institute's (NCI, USA) Laboratory of Medicinal Chemistry, then in Bethesda [109]. Open National Cancer Institute (NCI) Database contained huge data of synthetic and pure natural substances of potential anticancer agents [110].

MACCS and ISIS introduced similarity searching which transformed drug discovery scenario with finding structure similar to lead in chemical structure databases [6]. Docking methodology, which was introduced in 1980s, was employed more in 1990s with rising number of crystal structure of proteins which could be used for designing ligands [111]. Another approach widely explored was designing of an algorithm for a ligand that compiles to the structure of the receptor [112]. Virtual screening was explored widely for generating hypothetical structures of ligand, which have the potential to be synthesized and further experimentation. Combinatorial chemistry was adopted for computer-controlled robotic advancements for synthesizing new compounds [113]. Combinatorial chemistry reduced the workload on medicinal chemists as a single chemist could easily synthesize thousands of compounds per week when compared to the traditional approach of producing one compound in a week. High-throughput screening (HTS) allowed managing and analyzing data of thousands of compounds, easily with the help of computers and the advancements in the field of chemical informatics [114]. In 1997, Lipinski's "Rule of Five" transformed the drug discovery programs of pharmaceutical industries [115]. Pharmaceutical companies and software companies started weeding out compounds as per rule of five. The advancements in computational chemistry, combinatorial chemistry, and HTS reduced the drug discovery timelines considerably.

In 1995, three computational chemists, Paul Crutzen, Mario Molina, and F. Sherwood Rowland, won the Chemistry Nobel for computational work on developing models based on chemistry and thermodynamic laws to explain formation and decomposition of ozone in the earth's atmosphere [116]. Computational chemistry got its own distinct reputation with the works on computational methods in quantum chemistry and density functional theory by Walter Kohn and John Pople in 1998, when they won the Chemistry Nobel [117].

Artificial neural networks (ANNs) were widely used for QSAR in the 1990s [118]. In one study, ANNs with nine hidden layers were developed for predicting the mechanism of action of NCI database indexed anticancer drugs [119]. In another study, 1994, rational designing using trained ANNs was reported for simulating molecular evolution [120, 121]. These ANN studies were some initial instances when ML/DL approaches were explored for their ability to solve problems, learn from their experiences, and to predict new environments. Developments in softwares and hardwares for 1990s are summarized in Table 1.4.

Table 1.4 Some significant developments of 1990s

Interesting development	Application	Developer/team/ organization	Year	Ref.
Linux	Operating system	Linus Torvalds at the Helsinki University of Technology, Finland; Free Software Foundation (FSF)	1990s	[100]
Integrated Scientific Information System/Draw (ISIS/Draw)	Software for 2D drawing of structures and equations; reaction validation features and could calculate formula and molecular weight	Molecular Design Limited (MDL)	1991	[106]
Windows NT	Processor-independent, multiprocessing, and multiuser operating system	Microsoft	1993	[62]
JAVA	Platform-independent programming language	James Gosling, Patrick Naughton, Chris Warth, Ed Frank, and Mike Sheridan at Sun Microsystems	1995	[102]
R	Programming language used for statistical computing and data analysis	Robert Gentleman and Ross Ihaka	1995	[103]
GastroPlus®	First commercial PBPK modeling based on advanced compartmental absorption and transit (ACAT) model	Simulations Plus	1998	[107]

The emergence of high-throughput screening and combinatorial chemistry downsized the role of CADD. Further, the designing of billions of molecules without screening for activity further depressed the growth of CADD. In the early 1990s, hiring of computational/CADD scientists in the pharmaceutical industry reduced drastically. However, SBDD in the late 1990s again catapulted the interest of CADD with the discovery of the HIV protease inhibitors. Many HIV-1 protease inhibitors like saquinavir (developed by Roche and approved by FDA in the year 1996), indinavir (developed by Merck and approved by FDA in the year 1996), ritonavir (developed by AbbVie and approved by FDA in the year 1996), nelfinavir (developed by Agouron Pharmaceuticals and approved by FDA in the year 1997), and amprenavir (approved by FDA in 1999) reached market [122]. Dorzolamide, first SBDD-based drug molecule, developed by Merck and approved by FDA in 1995, was a carbonic anhydrase inhibitor for the treatment of glaucoma and ocular hypertension [123]. Zanamavir and oseltamivir, both neuraminidase inhibitors for the treatment of influenza, developed by Biota Holdings and approved by FDA in 1999, were the outcome of extensive docking and pharmacophore-based drug design, X-ray crystallography, and structural analysis [124]. Tirofiban, an antiplatelet drug for coronary artery, is a glycoprotein IIb/IIIa receptor inhibitor.

Table 1.5 Marketed drugs, whose discovery was assisted through computational chemistry/CADD, of 1990s

Drugs with therapeutic application	Computer application	Year of approval	Ref.
Dorzolamide: carbonic anhydrase inhibitor for the treatment of glaucoma	SBDD-based first drug	1995	[123]
Saquinavir: HIV-1 and HIV-2 protease inhibitor for the treatment of AIDS	Comparative QSAR	1996	[122]
Indinavir: HIV-1 protease inhibitor for the treatment of AIDS	Comparative QSAR	1996	[122]
Ritonavir: HIV-1 protease inhibitor for the treatment of AIDS	Comparative QSAR	1996	[122]
Nelfinavir: HIV-1 protease inhibitor for the treatment of AIDS	Comparative QSAR	1997	[122]
Tirofiban: glycoprotein IIb/IIIa receptor inhibitor for the treatment of coronary disease (antiplatelet drug)	Pharmacophore-based virtual screening	1998	[125]
Zanamivir: neuraminidase inhibitor for the treatment of influenza	Docking, X-ray crystallographic structural analyses, and pharmacophore-based virtual screening (LBDD and SBDD combined)	1999	[124]
Oseltamivir: neuraminidase inhibitor for the treatment of influenza	Docking, X-ray crystallographic structural analyses, and pharmacophore-based virtual screening (LBDD and SBDD combined)	1999	[124]
Amprenavir: HIV-1 protease inhibitor for the treatment of AIDS	Comparative QSAR	1999	[122]

The drug was developed by Merck using a pharmacophore-based virtual screening program and it was approved by FDA in 1998 [125]. Some marketed drugs developed through computational chemistry and CADD are briefed in Table 1.5.

1.6 CADD and Big Data: The 2000s

In 2000s, R and Python were gaining popularity among data scientists and chemists. The simplicity of python and other benefits, like simplified preprocessing, easy to maintain, and productivity for writing code, were being accepted well. R allows data to be preprocessed with assembly code or any language and is good at visualizations

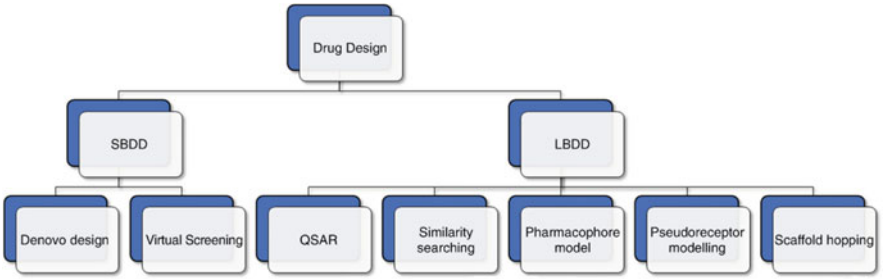


Fig. 1.2 SBDD and LBDD approaches for CADD. Copyright with Springer Nature. Reproduced from [131]

and statistical applications. R can also be used to create graphs, finalize reports, check for accuracy, and validate reports, which made R a vast package ecosystem. Several ML techniques like support vector machine (SVM) and random forest (RF) arose in the early 2000s [118]. From 2001 onwards, QSAR Gordon Research Conference became QSAR CADD with the rising acceptance of CADD term among scientific fields [48].

Trend of in silico experiment became a routine in 2000s. Before proceeding to the preclinical studies, researchers were keen to apply in silico methods alongside in vitro data to create the model and to test it [126]. *Materials Studio*, a PC-based molecular simulating software, was launched in 2000 by Molecular Simulations (Later Accelrys and now Biovia) for materials science research [62, 127]. Materials studio was a modular system having separate modules for molecular modeling, force field, simulation of molecules and materials, and simulation of X-ray/neutron/electron powder diffraction patterns. Later, *Discovery Studio* was launched as a comprehensive software suite to cater the requirements of both SBDD and LBDD [128]. *Glide* was introduced by Schrödinger in 2004 for virtual screening of small molecules and prediction of ligand-protein binding [129, 130]. This decade blurred the distinctions among scientists of various traditional disciplines, and drug research became more interdisciplinary and, thereby, collaborative intensive. Various approaches under SBDD and LBDD are illustrated in Fig. 1.2.

PubChem, ChemSpider, Zinc, Open National Cancer Institute Database, Binding DB, ChemBL, Human Metabolome Database, Drug Bank, and Therapeutic Target Database were some databases of interest to the pharmaceutical and life science [132]. Drug design packages, Discovery Studio (Accelrys Inc.), MOE (Chemical Computing Group), the Schrödinger package (Schrödinger Inc.), and SYBYL (Tripos Inc.) provided most comprehensive tool sets. AutoDock (Scripps Research Institute), DOCK (University of California), FlexX (BioSolveIT GmbH), FRED (OpenEye Scientific Software), Glide (Schrödinger, Inc.), GOLD (Cambridge Crystallographic Data Centre), ICM (Molsoft LLC), and SurflexDock (Tripos) were most popular docking tools of this decade [132]. Catalyst (Integrated in Discovery Studio, Accelrys), DISCOtech (Integrated in SYBYL, Tripos), LigandScout (Ligand), MOE (MOE), and PHASE (Integrated in Schrödinger, Schrödinger) were some popular

pharmacophore modeling programs [132]. For details on open-source drug discovery packages of mid-2000s, authors are directed to a review [133].

Human genome project, a project which started in the 1990s, was completed in 2003 [134]. Genetic studies got a boost for identifying disease risk genes. Additional, new scientific disciplines, termed as "omics," for genomics, proteomics, and metabolomics came into existence. Science of disease progressed further deeper into DNA, RNA, proteins, or other molecules for finding reasons of altered physiology in diseases or disorders. The concept of personalization of drug delivery and precision medicine originated. Pharmaceutical companies (e.g., AstraZeneca) started focusing their small-molecule drug projects on five Rs: the right target, the right patient, the right tissue, the right safety, and the right commercial potential [135]. Out of these five, first four are part of precision medicine. However, the technology limitation for the analysis of new genomic data and cost of analysis limited significant achievements in the 2000s. The future of the pharmaceutical industry goes beyond traditional competencies, which have impacted mergers, acquisitions, and collaborations among pharmaceutical, technology firms, and life science industries [136]. Some significant developments of 2000s are summarized in Table 1.6.

Big data navigated to chemical space [137] with the release of "NCI Open Database" in 1999 containing about 250,000 molecules. This database helped validation of cheminformatics methods and virtual screening techniques. The advent of PubChem and later ChEMBL databases subsequently increased the amount of free data for training and validation requirements. Subsequently, some other databases like Zinc, DrugBank, and ChemSpider followed the trend by providing free access to 2D and 3D structures and biological activity information.

Computational QSAR led to the discovery of rennin inhibitor aliskiren and HIV-1 protease inhibitors like tipranavir and darunavir (Table 1.7). Tipranavir (Aptivus) was launched by Boehringer Ingelheim after its approval by FDA in 2005 [146]. Darunavir as combination therapy with low-dose ritonavir was approved by FDA in 2005 for the treatment of both antiretroviral naïve and multiclass experienced patients [147]. Aliskiren acts as antihypertensive by blocking the first step in the rennin-angiotensin system and it was approved by FDA in 2007 [148]. A summary of drugs approved through computational QSAR is presented in Table 1.7.

1.7 AI Integration with CADD: The 2010s to Present

PCs with high computing powers have addressed the computing speed requirements of scientific softwares and databases. Dedicated workstations are still in wide use for molecular modeling. In the mid of 2010s, Graphics Processing Units' (GPUs) and Google's Tensor Processing Units' (TPUs) inclusion in modern computers have significantly improved computation speed (RAM) to meet the requirements of complex neural networks (deep learning) [149]. Quantum simulation will further increase the computing speed and accuracy of characterizations of molecular systems [150]. Deep neural networks (DNNs) are gaining wide popularity with their better prediction ability than ML algorithms. At the famous Kaggle challenge,

Table 1.6 Some significant developments of 2000s

Interesting development	Application	Developer/team/ organization	Year	Ref.
Materials Studio	Module-based molecular simulating software	Accelrys (now Biovia)	2000	[127] [62]
Discovery Studio	Module-based comprehensive software suite for drug discovery	Accelrys (now Biovia)	2002	[128]
MOE	Integrated platform for computer-aided molecular design and drug discovery	Chemical Computing Group	Around 2000	[138]
BioSuite	Modular software for genome/proteome sequencing, 3D modeling, molecular dynamics simulation, and drug design	CSIR and TCS	2004	[139]
Glide	Virtual screening and molecular docking	Schrödinger, Inc.	2004	[129, 130, 140]
PubChem	Freely accessible database of chemical molecules and their biological activity	National Center for Biotechnology Information	2004	[132]
ZINC	Freely accessible database of biologically relevant and 3D form of the molecule for virtual screening and docking	Irwin and Shoichet Labs, University of California	2004	[141]
Surflex-Dock	Virtual screening software and docking	Tripos Inc. (Now Certara)	2006	[142]
DrugBank	Freely accessible database of drugs, drug targets, and drug actions	Dr. David Wishart, University of Alberta, and The Metabolomics Innovation Centre, University of Alberta	2006	[143]
Lead Finder	Molecular docking	Mol Tech Ltd.	2007	[144]
ChemSpider	Database	Royal Society of Chemistry	2007	[132]
DOCK Blaster	Connects ZINC databases with DOCK to ascertain suitable ligand for protein	University of California	2009	[145]

a team from Toronto used DNNs as a method for QSAR with datasets obtained from Merck's drug discovery efforts [118]. NIH Tox21 challenge for toxicity prediction in 2014 further established the supremacy of DNNs in predicting toxicity of about

Table 1.7 List of drugs discovered through computational QSAR in 2000s

Drug with therapeutic application	Development assisted by	Year of approval	Ref.
Tipranavir: HIV-1 protease inhibitor used in HIV infection	Computational QSAR	2005	[146]
Darunavir: HIV-1 Protease inhibitor for the treatment of both antiretroviral naïve and multiclass experienced patients	Computational QSAR	2006	[147]
Aliskiren: renin inhibitor used as antihypertensive agent	Computational QSAR	2007	[148]

12,000 drugs and environmental chemicals [151]. Machine learning ruled the 1990s to mid of 2000s, but now in 2010s, DL had started outperforming ML in image analysis, molecular design, reaction prediction, and bioactivity prediction [152].

R and Python are two most preferred open-source programming languages for data science applications. R-based package for creating superior and publication-ready graphics is ggplot2, and for R- and Python-based packages are matplotlib, PyX, seaborn, and NetworkX [153]. Softwares like SciPy [153], Open Drug Discovery Toolkit [154], and TeachOpenCADD [155] are open-source computational resources for handling big data. The rise of python is inevitable due to its open-source availability, simplicity, and capability for algorithmic development and exploratory data analysis through a wide variety of machine learning algorithms, both supervised and unsupervised [156]. Computational chemists, data scientists, statisticians, and biologists prefer languages like Python for general purpose or R/MATLAB for predictive modeling, statistical, and machine learning. Python, R, or MATLAB is easy to learn and suits better for interactive execution. Python, R, or MATLAB is either mostly written in C and C++ or either their significant parts are programmed in C or C++. Directly or indirectly, C or C++ is the dominating programming language even now [32].

Chembridge, ZINC, Protein Data Bank (PDB), PubMed, and PubChem are popular databases for structure- or physicochemical property-related computational requirements of scientists [157–159]. CHARM, AMBER, NAMD, GROMACS, and OpenMM are preferred for molecular dynamic simulations [159]. FRED, DOCK, GLIDE, EUDOC, FLOG, SLIDE, ADAM, FlexX, and eHiTS are applied in CADD for systematic searching [159]. AutoDock, AutoDock Vina, GOLD, Cdocker, Mol Dock, Ligand Fit, PLANTS, Molegro Virtual, Docker, and ICM are useful for random searching [159, 160]. Homology modeling is attempted with MODELLER, Deep View, PROCHECK, ERRAT, and SWISS-MODEL [161].

In general, $1.8 billion are now required for developing a new drug [162, 163]. The time required for drug discovery has dropped from 20 years to 10 years in the last 30 years. The credit of this fast pace of drug discovery goes to computers, softwares, Internet, CADD, big data, artificial intelligence, and automation. Biologists, biophysicists, and computational scientists are now working in close collaborations for developing new drug molecules. Both commercial and academic CADD platforms have shifted the focus on faster drug discovery through

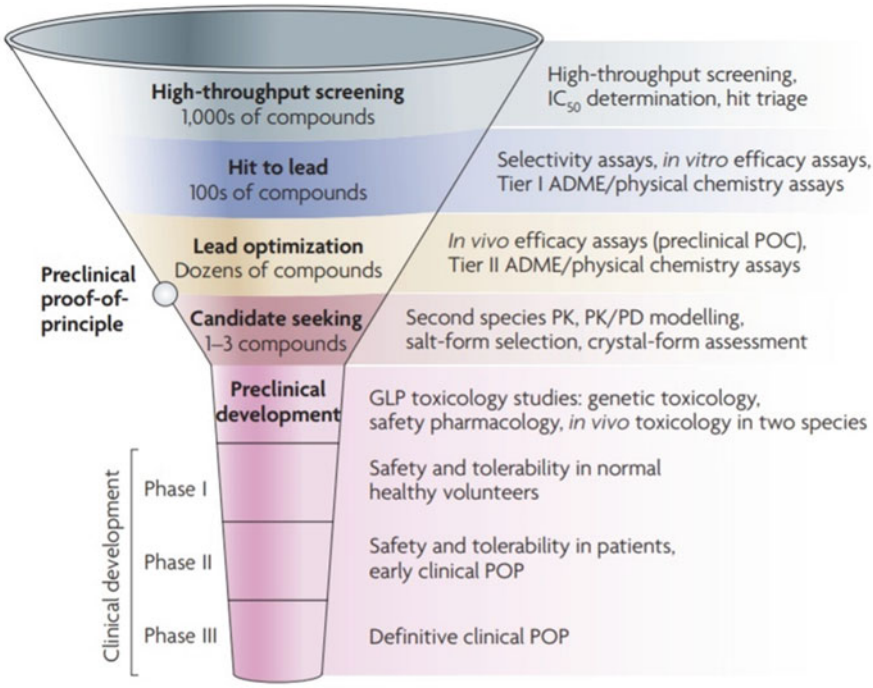


Fig. 1.3 The funnel of drug discovery. Copyright with Springer Nature. Reproduced from [166]

two approaches SBDD or LBDD [164]. AI-based drug development technology is well spun within CADD platforms for enhancing the accuracy and prediction. The readers are directed to a review article for more details on AI-based computational tools for drug discovery [165]. Quantum computing with noisy intermediate-scale quantum (NISQ) devices and fault-tolerant quantum computing devices (FTQC) could transform the molecular simulations with their high speed and increases prediction accuracy. CADD platforms act as primary screen on compound libraries received from HTS, and secondary CADD screens predicts PK/PD and toxicity prior to the conduct of preclinical studies. Clinical development, see the bottom part of funnel of drug discovery in Fig. 1.3, is also being accelerated with the use of softwares and databases for data collection, management, and analysis.

In 2014, Okada et al. performed a large-scale genomic-wide association studies on rheumatoid arthritis by analyzing samples obtained from humans of various ethnic backgrounds [167]. This study highlighted more than 100 disease susceptibility loci, of which 42 were novel. Motivated with this outcome, pharmaceutical companies like AstraZeneca started working on the development of CDK4/6 inhibitors as a potential indication for rheumatoid arthritis [162]. This study generated huge interest of the pharmaceutical companies to work in the area of pharmacogenomics. With the availability of human genome data in public domain, advances in genomics and proteomics have helped to explore potential drug targets

Table 1.8 List of drugs discovered using SBDD and LBDD approaches in 2010s

Drug	Computer application	Year of approval	Reference
Vemurafenib: for stage III BRAF V600 mutation-positive melanoma	First FBDD approved drug; took only 6 years from development to approval	2011	[171]
Crizotinib: for metastatic non–small cell lung cancer	SBDD	2016	[174]
Ventoclax: for the treatment of chronic lymphocytic leukemia	FBDD	2016	[172, 173]
Tagraxofusp-erzs: for blastic plasmacytoid dendritic cell neoplasm	Molecular docking	2018	[176]
DSP-1181: for the treatment of obsessive compulsive disorder	Algorithm-based drug development	Currently in clinical trial	[175]

that are subjected to validation studies for a significant outcome [163]. The driving forces for pharmacogenomics are the high cost of genomic analysis and the availability of open access big data. The cost for generating the initial "draft" of human genome sequence was $ 2.7 billion, which was dropped to $300 in 2020 [168]. The open domain availability of research data is now supported by most of the governments and funding agencies with their strict policies to publish papers in open access journals. Open access and open science are contributing to research by ensuring the free availability of research literature.

FBDD has now become a popular CADD with the entry of vemurafenib and ventoclax in the market. FBDD was referred to as SAR by NMR in 1990s by Abbott [169]. FBDD approach has the advantage over high-throughput screening for higher hit rates and hit compounds with low molecular weights, good solubility, and high affinity [170]. Vemurafenib, Zelboraf™, discovered at Plexxikon in partnership with Roche and later acquired by Daiichi Sankyo, was the first molecule developed with FBDD approach and it took around 6 years from fragment to approval [171]. Vemurafenib targets Ser-Thr protein kinase, B-Raf (V600E), mutation, and increases survival by approximately 6 month in late-stage melanoma patients. Ventoclax (ABT-199), Venclexta™, discovered at AbbVie (previously Abbott), was another outcome of FBDD and it took around two decades from initial 3D structure to approval [172, 173]. Ventoclax targets Bcl-2, a regulatory protein of apoptosis, and was approved by FDA in 2016 for the treatment of chronic lymphocytic leukemia (CLL) in patients with relapsed CLL, who have 17p deletion (mutation). In 2016, the US FDA approved crizotinib capsules (Xalkori, Pfizer, Inc.) for the treatment of patients with metastatic non–small cell lung cancer (NSCLC), whose tumors are ROS1 positive [174]. Crizotinib is an inhibitor of cMet (hepatocyte growth factor receptor) and developed by SBDD.

Britain's Exscientia and Japan's Sumitomo Dainippon Pharma have collaborated for developing an investigational drug DSP-1181 for obsessive compulsive disorder [175]. Usually, the drug takes about 4–6 years of time before it reaches clinical trials

but Exscientia's AI technology has extensively reduced this development time to less than a year. Table 1.8 summarizes various drugs developed recently through CADD/computational chemistry.

Many expectations of the 2000s are now fulfilled with the advancements in CADD tools [48, 177]. Virtual screening is now a routine in the pharmaceutical industry. We have learnt and applied the concepts of computational thermodynamics (free energy perturbation) to anticipate affinity of compounds. Lots of PK and ADME data are now available to help conversion of potent ligands into drugs. CADD tools are so dispersed with the advancements in technology that they are now more affordable and easy to use. We have gained quality knowledge for computational handling of chemical compounds and biological molecules. On the other side, the advent of big data with omics has brought some great challenges to deal with biological molecules. We are learning with new molecules and therapies like antibodies, gene therapy, and antisense oligonucleotides for new targets, better selection of drug targets, and investigating their novel molecular mechanisms. Gene expression, protein modifications, equilibria and biased signaling, and receptor conformational changes are some of these challenges which are expected to be explored in the near future. Complexity of biological molecules makes less accurate predictions, wherein the potential role of AI in forthcoming years cannot be ignored.

1.8 Computer-Aided Pharmaceutics and Drug Delivery

Computational pharmaceutics is a branch of pharmaceutical sciences which focuses on computer simulations for solving the problems of pharmaceutical formulation and drug delivery issues. Computational pharmaceutics involves applications of molecular modeling techniques to study pharmaceutical formulations at molecular level to develop better formulations that can improve the process of drug delivery. The term "Computational Pharmaceutics" was first used by Defang Ouyang and Sean C. Smith in their book "*Computational Pharmaceutics: Application of Molecular Modeling in Drug Delivery*" published by John Wiley & Sons in 2015 [178]. Ouyang and Smith define "computational pharmaceutics involves the application of computational modeling to drug delivery and pharmaceutical nanotechnology" [179]. Pharmaceutical research, being much broader than the design of drugs alone, also involves development of pharmaceutical formulations for effective drug delivery. This field holds a great potential to explore mechanistic details on how drug molecules can interact with the molecules of excipients/polymers at both nano- and microlevel for developing rational pharmaceutical formulations. Earlier in 2005, ACS Journal "*Molecular Pharmaceutics*" started publishing cutting-edge research on molecular-level mechanistic understanding of pharmaceutical formulations and drug delivery systems [180]. Some expert reviews in the field of computational pharmaceutics have also highlighted the potential role of molecular dynamics simulations and other computational techniques in predicting solubility and permeability of drugs [115, 181].

The approach of this book is not to limit the computer applications to pharmaceutical formulations. Rather, it further extends computer applications to different subdisciplines of pharmaceutics such as pharmaceutical manufacturing (technology), biopharmaceutics, pharmacokinetics, dispensing pharmacy, and physical pharmacy. Thus, this book covers a broader area of computer applications in pharmaceutics like development and optimization of pharmaceutical formulations, pharmaceutical manufacturing, biopharmaceutical characterization, prediction of physicochemical properties, ADMET simulation, PK/PD prediction, AI, automation, robotic technologies, information retrieval, modeling of drug-polymer interactions, and nanoparticles.

Chapter 2 focuses on historical developments of computer applications in pharmaceutics; in particular, formulation optimization, robotics, artificial intelligence, 3D printing, and pharmacokinetics.

Chapter 3 discusses various Design of Experiments (DoE) for optimizing pharmaceutical formulations. Two case studies have also been included to understand the concepts and applications of computer-based optimization.

Chapter 4 explains the emergence of the QbD regime through various ICH, The International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use, and guidelines. Further key elements of QbD are presented with some examples.

Chapter 5 presents a step-by-step protocol of an experimental designing and optimization of a pharmaceutical formulation using Design-Expert software.

Spritam (levetiracetam), first 3D-printed, orally disintegrating tablet for the treatment of seizures in epilepsy, was granted marketing permission in 2015 by FDA [182]. In 2020, FabRx has introduced a 3D printer MediMaker for manufacturing of personalized medicine [183]. These key developments are attracting the attention of pharmaceutical manufactures to 3D printing of medicines.

Chapter 6 gives a detailed account on 3D-printing technologies and their applications in making medicines. A section of Chap. 2 highlights key historical developments on 3D printing.

Chapter 7 deals with basics and recent advances along with applications and limitations of commonly used in silico and computational models for biopharmaceutical characterization, particularly the ACAT model-based GastroPlus™ software package.

Chapter 8 discusses the role, advancement, and development of biocomputational tools for simulating pharmacokinetics and pharmacodynamics of drugs.

Chapter 9 explores the history, regulatory aspects, and various components of physiological-based pharmacokinetics (PBPK), along with the general workflow and approaches for model development and salient applications of PBPK modeling in drug development.

Chapter 10 describes the role of computers in data collection and management for clinical trials. Most current e-technologies for data collection like softwares, digital wearables, and healthapps are also discussed. Pharmaceutical companies are also using these e-technologies even for their marketed products, for example, digital

pills are being used for collection of data from patients for improving the patient's adherence to the dosage regimen [184].

Chapter 11 discusses the role of ML and DL techniques in drug discovery, formulation development, and healthcare.

Chapter 12 gives a most recent update on robotic innovations for automation of manufacturing, packaging, warehousing, and laboratory processes of pharmaceutical industries. Industrial robots, cleanroom robots, robotic factory for production of personalized medicines [185], collaborative robots [186], laboratories assistants [187], and other robotic innovations are described.

Chapter 13 reviews drug delivery attempts with soft robots. A commercialized robotic pill [188] and other soft robots are described for targeting drugs to a particular site and sustained delivery of drugs.

Comprehensive literature searching has received a good impetus with Budapest Open Access Initiative [189] leading to rise of open access resources, repositories, and public. Various databases and literature resources, both commercial and open access, for comprehensive literature searching are described in Chap. 14. Guidance on comprehensive literature searching and electronic searching of databases is also provided to novice researchers and postgraduate students.

Chapter 15 gives a comprehensive coverage on systematic review of patent information in connection to the patent classification systems, types of patent searches, information sources, various search engines, and their peculiar characteristics.

Chapter 16 focuses on the different computational approaches such as molecular docking, virtual screening, homology modeling, pharmacophore development, and QSAR, with some illustrative case studies on clinically approved drugs.

Chapter 17 explains the applications of QSPR models in addressing critical issues of solubility and permeability of drugs in the development of rational pharmaceutical formulations.

Chapter 18 explores in silico approaches to generate three-dimensional models of drugs and polymers for evaluating interactions between them as a basis for the rational design of pharmaceutical formulations.

Chapter 19 discusses the current state, computer-aided analysis for drug and drug products, in particular, chromatographic data systems, analytical method development, analytical QbD, and nanoparticle tracking analysis.

Chapter 20 explores telemedicine's background, its current state, and future perspectives as a new frontier in providing a quality health service.

Bioinformatics tools for drug design and delivery systems are discussed in Chap. 21 with their applications in the treatment and personalized therapy.

Statistical aspects of modeling, namely univariate analysis, multivariate analysis, principal component analysis (PCA), variability analysis, probabilistic modeling, and support vector machines, are explained in Chap. 22.

In the modern era, nanotechnology is playing a pivotal role in drug delivery through various nanoparticulate drug carriers like dendrimer, quantum dots, silver nanoparticles, copper nanoparticles, etc. [190].

Chapter 23 explains various computational microscopic, mesoscopic, and other models for their applications in developing nanoparticulate formulations.

In Chap. 24, authors introduce "Pharmaceutics Informatics" to compass the application of both the bioinformatics and chemoinformatics tools in drug delivery. Data mining, computing physicochemical descriptors, machine learning methods, molecular dynamics, and docking experiments are explained for in silico preparation and characterization of drug-loaded delivery systems.

A practical approach on the development of a model of human extrathoracic airway is described in Chap. 25 for inhalation drug delivery. Experimental studies using 3D-printed hollow casts are presented to cross-validate the predictive dosimetry models.

1.9 Conclusion and Future Prospects

This chapter has brought forward a detailed description of how historical developments in the fields of hardware, software, and programming language have driven drug discovery and development. Dominance of QCPE for more than 25 years in twentieth century proved that commercial softwares are generally not widely used among the scientific community. Early, twenty-first century has also witnessed wide use of open/free access softwares and databases in the field of drug discovery and development. Scientific programming is governed by the execution of code rather than the total number of lines of a code. Computer languages like R, MATLAB, and Python are dominating languages of the early twenty-first century due to their interactive execution and, thereby, these languages are in extensive use for developing softwares and databases for both commercial and free access. Computational chemistry arises as an offshoot of physical chemistry and CADD developed gradually as a branch of computational chemistry. Current AI developments are more demanding in terms of computing speed which will obviously bring in quantum computing in a more aggressive manner. The quest of supremacy of ML or DL is not yet over as both are continuously evolving. However, the role of both ML and DL in the development of new algorithms for understanding the complexity of biological molecules will inevitably rise. Molecular modeling will see newer developments to address the complexity of biological molecules.

References

1. Wilkes MV (2000) Introduction to "Babbage's analytical engine plans 28 and 28a—the Programmer's Interface" By Allan Bromley. IEEE Ann Hist Comput 22:4–5. https://doi.org/10.1109/MAHC.2000.887985
2. Daylight EG (2015) Towards a historical notion of 'turing—the father of computer science'. Hist Philos Log 36:205–228. https://doi.org/10.1080/01445340.2015.1082050
3. Copeland BJ, Proudfoot D (2021) Alan Turing father of modern computer. RutherfordjournalOrg. http://www.rutherfordjournal.org/article040101.html#sdfootnote5269sym. Accessed 7 May 2021

4. History. UNIVAC, the first commercially produced digital computer, is dedicated 1951. https://www.history.com/this-day-in-history/univac-computer-dedicated. Accessed 7 May 2021
5. IBM. IBM 701 (2021) https://www.ibm.com/ibm/history/exhibits/701/701_intro.html. Accessed 5 May 2021
6. Boyd DB, Marsh MM (2006) History of computers in pharmaceutical research and development: a narrative. In: Ekins S (ed) Comput. Appl. Pharm. Res. Dev. Wiley, Hoboken, pp 1–50. https://doi.org/10.1002/0470037237.ch1
7. Iverson KE (1962) A programming language. In: AFIPS Proc.—1962 Spring Jt. Comput. Conf. AFIPS 1962. Association for Computing Machinery, Inc, New York, pp 345–351. https://doi.org/10.1145/1460833.1460872
8. Mali P, Panchal S (2011) A review on worldwide essential software resources in pharmacy. Chronicles Young Sci 2:11–11
9. Jamkhande PG, Ghante MH, Ajgunde BR (2017) Software based approaches for drug designing and development: a systematic review on commonly used software and its applications. Bull Fac Pharm Cairo Univ 55:203–210. https://doi.org/10.1016/j.bfopcu.2017.10.001
10. Shaffer PA, Schomaker V, Pauling L (1946) The use of punched cards in molecular structure determinations I. Crystal structure calculations. J Chem Phys 14:648–658. https://doi.org/10.1063/1.1724081
11. IBM (2021) 7094 Data Processing System. https://www.ibm.com/ibm/history/exhibits/mainframe/mainframe_PP7094.html. Accessed 3 May 2021
12. Dykstra CE, Frenking G, Kim KS, Scuseria GE (2005) Computing technologies, theories, and algorithms: the making of 40 years and more of theoretical and computational chemistry. In: Theory and applications of computational chemistry. Elsevier, pp 1–7. https://doi.org/10.1016/B978-044451719-7/50044-5
13. Alfonseca M (1979) APL continuous system modelling program: an interactive simulation language. Adv Eng Softw 1:73–76. https://doi.org/10.1016/0141-1195(79)90004-4
14. Quinn C (2016) Doug Engelbart: more than the inventor of the mouse. Learn Solut. https://learningsolutionsmag.com/articles/1957/doug-engelbart-more-than-the-inventor-of-the-mouse. Accessed 3 May 2021
15. YouTube (1968) The Mother of All Demos, presented by Douglas Engelbart. https://www.youtube.com/watch?v=yJDv-zdhzMY. Accessed 5 May 2021
16. Buxton B (2008) The long nose of innovation. Bus Week. http://www.businessweek.com/stories/2008-01-02/. Accessed 3 May 2021
17. Sammet JE (1993) The beginning and development of FORMAC (FORmula MAnipulation Compiler). ACM SIGPLAN Not 28:209–230. https://doi.org/10.1145/155360.155372
18. Dahl OJ, Nygaard K (1966) SIMULA: an ALGOL-based simulation language. Commun ACM 9:671–678. https://doi.org/10.1145/365813.365819
19. Griswold RE (1978) A history of the SNOBOL programming languages. ACM SIGPLAN Not 13:275–308. https://doi.org/10.1145/960118.808393
20. Ozgur C, Kleckner M, Li Y (2015) Selection of statistical software for solving big data problems: a guide for businesses, students, and universities. SAGE Open 5:215824401558437. https://doi.org/10.1177/2158244015584379
21. Boyd DB (2013) Quantum chemistry program exchange, facilitator of theoretical and computational Chemistry in pre-internet history. ACS Symp Ser 1122:221–273. https://doi.org/10.1021/bk-2013-1122.ch008
22. History of Visualization of Biological Macromolecules. ORTEP (2021) https://www.umass.edu/molvis/francoeur/ortep/ortep.html#report1. Accessed 3 May 2021
23. Lewars EG (2011) Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-90-481-3862-3

24. Hansch C, Fujita T (1964) ρ-σ-π analysis. A method for the correlation of biological activity and chemical structure. J Am Chem Soc 86:1616–1626. https://doi.org/10.1021/ja01062a035
25. Free SM, Wilson JW (1964) A mathematical contribution to structure-activity studies. J Med Chem 7:395–399. https://doi.org/10.1021/jm00334a001
26. Wachtel RE (1988) Use of BMDP statistical package to generate maximum likelihood estimates for single channel data. J Neurosci Methods 25:121–128. https://doi.org/10.1016/0165-0270(88)90148-3
27. Vueghs P, De Koning HP, Pin O, Beckers P (2010) Use of geometry in finite element thermal radiation combined with ray tracing. J Comput Appl Math 234:2319–2326. https://doi.org/10.1016/j.cam.2009.08.088
28. Burnett MN, Johnson CK (2021) ORTEP-III. Oak Ridge National Laboratory. https://ornl-ndav.github.io/ortep/ortep.html. Accessed 3 May 2021
29. Caskie REM, Mason REA (1973) Some design features of continuous system modelling program III. INFOR Inf Syst Oper Res 11:125–139. https://doi.org/10.1080/03155986.1973.11731543
30. Babatz R, Pankoke-Babatz U, Santo H, Theidig G (1983) Experimental use of KOMEX in the GMD. Comput Compact 1:83–88. https://doi.org/10.1016/0167-7136(83)90006-9
31. IBM100 (2021) The floppy disk. https://www.ibm.com/ibm/history/ibm100/us/en/icons/floppy/. Accessed 3 May 2021
32. Rassokhin D (2020) The C++ programming language in cheminformatics and computational chemistry. J Cheminform 12:10–26. https://doi.org/10.1186/s13321-020-0415-y
33. Chonacky N, Winch D (2005) Reviews of maple, mathematica, and matlab: coming soon to a publication near you. Comput Sci Eng 7:9–10. https://doi.org/10.1109/MCSE.2005.39
34. Enuke (2021) Which is best programming language for healthcare apps? https://www.enukesoftware.com/blog/which-is-best-programming-language-for-healthcare-apps.html. Accessed 28 Apr 2021
35. Krdzavac N, Mosbach S, Nurkowski D, Buerger P, Akroyd J, Martin J et al (2019) An ontology and semantic web service for quantum chemistry Calculations. J Chem Inf Model 59:3154–3165. https://doi.org/10.1021/acs.jcim.9b00227
36. Gaussian.com (2021) Expanding the limits of computational chemistry. https://gaussian.com/. Accessed 5 May 2021
37. Löwdin PO (ed) (1974) Proceedings of the international symposium on quantum biol-ogy and quantum pharmacology, Held at Sanibel Island, Florida, January 17–19, 1974. Int J Quantum Chem, Quantum Biol Symp No. 1, Wiley, New York
38. Graham Richards W (1985) Computer-aided molecular design. Comput Des 17:215–218. https://doi.org/10.1016/0010-4485(85)90072-7
39. RCSB PDB (2021) PDB history. https://www.rcsb.org/pages/about-us/history. Accessed 3 May 2021
40. The Cambridge Crystallographic Data Centre (CCDC) (2021) The Cambridge Structural Database (CSD). https://www.ccdc.cam.ac.uk/solutions/csd-core/components/csd/. Accessed 3 May 2021
41. Görlich D, Faldum A (2019) Implementing online interventions in ICare: a biostatistical perspective. Internet Interv 16:5–11. https://doi.org/10.1016/j.invent.2018.12.004
42. Chun-rong W, Chun-nuan J, Hai-yan W, Rong-jun Q, Chun-hua W (2021) Application of Minitab 15.0 Software to the preparation of phenolic resin-typed microspheres. CNKI. https://en.cnki.com.cn/Article_en/CJFDTotal-WOOD200903014.htm. Accessed 3 May 2021
43. Adnanhatem O, Suhail FSA, Juda AM (2020) Optimized programs and methods required for the computational study of beta blockers. In: J. Phys. Conf. Ser., vol. 1664, IOP Publishing Ltd, p 012088. https://doi.org/10.1088/1742-6596/1664/1/012088
44. Eichenberger AP, Allison JR, Dolenc J, Geerke DP, Horta BAC, Meier K et al (2011) GROMOS++ software for the analysis of biomolecular simulation trajectories. J Chem Theory Comput 7:3379–3390. https://doi.org/10.1021/ct2003622

45. Frenz B (1988) Computers and crystallography: joint progress. Comput Phys 2:42. https://doi.org/10.1063/1.4822724
46. Fernbach S, Taub AH (1970) Computers and their role in the physical sciences. Gordon and Breach
47. Ekins S (2006) Computer applications in pharmaceutical development, 1st edn. Wiley
48. Drie JH (2007) Computer-aided drug design: the next 20 years. J Comput Aided Mol Des 21: 591–601. https://doi.org/10.1007/s10822-007-9142-y
49. Martin YC (1978) Quantitative drug design: a critical introduction (medicinal research: 8), vol 8. Marcel Dekker, Inc., New York. https://doi.org/10.1002/jps.2600680152
50. Matthews DA, Alden RA, Bolin JT, Filman DJ, Freer ST, Hamlin R et al (1978) Dihydrofolate reductase from Lactobacillus casei. X-ray structure of the enzyme methotrexate. NADPH complex. J Biol Chem 253:6946–6954
51. Matthews DA, Alden RA, Freer ST, Xuong N, Kraut J (1979) Dihydrofolate reductase from Lactobacillus casei. Stereochemistry of NADPH binding. J Biol Chem 254:4144–4151. https://doi.org/10.1016/s0021-9258(18)50708-0
52. Spicer D (2016) Raymond Tomlinson: Email Pioneer, Part 1. IEEE Ann Hist Comput 38:72–79. https://doi.org/10.1109/MAHC.2016.25
53. Ritchie DM, Thompson K (1983) The UNIX time-sharing system. Commun ACM 26:84–89. https://doi.org/10.1145/357980.358014
54. Ojha M (2018) Chronic risk and disease management model using structured query language and predictive analysis. South Dakota State University. https://openprairie.sdstate.edu/etd/2480. Accessed 3 May 2021
55. Sterling T, Anderson M, Brodowicz M (2018) Performance monitoring. In: High performance computing. Elsevier Inc., pp 383–420. https://doi.org/10.1016/B978-0-12-420158-3.00013-7
56. Liskov B, Snyder A, Atkinson R, Schaffert C (1977) Abstraction mechanisms in CLU. ACM SIGPLAN 20:564–576. https://doi.org/10.1145/390017.808322
57. Melli R (1988) Automated software project planning and control. IFAC Proc Vol 21:1–7. https://doi.org/10.1016/S1474-6670(17)53672-1
58. Karanassios V, Horlick G (1985) Smart backplanes-I. The apple II. Talanta 32:601–614. https://doi.org/10.1016/0039-9140(85)80156-9
59. Bergin TJ (2006) The origins of word processing software for personal computers: 1976-1985. IEEE Ann Hist Comput 28:32–47. https://doi.org/10.1109/MAHC.2006.76
60. Kokol P (1987) Spreadsheet programs in software engineering. ACM SIGSOFT Softw Eng Notes 12:45–50. https://doi.org/10.1145/29934.29941
61. Ivakhnenko A, Lapa V (1967) Cybernetics and forecasting techniques. American Elsevier Pub. Co., New York
62. Knapman K (2000) Computational chemistry on the desktop PC. Chem Innov 30:38–44
63. Carr JR (2021) Computer programming in the 21st century: quo vadis. Univ Nevada. https://www.uh.edu/~jbutler/anon/carr.html. Accessed 3 May 2021
64. The R Foundation (2021) R: The R Project for Statistical Computing. https://www.r-project.org/. Accessed 3 May 2021
65. Haigh T (2008) Cleve Moler: mathematical software pioneer and creator of Matlab. IEEE Ann Hist Comput 30:87–91. https://doi.org/10.1109/MAHC.2008.2
66. Wiley (2021) Journal of Computational Chemistry. https://www.wiley.com/en-us/Journal+of+Computational+Chemistry-p-9780471541356. Accessed 3 May 2021
67. Halflord B (2014) Reflections on ChemDraw. Chem Eng News Arch 92:26–27. https://doi.org/10.1021/cen-09233-scitech1
68. Nguyen A, Huang YC, Tremouilhac P, Jung N, Bräse S (2019) ChemScanner: extraction and re-use(ability) of chemical information from common scientific documents containing ChemDraw files. J Cheminform 11:77. https://doi.org/10.1186/s13321-019-0400-5
69. Buntrock RE (2014) Chemical information: from print to the internet. ACS Symp Ser 1164: 19–42. https://doi.org/10.1021/bk-2014-1164.ch002

70. Stewart Computational Chemistry (2007) MOPAC®. http://openmopac.net/. Accessed 7 May 2021
71. Weiner PK, Kollman PA (1981) AMBER: assisted model building with energy refinement. A general program for modeling molecules and their interactions. J Comput Chem 2:287–303. https://doi.org/10.1002/jcc.540020311
72. Salomon-Ferrer R, Case DA, Walker RC (2013) An overview of the Amber biomolecular simulation package. Wiley Interdiscip Rev Comput Mol Sci 3:198–210. https://doi.org/10.1002/wcms.1121
73. Goodsell DS, Olson AJ (1990) Automated docking of substrates to proteins by simulated annealing. Proteins Struct Funct Bioinforma 8:195–202. https://doi.org/10.1002/prot.340080302
74. Leo A (2021) The history of the development of CLOGP. BioByte Corp. https://www.daylight.com/meetings/summerschool99/course/basics/ref/leo/clogp_history.html. Accessed 3 May 2021
75. Enslein K (1984) Estimation of toxicological endpoints by structure-activity relationships. Pharmacol Rev 36:131S–135S
76. Patlewicz G, Rodford R, Walker JD (2003) Quantitative structure-activity relationships for predicting mutagenicity and carcinogenicity. Environ Toxicol Chem 22:1885–1893
77. BIOVIA—Dassault Systèmes® (2021) BIOVIA Databases. https://www.3ds.com/products-services/biovia/products/scientific-informatics/biovia-databases/. Accessed 3 May 2021
78. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C et al (1990) Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. J Comput Chem 11:440–467. https://doi.org/10.1002/jcc.540110405
79. Cox NJ (2005) A brief history of Stata on its 20th anniversary. Stata J 5:2–18. https://doi.org/10.1177/1536867x0500500102
80. Stat-Ease. Design-Expert®. https://www.statease.com/about-us/history/. Accessed 9 May 2021
81. Rachlin H (2014) The escape of the mind. Oxford Univ Press. https://doi.org/10.1093/acprof:oso/9780199322350.001.0001
82. Johnson J, Roberts TL, Verplank W, Smith DC, Irby CH, Beard M et al (1989) Xerox Star: a retrospective. Computer (Long Beach Calif) 22:11–26. https://doi.org/10.1109/2.35211
83. Seiler WM (1991) SUPERCALC—a REDUCE package for commutator calculations. Comput Phys Commun 66:363–376. https://doi.org/10.1016/0010-4655(91)90082-V
84. NobelPrize.org (2021) The nobel prize in chemistry 1985. Nobel Media AB 2021. https://www.nobelprize.org/prizes/chemistry/1985/summary/. Accessed 3 May 2021
85. Koga H, Itoh A, Murayama S, Suzue S, Irikura T (1980) Structure-activity relationships of antibacterial 6, 7-and 7, 8-disubstituted 1-Alkyl-1, 4-dihydro-4-oxoquinoline-3-carboxylic acids. J Med Chem 23:1358–1363. https://doi.org/10.1021/jm00186a014
86. Ito A, Hirai K, Inque M (1980) In vitro antibacterial activity of AM-715, a new nalidixic acid analog. Antimicrob Agents Chemother 17:103–108. https://doi.org/10.1128/AAC.17.2.103
87. Taylor JS, Garrett DS, Cohrs MP (1988) Solution-state structure of the dewar pyrimidinone photoproduct of thymidyly1-(3'→5/)-thymidine. Biochemistry 27:7206–7215. https://doi.org/10.1021/bi00419a007
88. Hyde RM, Livingstone DJ (1988) Perspectives in QSAR: computer chemistry and pattern recognition. J Comput Aided Mol Des 2:145–155. https://doi.org/10.1007/BF01532089
89. Richard WB (1985) Digital technical journal. Digital Equipmt:nt Corporation
90. New products (1987) Computer (Long Beach Calif) 20:116–122. https://doi.org/10.1109/MC.1987.1663573
91. Bride E (2011) The IBM personal computer: a software-driven market. Computer (Long Beach Calif) 44:34–39. https://doi.org/10.1109/MC.2011.193
92. Pirhadi S, Sunseri J, Koes DR (2016) Open source molecular modeling. J Mol Graph Model 69:127–143. https://doi.org/10.1016/j.jmgm.2016.07.008

93. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S et al (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun 91:1–41. https://doi.org/10.1016/0010-4655(95)00041-D
94. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4:187–217. https://doi.org/10.1002/jcc.540040211
95. Sisodiya D (2012) Drug designing softwares and their applications in new drug discover. J Pharm Res 5:124–126
96. Davison P (2014) Because of the pixels: on the history, form, and influence of MS paint. J Vis Cult 13:275–297. https://doi.org/10.1177/1470412914544539
97. Schwieters CD, Bermejo GA, Clore GM (2018) Xplor-NIH for molecular structure determination from NMR and other data sources. Protein Sci 27:26–40. https://doi.org/10.1002/pro.3248
98. Enslein K, Blake BW, Borgstedt HH (1990) Prediction of probability of carcinogenicity for a set of ongoing NTP bioassays. Mutagenesis 5:305–306. https://doi.org/10.1093/mutage/5.4.305
99. Tulchak LV, Marchuk AO (2016) History of python. Vinnytsia National Technical University. https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/10471/461.pdf?sequence=3. Accessed 3 May 2021
100. Bokhari SH (1995) The Linux operating system. Computer (Long Beach Calif) 28:74–79. https://doi.org/10.1109/2.402081
101. Computer History Museum (2021) Floppy disks. https://www.computerhistory.org/revolution/memory-storage/8/261. Accessed 3 May 2021
102. Schildt H, Java TM (2007) The complete reference, 7th edn. The McGraw-Hill Companies, New York
103. Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. J Comput Graph Stat 5:299. https://doi.org/10.2307/1390807
104. Tippmann S (2015) Programming tools: adventures with R. Nature 517:109–110. https://doi.org/10.1038/517109a
105. Harkness R, Crook M, Povey D (2007) Programming review of visual Basic.NET for the laboratory automation industry. J Assoc Lab Autom 12:25–32. https://doi.org/10.1016/j.jala.2006.10.014
106. Dalby A, Nourse JG, Hounshell WD, Gushurst AKI, Grier DL, Leland BA et al (1992) Description of several chemical structure file formats used by computer programs developed at molecular design limited. J Chem Inf Comput Sci 32:244–255. https://doi.org/10.1021/ci00007a012
107. Hens B, Bolger MB (2019) Application of a dynamic fluid and pH model to simulate intraluminal and systemic concentrations of a weak base in GastroPlus™. J Pharm Sci 108:305–315. https://doi.org/10.1016/j.xphs.2018.10.041
108. Kato T, Nakagawa H, Mikkaichi T, Miyano T, Matsumoto Y, Ando S (2020) Establishment of a clinically relevant speci fication for dissolution testing using physiologically based pharmacokinetic (PBPK) modeling approaches. Eur J Pharm Biopharm 151:45–52. https://doi.org/10.1016/j.ejpb.2020.03.012
109. NCI/CADD Group (2020) National Cancer Institute. https://cactus.nci.nih.gov/ncicadd/about.html. Accessed 3 May 2021
110. Monga M, Sausville EA (2002) Developmental therapeutics program at the NCI: molecular target and drug discovery process. Leukemia 16:520–526. https://doi.org/10.1038/sj.leu.2402464
111. Torres PHM, Sodero ACR, Jofily P, Silva FP Jr (2019) Key topics in molecular docking for drug design. Int J Mol Sci 20:4574. https://doi.org/10.3390/ijms20184574

112. Guedes IA, de Magalhães CS, Dardenne LE (2014) Receptor-ligand molecular docking. Biophys Rev 6:75–87. https://doi.org/10.1007/s12551-013-0130-2
113. Geysen HM, Schoenen F, Wagner D, Wagner R (2003) Combinatorial compound libraries for drug discovery: an ongoing challenge. Nat Rev Drug Discov 2:222–230. https://doi.org/10.1038/nrd1035
114. Pereira DA, Williams JA (2007) Origin and evolution of high throughput screening. Br J Pharmacol 152:53–61. https://doi.org/10.1038/sj.bjp.0707373
115. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26. https://doi.org/10.1016/S0169-409X(00)00129-0
116. American Chemical Society (2021) Computational chemistry. https://www.acs.org/content/acs/en/careers/college-to-career/chemistry-careers/computational-chemistry.html. Accessed 3 May 2021
117. NobelPrize.org (2021) The nobel prize in chemistry 1998. Nobel Prize Outreach AB. 2021. https://www.nobelprize.org/prizes/chemistry/1998/summary/. Accessed 3 May 2021
118. Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V (2015) Deep neural nets as a method for quantitative structure-activity relationships. J Chem Inf Model 55:263–274. https://doi.org/10.1021/ci500747n
119. Weinstein JN, Kohn KW, Grever MR, Viswanadhan VN, Rubinstein LV, Monks AP et al (1992) Neural computing in cancer drug development: predicting mechanism of action. Science 258:447–451. https://doi.org/10.1126/science.1411538
120. Schneider G, Wrede P (1994) The rational design of amino acid sequences by artificial neural networks and simulated molecular evolution: de novo design of an idealized leader peptidase cleavage site. Biophys J 66:335–344. https://doi.org/10.1016/S0006-3495(94)80782-9
121. Schneider G, Schuchhardt J, Wrede P (1994) Artificial neural networks and simulated molecular evolution are potential tools for sequence-oriented protein design. Bioinformatics 10:635–645. https://doi.org/10.1093/bioinformatics/10.6.635
122. Bhhatarai B, Garg R (2008) Comparative QSAR as a cheminformatics tool in the design of dihydro-pyranone based HIV-1 protease inhibitors. Curr Comput Aided Drug Des 4:283–310. https://doi.org/10.2174/157340908786786029
123. Talele T, Khedkar S, Rigby A (2010) Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. Curr Top Med Chem 10:127–141. https://doi.org/10.2174/156802610790232251
124. Mallipeddi P, Kumar G, White S, Webb T (2014) Recent advances in computer-aided drug design as applied to anti-influenza drug discovery. Curr Top Med Chem 14:1875–1889. https://doi.org/10.2174/1568026614666140929153812
125. Hartman GD, Egbertson MS, Halczenko W, Laswell WL, Duggan ME, Smith RL et al (1992) Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. J Med Chem 35:4640–4642. https://doi.org/10.1021/jm00102a020
126. Ekins S, Mestres J, Testa B (2007) In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. Br J Pharmacol 152:9–20. https://doi.org/10.1038/sj.bjp.0707305
127. Meunier M, Robertson S (2021) Materials Studio 20th anniversary. Mol Simul. https://doi.org/10.1080/08927022.2021.1892093
128. Constans A (2002) Building an informatics bridge. Science 16:54–55
129. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT et al (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47:1739–1749. https://doi.org/10.1021/jm0306430
130. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT et al (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem 47:1750–1759. https://doi.org/10.1021/jm030644s

131. Kumar S (2021) Online resource and tools for the development of drugs against novel coronavirus. In: Methods Pharmacol. Toxicol. Springer, New York. https://doi.org/10.1007/7653_2020_53
132. Liao C, Sitzmann M, Pugliese A, Nicklaus MC (2011) Software and resources for computational medicinal chemistry. Future Med Chem 3:1057–1085. https://doi.org/10.4155/fmc.11.63
133. Geldenhuys WJ, Gaasch KE, Watson M, Allen DD, Van Der Schyf CJ (2006) Optimizing the use of open-source software applications in drug discovery. Drug Discov Today 11:127–132. https://doi.org/10.1016/S1359-6446(05)03692-5
134. Sawicki MP, Samara G, Hurwitz M, Passaro E (1993) Human Genome Project. Am J Surg 165:258–264. https://doi.org/10.1016/S0002-9610(05)80522-7
135. Cook D, Brown D, Alexander R, March R, Morgan P, Satterthwaite G et al (2014) Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. Nat Rev Drug Discov 13:419–431. https://doi.org/10.1038/nrd4309
136. Reiss T (2001) Drug discovery of the future: the implications of the human genome project. Trends Biotechnol 19:496–499. https://doi.org/10.1016/S0167-7799(01)01811-X
137. Brown N, Ertl P, Lewis R, Luksch T, Reker D, Schneider N (2020) Artificial intelligence in chemistry and drug design. J Comput Aided Mol Des 34:709–715. https://doi.org/10.1007/s10822-020-00317-x
138. Vilar S, Cozza G, Moro S (2008) Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. Curr Top Med Chem 8:1555–1572. https://doi.org/10.2174/156802608786786624
139. Vidyasagar M, Mande S, Rajgopal S, Gopalkrishnan B, Srinivas STPT, Uma Maheswara Rao C et al (2007) BioSuite: a comprehensive bioinformatics software package (a unique industry-academia collaboration). Curr Sci 92:29–38
140. Schrödinger, Inc. (2021) Glide. https://www.schrodinger.com/products/glide. Accessed 30 Apr 2021
141. Irwin JJ, Shoichet BK (2005) ZINC—a free database of commercially available compounds for virtual screening. J Chem Inf Model 45:177–182
142. Jain AN (2003) Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem 46:499–511. https://doi.org/10.1021/jm020406h
143. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D et al (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 36:D901–D906. https://doi.org/10.1093/nar/gkm958
144. Stroganov OV, Novikov FN, Stroylov VS, Kulkov V, Chilov GG (2008) Lead finder: an approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening. J Chem Inf Model 48:2371–2385. https://doi.org/10.1021/ci800166p
145. Irwin JJ, Shoichet BK, Mysinger MM, Huang N, Colizzi F, Wassam P et al (2009) Automated docking screens: a feasibility study. J Med Chem 52:5712–5720. https://doi.org/10.1021/jm9006966
146. Back NKT, Van Wijk A, Remmerswaal D, Van Monfort M, Nijhuis M, Schuurman R et al (2000) In-vitro tipranavir susceptibility of HIV-1 isolates with reduced susceptibility to other protease inhibitors. AIDS 14:101–102. https://doi.org/10.1097/00002030-200001070-00019
147. Wolfe C, Wolfe C (2009) Profile of darunavir in the management of treatment-experienced HIV patients. HIV/AIDS—Res Palliat Care 1:13–21. https://doi.org/10.2147/HIV.S4842
148. Wal P, Wal A, Rai A, Dixit A (2011) Aliskiren: an orally active renin inhibitor. J Pharm Bioallied Sci 3:189–193. https://doi.org/10.4103/0975-7406.80764
149. Gawehn E, Hiss JA, Brown JB, Schneider G (2018) Advancing drug discovery via GPU-based deep learning. Expert Opin Drug Discov 13:579–582. https://doi.org/10.1080/17460441.2018.1465407
150. Ying M (2010) Quantum computation, quantum theory and AI. Artif Intell 174:162–176. https://doi.org/10.1016/j.artint.2009.11.009

151. Mayr A, Klambauer G, Unterthiner T, Hochreiter S (2016) DeepTox: toxicity prediction using deep learning. Front Environ Sci 3:80–95. https://doi.org/10.3389/fenvs.2015.00080
152. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T (2018) The rise of deep learning in drug discovery. Drug Discov Today 23:1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039
153. Glicksberg BS, Li L, Chen R, Dudley J, Chen B (2019) Leveraging big data to transform drug discovery. In: Methods Mol. Biol., vol 1939. Humana Press Inc., pp 91–118. https://doi.org/10.1007/978-1-4939-9089-4_6
154. Wójcikowski M, Zielenkiewicz P, Siedlecki P (2015) Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field. J Cheminform 7:1–6. https://doi.org/10.1186/s13321-015-0078-2
155. Sydow D, Morger A, Driller M, Volkamer A (2019) TeachopenCadd: a teaching platform for computer-aided drug design using open source packages and data. J Cheminform 11:29. https://doi.org/10.1186/s13321-019-0351-x
156. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos Pedregosa, Varoquaux, Gramfort et al. Matthieu Perrot. J Mach Learn Res 12:2825–2830
157. Sliwoski G, Kothiwale S, Meiler J, Lowe EW (2014) Computational methods in drug discovery. Pharmacol Rev 66:334–395. https://doi.org/10.1124/pr.112.007336
158. Ghani SS (2020) A comprehensive review of database resources in chemistry. Eclet Quim 45:57–68. https://doi.org/10.26850/1678-4618eqj.v45.3.2020.p57-68
159. Protheragen (2021) CADD Platform. https://aimed.protheragen.com/cadd-platform.html. Accessed 5 May 2021
160. Singh N, Chaput L, Villoutreix BO (2021) Virtual screening web servers: designing chemical probes and drug candidates in the cyberspace. Brief Bioinform 22:1790–1818. https://doi.org/10.1093/bib/bbaa034
161. Skariyachan S, Garka S (2018) Exploring the binding potential of carbon nanotubes and fullerene towards major drug targets of multidrug resistant bacterial pathogens and their utility as novel therapeutic agents. In: Fullerenes, graphenes and nanotubes: a pharmaceutical approach. Elsevier, pp 1–29. https://doi.org/10.1016/B978-0-12-813691-1.00001-4
162. Sakaue S, Okada Y (2017) Human genetics contributes to the understanding of disease pathophysiology and drug discovery. J Orthop Sci 22:977–981. https://doi.org/10.1016/j.jos.2017.07.009
163. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR et al (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov 9:203–214. https://doi.org/10.1038/nrd3078
164. Cao Y, Romero J, Aspuru-Guzik A (2018) Potential of quantum computing for drug discovery. IBM J Res Dev 62:6:1–6:20. https://doi.org/10.1147/JRD.2018.2888987
165. Chan HCS, Shan H, Dahoun T, Vogel H, Yuan S (2019) Advancing drug discovery via artificial intelligence. Trends Pharmacol Sci 40:592–604. https://doi.org/10.1016/j.tips.2019.06.004
166. Kramer JA, Sagartz JE, Morris DL (2007) The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. Nat Rev Drug Discov 6:636–649. https://doi.org/10.1038/nrd2378
167. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K et al (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506:376–381. https://doi.org/10.1038/nature12873
168. Nebula Genomics (2021) Whole genome sequencing DNA test. https://nebula.org/whole-genome-sequencing-dna-test/. Accessed 3 May 2021
169. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274:1531–1534. https://doi.org/10.1126/science.274.5292.1531

170. Sheng C, Zhang W (2013) Fragment informatics and computational fragment-based drug design: an overview and update. Med Res Rev 33:554–598. https://doi.org/10.1002/med.21255
171. Bollag G, Tsai J, Zhang J, Zhang C, Ibrahim P, Nolop K et al (2012) Vemurafenib: the first drug approved for BRAF-mutant cancer. Nat Rev Drug Discov 11:873–886. https://doi.org/10.1038/nrd3847
172. Deeks ED (2016) Venetoclax: first global approval. Drugs 76:979–987. https://doi.org/10.1007/s40265-016-0596-x
173. Park CM, Bruncko M, Adickes J, Bauch J, Ding H, Kunzer A et al (2008) Discovery of an orally bioavailable small molecule inhibitor of prosurvival B-cell lymphoma 2 proteins. J Med Chem 51:6902–6915. https://doi.org/10.1021/jm800669s
174. US Food and Drug Administration (2016) FDA Approves Crizotinib Capsules. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-crizotinib-capsules. Accessed 3 May 2021
175. Burki T (2020) A new paradigm for drug development. Lancet Digit Heal 2:e226–e227. https://doi.org/10.1016/s2589-7500(20)30088-1
176. Bajracharya R, Song JG, Back SY, Han HK (2019) Recent advancements in non-invasive formulations for protein drug delivery. Comput Struct Biotechnol J 17:1290–1308. https://doi.org/10.1016/j.csbj.2019.09.004
177. Bender A, Cortés-Ciriano I (2021) Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: ways to make an impact, and why we are not there yet. Drug Discov Today 26:511–524. https://doi.org/10.1016/j.drudis.2020.12.009
178. Ouyang D, Smith SC, Douroumis D, Fahr A, Siepmann J, Snowden MJ et al (2015) Computational pharmaceutics: application of molecular modeling in drug delivery. Wiley
179. Ouyang D, Smith SC (2015) Introduction to computational pharmaceutics. Comput Pharm Appl Mol Model Drug Deliv 1–5. https://doi.org/10.1002/9781118573983.ch1
180. ACS Publications (2021) About the journal. https://pubs.acs.org/page/mpohbp/about.html. Accessed 10 May 2021
181. Hossain S, Kabedev A, Parrow A, Bergström CAS, Larsson P (2019) Molecular simulation as a computational pharmaceutics tool to predict drug solubility, solubilization processes and partitioning. Eur J Pharm Biopharm 137:46–55. https://doi.org/10.1016/j.ejpb.2019.02.007
182. Clinical Trials Arena (2021) Spritam (levetiracetam) for the treatment of epilepsy. https://www.clinicaltrialsarena.com/projects/spritam-levetiracetam-epilepsy/. Accessed 30 April 2021
183. FabRx Ltd (2021) The future of medicine is 3D. https://www.fabrx.co.uk/. Accessed 30 Apr 2021
184. Associated Press (2017) Digital pill to track medication adherence is coming. Mod Healthc. https://www.modernhealthcare.com/article/20171114/NEWS/171119948/digital-pill-to-track-medication-adherence-is-coming. Accessed 30 Apr 2021
185. Bright Z (2019) Lights-out pharma factory: why the future of pharma production is robotic. Multiply Labs. https://medium.com/multiply-labs/lights-out-pharma-factory-why-the-future-of-pharma-production-is-robotic-ee1c5297bb8a. Accessed 3 May 2021
186. Universal Robotics (2017) Press kit. https://www.universal-robots.com/media/downloads/. Accessed 12 Apr 2021
187. Boyd J (2002) Tech.Sight. Robotic laboratory automation. Science 295:517–518. https://doi.org/10.1126/science.295.5554.517
188. Drucker DJ (2020) Advances in oral peptide therapeutics. Nat Rev Drug Discov 19:277–289. https://doi.org/10.1038/s41573-019-0053-0
189. Budapest Open Access Initiative (2021). https://www.budapestopenaccessinitiative.org/. Accessed 10 May 2021
190. Bayda S, Adeel M, Tuccinardi T, Cordani M, Rizzolio F (2020) The history of nanoscience and nanotechnology: from chemical-physical applications to nanomedicine. Molecules 25:112. https://doi.org/10.3390/molecules25010112

Historical Developments on Computer Applications in Pharmaceutics

2

Nidhi Nainwal, Richa Bahuguna, Surojit Banerjee, and Vikas Anand Saharan

Contents

N. Nainwal · R. Bahuguna · S. Banerjee · V. A. Saharan (✉)
Department of Pharmaceutics, School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Dehradun, Uttarakhand, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_2

Abstract

A lot of mathematical and statistical calculations are involved in optimization of pharmaceutical formulations. Computations involving correlating variables with responses, regression analysis, predictions, and simulations are now better handled with computers equipped with specially designed softwares. These softwares have been developed from simple programs to artificial intelligence (AI) integrated packages. In addition to these, computers are playing a larger role in designing internal architecture and outer appearance of dosage forms with the advancements in computer-aided designing and 3D printing technologies. Initial robots were made to automate pick and place operations but significant developments of programming softwares and AI have gradually advanced robots with more flexibility, adaptability, and intelligence. Soft robots with greater degrees of freedom and locomotion abilities have a potential role in delivering drugs to targeted locations. Software packages for pharmacokinetics perform tedious calculations, data analysis, and modeling to enhance the speed of drug discovery and formulation development programs. This chapter focuses on some significant historical developments on different applications of computers in the field of formulation optimization, robotics, artificial intelligence, 3D printing, and pharmacokinetics.

Keywords

History · Computer · Optimization · Pharmaceutical development · Pharmacokinetics · Artificial intelligence · Robot · 3D printing

Chapter Objectives
Upon reading of this chapter, the reader will be able to understand:

- Historical developments in optimization of pharmaceutical formulations
- Dated innovations in the developmental history of 3D printing
- Developments in AI with time for formulation optimization
- A brief historical account on pharmacokinetic softwares
- History of industrial robots and soft robots

2.1 Introduction

Computers are an integral part of pharmaceutical research and development. Computer's ability to store a large amount of data, computations at high speed, and data analysis at unprecedented speeds have significantly reduced the time for

drug discovery and development. A few decades back, we cannot imagine the idea of a computer on the desk of every researcher and company manager. But at present computers have become a necessity for generation, management, and transmission of information in almost every work environment. The development of computers in the pharmaceutical field had begun around the 1940s, mainly for payroll and accounting of employees. Pharmaceutical scientists did eventually gain access to computers either by company or from educational institutions via contractual agreements [1]. In pharmacy, computers are used to get the information about drug data, records, files, and drug management. Receiving the details, storing it, and processing it and its dissemination are the main roles of computers and this continuous flow of information shows effective functioning of any system [2].

Application of in silico methods for optimization started appearing in the 1990s. Since then, formulation scientists are relying on computers and softwares for performing computations involved in designing pharmaceutical formulations. For example, to prepare a good-quality emulsion, one needs to know the desired density, droplet size, viscosity of emulsion, conductivity, stability issues, etc., which are referred to as dependent variables. For achieving a best formulation, formulation scientists need to note different independent variables like stirring speed, hydrophilic-lipophilic balance (HLB), value of emulsifier, and mixing time along with their respective levels. In silico optimization requires details of such independent variables and a suitable optimization design to be fed in the software (like SAS, SPSS, Design-Expert, etc.) for generation of minimum experimental runs. As per these experiments, formulations are prepared and evaluated for check on quality, which is assessed in terms of dependent variables or responses. The results of quality check of dependent variables are fed to the software for obtaining regression equations that correlate dependent and independent variables. Further, optimization softwares also help in subsequent analysis of the optimization process through response surface graphs and contour plots. The outcome of the optimization process leads to better formulations, which match with the set quality benchmarks (desired dependent variables). Further, regression equations allow prediction of dependent variables for any new formulation eliminating the need for its preparation and assessment of the formulation. Different research reports have described the use of computers in preparing emulsions and other formulations [3, 4]. Formulation scientists have also adopted artificial neural networks (ANN) and algorithms for optimizing pharmaceutical formulations like emulsions, multiple emulsions, etc. The roles of AI like ANN and algorithms are increasing the prediction power of optimization softwares/packages. In addition to this, AI-enhanced robots are now being employed as assistants in research laboratories reducing the human workforce requirements for boring and repetitive tasks. Computer-aided design of dosage forms coupled with additive manufacturing are enhancing the potential for targeted and personalized drug delivery. This chapter focuses on some historical developments of computer applications in pharmaceutics. The chapter discusses historical developments in the field of formulation development and optimization, artificial intelligence (AI), robotics, and pharmacokinetics.

2.2 Computer-Aided Formulation Development

The regulatory approval and patient acceptance of pharmaceutical products depend on the quality of the formulation. Drugs can be administered to humans via different routes and in different forms like tablets, capsule, oral liquids, parenteral, topical formulation, eye preparation, suppositories, pessaries, and inhalation preparations. All these dosage forms follow more or less similar processing steps starting from the selection of quality raw materials, preparation of a suitable dosage form, and subsequent quality checks. Selection of right additives or excipients is a critical step for obtaining required efficacy, stability, and safety. For instance, the excipients may be chemically or physically incompatible with the drug or they may exhibit batch-wise variability to such an extent that at the extremes of their specification they may cause failure in achieving the desired therapeutic effect. In addition, some excipients, especially those that are hygroscopic, may be contraindicated if the formulation is to be manufactured for tropical countries. Hence, formulators must work in a design space that is multidimensional in nature and sometimes difficult to realize and conceptualize. Continuous and effective hard work of scientists across various disciplines has led to advanced computing techniques like expert-/knowledge-based systems, neural computing, and computer simulations for developing on time, cost-effective pharmaceutical formulations [5].

2.2.1 Design of Experiment (DOE)

Design of experiments is a mathematical approach that is used to plan, perform the experiments, and to analyze and interpret the data obtained through experiments. It is a branch of applied statistics that is used for conducting scientific studies of a system, process, or product, in which input variables (Xs) are manipulated to investigate its effects on response variable (Y) [6]. The concept of DOE dates back to early 300 BC, when Aristotle, the Greek philosopher, first wrote about it. He started the scientific method basically by putting a hypothesis to the method, testing it, and then validating the hypothesis or invalidating it, or so forth. Over the past several decades, the use of DOE for improving the quality of products or processes is widely employed in pharmaceutical industries [7, 8]. DOE has also been accepted well in other areas like administration, marketing, hospitals, food industry, energy, architecture, and analytical assessments like chromatography [9–12]. DOE is applicable to physical processes as well as computer simulation models [13]. The year-wise development of design of experiment is summarized in Fig. 2.1.

2.2.2 The Agricultural Origins (1918–1940s): The Emergence of Factorial Designs and ANOVA

Sir Ronald A. Fisher, a biostatistician from the UK, had laid the foundation of statistics and DOE for the development of experimentation methods for agriculture

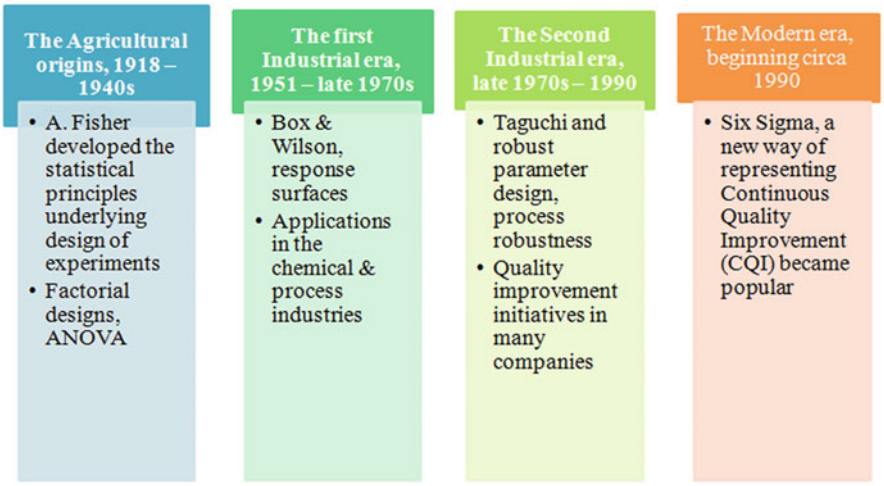


Fig. 2.1 History of DOE

[14–17]. Sir Fisher developed the statistical principles of DOE during his work at Rothamsted Experimental Station in the 1920s and 1930s. Fisher wanted to increase the crop yield during world war to feed soldiers and citizens. He needed a better way to plan experiments and get data. That is why he integrated the fundamental DOE concepts of randomization, replication, and blocking into a single unified approach. Fisher realized from his experience of 70 years of planting, harvesting, and analyzing data that the amount of information extracted from data can never exceed what was inherently contained in the data. Later, he realized and shifted his focus on how the data were generated [18]. From this point on, Fisher focused on understanding what types of experimental structures yielded the most fruitful outcomes for a given expenditure in time, money, and labor.

Fisher along with his colleague Frank Yates developed many basic concepts like orthogonal designs and Latin squares during the 1920s through the 1940s [19]. Fisher restated that the analysis of variance is not a mathematical theorem, but rather a convenient method of arranging the arithmetic [20]. The formula for ANOVA ($F = MST/MSE$) uses Fisher ratio (F-ratio) as ANOVA coefficient. DOE had not signified a strong expansion as there were no software packages that would foster its application. It took about another 40–50 years, the late 1960s and 1970s, for the DOE to achieve significant applications in the research. The use of DOE in research over various scientific areas further raised sharply in the 1990s and later due to education and software development [21].

2.2.3 The First Industrial Era (1951–1970s): Response Surface Methodology

Immediately following World War II, the first industrial era marked resurgence in the use of DOE. At this time, Box and Wilson (1951) wrote about response surface designs thinking of the output as a response function and trying to find the optimum conditions for this function [22]. They finally well established the idea of response surface methodology (RSM) that began in the early 1930s. RSM is defined as a collection of statistical design and numerical optimization techniques for empirical model building and model exploitation used to optimize processes and product design [23, 24]. The use of experimental design methods in the chemical industry was promoted in the 1950s by the extensive work of Box and his collaborators on response surface designs [25]. Probably, the first application of DOE in formulation development was credited to Marlow and Shangraw [26]. Marlow and Shangraw used factorial designing and applied ANOVA to assess the effect of formulation variables on dissolution of sodium salicylate tablets prepared by wet granulation.

2.2.4 The Second Industrial Era (Late 1970s to 1990): Quality Revolution

The quality revolution started in Japan in the 1950s with works of W. Edward Deming in statistical quality control. Total quality management (TQM) and continuous quality improvement (CQI) were management techniques that came out as a result of statistical quality revolution. Genechi Taguchi, a Japanese engineer, discovered and published a lot of statistical techniques using an independent development of what he referred to as orthogonal arrays [27]. Taguchi efforts were later recognized as fractional factorial designs. Taguchi is also known for *robust parameter design,* which is an experimental design used for finding interaction of uncontrollable and controllable variables and thereby minimizing response variation from uncontrollable variables.

2.2.5 The Modern Era (1990s to 2000): Continuous Quality Improvement (CQI)

In 1990, a new way of representing continuous quality improvement (CQI) became popular as a Six Sigma concept. This is a technique that uses statistics to make decisions based on quality and feedback loops. It includes a lot of previous statistical and management techniques [28].

2.2.6 The Present Era (2000–2020): The Era of QbD

About 800 studies were found with words "Design of Experiment" and "pharmaceutical" in 2016 [29], in Scopus database. The number of studies was 281 in 2010, 24 in 2005, and 29 in 2000. The trend of applying DOE has considerably gone higher with ICHQ8, ICHQ9, and ICHQ10 (QbD tripartite guidelines) guidelines and their adoption in United States, Europe, Japan, and other countries since 2004 [30]. As of now, DOE as an element of QbD supports pharmaceutical industries in designing drug products that are well accepted by drug regulatory agencies. DOE applications in formulation development have been made considerably easier with the availability of so many softwares (Table 2.1).

2.3 Artificial Intelligence

Artificial intelligence (AI) simulates human intelligence in machines. AI has made machines capable of thinking and to act like humans. AI is assisting humans in not only various daily tasks but also increasing analytical and computational abilities in research and development. It is like upgradation of human intelligence by involving computer systems performing tasks. AI is extremely useful in managing big data and analyzing results, which in turn promotes better decision-making and helps in saving human effort and time. Exploration of AI seems never ending as it grows enormously with computing speeds, cloud technologies, and networking. Many pharmaceutical companies are exploring AI through various collaborations with technological companies for their applications in drug discovery and development.

Artificial intelligence techniques are getting mainstream by developing models that can relate to alteration or modification in components and processing conditions in order to get improved product formulations with desirable properties [5].The types of computer systems and their application in pharmaceutical sciences are given in the Table 2.2.

2.3.1 Expert- and Knowledge-Based Systems

An expert system is a knowledge-based system that is designed to solve complex problems of various fields that are normally solved by humans [33]. It is not surprising, considering the widespread use of tablets and capsules that these domains have received most attention for the development of expert systems by both companies and academic institutions. However, it should be noted that other domains such as inhalation preparations, topicals, and parenterals have also been investigated. The galenical development system, developed at the University of Heidelberg, Germany, has been found effective in the development of a variety of dosage forms starting from the chemical and physical properties of a drug. The project was initiated in 1990 and has been extensively revised and enhanced in the interim [34].

Table 2.1 Softwares for design of experiments (DOE)

Software, latest version, and its release date	Brief description	URL
Design-Expert Version 13, January 2021	First launch in 1988; provides 3D plots that can be rotated to visualize response surfaces; numerical and graphical optimization; and access by subscription	http://www.statease.com/
JMP Version 16; March 2021	JMP (John's Macintosh Project) was developed in 1989; Features: control charts, elementary design of experiments (DOE), survival features, graph builder, dynamic bubble plots, data mining, predictive analytics, and automated model building; orthogonal supersaturated design; and access by subscription	https://www.jmp.com/
Minitab 20.2.0; April 2021	First launch in 1972; powerful DOE software used for automated data analysis, graphic, and help features, including MS-Excel compatibility and almost all designs of RSM; and access by subscription	https://www.minitab.com/
XLSTAT version 2021.2; April 2021	First launch in 1993; flexible Excel data analysis add-on software; helps in selection of an experimental design; DOE for screening, response surface, and mixture designs and their analysis; and access by subscription	https://www.xlstat.com/en/
Statistica 14.0; December 2020	First launch in Mid-1980s; Statistica provides data analysis, data management, statistics, data mining, machine learning, text analytics, and data visualization procedures; access by subscription	https://docs.tibco.com/products/tibco-statistica-14-0-0
SPSS Statistics 27; November 2020	Comprehensive statistical software with features implementing experimental design	https://www.ibm.com/in-en/products/spss-statistics
MODDE® 12; February 2017	Use for evaluation of fitting of model and suitable for response surface modeling; access by subscription	https://www.sartorius.com/en/products/process-analytical-technology/data-analytics-software/support/knowledge-base/modde-12-550758
Prism 9.1.0 (Graph Pad Software); March 2021	First launch in 1989; used for data analysis, statistics, and graphing	https://www.graphpad.com/

(continued)

Table 2.1 (continued)

Software, latest version, and its release date	Brief description	URL
Statgraphics Centurion 19; 2020	First launch in 1980; Statgraphics provide extensive catalog of screening, response surface, optimal designs, mixture, and RPD experiments; access by subscription	http://statgraphics.com/

Table 2.2 Types of computer systems and their application in pharmaceutical sciences

Advanced computing techniques in pharmaceutical formulations	Application	Ref.
Expert- and knowledge-based systems	Generation of initial formulations and processing conditions ab initio	[31]
Neural computing	For modeling formulation and process data to explore relationships within the dataset and optimize the formulation	[5]
Computer simulation	For the development of mathematical models for the interactions between the ingredients of formulation and the manufacturing process to predict outcomes	[32]

2.3.1.1 Tablet Formulations

In tablet formulation, expert systems have been developed by many pharmaceutical companies and organizations such as ICI/Zeneca/AstraZeneca (UK), Cadila Laboratories Ltd. (India), and a pool of pharmaceutical companies from Japan.

This system, initiated in 1988, was implemented with the Formulogic shell and knowledge acquisition through interview, and structured with frames, objects, and production rules [35]. All the physicochemical and mechanical properties of the drug along with the dose required were entered by the user. The system proposes a target tablet weight, selects the excipients, and calculates their concentrations as per the predetermined constraints based on the manufacturability of the formulation. If the recommendations are not satisfactory, the formulator has a choice to override the system. The system also has a formulation optimization procedure that includes tablet hardness, weight variation, disintegration time, tablet defects, etc.

In 1992, the Cadila system was designed to formulate tablets for drugs based on their physical, chemical, and biologically interrelated (dissolution rate) properties [36]. The system was capable of identifying the desirable properties of excipients for optimum compatibility with the drug, selecting those excipients that have the required properties, and then suggesting tablet formulations that contain at least a binder, a disintegrant, and a lubricant. Other excipients such as fillers or glidants were then added as per formulation requirements. The data to be fed in expert- and knowledge-based system for tablet formulations are shown in Fig. 2.2.

The system was menu driven, interactive with the user, and written in PROLOG (logic programming). The prototype system when first implemented had 150 rules,

Fig. 2.2 The data to be fed in expert- and knowledge-based system for tablet formulations

Fig. 2.3 Steps involved in statistical designing of experiment on capsule formulation

but to increase the reliability of the system, it expanded to over 300 rules. It was reported to reduce 35% of the development time for a new tablet formulation and had proven to be beneficial in planning the purchase and stocking of excipients.

Knowledge was acquired by questionnaires and discussions with experts. After this, on the basis of majority decision, the system was developed. The system operates on relational database. The decision trees are used for selection of excipients, which depends on the flow properties, compression characteristics, disintegration, and solubility of the drug. In 2001, Pfizer had reported a prototype system implemented with the Formulogic shell [37]. The system was designed in such a way that it can use preformulation data of new drugs for recommending early formulations, predicting properties of products, and selecting processing conditions that can be most suitable for scale-up.

2.3.1.2 Capsule Formulations

The system was developed at Sanofi that uses the Formulogic shell for recommending first-pass clinical capsule formulations to accommodate an experimental design [34]. In addition to the formulation, the system was capable of providing suggestions on the milling of the drug, blending procedure to be used, details of the capsule shell that can be used, as well as the reason for selecting any or all the components.

Expert systems for capsule formulation were developed originally as a part of Ph. D. program at the University of London, but later the work progressed as a team with Capsugel and Sanofi Research Division in Philadelphia [38]. The knowledge base of the system was broad and it contained information on a large number of excipients, and frequently updated database of marketed formulations from Germany, Italy, Belgium, France, and United States. An information base of literature references related to capsule formulation was updated through monitoring of current literature. Furthermore, it contained the experience and nonproprietary knowledge of various global experts and also the outcomes from statistically designed experiments on capsule formulations. The framework utilizes decision trees and production rules for information portrayal. The steps involved in statistical designing of experiment on capsule formulations are shown in Fig. 2.3.

Before recommending a formulation with any necessary processing conditions, the system uses an input questionnaire to collect the data on a new drug and subsequently utilizing different techniques to anticipate properties of the drug with different excipients. The system also provides a statistical design to improve the formulation. It has a semiautomatic learning tool that monitors user habits and collects information on the use of excipients. This, along with the outcomes from user questionnaires, gives the background to further upgrade the formulation [39].

2.3.1.3 Other Formulations

In addition to the above described formulations, the expert systems have also been reported for parenteral, film coatings, and topical formulations [36, 40].

2.3.2 Neural Computing

Neuron is the basic unit of both the mammalian nervous system and neural computing. To mimic mammalian intelligence and learning, artificial neural networks (ANNs) were first investigated in the 1940s [41]. From a moderate couple of uses in the early-to-mid 1990s, the utilization of neural computing is presently acquiring acknowledgment worldwide in various industry areas. The new age of formulators can hope to utilize this strategy regularly, making it ideal for educators to know about this emerging new field. In a survey of the utilization of 93 neural computing applications in 75 UK organizations covering practically all business areas, the significant advantages recognized were improved quality, improved response times, and increased efficiency. The best advantages are accomplished for multidimensional issues, where it is difficult to communicate any analytical model and hard to abridge the standards by some other system than neural computing. It helps if the issue is of practical significance, is important for the association's essential activity, and meets a genuine business need. Pharmaceutical formulation meets these standards well, and neural computing has given significant advantages in the pharmaceutical industry.

The performance of a product not only depends on the ratio of the ingredients used but also on the processing conditions. The relationship between the product performance, ingredient level, and the processing conditions cannot be quantified and are not evidence based. The traditional methods of designing the formulation by statistical techniques can be misleading in the case of complex formulations. Recent advances in mathematics and computer science have resulted in the development of three different techniques of neural computing–neural networks, genetic algorithms (Holland in the 1970s), and fuzzy logic as shown in Fig. 2.4 for the domain of product formulation [42, 43].

Preceding 1995, there were not many announced application of neural computing in product formulation. From that point forward, the number has increased quickly with applications being accounted for in adhesives, dyes, paints, pharmaceuticals, and a lot more fields. In the course of the last many years, neural networks have acquired acknowledgments in modeling pharmaceutical formulations than in some

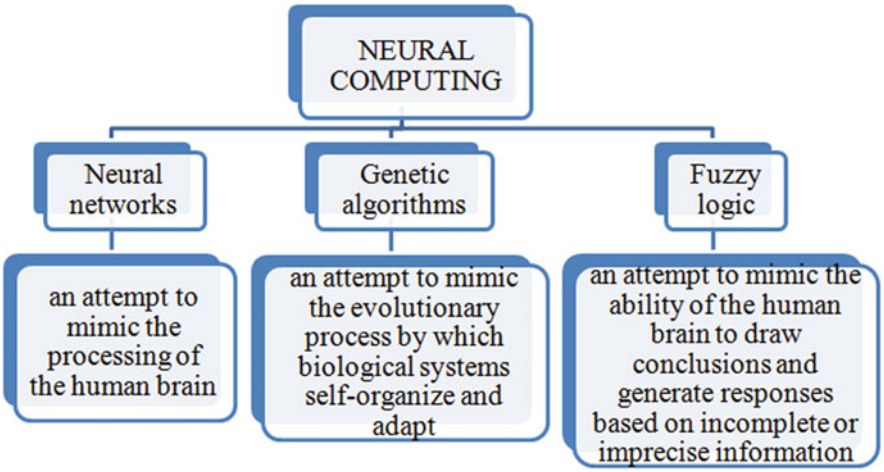


Fig. 2.4 Types of neural computing

other field. Applications currently exist for immediate and controlled released tablets, skin creams, hydrogel ointments, emulsions, and film coatings. Over the past few decades, neural networks have gained more acceptances in modeling pharmaceutical formulations than in any other field. Applications now exist for immediate and controlled release tablets, skin creams, hydrogel ointments, emulsions, and film coatings.

2.3.2.1 Tablet Formulations

Immediate-release tablet formulations of hydrochlorothiazide were modeled, trying to increase tablet strength and select the best lubricant, though something similar for caffeine was modeled to relate both formulation and processing variables with granule and tablet properties [44]. Both of these studies were fruitful in exhibiting that neural networks performed better compared to conventional statistical techniques.

The merits and demerits of neural networks for tablets have been featured by work force from Trinity College (Dublin, Ireland), Novartis and the University of Basel in Switzerland [45, 46]. Different studies showed that neural networks have been discovered valuable in modeling tablet formulations of antacids, plant extracts, theophylline, and diltiazem [47–49].

Researchers have modeled and optimized pigmented film coating formulations from three distinctive immediate-release tablet formulations to improve opacity and diminish film cracking with neural networks combined with genetic algorithms just as being studied with neuro-fuzzy logic [50].

In controlled-release domain, initial studies were carried out in the early 1990s by Hussain and colleagues at the University of Cincinnati [51]. They designed the in vitro release attributes of various drugs from matrices containing various hydrophilic polymers and found that in vast majority of the cases, neural networks having

single hidden layer have a good performance in anticipating drug release profiles. Afterward, researches with similar formulations have avowed these discoveries [52, 53].

2.3.2.2 Topical Formulations

Topical formulations are usually multicomponent by their nature, and hence, neural networks have been applied to manage this intricacy. In 1997, the first work of neural networks was performed on hydrogel formulations containing anti-inflammatory drugs (Japan), followed up by additional investigations in 1999 and in 2001, which concludes the prevalence of neural networks over conventional statistics [53, 54].

2.3.2.3 Other Formulations

Neural networks were applied to the modeling of pellet formulations to control the release of theophylline and rate of degradation of omeprazole [55, 56]. They had likewise been applied to the preparation of acrylic microspheres to model insulin release from an implant and hydrocortisone release from a biodegradable matrix [57–59].

2.3.3 Computer Simulations

Simulation is the process of translating a real system into a working model using mathematics to express cause-and-effect relationships that determine the behavior of the system. In 1987, the computer simulation of the tablet compaction process with finite elements was first attempted and further refined [60]. This methodology is based on the assumption that constitutive equations can be used to define the properties of a tablet. The assumption worked well for tablet formulations consisting of a single ingredient but had little relevance to multicomponent formulations. A combined finite-discrete element method for simulating multicomponent pharmaceutical powder tableting was proposed [61].

A prerequisite of tablet compaction is the filling of the tablet die with powder. Powder packing is one process that has received a great deal of attention (Macro Pac, Intelligensys Ltd., UK). The commercial software are able to simulate the packing of multicomponent formulations of particles of any shape and size with a Monte Carlo technique and is ideal for the simulation of the packing of pharmaceutical formulations into both tablet dies and hard gelatin capsule shells. These software were capable to simulate the packing of multicomponent formulations of particles of any shape and size with a Monte Carlo technique and were also good for simulating the packing of pharmaceutical formulations into tablet dies and hard gelatin capsule shells [61].

To improve the color and/or opacity of tablet film coatings, solid inclusions (in the form of pigments) are often used. But localized cracking around the individual particles or aggregates can be seen, which can affect the release control of the active drug. A simulation of crack propagation in such systems was developed,

allowing the investigation of such effects of the addition of a second population of pigments, pigment particle size and size distribution, polymer molecular weight, addition of plasticizers, and many other factors affecting the film coating formulation [62]. The developed simulation was made available as MacroCrack from Intelligensys Ltd., UK.

2.4 Three-Dimensional (3D) Printing

Three-dimensional printing (3DP) is perhaps the most impressive and progressive revelations for humankind, particularly in the field of drug innovation. This innovation is utilized for creating novel dosage forms, tissues, and organs engineering as well as disease modeling [63]. As per International Standard Organization (ISO), the term three-dimensional printing was defined as: "*manufacturing of articles through the deposition of a material using a print head, nozzle, or another printer technology*" [64]. 3D printing belongs to a group of methods collectively known as additive manufacturing. In additive manufacturing, the parts are prepared from 3D model data in the process of joining materials layer by layer. This pragmatic approach of additive manufacturing is known as rapid prototyping. As compared to subtractive and formative manufacturing methodologies, additive manufacturing has several advantages like reduction in time and costs of prototyping, individualized product structure or series, and easy modifications of a product. Three-dimensional printing can be interpreted as repeated and coordinated two-dimensional printing and it is one of the effective strategies to overcome challenges of a conventional pharmaceutical unit operation [65]. It has given extraordinary adaptability in the design and manufacturing of complex objects, which can be used in customized and programmable medication [66].

The idea of 3D printing dated back to early 70s of the twentieth century. Pierre A. L. Ciraud described it for the first time [67]. In his method, he described 3D printing as application of powdered material and subsequent solidification of each layer via action of a high energy beam. Major methods for 3D printing are based on: powder solidification, liquid solidification, and extrusion. Some significant achievements in 3D printing in pharmaceutical applications are presented in Fig. 2.5.

2.4.1 Emergence of Main Techniques of 3D Printing: The 1980s

Late 1980s was the time when the earliest rapid prototyping (RP) technologies were first attempted. Dr. Kodama was the first to attempt 3D printing by developing a rapid prototyping technique. He was the first to describe a polymerization of a photosensitive resin using UV light. In 1984, 3D printing technology was first used by Charles Hull by inventing stereolithography. In 1986, the concepts of stereolithography (SLA) were patented. He also founded the 3D Systems Corporation [68]. In 1986, Carl Deckard and Joe Beaman, at the University of Texas, introduced selective laser sintering (SLS). This technique offers the 3D printing of

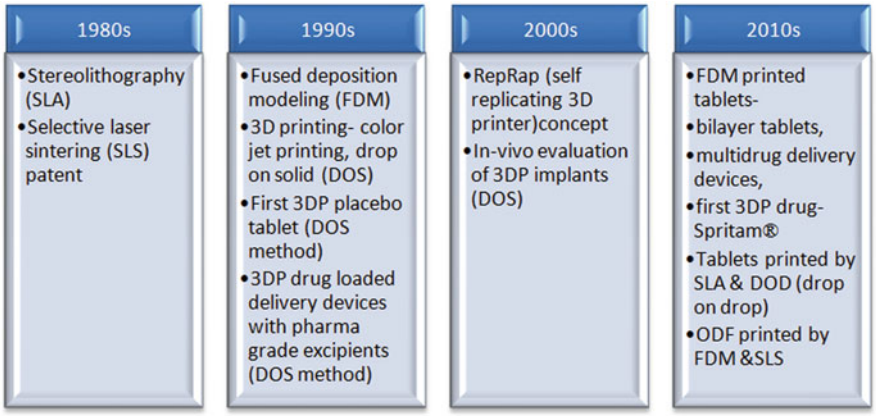


Fig. 2.5 Some milestones achieved in 3D printing in the field of pharmaceutical technology

metal materials as an alternative to the photosensitive resin used in SLA systems. In 1988, first commercial product, that is, SLA-1, was released. In 1988, Carl Deckard brought a patent for the SLS technology, in which laser was used to fuse powder grains together.

In 1989, Scott Crump (a fellow benefactor of Stratasys, Inc.) documented a patent on 3D printing innovation for fused deposition modeling, in which the expelled polymer filaments were heated into a semi-liquid state, expelled through a heated nozzle and saved onto a form stage layer by layer and left to solidify [69].

2.4.2 Evolution of 3D Printer Manufacturer and Modeling Tool: The 1990s

By the 1990s, the basics of 3D printing were established, the main 3D printers manufacturers were emerging, and 3D modeling tools were about to be developed. All these developments were carrying additive manufacturing to the higher level. The first 3D printing technique, that is, inkjet printing, in the field of pharmaceutics was used in the early 1990s at the Massachusetts Institute Technology (MIT), invented and patented by Sachs et al. [67]. A binder solution was placed onto a powder bed, thus, binding the particles together. This process was repeated until the last wanted design was acquired.

MIT licensed its inkjet-based technology to several companies. One such company is Z Corporation, which commercialized its Z Printer [68]. The fused deposition modeling patent was issued to Stratasys in 1992. Stratasys developed 3D printers based on both of these technologies. From 1993 to 1999, several significant milestones emerged in 3D printing sector. A few examples include creation of the Z402 by ZCorp and binder jetting, Arcam MCP technology and Selective Laser Melting, and the creation of Sanders prototype (now known as Solidscape).

The year 1999 was when the first lab-grown organ was implanted in young patients who underwent urinary bladder augmentation using a 3D synthetic scaffold, which was coated with patient's own cells [70].

2.4.3 Revolutionary Developments in 3D Printing: The 2000s

From the year 2000 onwards, revolutionary developments were observed in 3D printing, which changed the way in which manufacturing was performed. During this time, a 3D inkjet printer was developed, which was based on inkjet deposition of a photopolymer and a UV light source. It was named "Quadra" by Object Geometries of Israel [68]. In 2005, the very first high-definition color 3D printer, Spectrum Z510, was launched by ZCorp. Sculpteo developed the first online 3D printing service in 2009.

In 2002, a functioning kidney was developed which was printed directly using a "bioink" [71]. A new concept of "RepRap—self-replicating 3D printer" came into limelight in 2008. It was described as a fused filament fabrication (FFF) 3D printer that was capable of printing most of its own components. Early 2010s marked the arrival of the world's first 3D-printed prototype car, "Urbee," and the first 3D printed aircraft at the University of Southampton [72].

2.4.4 Entry of 3D Printing in Drug Development: The 2010s

In 2014, Charles Hull was granted the European Inventor Award in the Non-European nations' class by the European Patent Office. The first 3D printed prescription drug, approved by the Food and Drug Administration (FDA) in 2016, was oral Spritam (levetiracetam) tablets. Aprecia Pharmaceuticals uses ZipDose Technology (inkjet printing method) to produce Spritam [69]. Various steps involved in preparation of tablets by Zip Dose Technology are shown in Fig. 2.6 [73].

GlaxoSmithKline as of late finished an investigation where inkjet 3D printing and ultraviolet (UV) curing were utilized to make tablets to treat Parkinson's sickness [74]. Researchers recently made a 3D-manufactured ingestible electronic pill capsule that is connected to the user's smartphone and can be controlled using Bluetooth wireless technology. The capsule can reside in the stomach for at least a month. It can be customized to deliver drugs and sense environmental conditions.

In 2016, another idea of 3D bioprinting device called Tissue and Organ Printing System (ITOP) was created at the Wake Forest Institute under the supervision of Dr. Anthony Atala [75]. From 2017, different US FDA-approved 3D printed devices were introduced. By this time, more than 100 medical 3D printed devices are approved by US FDA, which include simple tools for surgery as well as complex implantables like cranial implant device, vertebral interbody cages, etc. [76].

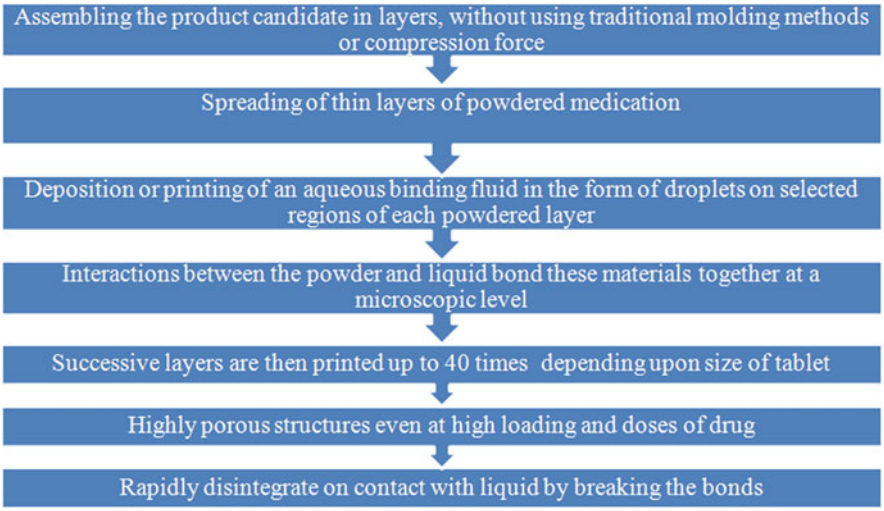


Fig. 2.6 Steps in preparing tablets with ZipDose technology

2.4.5 Advancement of 3D Printing: 2020s

Developers are working in developing high-performance materials with extreme resistance, rigidity, and more sustainability using bio-based materials like Nylon PA11 material series. A portion of these materials are offering thermal resistance, chemical resistance, and even heat resistance for the most requesting applications [68]. 3D printing is more developed in the fields of tissue designing, biomedical, automobile, and aviation, than in the drug business where it is as yet in its initial stage. FDA advances the improvement of manufacturing technologies, including 3D printing, using risk-based approaches [77]. The advancement course of events of 3D printing innovation is represented in the Figs. 2.7 and 2.8 [78].

2.5 Computer-Aided Pharmacokinetics

Pharmacokinetics (PK) generally describes what the body does to the drug, whereas pharmacodynamics (PD) pertains to what the drug does to the body. Over the past several decades, various pharmacokinetic software programs have been designed for the clinicians and researchers to help in the analysis, interpretation, and reporting of pharmacokinetic and pharmacodynamic data. Clinical pharmacokinetic software programs were first developed in late 1967 by Roger Jelliffe [79] and colleagues at the University of Southern California. The initial program, USC*Pack, was created before the introduction of the personal computer (PC) and was developed using a mainframe computer system. As the PC became more commonplace, USC*PACK was later translated to a PC-based application. In the early 1980s, other PC-based programs (viz., SIMKIN and DATAMED) were developed and released

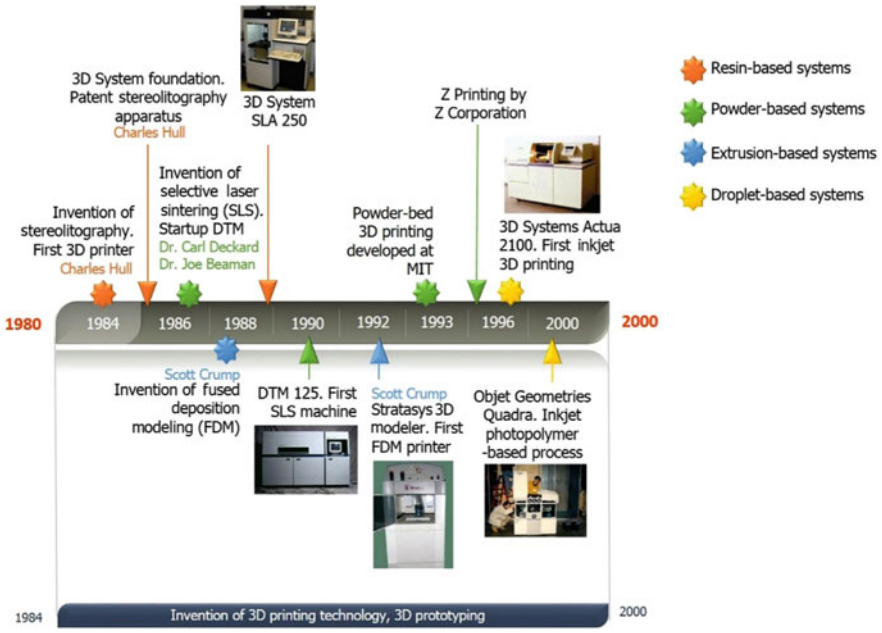


Fig. 2.7 Some significant developments in 3D printing from its birth till 2000. (Reproduced with permission from Ref. [78])

commercially. Schentag and Adleman developed a microcomputer program in 1983 for the analysis of serum tobramycin concentration, which was written for the Apple II Plus computer [80]. Since 1987, totally 13 new clinical pharmacokinetic software programs have been commercially available in the United States. The marketing of two clinical pharmacokinetic software systems was discontinued as no further updates and support were available. These softwares were computer-assisted dosing and cedars pharmacokinetics system [81].

Most widely used PK/PD modeling softwares WinNonlin, NONMEM, Gastroplus, Simcyp, PK-Sim, and PKBugs were developed in the 1990s. All these PK/PD modeling softwares are briefly described in Table 2.3. Most of these softwares were commercial and thereby they still sustain with continuous improvements over time for the last 30 years. Most recent developing attempts on these softwares are focused to make them more user friendly. These softwares are now not only helping scientists but also the regulators for predicting PK/PD for taking decisions on regulatory clearances for the approval of new drugs and drug products. In mid-1990s, Bayer's PBPK project started, which led to the development of a commercial software package called PK-Sim™. Now it has been integrated in Open Systems Pharmacology Suite that provides free access.

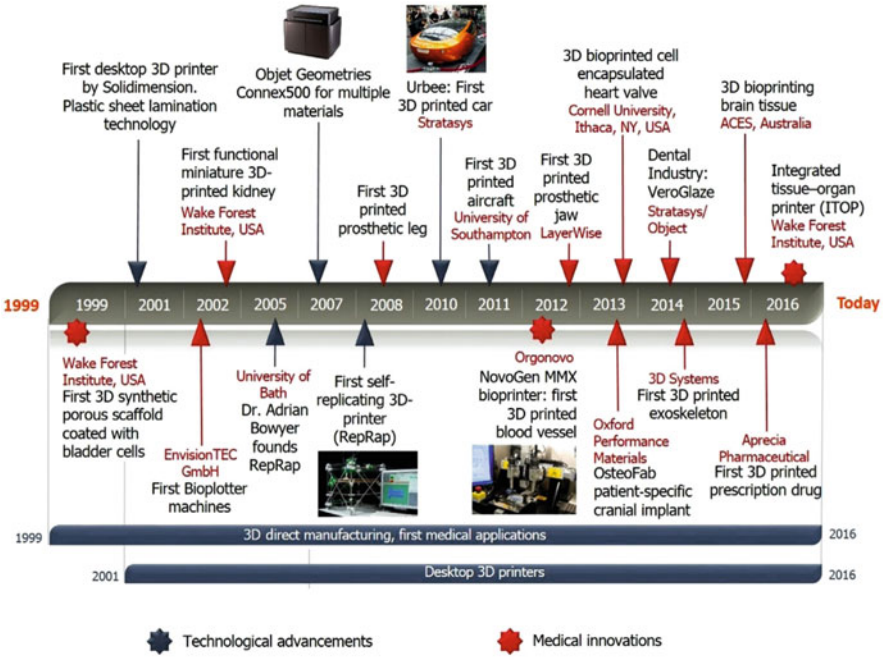


Fig. 2.8 Some milestone developments of 3D printed in twenty-one century. (Reproduced with permission from Ref. [78])

2.6 A Brief History on Robotic Developments

Robotics Institute of America (1979) has defined robots as the "Re-programmable, multi-functional manipulator-designed to move materials, parts, tools, or specialized devices, through variable programmed motions for the performance of a variety of tasks" [82]. Since 1950, scientists were developing robots primarily for automation of industries. Today's industrial robots, collaborative and humanoids, are the results of their evolution with time with the development of robotic technologies, computational abilities, sensor and vision technologies, and artificial intelligence. Soft robots are an important subclass of robots, which deals with fabricating robots with materials having appearance and feel similar to animals or human tissues. These are also referred to as bioinspired systems which mimic animal or human capabilities. Starting from the 1980s, soft robots have been developed for simulation of movements and locomotion of animals, insects, or humans. From 2017, researchers are also leveraging the abilities of soft robots to deliver drugs at specific sites in the human body. Developments with time in robotics are presented in Fig. 2.9.

Table 2.3 Various pharmacokinetic softwares, their features, latest version, and platform to use

Software; latest version; release date	Brief description	Date of release of first version	URL
Phoenix; WinNonlin 8.3; June 2020	Noncompartmental analysis (NCA), PK/PD modeling, and toxicokinetic (TK) modeling; runs on Windows	1990	https://www.certara.com/
GastroPlus® 9.8; Oct. 2020	Available as ten different modules, viz. Drug Drug Interaction, PBPKPlus™, ADMET Predictor®, Additional Dosage Routes, Metabolism and Transporter, Biologics, Optimization, PDPlus™, PKPlus™, and IVIVCPlus™; PBPK modeling, PBBM modeling, in silico PK prediction, IVIVC prediction, virtual trials, prediction of drug-drug interactions food effects and pharmacodynamics, optimization of generic formulations, and population pharmacokinetic; runs on Windows	August 1998	https://www.simulations-plus.com/
NONMEM®	Nonlinear mixed-effect modeling software program (NONMEM®), Population PK and PK/PD modeling (PREDPP); runs on Windows or Linux; access by subscription	1989	https://www.iconplc.com/innovation/nonmem/
MATLAB Simbiology	PBPK, PKPD, quantitative systems pharmacology, model building either MATLAB programming or Simbiology block diagram editor; noncompartment analysis	Late 1970 (MATLB)	https://www.mathworks.com/
Simcyp™ PBPK Simulator version 20	Physiologically based pharmacokinetic (PBPK) modeling and simulation; models for maternal health, biowaivers, and long-acting injectable drugs; separate modules for animals, pediatric, cardiac safety, long acting, injectables, and	1999	https://www.certara.com/

(continued)

Table 2.3 (continued)

Software; latest version; release date	Brief description	Date of release of first version	URL
	lactation; SimCyp founded in 1999		
PK-Sim™ (Part of OSP Suite. Version 9.1; July 2020)	PK-Sim™ allows PBPK modeling and PBPK/PD modeling; based on compartmental gastrointestinal (GI) transit model; communication and exchange of data via Matlab®, R, and MS Excel; an open access suite; both PK-Sim™ and modeling software tool MoBi® are integrated into OSP; run on Windows	Mid-1990s	http://www.open-systems-pharmacology.org/
PKBugs version 2; WinBUGS 1.4; 2007	Complex population PK/PD modeling; Markov Chain Monte Carlo methods to be applied to PK/PD analysis; no limitation of dimensional array and size; free to download; run on Windows	1997	https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/winbugs-development/
MEDICI-PK™	Whole-body PK modeling; virtual PK laboratory based on modules; in silico comparative PK studies of different species, different individuals, different compounds, and/or different models; bidirectional interface to Microsoft Excel; access by subscription	Mid-2000s (2005–2006)	https://www.cit-wulkow.de/products/medici-pk
Kinetica 5.1 2014	Noncompartment, compartment, and population PK/PD modeling; bioequivalence, protein, and enzyme kinetics computations; data communication with Thermo Scientific Watson LIMS™, MS Excel, Oracle, and others; access by subscription		https://www.thermofisher.com/

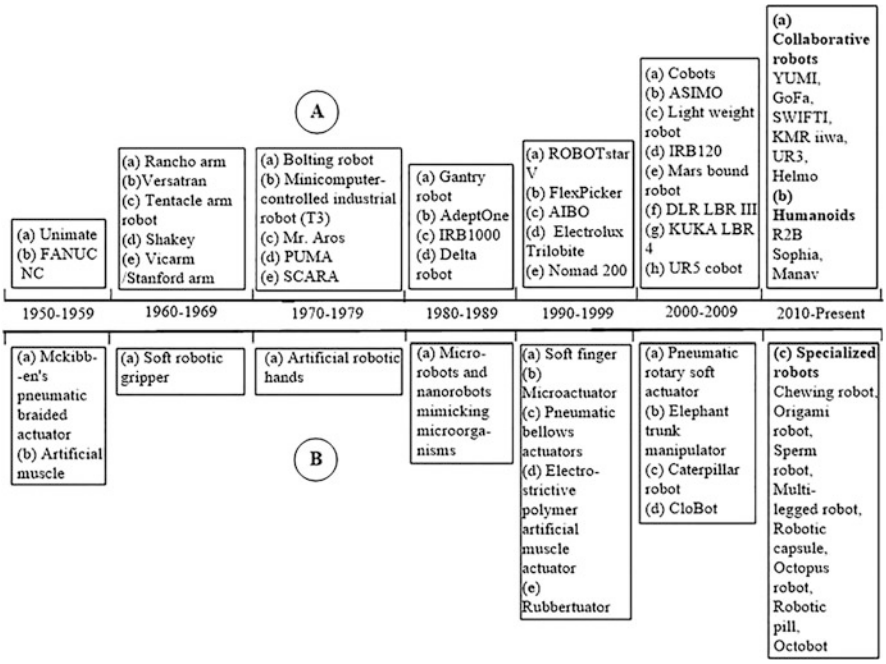


Fig. 2.9 Development history of (A) industrial robots and (B) soft robots

2.6.1 1950–1959: Birth of First Industrial Robot and Numerical Control System

This era marked the birth of robots with *Unimate* as the first industrial robot. George Charles Devol, the father of robotics, developed *Unimate* in 1954 [83] and later a patent on this innovation was granted to him in 1961 [84]. *Unimate* was actually a mechanical arm. In 1956, a Japanese company FANUC developed a numerical control (NC) system for the automation in manufacturing division [85]. It was named after the company name. FANUC represents *Fuji Automatic Numerical Control*. In 1972, it was modified further as computer numerical control (CNC).

Simulation of body parts started in 1950 with the invention of pneumatic artificial muscle (McKibben's braided pneumatic actuators) [86]. It was developed by Joseph Laws McKibben for polio patients. It was contracted and extended using a pneumatic bladder containing pressurized air. In 1958, R.H. Gaylord invented an artificial muscle (pneumatic actuator), which was made up of pure latex [87]. Artificial muscle was easily expandable and its outer part was covered using double helical braided wire. It was developed for helping functionally disabled persons who cannot walk, sit, or stand.

2.6.2 1960–1969: Evolution of Industrial Robots with More Degrees of Freedom and Robotic Grippers

This decade marked some major modifications in industrial robots with attempts to increase their degrees of freedom to get greater work envelope. In 1963, *Rancho arm* robot was developed by Rancho Los Amigos Hospital, California [88]. This robotic arm had six joints and was mainly developed to assist handicapped persons. *Versatran* was developed by Harry Johnson and Velijko Milenkovicat of American Machine and Foundry in 1962 [89]. It was a programmed cylindrical frame robotic arm. Model 102 and model 212 were simultaneously developed with this robotic arm. *Tentacle arm robot*, developed by Marvin Minsky in 1968 [88], had 12 joints and a structure similar to a mechanical arm. *Tentacle arm robot* had the ability to lift a person. *Shakey* was a mobile robot, developed by Stanford Research Institute, California, in 1968 [90]. Shakey was equipped with vision capability and was controlled by a computer. *Vicarm/Stanford arm,* developed by Viktor Scheinmann at Stanford University in 1969 [91], had six degrees of freedom. Unlike Unimate, it was smaller but it was not able to carry such heavy loads. Soft robotic grippers were first developed by James I. Baer who received a patent (US3343864A) in 1967 [92]. Grippers consisted of curling flexible fingers fitted into gripping heads and other end effectors. Soft robotic grippers enable robots to pick up and hold an object, which is an essential task in robotic automations in industries.

2.6.3 1970–1979: Development of Robots for Special Purposes

Special purpose robots started appearing in the 1970s. *Bolting robot* was developed by Hitachi in 1973 mainly for concrete pile and pole industries [93, 94]. Bolting robot was equipped with dynamic sensors for vision to identify moving objects, a unique feature at that time [95]. Japanese company Kawasaki developed the first *arc welding robot* in 1974 for preparing frames of motorcycles [93]. In 1975, Hitachi came up with an arc welding robot, *Mr. Aros*, equipped with sensors and a microprocessor for detecting locations. A minicomputer-controlled industrial robot *T3* was introduced in 1974 by Richard Hohn at Cincinnati Milacron Corporation [93]. First, computer-type, programming language, namely *WAVE,* for robots was developed in 1973 by SRI International, USA [96]. Victor Scheinman from Unimation developed another industrial robotic arm, namely *PUMA* (Programmable Universal Machine for Assembly, or Programmable Universal Manipulation Arm), in 1978. In 1979, *SCARA* (Selective Compliance Assembly Robot Arm) type robot was first developed at Yamanashi University, Japan. Artificial robotic hands were developed by Stanley R. Rich in 1971 (US3561330A) [97] for simulating human hands. Artificial robotic hand, consisting of five rubber tubes, was able to deform longitudinally with changes in fluid pressure in the tubes.

2.6.4 1980–1989: Growth of Modern Industrial Robots

The 1980s marked some of the significant developments of robots, which we can still observe in application. Some of currently used industrial robots have their roots in this era. American company PaR Systems developed and launched the first *gantry robot* in 1981 [98]. It had a high range of motion and capacity to replace many robots. *AdeptOne* was mainly a SCARA robot, developed by Omron Adept Technology, Inc., in 1984 [99]. This robot was highly accurate and capable to work at a very high speed with its 6 kg load carrying capacity. In 1984, ABB of Sweden developed *IRB1000,* which was a like an extended arm of a human in a vertical direction [93]. IRB1000 was around 50% faster than any of the traditional arm robots of that time. *Delta robots* are widely used nowadays in different divisions of pharmaceutical industries. *Delta robot* was first developed by Professor Reymon Clavel in 1985 at The École polytechnique fédérale de Lausanne, Switzerland [100]. It was a specific type of parallel robot consisting of three arms joined at a common point, which gave appearance similar to a spider.

Microorganisms have the ability of self-assembling and moving in different media. Soft robots capable of simulating microorganisms were developed in the 1980s [101, 102]. For mimicking micoorganisms, it was very difficult to miniaturize the conventional robots. Hence, miniaturized soft microrobots and nanorobots were developed. Soft robots of this era contained sensors, processors, and actuators for controlling the robot externally.

2.6.5 1990–1999: The Decade of Fastest Industrial Robots and Actuators

This era had some of the remarkable developments in robotics as the use of robots in industries rose and expanded to different divisions [103]. *ROBOTstar V* was developed by the German-based automation company, namely Reis Robotics, in 1998 [104]. It was mainly a robot controlling generation. The famous German-based automation company KUKA had taken 51% of share of Reis Robotics in 2013. *FlexPicker*, a delta robot, was developed by ABB of Sweden in 1998 [93]. It had a great advantage of fastest picking from manufacturing/packing lines. In 1996, Professor J. Edward Colgate and Professor Michael Peshkin at Northwestern University, Illinois, created physical interaction between a human and a robot with a system controlled by a computer which promulgated the concept of *Cobots or collaborative robots* [105].

The decade of 1990s also marked the developments of different actuators for their use in soft robots. Researchers developed *soft finger* in 1990 [106]. It was developed for controlling the contact forces, originated when the finger is close on an object. Suzumori developed a flexible *microactuator* in 1996 [107]. It was actually a pneumatic rubber actuator, made up of silicone rubber reinforced with nylon fiber. It was widely used in robotic arm, multifingered robot hand, walking robots, etc. The *pneumatic bellows actuator* was developed by G Robinson, John Bruce Clayfield

Davies at Heriot-Watt University, Scotland in 1998 [108]. Bellows were attached with the actuators to create pressure and rise temperature at a continuous rate. *Electrostrictive polymer artificial muscle actuator* from polyurethane and silicone was developed by R. Kornbluh, R. Pelrine, J. Eckerle, and J. Joseph in 1998 [109]. The actuator was able to create pressure up to 1.9 MPa with an efficiency and response faster than normal muscle. A light weight, high power, and spark-free *Rubbertuator,* a pneumatic robot, was developed in 2000 [110].

Honda started research on the bipedal function of robots in the 1990s with its P series robots [111]. Their motto was to develop robots, which can walk and function like human. In the P series, Honda first introduced the P1 prototype bipedal model (1993–1997), which contained upper limbs and the body. Prototype P1 was modified to P2 (1996) for independent walking on ground, walking up and down on stairs, and other activities in a manner similar to humans. The year 1997 marked the completion of the first bipedal humanoid walking robot.

2.6.6 2000–2009: The Birth of Collaborative Robots and Humanoids and Evolution of Bioinspired Robots

Honda's *ASIMO* (Advanced Step in Innovative Mobility) was the world's first humanoid robot which was launched in 2000 [112]. *ASIMO* is an autonomous robot and can adapt any situation intelligently and physically. Its latest version was released in 2011. German-based automation company KUKA and Institute of Robotics and Mechatronics, Germany, together first developed *light weight robot* made of aluminum in 2006 [93]. It contained a powerful sensor and had a payload capacity of 7 kg. In 2009, ABB of Sweden developed a small and multipurpose robot named *IRB120* [93]. It weighed 25 kg and had a payload capacity of up to 3 kg. This era also marked the travel of robots in space missions. The highly successful Mars Exploration Rover Mission started in 2003. NASA sent two rovers (robots), namely S*pirit* and *Opportunity,* to explore the surface and geology of Mars [113]. S*pirit* and *Opportunity* landed in 2004 at two different locations on Mars. Esben Østergaard, Kasper Støy, and Kristian Kassow from University of Southern Denmark (2001–2005) reinvented human robot interaction and a collaborative robot and, later in 2005, founded Universal Robots in Denmark [114]. *UR5 cobot*, the world's first collaborative robot, was launched in 2008 by Universal Robotics. Collaborative robots have the ability to work with humans safely. The base technology for the collaborative robots originated in the German Aerospace Center (DLR), where a lightweight, sensitive robot arm was developed for use in space [115]. In 2004, the technology of DLR LWR III was licensed to KUKA GmbH, a Germany-based robotic company. Single-arm collaborative robot KUKA LBR 4, launched in 2008, was an outcome of refinement of DLR LWR III technology further.

Soft robots advanced further to excel in simulating motions of animals like elephant, caterpillar, etc. In 2001, Toshiro Noritsugu, Mitsuhiko Kubota, and Sadaharu Yoshimatsu at Okayama University, Japan, developed a *pneumatic rotary soft actuator* from silicone rubber for its subsequent use in preparing robotic fingers

and robotic hand [116]. M.W. Hman and I.D. Walker designed *elephant trunk manipulator* in 2001 at Clemson University [117]. It was actually an electronically controlled arm like a manipulator, having 32 degrees of freedom and a structure resembling the trunk of an elephant. *Caterpillar robot,* a soft robot capable of crawling through the ropes and wires, simulated locomotion of caterpillar. Barry A. Trimmer, Ann E. Takesian, and Brian M. Sweet at Trufts University (US) fabricated one *caterpillar robot* in 2006 [118]. G. Chen, M. T. Pharm, and T. Redarce at Unilever, UK, and Laboratoire Ampere, France, came up with *CloBot* in 2007 [119]. *CloBot* was a pneumatically controlled soft robotic manipulator for its possible use in medical surgery like colonoscopy.

2.6.7 2010–Present: The Rise of Collaborative Robots, Humanoids, and Soft Robots

The present era has observed advancements of cobots to laboratory robots, humanoids, and soft robots with significant technological advancements in artificial intelligence and sensor technologies. Collaborative robots have already been deployed in the assembly lines for performing a wide variety of simple tasks like picking, packing, palletizing, handling, and inspection. Further, collaborative robot applications in quality control purposes in industries, laboratories, and research institutions are also rising. ABB's *YUMI* (dual-arm collaborative robot launched in 2015), *GoFa* (launched in 2021), *SWIFTI* (launched in 2021) cobot families [120, 121], KUKA's *KMR-iiwa* (mobile robot launched in 2013) [122, 123], Stäubli's *Helmo* (launched in 2019) [124], and Universal Robots' *UR3* (launched in 2015) [125] are some of the recently launched collaborative robots that await their deployment in industries and research laboratories. NASA developed the first space bound humanoid robot *R2B (robonaut)* in 2011 [93].

Humanoid robots have received citizenship in some countries, like Saudi Arabia and Japan. *Sophia* is a well-known first social humanoid robot with artificial intelligence. It was developed by David Hanson at the Hong Kong-based company Hanson robotics in 2015–2016 [126, 127]. This robot was modeled after three persons, namely the queen of Egypt, Nefertiti, British actress, and humanitarian Audrey Hepburn and Amanda Hanson, wife of developer. In October, 2017, Sophia got citizenship in Saudi Arabia. This robot can show 62 facial expressions, change color of eyes with light, and socialize with humans.

Manav is the first 3D printed humanoid robot from India. It was developed by Diwakar Vaish at A-SET Training and Research Institutes, Uttar Pradesh, in the year of 2014 [128]. The outer frame of this robot is made of acrylonitrile butadiene styrene. It consists of processors and highly advanced sensors. It is equipped with 21 sensors, 2 mic on either side of the head, and 2 cameras located in eye sockets. With an open-source code, Manav may learn and respond similar to a human child, play football, and perform activities like headstands and push-ups.

Special purpose robots for assessment of drug delivery systems are also under investigation. A *chewing robot* has been recently developed by Kazem Alemzadeh,

Siân Bodfel Jones, Maria Davies, and Nicola West at the University of Bristol, Bristol, in 2020. This robot simulates human mastication for assessing the release of drug from chewing gum [129].

Soft robots have received considerable attention from scientists for delivering drugs at targeted locations in the human body. *Origami robot,* made from biodegradable and biocompatible materials in 2016, was an outcome of research efforts of researchers at MIT, the University of Sheffield, and the Tokyo Institute of technology [130]. Researchers from Institute for Integrative Nanosciences and Material Systems for Nanoelectronics of Germany fabricated *sperm robot* in 2018 for targeted drug delivery [131]. Sperm robot simulates the movements of human sperm. *Soft multi-legged robot,* made from soft materials like polyurethane, polyethylene terephthalate, polyethyleneoxide, etc., was developed at Beijing Institute of Technology and City University of Hong Kong, China, in 2018 [132]. *Arobotic capsule* was developed at Vanderbilt University, USA, and Politecnico di Milano, Italy, in 2015, for targeted drug delivery [133]. The robotic capsule consisted of a drug chamber, two magnetic coils, and the magnetic piston. The drug chamber was embedded in a cylindrical enclosure together with an axially magnetized cylindrical permanent magnet acting as a piston. An *octopus robot* was another bioinspired robot developed by researchers of The BioRobotics Institute, Italy, and Hebrew University of Jerusalem, Israel, in 2014 [134]. It had got its name from its eight arms in radial directions. Its front arm can grasp any object and other arms are for walking. Due to its neutral (approximately) buoyancy, it can walk on the water. Researchers at Harvard University developed the first autonomous, untethered, entirely soft robot, namely *octobot* [135]. Researchers have used 3D printed components to fabricate this robot. Even its fuel storage is also 3D printed. It is actually a pneumatic soft robot in which a small amount of hydrogen peroxide is required for its activation.

2.7 Conclusion

Formulation development, manufacturing of dosage forms, and pharmacokinetic assessment of dosage forms have widely observed the emerging use of computers, artificial intelligence, and robotics. ANN as simple as one hidden layer and complex ANN having multiple hidden layers, for example, 39, have been used for optimizing pharmaceutical formulations. The increase in computational abilities with time, advanced softwares equipped with AI, and more interactive human-machine interface have a tremendous role in the rising use of computer-based optimization of both pharmaceutical formulations and processes. Industrial robots equipped with AI-based advanced softwares are available to address the growing demand for automation of pharmaceutical industries. Soft robots equipped with soft computers and sensors are rapidly being developed to achieve better efficiency for spatial and temporal delivery of drugs. Computer-assisted 3D printing and other computer-assisted manufacturing techniques are gaining wide acceptance due to their great flexibilities for computer-aided designing of internal architecture and external features of various types of dosage forms, devices, tissues, and organs.

References

1. Ekins S (2006) Computer applications in pharmaceutical research and development. Wiley, USA, pp 1–442
2. Boyd DB, Marsh MM (2006) History of computers in pharmaceutical research and development: a narrative. In: Ekins S (ed) Computer applications in pharmaceutical research and development. Wiley, USA, pp 1–50. https://doi.org/10.1002/0470037237.ch1
3. Prinderre P, Piccerelle P, Cauture E, Kalantzis G, Reynier JP, Joachim J (1998) Formulation and evaluation of o/w emulsions using experimental design. Int J Pharm 163:73–79. https://doi.org/10.1016/S0378-5173(97)00368-2
4. Simovic S, Milic-Askrabic J, Vuleta G, Ibric S, Stupar M (1999) The influence of processing variables on performance of O/W emulsion gels based on polymeric emulsifier (Pemulen®TR-2NF). Int J Cosmet Sci 21:119–125. https://doi.org/10.1046/j.1467-2494.1999.183572.x
5. Rowe RC, Colbourn EA (2003) Neural computing in product formulation. Chem Educ 8:1–8
6. Niedz RP, Evens TJ (2016) Design of experiments (DOE)—history, concepts, and relevance to in vitro culture. Vitr Cell Dev Biol Plant 52:547–562. https://doi.org/10.1007/s11627-016-9786-1
7. Durakovic B, Basic H (2013) Textile cutting process optimization model based on six sigma methodology in a medium-sized company. Period Eng Nat Sci 1:39–46. https://doi.org/10.21533/pen.v1i1.15
8. Paulo F, Santos L (2017) Design of experiments for microencapsulation applications: a review. Mater Sci Eng C 77:1327–1340. https://doi.org/10.1016/j.msec.2017.03.219
9. Yu P, Low MY, Zhou W (2018) Design of experiments and regression modelling in food flavour and sensory analysis: a review. Trends Food Sci Technol 71:202–215. https://doi.org/10.1016/j.tifs.2017.11.013
10. Schlueter A, Geyer P (2018) Linking BIM and Design of Experiments to balance architectural and technical design factors for energy performance. Autom Constr 86:33–43. https://doi.org/10.1016/j.autcon.2017.10.021
11. Durakovic B, Torlak M (2017) Simulation and experimental validation of phase change material and water used as heat storage medium in window applications. J Mater Environ Sci 8:1837–1846
12. Hibbert DB (2012) Experimental design in chromatography: a tutorial review. J Chromatogr B Anal Technol Biomed Life Sci 910:2–13. https://doi.org/10.1016/j.jchromb.2012.01.020
13. Garud SS, Karimi IA, Kraft M (2017) Design of computer experiments: a review. Comput Chem Eng 106:71–95. https://doi.org/10.1016/j.compchemeng.2017.05.010
14. Fisher RA (1921) Studies in crop variation: I. An examination of the yield of dressed grain from broadbalk. J Agric Sci 11:107–135. https://doi.org/10.1017/S0021859600003750
15. Fisher RA (1925) Statistical methods for research workers. Oliver and Boyd, Edinburgh, pp 152–293. https://doi.org/10.1111/j.1744-7348.1926.tb04258.x
16. Fisher RA (1926) The arrangement of field experiments. J Minist Agric Gt Britain 33:503–513
17. Fisher RA (1937) The design of experiments. Oliver and Boyd, Edinburgh, pp 26–54
18. Fisher RA (1947) Development of the theory of experimental design. Proc Int Stat Conf 3: 434–439
19. The Pennsylvania State University (2021) A quick history of the design of experiments (DOE). https://online.stat.psu.edu/stat503/lesson/1/1.1. Accessed 27 Apr 2021
20. Gaither CC, Cavazos-Gaither AE (2012) In: Gaither CC, Cavazos-Gaither AE (eds) Gaither's dictionary of scientific quotations. Springer, New York, pp 489–838. https://doi.org/10.1007/978-1-4614-1114-7
21. Durakovic B (2017) Design of experiments application, concepts, examples: state of the art. Period Eng Nat Sci 5:421–439. https://doi.org/10.21533/pen.v5i3.145
22. Box GEP, Wilson KB (1951) On the experimental attainment of optimum conditions. J R Stat Soc Ser B 13:1–38. https://doi.org/10.1111/j.2517-6161.1951.tb00067.x

23. Mead R, Pike DJ (1975) A review of response surface methodology from a biometric viewpoint. Biometrics 31:803–851
24. Myers RH, Montgomery DC, Geoffrey Vining G, Borror CM, Kowalski SM (2004) Response surface methodology: a retrospective and literature survey. J Qual Technol 36:53–78. https://doi.org/10.1080/00224065.2004.11980252
25. Box GEP, Draper NR (2007) Response surfaces, mixtures, and ridge analyses. Wiley-Interscience, pp 152–289
26. Marlowe E, Shangraw RF (1967) Dissolution of sodium salicylate from tablet matrices prepared by wet granulation and direct compression. J Pharm Sci 56:498–504. https://doi.org/10.1002/jps.2600560415
27. Taguchi G, Chowdhury S, Wu Y (2005) Taguchi's quality engineering handbook. Wiley Blackwell, pp 3–58. https://doi.org/10.1002/9780470258354
28. Andersson R, Eriksson H, Torstensson H (2006) Similarities and differences between TQM, six sigma and lean. TQM Mag 18:282–296. https://doi.org/10.1108/09544780610660004
29. Politis SN, Colombo P, Colombo G, Rekkas DM (2017) Design of experiments (DoE) in pharmaceutical development. Drug Dev Ind Pharm 43:889–901. https://doi.org/10.1080/03639045.2017.1291672
30. Fukuda IM, Pinto CFF, Moreira CDS, Saviano AM, Lourenço FR (2018) Design of experiments (DoE) applied to pharmaceutical and analytical quality by design (QbD). Braz J Pharm Sci 54:e01006. https://doi.org/10.1590/s2175-97902018000001006
31. Leo Kumar SP (2019) Knowledge-based expert system in manufacturing planning: state-of-the-art review. Int J Prod Res 57:4766–4790. https://doi.org/10.1080/00207543.2018.1424372
32. Al-Khattat IM, Al-Hassani STS (1987) Towards a computer-aided analysis and design of tablet compaction. Chem Eng Sci 42:707–712. https://doi.org/10.1016/0009-2509(87)80030-1
33. Batanov D, Nagarur N, Nitikhunkasem P (1993) EXPERT-MM: a knowledge-based system for maintenance management. Artif Intell Eng 8:283–291. https://doi.org/10.1016/0954-1810(93)90012-5
34. Frank J, Rupprecht B, Schmelmer V (1997) Knowledge-based assistance for the development of drugs. IEEE Expert Syst Their Appl 12:40–48. https://doi.org/10.1109/64.577412
35. Dai S, Xu B, Shi G, Liu J, Zhang Z, Shi X et al (2019) SeDeM expert system for directly compressed tablet formulation: a review and new perspectives. Powder Technol 342:517–527. https://doi.org/10.1016/j.powtec.2018.10.027
36. Ramani KV, Patel MR, Patel SK (1992) An expert system for drug preformulation in a pharmaceutical company. Interfaces (Providence) 22:101–108. https://doi.org/10.1287/inte.22.2.101
37. Khan A, Iqbal Z, Rehman Z, Nasir F, Khan A, Ismail M et al (2014) Application of SeDeM Expert system in formulation development of effervescent tablets by direct compression. Saudi Pharm J 22:433–444. https://doi.org/10.1016/j.jsps.2013.07.002
38. Puchkov M, Tschirky D, Leuenberger H (2013) 3-D cellular automata in computer-aided design of pharmaceutical formulations: mathematical concept and F-CAD software. In: Aguilar JE (ed) Formulation tools for pharmaceutical development. Elsevier Ltd, pp 155–201. https://doi.org/10.1533/9781908818508.155
39. Zhang GGZ, Law D, Schmitt EA, Qiu Y (2004) Phase transformation considerations during process development and manufacture of solid oral dosage forms. Adv Drug Deliv Rev 56:371–390. https://doi.org/10.1016/j.addr.2003.10.009
40. Rowe RC, Wakerly MG, Roberts RJ, Grundy RU, Upjohn NG (1995) Expert systems for parenteral development. PDA J Pharm Sci Technol 49:257–261
41. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NAE, Arshad H (2018) State-of-the-art in artificial neural network applications: a survey. Heliyon 4:e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

42. Sbirrazzuoli N, Brunel D (1997) Computational neural networks for mapping calorimetric data: application of feed-forward neural networks to kinetic parameters determination and signals filtering. Neural Comput Appl 5:20–32. https://doi.org/10.1007/BF01414100
43. Zadeh LA (1968) Fuzzy algorithms. Inf Control 12:94–102. https://doi.org/10.1016/S0019-9958(68)90211-8
44. Kesavan JG, Peck GE (1996) Pharmaceutical granulation and tablet formulation using neural networks. Pharm Dev Technol 1(4):391–404. https://doi.org/10.3109/10837459609031434
45. Turkoğlu M, Özarslan R, Sakr A (1995) Artificial neural network analysis of a direct compression tabletting study. Eur J Pharm Biopharm 41:315–322
46. Bourquin J, Schmidli H, Van Hoogevest P, Leuenberger H (1997) Application of Artificial Neural Networks (ANN) in the development of solid dosage forms. Pharm Dev Technol 2: 111–121. https://doi.org/10.3109/10837459709022616
47. Rocksloh K, Rapp FR, Abu Abed S, Müller W, Reher M, Gauglitz G et al (1999) Optimization of crushing strength and disintegration time of a high-dose plant extract tablet by neural networks. Drug Dev Ind Pharm 25:1015–1025. https://doi.org/10.1081/DDC-100102264
48. Chen Y, Thosar SS, Forbess RA, Kemper MS, Rubinovitz RL, Shukla AJ (2001) Prediction of drug content and hardness of intact tablets using artificial neural network and near-infrared spectroscopy. Drug Dev Ind Pharm 27:623–631. https://doi.org/10.1081/DDC-100107318
49. Sathe PM, Venitz J (2003) Comparison of neural network and multiple linear regression as dissolution predictors. Drug Dev Ind Pharm 29:349–355. https://doi.org/10.1081/DDC-120018209
50. Ibric S, Djuric Z, Parojcic J, Petrovic J (2009) Artificial intelligence in pharmaceutical product formulation: neural computing. Chem Ind Chem Eng Q 15:227–236. https://doi.org/10.2298/CICEQ0904227I
51. Hussain AS, Yu X, Johnson RD (1991) Application of neural computing in pharmaceutical product development. Pharm Res 8:1248–1252. https://doi.org/10.1023/A:1015843527138
52. Hussain AS, Shivanand P, Johnson RD (1994) Application of neural computing in pharmaceutical product development: computer aided formulation design. Drug Dev Ind Pharm 20: 1739–1752. https://doi.org/10.3109/03639049409038390
53. Wu PC, Obata Y, Fujikawa M, Li CJ, Higashiyama K, Takayama K (2001) Simultaneous optimization based on artificial neural networks in ketoprofen hydrogel formula containing O-ethyl-3-butylcyclohexanol as percutaneous absorption enhancer. J Pharm Sci 90:1004–1014. https://doi.org/10.1002/jps.1053
54. Takahara J, Takayama K, Nagai T (1997) Multi-objective simultaneous optimization technique based on an artificial neural network in sustained release formulations. J Control Release 49:11–20. https://doi.org/10.1016/S0168-3659(97)00030-8
55. Peh KK, Lim CP, Quek SS, Khoh KH (2000) Use of artificial neural networks to predict drug dissolution profiles and evaluation of network performance using similarity factor. Pharm Res 17:1384–1388. https://doi.org/10.1023/A:1007578321803
56. Türkoğlu M, Varol H, Çelikok M (2004) Tableting and stability evaluation of enteric-coated omeprazole pellets. Eur J Pharm Biopharm 57:279–286. https://doi.org/10.1016/j.ejpb.2003.10.008
57. Yuksel N, Turkoglu M, Baykara T (2000) Modelling of the solvent evaporation method for the preparation of controlled release acrylic microspheres using neural networks. J Microencapsul 17:541–551. https://doi.org/10.1080/026520400417603
58. Surini S, Akiyama H, Morishita M, Nagai T, Takayama K (2003) Release phenomena of insulin from an implantable device composed of a polyion complex of chitosan and sodium hyaluronate. J Control Release 90:291–301. https://doi.org/10.1016/S0168-3659(03)00196-2
59. Reis MAA, Sinisterra RD, Belchior JC (2004) An alternative approach based on artificial neural networks to study controlled drug release. J Pharm Sci 93:418–430. https://doi.org/10.1002/jps.10569

60. Michrafy A, Ringenbacher D, Tchoreloff P (2002) Modelling the compaction behaviour of powders: application to pharmaceutical powders. Powder Technol 127:257–266. https://doi.org/10.1016/S0032-5910(02)00119-5
61. Lewis RW, Gethin DT, Yang XS, Rowe RC (2005) A combined finite-discrete element method for simulating pharmaceutical powder tableting. Int J Numer Methods Eng 62:853–869. https://doi.org/10.1002/nme.1287
62. Rowe RC, Roberts RJ (1992) Simulation of crack propagation in tablet film coatings containing pigments. Int J Pharm 78:49–57. https://doi.org/10.1016/0378-5173(92)90354-5
63. Jamróz W, Szafraniec J, Kurek M, Jachowicz R (2018) 3D printing in pharmaceutical and medical applications—recent achievements and challenges. Pharm Res 35:176. https://doi.org/10.1007/s11095-018-2454-x
64. ISO (2015) ISO/ASTM 52900:2015(en). Additive manufacturing—general principles—terminology. https://www.iso.org/obp/ui/#iso:std:iso-astm:52900:ed-1:v1:en. Accessed 28 Apr 2021
65. Alhnan MA, Okwuosa TC, Sadia M, Wan KW, Ahmed W, Arafat B (2016) Emergence of 3D printed dosage forms: opportunities and challenges. Pharm Res 33:1817–1832. https://doi.org/10.1007/s11095-016-1933-1
66. Norman J, Madurawe RD, Moore CMV, Khan MA, Khairuzzaman A (2017) A new chapter in pharmaceutical manufacturing: 3D-printed drug products. Adv Drug Deliv Rev 108:39–50. https://doi.org/10.1016/j.addr.2016.03.001
67. Sachs EM, Haggerty JS, Cima MJ, Williams PA (1989) Three-dimensional printing techniques. US5340656A. https://patents.google.com/patent/US5340656A/en. Accesssed 27 Apr 2021
68. Sculpteo (2021) The history of 3D printing: from the 80s to today. https://www.sculpteo.com/en/3d-learning-hub/basics-of-3d-printing/the-history-of-3d-printing/. Accessed 28 Apr 2021
69. Wang J, Goyanes A, Gaisford S, Basit AW (2016) Stereolithographic (SLA) 3D printing of oral modified-release dosage forms. Int J Pharm 503:207–212. https://doi.org/10.1016/j.ijpharm.2016.03.016
70. Serrano-Aroca Á, Vera-Donoso CD, Moreno-Manzano V (2018) Bioengineering approaches for bladder regeneration. Int J Mol Sci 19:1–26. https://doi.org/10.3390/ijms19061796
71. Whitaker M (2014) The history of 3D printing in healthcare. Bull R Coll Surg Engl 96:228–229. https://doi.org/10.1308/147363514x13990346756481
72. Lee Ventola C (2014) Medical applications for 3D printing: current and projected uses. P T 39:704–711
73. Jassim-Jaboori AH, Oyewumi MO (2015) 3D printing technology in pharmaceutical drug delivery: prospects and challenges. J Biomol Res Ther 4:4–7. https://doi.org/10.4172/2167-7956.1000e141
74. Clark EA, Alexander MR, Irvine DJ, Roberts CJ, Wallace MJ, Sharpe S et al (2017) 3D printing of tablets using inkjet with UV photoinitiation. Int J Pharm 529:523–530. https://doi.org/10.1016/j.ijpharm.2017.06.085
75. Kang HW, Lee SJ, Ko IK, Kengla C, Yoo JJ, Atala A (2016) A 3D bioprinting system to produce human-scale tissue constructs with structural integrity. Nat Biotechnol 34:312–319. https://doi.org/10.1038/nbt.3413
76. Jakus AE (2019) An introduction to 3D printing—past, present, and future promise. In: 3D print. orthop. surg. Elsevier, pp 1–15. https://doi.org/10.1016/b978-0-323-58118-9.00001-4
77. Prasad LK, Smyth H (2016) 3D Printing technologies for drug delivery: a review. Drug Dev Ind Pharm 42:1019–1031. https://doi.org/10.3109/03639045.2015.1120743
78. Provaggi E, Kalaskar DM (2017) 3D printing families: laser, powder, nozzle based techniques. In: 3D printing in medicine. Elsevier Inc., pp 21–42. https://doi.org/10.1016/B978-0-08-100717-4.00003-X
79. Jelliffe RW, Jelliffe SM (1972) A computer program for estimation of creatinine clearance from unstable serum creatinine levels, age, sex, and weight. Math Biosci 14:17–24. https://doi.org/10.1016/0025-5564(72)90003-X

80. Schentag JJ, Adelman MH (1983) A microcomputer program for tobramycin consult services, based on the two-compartment pharmacokinetic model. Drug Intell Clin Pharm 17:528–531. https://doi.org/10.1177/106002808301700706
81. Fuchs A, Csajka C, Thoma Y, Buclin T, Widmer N (2013) Benchmarking therapeutic drug monitoring software: a review of available computer tools. Clin Pharmacokinet 52:9–22. https://doi.org/10.1007/s40262-012-0020-y
82. Mannan A, Mubeen H (2018) Digitalization and automation in pharmaceuticals from drug discovery to drug administration. Int J Pharm Pharm Sci 10:1. https://doi.org/10.22159/ijpps.2018v10i6.24757
83. Robotic Industries Association (2021) Unimate—the first industrial robot. https://www.robotics.org/joseph-engelberger/unimate.cfm. Accessed 7 Apr 2021
84. Devol JGC (1954) Programmed article transfer. US2988237A. https://patents.google.com/patent/US2988237. Accessed 7 Apr 2021
85. FANUC (2021) FANUC Company history. https://www.fanuc.eu/dk/en/who-we-are/fanuc-history. Accessed 7 Apr 2021
86. Kazutoshi K, Shukushi S, Suzumori K, Shuichi W (2017) Mckibben artificial muscle. WO2017047208A1. https://patents.google.com/patent/WO2017047208A1/en. Accessed 14 Apr 2021
87. Krishna S, Nagarajan T, Rani AMA (2011) Review of current development of pneumatic artificial muscle. J Appl Sci 11:1749–1755. https://doi.org/10.3923/jas.2011.1749.1755
88. RobotWorx (2021) Industrial robot history. https://www.robots.com/articles/industrial-robot-history. Accessed 7 Apr 2021
89. Cyberneticzoo.com (2021) "VERSATRAN" industrial robot. http://cyberneticzoo.com/early-industrial-robots/1958-62-versatran-industrial-robot-harry-johnson-veljko-milenkovic/. Accessed 7 Apr 2021
90. Futura Automation (2021) A history timeline of industrial robotics. https://futura-automation.com/2019/05/15/a-history-timeline-of-industrial-robotics/. Accessed 7 Apr 2021
91. Robotics Online (2011) Automate 2011 honors 50 years of robotics. https://roboticsonline.wordpress.com/tag/vicarm/. Accessed 7 Apr 2021
92. Baer JI (1967) Material handling apparatus and the like. US3343864A. https://patents.google.com/patent/US3343864. Accessed 8 Apr 2021
93. International Federation of Robotics (2020) Robot history. https://ifr.org/robot-history. Accessed 7 Apr 2021
94. Nam H, Choi W, Ryu D, Lee Y, Lee SH, Ryu B (2007) Design of a bolting robot for constructing steel structure. In: ICCAS 2007—int. conf. control. autom. syst., pp 1946–1949. https://doi.org/10.1109/ICCAS.2007.4406667
95. Mo YH, Woo BW, Choe YG, Park JM, Lim MT (2010) Bolting robot assistance system using image processing. In: ICCAS 2010—int. conf. control. autom. syst., pp 2342–2345. https://doi.org/10.1109/iccas.2010.5669922
96. Groover MP (2012) Industrial robotics: technology, programming and application. McGraw-Hill Higher Education, New Delhi, pp 1–8. https://doi.org/10.1145/3284557.3284723
97. Rich SR (1971) Fluid operable motor. US3561330A. https://patents.google.com/patent/US3561330. Accessed 8 Apr 2021
98. Par Systems (2021) About PaR Systems. https://www.par.com/company/about-par-systems/. Accessed 7 Apr 2021
99. Newman WS, Patel JJ (1991) Experiments in torque control of the AdeptOne robot. Proc IEEE Int Conf Robot Autom 2:1867–1872. https://doi.org/10.1109/robot.1991.131897
100. Kuo YL (2016) Mathematical modeling and analysis of the Delta robot with flexible links. Comput Math Appl 71:1973–1989. https://doi.org/10.1016/j.camwa.2016.03.018
101. Palagi S, Fischer P (2018) Bioinspired microrobots. Nat Rev Mater 3:113–124. https://doi.org/10.1038/s41578-018-0016-9
102. Martel S (2016) Swimming microorganisms acting as nanorobots versus artificial nanorobotic agents: a perspective view from an historical retrospective on the future of medical

nanorobotics in the largest known three-dimensional biomicrofluidic networks. Biomicrofluidics 10:021301. https://doi.org/10.1063/1.4945734
103. Lai R, Lin W, Wu Y (2018) Review of research on the key technologies, application fields and development trends of intelligent robots. In: Chen Z, Mendes A, Yan Y, Chen S (eds) Lecture notes in computer science. Springer Nature, Switzerland, pp 449–458. https://doi.org/10.1007/978-3-319-97589-4_38
104. KUKA (2021) KUKA Service for Reis systems and robots. https://www.kuka.com/en-us/services/service-roboter-und-maschinen/servicing-of-kuka-industries-products. Accessed 7 Apr 2021
105. Colgate JE, Peshkin MA (1999) Cobots. US5952796A. https://patents.google.com/patent/US5952796A/en?q=cobot&oq=cobot. Accessed 14 Apr 2021
106. Akella P, Siegwart R, Cutkosky M (1991) Manipulation with soft fingers: contact force control. Proc IEEE Int Conf Robot Autom 1:652–657. https://doi.org/10.1109/robot.1991.131657
107. Suzumori K (1996) Elastic materials producing compliant robots. Rob Auton Syst 18:135–140. https://doi.org/10.1016/0921-8890(95)00078-X
108. BellowsTech (2021) Actuators. https://bellowstech.com/products/applications/actuators/. Accessed 8 Apr 2021
109. Kornbluh R, Pelrine R, Eckerle J, Joseph J (1998) Electrostrictive polymer artificial muscle actuators. Proc IEEE Int Conf Robot Autom 3:2147–2154. https://doi.org/10.1109/ROBOT.1998.680638
110. Ozkan M, Inoue K, Negishi K, Yamanaka T (2000) Defining a neural network controller structure for a rubbertuator robot. Neural Netw 13:533–544. https://doi.org/10.1016/S0893-6080(00)00020-4
111. Honda (2021) Robot development history. https://global.honda/innovation/robotics/robot-development-history.html. Accessed 12 Apr 2021
112. Honda (2021) Honda robotics. https://global.honda/innovation/robotics/ASIMO.html. Accessed 12 Apr 2021
113. NASA (2021) Mars exploration rovers. https://mars.nasa.gov/mer/. Accessed 7 Apr 2021
114. Robots U (2017) Press kit. https://www.universal-robots.com/media/1800138/ur_press_kit_en.pdf. Accessed 11 Apr 2021
115. Institute of Robotics and Mechatronics (2021) DLR—Institute of Robotics and Mechatronics—history of the DLR LWR. https://www.dlr.de/rm/en/desktopdefault.aspx/tabid-12464/21732_read-44586/. Accessed 11 Apr 2021
116. Noritsugu T, Kubota M, Yoshimatsu S (2001) Development of pneumatic rotary soft actuator. Trans Jpn Soc Mech Eng C 66:2280–2285. https://doi.org/10.1299/kikaic.66.2280
117. Hannan MW, Walker ID (2001) The "elephant trunk" manipulator, design and implementation. IEEE/ASME Int Conf Adv Intell Mechatronics, AIM 1:14–19. https://doi.org/10.1109/aim.2001.936423
118. Trimmer BA, Takesian AE, Sweet BM, Rogers CB, Hake DC, Rogers DJ (2006) Caterpillar locomotion: a new model for soft-bodied climbing and burrowing robots. In: 7th international symposium on technology and the mine problem, California
119. Chen G, Pham MT, Redarce T (2009) Sensor-based guidance control of a continuum robot for a semi-autonomous colonoscopy. Rob Auton Syst 57:712–722. https://doi.org/10.1016/j.robot.2008.11.001
120. ABB (2021) YuMi®—IRB 14000 | Collaborative Robots. https://new.abb.com/products/robotics/collaborative-robots/irb-14000-yumi. Accessed 12 Apr 2021
121. ABB (2021) ABB launches next generation cobots to unlock automation for new sectors and first-time users. https://new.abb.com/news/detail/74784/abb-launches-next-generation-cobots-to-unlock-automation-for-new-sectors-and-first-time-users. Accessed 12 Apr 2021
122. KUKA AG (2021) KMR iiwa. https://www.kuka.com/en-gb/products/mobility/mobile-robots/kmr-iiwa. Accessed 12 Apr 2021

123. KUKA AG (2021) Maintenance mode. https://static.kuka.com/?status=#/en-gb/products/mobility/mobile-robots/kmr-iiwa; https://www.dlr.de/rm/en/desktopdefault.aspx/tabid-12464/21732_read-44586/. Accessed 12 Apr 2021
124. Stäubli International AG (2021) HelMo the mobile robot system. https://www.staubli.com/en-in/robotics/product-range/helmo-mobile-robot-system/. Accessed 12 Apr 2021
125. Robots U (2017) Press kit. https://www.universal-robots.com/media/downloads/. Accessed 12 Apr 2021
126. Faraj Z, Selamet M, Morales C, Torres P, Hossain M, Chen B et al (2021) Facially expressive humanoid robotic face. HardwareX 9:e00117. https://doi.org/10.1016/j.ohx.2020.e00117
127. Prescott TJ, Robillard JM (2021) Are friends electric? The benefits and risks of human-robot relationships. IScience 24:101993. https://doi.org/10.1016/j.isci.2020.101993
128. Menezes B (2015) Meet Manav, India's first 3D-printed humanoid robot. Mint. https://www.livemint.com/Industry/rc86Iu7h3rb44087oDts1H/Meet-Manav-Indias-first-3Dprinted-humanoid-robot.html. Accessed 8 Apr 2021
129. Alemzadeh K, Jones SB, Davies M, West N (2021) Development of a chewing robot with built-in humanoid jaws to simulate mastication to quantify robotic agents release from chewing gums compared to human participants. IEEE Trans Biomed Eng 68:492–504. https://doi.org/10.1109/TBME.2020.3005863
130. Miyashita S, Guitron S, Yoshida K, Li S, Damian DD, Rus D (2016) Ingestible, controllable, and degradable origami robot for patching stomach wounds. In: Proc.—IEEE int. conf. robot. autom., pp 909–916. https://doi.org/10.1109/ICRA.2016.7487222
131. Xu H, Medina-Sánchez M, Magdanz V, Schwarz L, Hebenstreit F, Schmidt OG (2018) Sperm-hybrid micromotor for targeted drug delivery. ACS Nano 12:327–337. https://doi.org/10.1021/acsnano.7b06398
132. Lu H, Zhang M, Yang Y, Huang Q, Fukuda T, Wang Z et al (2018) A bioinspired multilegged soft millirobot that functions in both dry and wet conditions. Nat Commun 9:3944–3951. https://doi.org/10.1038/s41467-018-06491-9
133. Beccani M, Di Natali C, Aiello G, Benjamin C, Susilo E, Valdastri P (2015) A Magnetic drug delivery Capsule based on a coil actuation mechanism. Proc Eng 120:53–56. https://doi.org/10.1016/j.proeng.2015.08.564
134. Cianchetti M, Calisti M, Margheri L, Kuba M, Laschi C (2015) Bioinspired locomotion and grasping in water: the soft eight-arm OCTOPUS robot. Bioinspir Biomim 10:035003. https://doi.org/10.1088/1748-3190/10/3/035003
135. Tech Briefs TV (2021) Octobot: first autonomous, untethered, entirely soft robot—tech briefs. https://www.techbriefs.com/component/content/article/tb/tv/31492. Accessed 14 Apr 2021

Computer-Aided Formulation Development

3

Rania M. Hathout and Vikas Anand Saharan

Contents

Abstract

Quality, safety, and efficacy are all vital components for market success of a pharmaceutical product. The traditional concept of 'quality by QC' has now being transformed to the concept of 'Quality by Design' (QbD) with the adoption of QbD by most of the drug regulatory agencies across the globe. QbD ensures systematic adoption of certain approaches, such as DoE, risk assessment, and process analytical technology, for the development of pharmaceutical products to meet the stringent set targets of quality. In this context, DoE-based optimization

R. M. Hathout (✉)
Department of Pharmaceutics and Industrial Pharmacy, Faculty of Pharmacy, Ain Shams University, Cairo, Egypt
e-mail: rania.hathout@pharma.asu.edu.eg

V. A. Saharan
Department of Pharmaceutics, School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Dehradun, Uttarakhand, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_3

appears as a key player for ensuring quality of pharmaceutical products. Design of experiments (DoE) involves the selection of an appropriate design that can navigate the complete or the maximum possible space of the experimental runs and captures the correct mathematical relationships between the formulation or process inputs and the final quality parameters of a pharmaceutical product. Accordingly, an optimized pharmaceutical product possessing desired properties and high quality can be produced. This chapter addresses the role of computer-based experimental designs in the optimization and development of pharmaceutical formulations.

Keywords

QbD · Quality · Mathematical · Optimization · Pharmaceutical · Drug · Formulation · Formulation

Chapter Objectives

- Understanding the concept of optimization
- Shedding light on the "Design of Experiments" approach as an essential element of optimization
- Comparing different statistical designs and their utilities
- Gaining information on the computer-based optimization of pharmaceutical emulsions and microemulsions

3.1 Introduction

The concepts of optimization and quality by design (QbD) were introduced as holistic approaches in order to reach a high-quality and a robust pharmaceutical end product. Accordingly, the "computer-aided pharmaceutical development" approach is embraced by pharmaceutical industries to implement the aforementioned concepts using computer software and advanced statistical experimental designs. Previously only a limited number of factors were studied by taking one factor at a time (OFAT) to optimize the pharmaceutical formulations, which was time intensive and cumbersome. Figure 3.1 demonstrates an example of OFAT experiments that were used to study the preliminary factors affecting the preparation of the chitosan-sodium tripolyphosphate nanoparticles, where three factors: the acetic acid concentration, the stirring rate, and the chitosan:TPP volume ratio, were studied [1]. Two factors should be kept constant while varying the third as depicted by the Fig. 3.1. Accordingly, a lot of experiments were required to be performed, though covering a limited experimental design space. With the advent of novel and more complicated drug delivery systems, such as the vesicular systems [2], microemulsions [3], and lipidic, proteinaceous, and polymeric nanoparticles and platforms [4–7], the use of computerized optimization techniques becomes highly warranted.

The selection of the optimized formulation of a certain type of a drug carrier is a main objective of QbD, wherein the role and impacts of different critical material

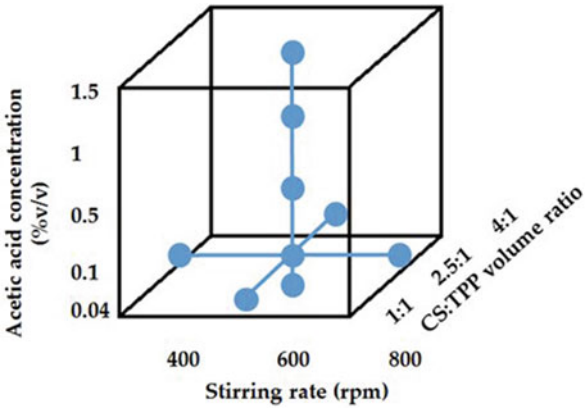


Fig. 3.1 OFAT experiments to investigate the effect of three factors on the preparation of chitosan (CS)-tripolyphosphate (TPP) nanoparticles

attributes (CMAs) and critical process parameters (CPPs) on CQAs are rigorously analyzed aiming for the reduction of manufacturing variability, controlling the manufacturing cost, and improving the safety and the quality of the developed formulation [8]. DoE is one significant element of QbD which ensure quality of the finished products of pharmaceutical industries. Drug regulatory agencies, like ICH, FDA, EMA, etc., across the globe, rely on QbD and DoE concepts for assuring quality of developed products.

DoE involves a statistical planning of experiments through appropriate selection of independent and dependent variables, subsequent analysis of impact of variables on the quality of product/process, and generation of suitable mathematical models, which may be subsequently used for predicting interactions among variables and the quality of the product. DoE is a computational effort to find cause-and-effect relationships among independent variables (composition of formulation, quality of key ingredients, process parameters, etc.) and dependent factors (response and quality attributes of product/process) with the help of software tools and/or certain algorithms. DoE brings out more knowledge (design space) about the product/process such as the quality of raw materials or input process parameter and the interaction among these. This increased knowledge about design space is used to fine-tune influential independent variables to achieve optimized product/process. DoE is used for three different purposes, namely screening, optimization, and robustness testing and, thereby, designs also derive their names from these purposes. Screening designs are used for identifying most influential factors out of large number of factors. Optimization designs allow investigating best combination of factors for one or more optimal response. Three-dimensional contour plots or contour lines on two dimensions are graphical representations (response surfaces) of mathematical relationships among factors and responses. Mathematical models of optimization are algebraic expressions that define relationships among the responses and the factors. Mathematical models and response surface are obtained nowadays with the use of specially designed software tools. Robustness testing of an optimization involves the assessment of the optimized formulation/process for ensuring

response (quality) within specification, that is, as per the predefined objectives of the optimization process. Robustness testing successfully allows the use of developed models and response surfaces for predicting the responses (quality) of the product/process.

3.2 The Concept of Optimization and Optimization Parameters

In its simplest form, the concept of optimization can be regarded as finding the "best available" values of an objective function given a defined domain (or input), including a variety of different types of objective functions and different types of domains. More specifically, the mathematical optimization or mathematical programming is the selection of a best element, with regards to some criterion, from a set of available alternatives.

The optimization parameters usually involve two main classes; one for the inputs which are the causes or independent factors of the materials represented by formulation parameters and the critical process parameters (CPPs). On the other hand, the outputs comprising the effects or responses or critical quality attributes (CQAs) are represented by evaluation parameters either for the formulation or the process [7]. Figure 3.2 summarizes these parameters.

Usually, the statistical designs used for optimization deal with parameters such as the factors and levels. The factors are the independent variables affecting a certain response (dependent variable). The factors can vary over different values. Statistically, these values are usually named as levels. For example, a pH factor can be varied over three levels (values): 3, 5, and 7. The effects of the studied factors

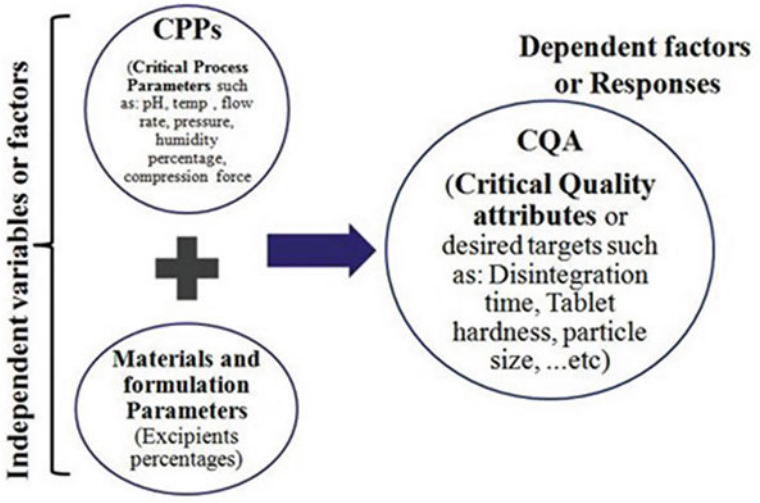


Fig. 3.2 The main optimization parameters in pharmaceutical development

independently on the investigated response are called the main effects (with no synergism or antagonism). On the other hand, the factors interactions are the synergistic or antagonistic obtained effects due to the combined action of the two or more factors together.

3.3 Statistical Experimental Designs

Statistical experimental designs may be grouped into four different types based on their application in formulation optimization. These are screening designs, optimization designs, and mixture designs. A fourth group of statistical designs are full factorial designs and fractional factorial designs than can be used for both screening and optimization. Software packages like Design-Expert, Minitab, JMP, XLSTAT, etc. are widely used for experimental designing for screening of factors, formulation optimization, and prediction of responses through developed models. Table 2.2 in Chap. 2 may be referred for brief features details on these software tools.

3.3.1 Screening Designs

Screening designs are efficient means to segregate most influential factors among a large number of potential factors. These designs are primarily used to study the effect of influential or significant factors, rather than the effect of interaction of factors. They are considered low-resolution designs (III or IV). Among these designs are full factorial designs, fractional factorial designs, Taguchi designs (arrays), and Plackett-Burman designs. The main advantage of using a screening design is to generate minimum experimental runs to find out the most influential factors, which ultimately helps in saving resources and time [9].

3.3.1.1 Taguchi Designs

These designs are named after a Japanese engineer, Genichi Taguchi, who was the first to introduce these screening methods [10]. It is a method that is normally used for implementing good performance in the designing phase of drugs products or manufacturing. This design mainly depends on the use of orthogonal arrays, which provide a set of minimum experiments that serve as objective functions for optimization. The Taguchi design starts with a minimum of three factors and produces four experimental runs at two levels lacking any center point runs [11]. They are also referred to as "off-line quality control" experimental screening designs. Figure 3.3 displays the choice of points (runs) in Taguchi designs. As demerits of these designs, these arrays were criticized for being inefficient and often ineffective [9]. Table 3.1 displays the actual experimental runs according to three factors and three levels.

3.3.1.2 Plackett-Burman Designs

Plackett-Burman designs (PBDs) are useful for investigating huge number of potentially influential factors. PBDs identify main effects and generate minimal experimental runs. They are also referred to as "saturated designs". They are characterized

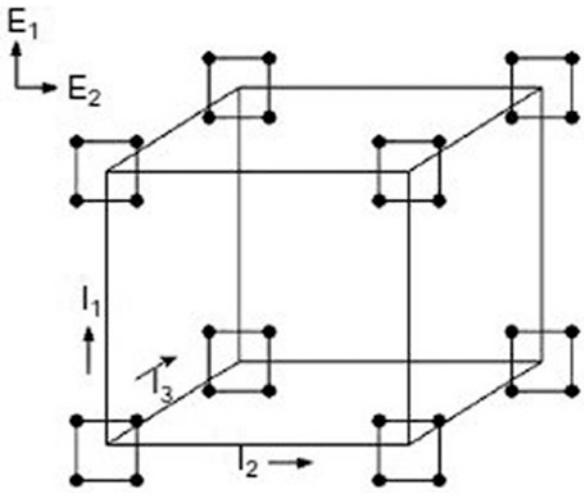


Fig. 3.3 The Taguchi arrays design

Table 3.1 Experimental runs in a three-factor simple Taguchi array

	Control factors and levels		
Experiments	A	B	C
1	1	1	1
2	1	2	2
3	1	3	3
4	2	1	2
5	2	2	3
6	2	3	1
7	3	1	3
8	3	2	1
9	3	3	2

Table 3.2 Experimental runs in Plackett-Burman design

	Factors						
Experiments	A	B	C	D	E	F	G
1	1	1	1	−1	1	−1	−1
2	−1	1	1	1	−1	1	−1
3	−1	−1	1	1	1	−1	1
4	1	−1	−1	1	1	1	−1
5	−1	1	−1	−1	1	1	1
6	1	−1	1	−1	−1	1	1
7	1	1	−1	1	−1	−1	1
8	−1	−1	−1	−1	−1	−1	−1

by the lack of the two-factor interactions. PBDs are also considered as a special two-level fractional factorial designs that are used generally for the screening *of* *f* factors, that is, $N - 1$ factors, where N is a multiple of 4. PBDs start with a minimum of seven factors where they produce a total of eight experimental runs

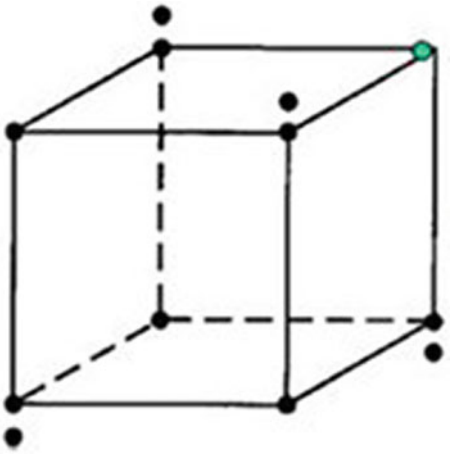


Fig. 3.4 The Plackett-Burman design

Table 3.3 Experimental runs for a 2^3 full factorial design

	Factors		
Experiments	A	B	C
1	−1	−1	−1
2	1	−1	−1
3	−1	1	−1
4	1	1	−1
5	−1	−1	1
6	1	−1	1
7	−1	1	1
8	1	1	1

(Table 3.2) without the employment of any center point runs [11]. PBDs are applied for screening a very large number of factors, where even the fractional factorial designs lead to a huge number of experimental runs [9].

Their disadvantages include their complexity and the presence of confounding (perplexing) effects due to the few numbers of experiments. Figure 3.4 demonstrates the points' selection on PBDs.

3.3.2 Full Factorial and Fractional Factorial Designs

A full factorial design is a one that starts with two or more factors and takes in its considerations all the possible combinations of the factors and their all available levels. These designs can be represented by 2^k, where k is number of factors. More generally, the total number of runs equals the number of levels raised to a power representing the number of the factors; l^k, where l is the number of levels (Table 3.3). A full factorial design is sometimes named as the "fully crossed design" as it allows the study of effect of each factor on response. It usually leads to a polynomial

Table 3.4 Determination of factor interactions in a 2^3 full factorial design

Experiments	Contrast coefficients			
	AB	AC	BC	ABC
1	1	1	1	−1
2	−1	−1	1	1
3	−1	1	−1	1
4	1	−1	−1	−1
5	1	−1	−1	1
6	−1	1	−1	−1
7	−1	−1	1	−1
8	1	1	1	1

Table 3.5 Experimental runs of fractional factorial design (2^{4-1})

Experiments			Factors	
	A	B	C	D
1	−1	−1	−1	−1
2	1	−1	−1	1
3	−1	1	−1	1
4	1	1	−1	−1
5	−1	−1	1	1
6	1	−1	1	−1
7	−1	1	1	−1
8	1	1	1	1

mathematical model. Full factorial designs contain center and corner points and capture the effects of each factor on the response and allow study of interactions of factors (Table 3.4) [12]. With less number of factors and their corresponding levels (e.g., 2 or 3), full factorial designs are recommended as screening designs. More number of factors and levels result in vast number of experimental runs that become practically unfeasible [9].

For full factorial design, factor effects are used to build the regression model given in Eq. 3.1, where y is the response (dependent variable), β_0 the intercept, and β_i the regression coefficients (regression coefficients stand for factor effects). Interactions are calculated from the columns of contrast coefficients.

$$y = \beta_0 + \sum_{i=1}^{f} \beta_i x_i \tag{3.1}$$

With a large number of factors and/or levels, a huge number of experimental runs required as per full factorial designs are significantly reduced by adopting fractional factorial designs, which are a finite fraction of full factorial design. Experimental runs for fractional factorial designs may be calculated with l^{k-v}, where a number (v) is subtracted from the power of levels or the factor (k) of the full factorial design runs equation in order to reduce the number of performed experiments. l^{-v} $(1/l^v)$ represents the fraction of experiments from the full factorial designs that are omitted (*if* $v = 1$ *and* $l = 2$, the design is ½, i.e., *half of full factorial design*) (Table 3.5).

However, the disadvantages include bad resolutions, difficulty to construct, and the effects being confounded by the interactions. They sometimes also lack any center point [11].

3.3.3 Response Surface Designs

Response surface designs provide extensive information of the factors effects together with their interactions on the investigated responses and draw conclusions covering the whole experimental design space. Generally, the number of factors is restricted to two or three. However, the levels are three at the minimum. Qualitative factors (discrete) cannot be employed in response surface designs. Main purpose of using these designs is to optimize the formulation or the process. Response surface designs can have symmetrical or asymmetrical shape. The central point is often replicated several [3–5] times to determine experimental error. Some of the most often used symmetrical designs are three-level full factorial, central composite design, Box-Behnken design (BBD). Asymmetrical designs are represented by D-optimal design.

The regression model (for two-factor studies) for response surface designs is depicted in Eq. 3.2, where y represents the response, x_i and x_j are two factors, β_0 is the intercept, β_i, β_{ij}, *and* β_{ii} are the main coefficient, interaction coefficient of x_i and x_j factors the quadratic coefficient for two factors, respectively [13].

$$y = \beta_0 + \sum_{i=1}^{f} \beta_i x_i + \sum_{1 \le i < j}^{f} \beta_{ij} x_i x_j + \sum_{i=1}^{f} \beta_{ii} x_i^2 \tag{3.2}$$

Interactions among more than two factors are generally ignored, and accordingly, the terms which are nonsignificant in a regression model are removed after statistical analysis.

3.3.3.1 Three-Level Factorial Design

The three-level factorial design is also considered a response surface design, where it can provide a complete picture of the whole space of an experiment involving three different levels (values) for any two or more factors. Figure 3.5 depicts the experimental points of an experiment comprised of 3 factors, each varied over 3 levels to give a total of 27 experiments. It is usually avoided due to the higher number of experimental runs.

3.3.3.2 Central Composite Designs

Central composite designs (CCD) comprise either a full factorial or fractional factorial design. These are also known as Box and Wilson designs. The points of full factorial design are at ± 1 level, center point at 0, and points of star design at the level of $\pm\alpha$, where $|\alpha| \geq 1$ [14, 15]. The value of α depends on the required precision of the estimation of the surface and the calculation possibilities. Precision of the

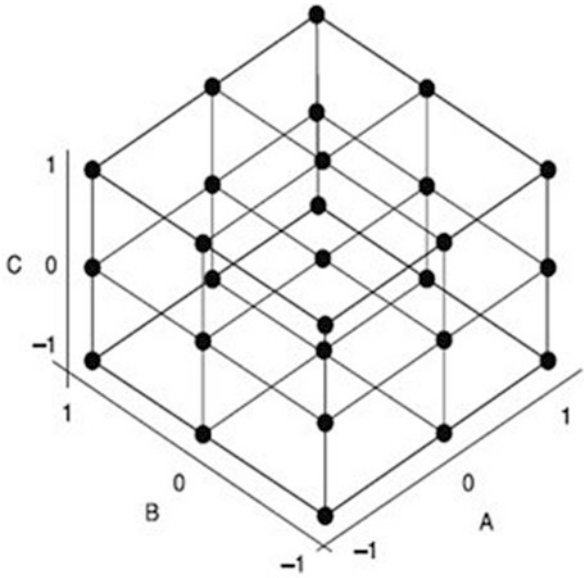


Fig. 3.5 Three-level full factorial design for 3 factors, 3^3 design, with 27 experiments

Table 3.6 Experimental runs required for a CCD design for three factors

		Factors	
Experiments	A	B	C
1	−1	−1	−1
2	1	−1	−1
3	−1	1	−1
4	1	1	−1
5	−1	−1	1
6	1	−1	1
7	−1	1	1
8	1	1	1
9	$-\alpha$	0	0
10	$+\alpha$	0	0
11	0	$-\alpha$	0
12	0	$+\alpha$	0
13	0	0	$-\alpha$
14	0	0	$+\alpha$
15 (+ replicates)	0	0	0

estimation is affected with the number of trials at the center point and the value of α. The star points, which lie outside the experimental domain, and the center point make it possible to assess the curvature of the response surface. CCD for three factors consist of a two-level full factorial design (2^k experiment), a center point (1 experiment) and a star design ($2f$ experiment) [13]. Hence, $N = 2^k + 2k + 1$ experiments are required to investigate k factors (Table 3.6).

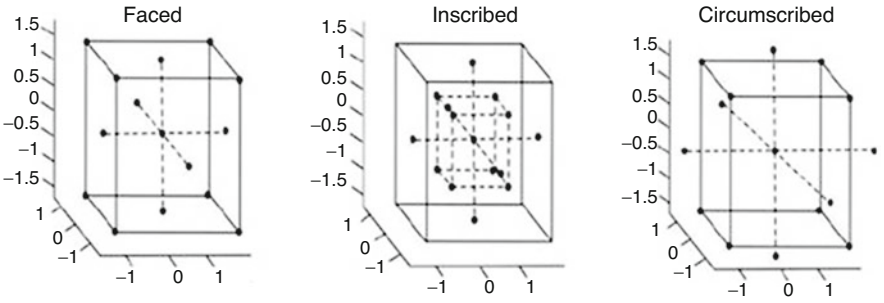


Fig. 3.6 FCCD/CCF, CCI, and designs for two factors

There are three types of CCD designs (Fig. 3.6), namely central composite circumscribed (ccc), face-centered composite/central composite face centered (FCCD/CCF), and central composite inscribed (CCI) [14]. In CCC design, the star points lie outside the experimental domain (distance α). CCI design requires minimum five levels to be studied for each factor. In FCCD/CCF, the star points are present on the faces of the experimental domain (i.e., $\alpha = \pm 1$). Three levels per factors are required for FCCD. CCI is a simply reduced CCC design to fit in the experimental domain. This plan also requires five levels per factor.

CCD is usually recommended when a curvature in the data is recognized and, hence, a quadratic model will best represent the experimental space. It is a highly accurate and robust design. However, it is unsuitable for delicate structures where the increase in the experimental space by the star points would lead to loss of structures or losing the whole system.

3.3.3.3 The Box-Behnken Design

The Box-Behnken design (BBD) is used for designing experiments with greater than three or more factors and the optimum is expected to be somewhere in the middle of the range of factors (Fig. 3.7). For three factors, the number of experiments may be calculated from $N = (2k(k-1)) + c_0$ experiments, where k is the number of factors and c_0 is the number of center points [16]. BBD is regarded as one of the most common substitute to the CCD.

BBD for three factors is a second-order design and based on three-level incomplete factorial design [13]. The first four experiments from the Table 3.7 represent a full 2^2 design for factors A and B, keeping the third factor C at 0 level, that is, constant level. Similarly, experiments from 5 to 8 represent a full 2^2 design for factors A and C, keeping factor B at level 0. The third block of experiments from 9 to 12 represents a full 2^2 design for factors B and C and, here, factor A is at level 0 (Tables 3.7 and 3.8). In a similar manner, in a BBD for 5 or more factors, Box-Behnken uses full 2^3 designs and keeping rest two factors at constant level.

BBD usually takes factors at three levels and all the design points fall within what is called the safe operating zone, where (in case of three factors), in each run, the levels of two factors are held at their extreme levels while the third is either held at its

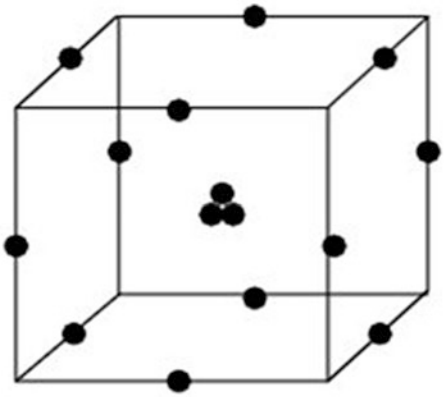


Fig. 3.7 The Box-Behnken design (BBD)

Table 3.7 Experiments in a BBD design for three factors, with center point in triplicate

		Factors	
Experiments	A	B	C
1	−1	−1	0
2	1	−1	0
3	−1	1	0
4	1	1	0
5	−1	0	−1
6	1	0	−1
7	−1	0	1
8	1	0	1
9	0	−1	−1
10	0	1	−1
11	0	−1	1
12	0	1	1
13	0	0	0
14	0	0	0
15	0	0	0

Table 3.8 A simplified representation of experiments of BBD design for three factors

		Factors	
Experiments	A	B	C
1–4	±1	±1	0
5–8	±1	0	±1
9–12	0	±1	±1
Replicates	0	0	0

middle or low level, thereby eliminating the experimental runs performed with all the highest or lowest conditions at the same time. This design is useful in case of delicate systems, where it is unfavorable to expose these systems to extremities of harsh conditions (highest or lowest levels of the studied factors) all at the same instances such as the liposomes or proteins that can be easily denatured and lose their tertiary structures. However, this implies poor coverage due to the lack of corner points. Hence, BBD may introduce errors when extrapolated to extremes of the design space [1].

3.3.3.4 The Doehlert Design

Doehlert (uniform shell) designs have uniform distances from all neighboring design points [13]. Factors are varied at different number of levels like one factor can have five levels while the other factors can be varied at three levels. This provides flexibility to the experimenter to choose the number of levels for each factor based on the desired experimental space and the nature of factors (Table 3.9). Experiments as per Doehlert design for two and three factors are shown in Tables 3.10 and 3.11, respectively. This design has spherical experimental domain and uniform distances among design points allow uniform space filling. The total number of experiments in

Table 3.9 Doehlert design matrix for two factors

	Factors	
Experiments	A	B
1	0	0
2	1	0
3	0.5	0.866
4	−1	0
5	−0.5	−0.866
6	0.5	−0.866
7	−0.5	0.866

Table 3.10 Doehlert design matrix for three factors

		Factors	
Experiments	A	B	C
1	0	0	0
2	1	0	0
3	0.5	0.866	0
4	0.5	0.289	0.817
5	−1	0	0
6	−0.5	−0.866	0
7	−0.5	−0.289	−0.817
8	0.5	−0.866	0
9	0.5	−0.289	−0.817
10	−0.5	0.866	0
11	0	0.577	−0.817
12	−0.5	0.289	0.817
13	0	−0.577	0.817

Table 3.11 Some published researches on computer-based optimization of emulsifying/microemulsifying systems

Drug/formulation	Optimization design	Factors/ levels	Used software	Developed mathematical model	How optimized formulation achieved	Reference
Doxycycline HCl/emulsions	Full-factorial	3/3	Design Expert®	Quadratic	–	[24]
Doxycycline HCl/emulsions	D-optimal	3/3	Design Expert®	Quadratic	Desirability function	[24]
Methyl dihydrojasmonate/ microemulsions	Simplex lattice mixture	3/2	Design Expert®	Quadratic	Graphical optimization	[27]
Aspirin/nanoemulsion	Central composite	6/5	Design Expert®	Linear	Numerical and graphical optimization using the "optimization" function of the software	[32]
Telmisartan/SMEDDS	Box-Behnken	3/3	Design Expert®	Quadratic	Desirability function	[33]
Ketoprofen/ microemulsions	Simplex lattice	3/2	Design Expert®	Quadratic	Desirability function	[34]
Irbesartan/SNEDDS	Full factorial	2/3	Design Expert®	Quadratic	Principal component analysis was used to determine the main dependent variables followed by numerical optimization	[35]
HL235 (cathepsin K inhibitor)/SMEDDS	D-optimal mixture	3/2	Design Expert®	Quadratic	Desirability function	[36]
Brazilian red propolis benzophenones/ nanoemulsions	Box-Behnken	3/3	Minitab®	Quadratic	Mathematically using the Minitab® software	[37]
Lycopene/emulsions	Central-composite	3/5	Design Expert®	Quadratic	Graphical and numerical optimization	[38]
Sparfloxacin/ microemulsions	Simplex lattice	3/2	Minitab®	Cubic	Desirability function	[39]

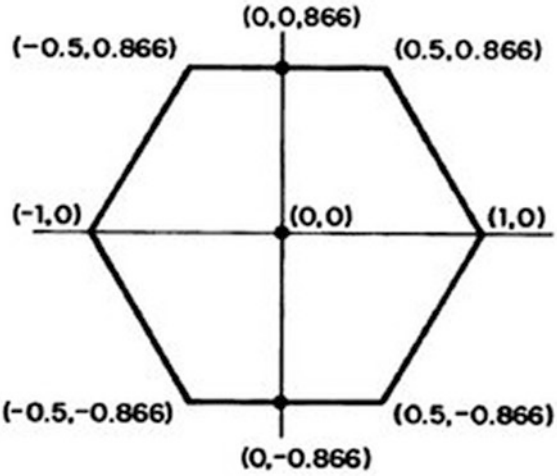


Fig. 3.8 Doehlert design

a Doehlert design is calculated from $k^2 + k + C_0$, where k is the number of factors and C_0 is the number of center points [17].

In case of two factors, the design takes the shape of a regular hexagon with one center point and rest all design points the perimeter of the circle (Fig. 3.8). Factor which has stronger effect shall be studied at more levels (5 in this case) to obtain more information from this design. Doehlert matrix and BBD are very efficient in comparison to 3^k full factorial designs, while they have slightly better efficiency than the CCD design. In statistics, an efficient experimental design is one that produces your desired experimental results with the minimum amount of resources (number of runs). Relative efficiency of a design may be estimated from Eq. 3.3 [13].

$$R_{\text{eff}} = \frac{P}{N} \leq 1 \tag{3.3}$$

where P is the number of estimated coefficients in the estimated model divided by N, which is the number of runs.

3.3.3.5 The D-Optimal Design

D-optimal designs (DODs) are asymmetrical designs that are used for investigating asymmetrical experimental domains [13]. DODs are custom-made designs, which vary from case to case. These computer-generated designs allow more flexibility in the specifications of the each problem and especially useful in cases when there is a constraint in a region and no other classical design is applicable.

DODs are used for irregular experimental regions, screening of qualitative factors at multiple levels, optimization designs with qualitative factors, when less number of experimental runs are required than other classical designs, combining mixture and process factors in a single optimization plan, and updating of model.

The optimal designs are also used for investigating three-factor experiments. The D-optimal is the most commonly used of this category. The DOD is a

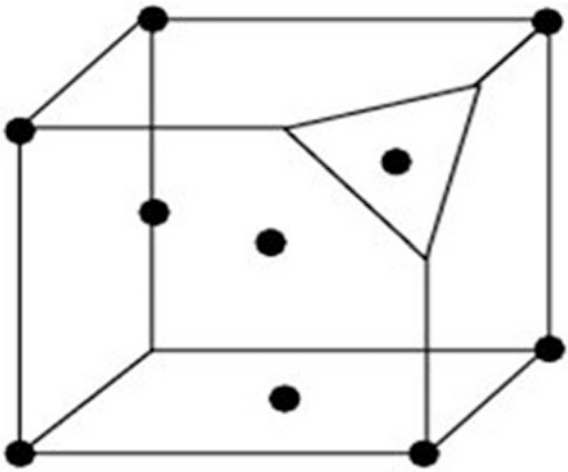


Fig. 3.9 The D-optimal design (BBD)

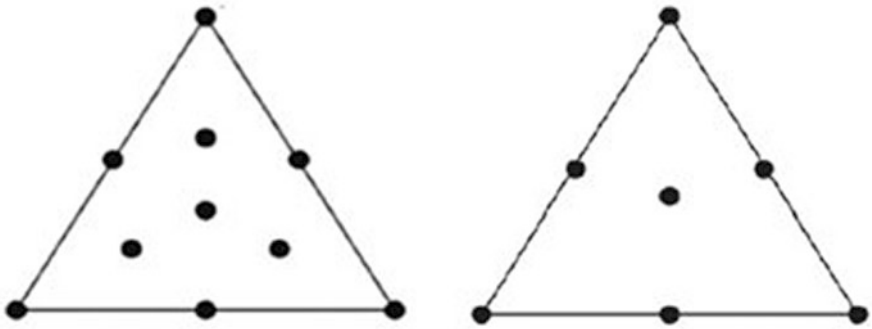


Fig. 3.10 Possible points selection in mixture designs

computer-aided design that is composed of the best subset of all possible experiments. It works by space filling sampling. Briefly, the design works by maximizing the determinant obtained from the information matrix generated from all the possible combinations of the involved factors. This explains the use of the letter "D" in the D-optimal designs. The design points can be corner, central, axial, or in any position in the experimental domain (Fig. 3.9). The D-optimal was found to be highly sensitive and more predictive with the least percentage bias [1, 18].

3.3.4 Mixture Designs

Mixture designs are used to study mixture variables such as quantity (%) of excipients in a formulation (Fig. 3.10). All mixture components are examined in one design provided that the design factors (mixture variables) represent components

that make a total of 100% (i.e., constitute a mixture) [13]. In a mixture, the quantity of one component cannot be altered independently. Unlike other unconstrained experimental designs, mixture designs cannot be represented or visualized as cubes, squares, or hypercubes. If there are k components in a mixture, the resulting mixture design will have $k - 1$ dimensions like three components can be represented in two dimensions with a triangle. In this context, four components will require representation in three dimensions within tetrahedron. With increase in components, the data analysis becomes more complicated because the mixture factors are correlated among themselves.

When constraints are defined for one or more components with regard to the amount/%wt., the experimental points do not form the part of a triangle/tetrahedron, in cases of three or four component formulations. With such constraints, regular mixture designs are not applicable, rather experimental domains of different shape within the triangle/tetrahedron are selected. Such irregularity due to constraints is best addressed with DODs.

Examples of these designs include the simplex lattice, simplex centroid, and the D-optimal mixture design. The simplex centroid is the least accurate and its usage is now obsolete [19]. The simplex lattice design is usually used for areas comprising a triangular domain [20] while the D-optimal of the mixture type is used for irregular domains and lead to highly accurate results [21, 22]. These designs are specifically useful in optimizing microemulsions and other similar systems such as the surfactants-based vesicular systems [23] and lipid nanocapsules [21]. Figure 3.14 depicts different possible points combinations in the mixture designs.

3.4 Computer-Aided Pharmaceutical Formulation

By time, the use of computers in the development of various drug delivery systems and dosage forms is gaining more grounds. The applications vary from traditional delivery systems to the more advanced ones. In this section, an example of the development of a conventional delivery system, namely Emulsions, will be discussed followed by an example of the advanced form of this delivery system; the microemulsion system.

3.4.1 Development of Pharmaceutical Emulsions

The emulsions are liquid dispersed systems that consist of two immiscible liquids, mainly aqueous and oily, where one of them is dispersed in the other in presence of an emulsifier (surfactant or multimolecular adsorbed hydrophilic colloids or finely divided solids). They are usually present in two main types: water-in-oil (w/o) and oil-in-water (o/w) emulsions. These systems are usually characterized as thermodynamically unstable and they are subjected to several physical instability problems, such as creaming, flocculation, phase inversion, and coalescence. Many factors and conditions affect the stability of the emulsions; therefore, attempts for the

formulation and process optimization methods have never ceased. The use of computers and in silico approaches in the emulsions optimization has started by the end of the last century and still ongoing.

An example of optimizing pharmaceutical emulsions regarding their phase stability was carried out in 2014, where two designs, 3^3 factorial and D-optimal, were utilized to optimize emulsions to be used as the liquid domain in electrospinning [24].

A w/o emulsion thickened with polycaprolactone was prepared where it contained a drug: doxycyline HCl. Span 60 and sodium lauryl sulfate were used as the emulsifier blend. The goal of the study was to obtain a high-quality electrospinnable emulsion system that would allow for the development of desired hydrophilic drug-loaded nanofibers (Fig. 3.11). The studied factors were the ratio of organic:aqueous phase, composition of emulsifier blend, and the concentration of polymer. The critical quality attributes (CQAs or responses) to be optimized were the stability, the viscosity, and the conductivity. The full factorial design was utilized to optimize the stability while the D-optimal design was used for the other two CQAs.

The stability of emulsion phase was affected significantly by all three investigated independent factors at $p < 0.05$. The ratio of Span 60:SLS ($X1$) exerted the most obvious effect as demonstrated in Fig. 3.12.

Regarding the emulsion conductivity, it was also significantly affected by the three independent variables. Though the Span 60:SLS ratio had a prominent negative effect on conductivity, the Org:Aq ratio and polymer concentration led to a net positive effect with the latter being more obvious. The positive effects of $X2$ and $X3$ on conductivity can be attributed to the inherent effect of SLS in the organic polymer phase and due to the presence of doxycycline and SLS in the aqueous phase,

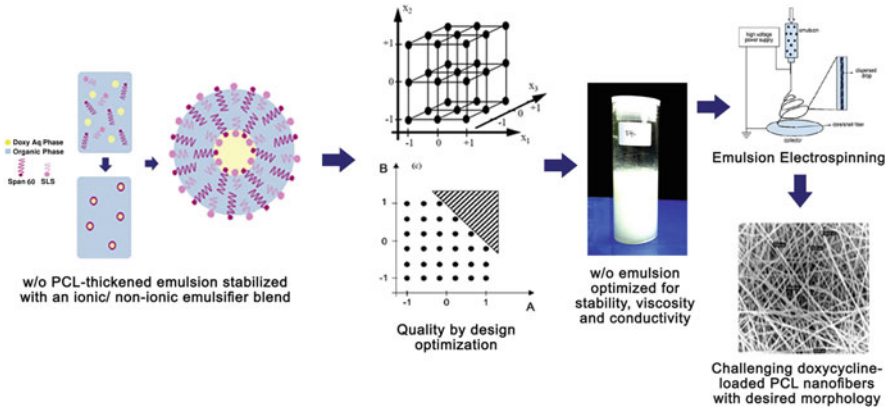


Fig. 3.11 Optimizing a polycaprolactone-thickened emulsion system suitable for electrospinning. Modified from Badawi and El-Khourdagui, 2014 [24], and reprinted under the RightsLink license number 5055960051761

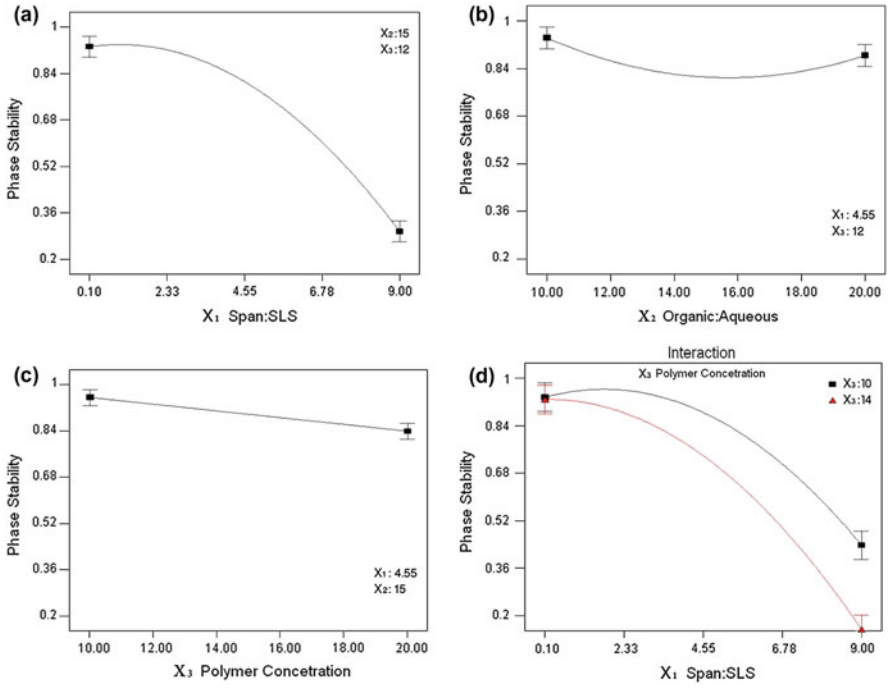


Fig. 3.12 Full factorial design generated one-factor plots of (**a**) the ratio of Span 60:SLS (*X*1), (**b**) the volume ratio of organic:aqueous phases (*X*2), (**c**) the concentration of polymer (*X*3) on stability of emulsion phase, and (**d**) interaction plot for factors *X*1 and *X*3. (Modified from Badawi and El-Khourdagui, 2014, and used with permission under the RightsLink license number 5055960051761)

respectively. As a conclusion, the concentration of sodium lauryl sulfate was the primary determinant of the emulsion system conductivity as depicted in Fig. 3.13.

As a conclusion from this study, a quadratic model was obtained as follows:

$$Y(\text{Emulsion conductivity}) = 3.36 - 0.78X1 + 0.068X2 + 0.4X3 - 0.40X1X2 - 0.41X1X3 - 0.39X2 \quad (3.4)$$

where *X*1 is the Span 60:SLS ratio, *X*2 organic:aqueous phase volume ratio, and *X*3 is polymer concentration.

Accordingly, an optimized formulation producing electrospinnable emulsions with the desired critical quality attributes possessed the following composition: Org:Aq phase volume ratio of 20:1, emulsifier blend in 3:1 ratio of Span 60:SLS (total amount 3.5% w/w), and polycaprolactone concentration of 12% w/v. The synthesized nanofibers produced with the help of this optimized emulsion formulation were defect-free, nanosized, continuous, and uniform.

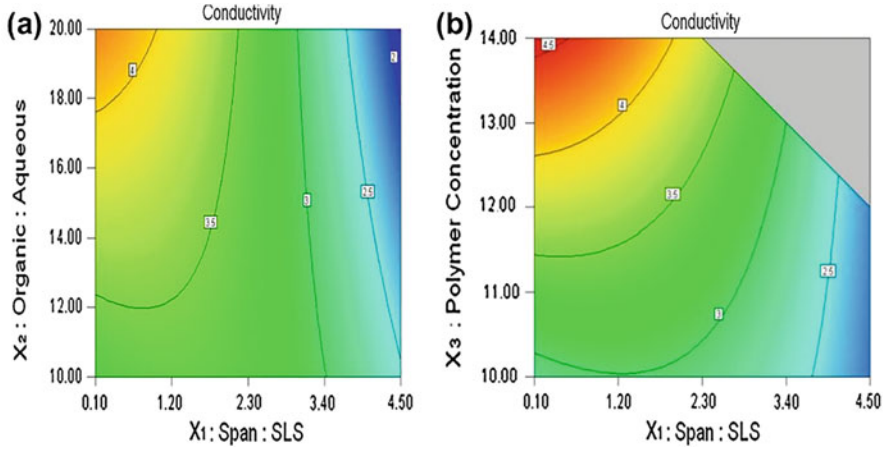


Fig. 3.13 D-optimal design generated 2D contour plots for (**a**) the interaction effect of (X_1) and (X_2) and (**b**) the interaction effect of (X_1) and (X_3), on the emulsion system conductivity (Y_2). Moving from blue to red colors indicates increase in the response (conductivity) as modified from Badawi and El-Khourdagui, 2014, and reprinted under the RightsLink license number 5055960051761

3.4.2 Development of Pharmaceutical Microemulsions

Microemulsions are one of the most powerful advanced drug delivery systems, specifically for the transdermal delivery of drugs. Unlike the conventional emulsions, they are thermodynamically stable, transparent, or translucent and, usually, comprise aqueous and oily phases, surfactant, and a cosurfactant to decrease the interfacial tension between the former immiscible phases to the lease minimum values [25, 26].

In a study conducted in 2017 [27], methyl dihydrojasmonate oil, as a potent anticancer agent, was integrated in the oily phase of transdermal microemulsion systems composed of MDHJ and Capryol 90® (propylene glycol caprylate) or oleic acid as an oily phase, Labrasol® (PEG-8 caprylic/capric glycerides) as a surfactant, Transcutol® (diethylene glycol monoethyl ether) as a cosurfactant, and water as the aqueous phase in order to target solid tumors such as the breast cancer. A schematic summary of the study is presented in Fig. 3.14.

Two pseudo-ternary phase diagrams, one for each of the two investigated microemulsion systems, were constructed using the water titration method (Fig. 3.15).

A triangular domain was recognized in each of the two microemulsion systems formation areas of the constructed pseudophase diagrams. Hence, a simplex lattice mixture design was selected to optimize the microemulsion formulations in these

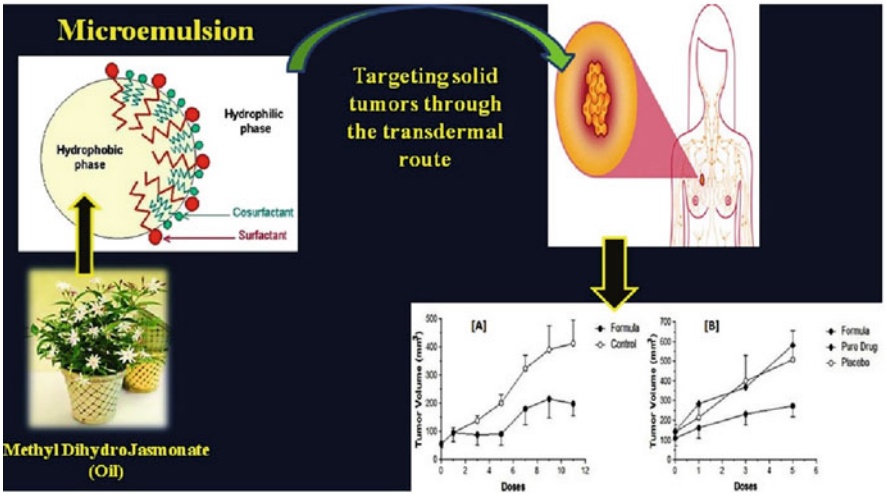


Fig. 3.14 Integrating an anticancer oil (methyl dihydrojasmonate) in the oily phase of transdermal microemulsion systems targeting solid tumors such as the breast cancer. Modified from Yehia et al., 2017, and reprinted under the RightsLink license number 5055961022282

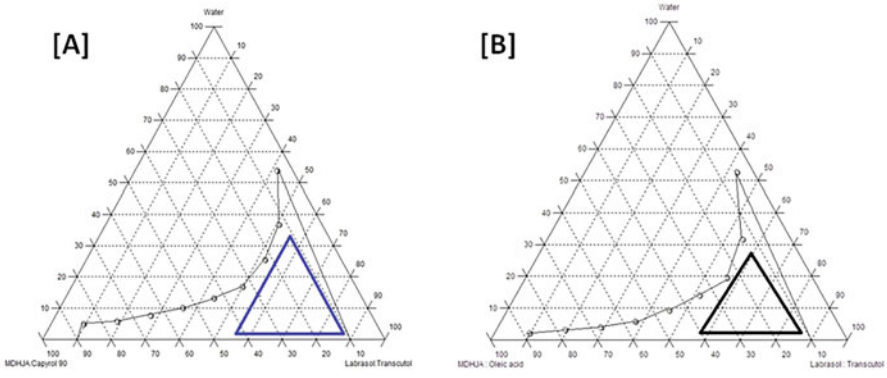


Fig. 3.15 Construction of pseudoternary phase diagrams of (**a**) methyl dihydrojasmonate/Capryol 90®/Labrasol®/Transcutol®/water and (**b**) methyl dihydrojasmonate/oleic acid/Labrasol®/Transcutol®/water. Modified from Yehia et al., 2017, and reprinted under the RightsLink license number 5055961022282

two areas with respect to droplet size. A smaller droplet size usually leads to better skin penetration and higher drug flux through the skin. The design proposed 13 formulations in each area and, subsequently, model analysis of their droplet size results was performed.

Two quadratic models were generated as follows for the two prepared systems, respectively:

$$\begin{aligned}\text{Logdropletsize}(\log(\text{D.S.})) = {} & +0.034241 \times \text{MDHJ} + \text{Capryol90} - 0.18019 \\ & \times \text{Water} + 0.021683 \times \text{Labrasol} + \text{Transcutol} - 2.81582\text{E} \\ & - 003 \times \text{MDHJ} + \text{Capryol90} \times \text{Water} + 3.35903\text{E} - 003 \\ & \times \text{Water} \times \text{Labrasol} + \text{Transcutlol} \end{aligned} \tag{3.5}$$

and

$$\begin{aligned}\text{Dropletsize} = {} & -32.72795 \times \text{MDHJ} + \text{Oleic} + 20.30559 \times \text{Water} \\ & - 1.7574 \times \text{Labrasol} + \text{Transcutol} - 0.33297 \times \text{MDHJ} + \text{Oleic} \\ & \times \text{Water} + 0.64790 \times \text{MDHJ} + \text{Oleic} \times \text{Labrasol} + \text{Transcutol} \\ & - 0.23852 \times \text{Water} \times \text{Labrasol} + \text{Transcutol} \end{aligned} \tag{3.6}$$

The two generated models possessed *r*-squared, adjusted *r*-squared (after removing the insignificant terms according to ANOVA), and predicted *r*-squared with values above 0.8 for the two models indicating a high fit-to-data model together with a high predicting power for the uncarried experiments [28, 29]. The adequate precision (indicating the signal-to-noise ratio) was 9.806 and 28.289 for the two models, respectively, implying sufficiency of the two models to navigate the whole spaces of the experiments and also confirming that the results originate from real differences between the formulations and are not random results [30, 31]. Figure 3.16 demonstrates the contour and 3D plots generated from the two obtained models of the two studied microemulsion systems.

A formulation (39% oleic acid: MDHJ, 1% water, and 60% Labrasol®:Transcutol P®) was selected and compared regarding cytotoxicity with the drug alone (MDHJ) on MCF-7 breast cancer cell line, where it scored an IC_{50} of 42.2 μL/mL equivalent to 8.3 μL/mL compared to 3.85 μL/mL corresponding to the IC_{50} of the pure drug (MDHJ). This result is very promising as the transdermal microemulsion system guarantees a controlled-release manner and solves the drug problems of poor solubility that consequently leads to low bioavailability. This optimized formulation scored the highest steady-state flux (J_{ss}) reaching 0.07 μL cm^{-2} h^{-1}. This high flux was reflected in the in vivo studies performed on Ehlrich carcinoma cells that were induced subcutaneously as a model solid tumor where the selected formulation applied transdermally scored a tumor inhibition percentage reaching 54% compared to the control.

The developed microemulsion systems were introduced as successful delivery carriers for MDHJ that could be applied through a transdermal patch on a solid breast carcinomas to compliment any chemotherapeutic or radiological therapy.

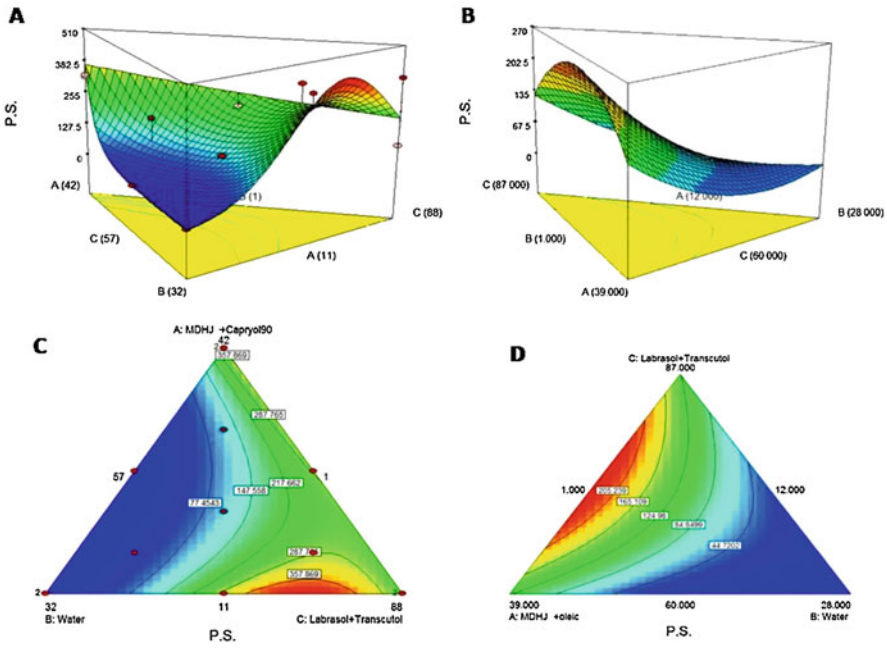


Fig. 3.16 3D surface plots (**a**, **b**) and contour plots (**c**, **d**) generated according to the simplex lattice model of (**a**) methyl dihydrojasmonate/Capryol 90®/Labrasol®/Transcutol®/water and (**b**) methyl dihydrojasmonate/Oleic acid/Labrasol®/Transcutol®/water microemulsion systems, respectively. Moving from blue to red colors indicate an increase in the response (particle size). Modified from Yehia et al., 2017, and reprinted under the RightsLink license number 5055961022282

Other attempts were also conducted to optimize emulsions and microemulsions systems (and their similar drug delivery systems such as the self-emulsifying and self-microemulsifying systems) as summarized in Table 3.11.

3.5 Conclusions

There is a high and urgent demand for the use of computers and computer-aided approaches in optimizing pharmaceutical formulations and modern industrial processes. The implementation of computer-aided DoE and optimization would save the huge resources of formulation scientists, experimentation time and efforts, and would lead to better formulations and, hence, better in vitro quality which will subsequently improve in vivo efficacy. In this context, computers software are becoming indispensable tools for formulation scientists just as the test tubes and the other glassware are needed for wet-lab experimentation. With good prediction power and robust optimized models, lot of wet-lab experiments may be avoided.

References

1. Abdel-Hafez SM, Hathout RM, Sammour OA (2014) Towards better modeling of chitosan nanoparticles production: screening different factors and comparing two experimental designs. Int J Biol Macromol 64:334–340. https://doi.org/10.1016/j.ijbiomac.2013.11.041
2. Guinedi AS, Mortada ND, Mansour S, Hathout RM (2005) Preparation and evaluation of reverse-phase evaporation and multilamellar niosomes as ophthalmic carriers of acetazolamide. Int J Pharm 306:71–82. https://doi.org/10.1016/j.ijpharm.2005.09.023
3. Metwally AA, Hathout RM (2015) Replacing microemulsion formulations experimental solubility studies with in-silico methods comprising molecular dynamics and docking experiments. Chem Eng Res Des 104:453–456. https://doi.org/10.1016/j.cherd.2015.09.003
4. Kunjan P, Marika N, Rupanjali BS, Hemanta KS (2016) Nanosized drug delivery systems for direct nose to brain targeting: a review. Recent Pat Drug Deliv Formul 10:156–164. https://doi.org/10.2174/1872211310666160321123936
5. Abozeid SM, Hathout RM, Abou-Aisha K (2016) Silencing of the metastasis-linked gene, AEG-1, using siRNA-loaded cholamine surface-modified gelatin nanoparticles in the breast carcinoma cell line MCF-7. Colloids Surf B Biointerfaces 145:607–616. https://doi.org/10.1016/j.colsurfb.2016.05.066
6. Ossama M, Hathout RM, Attia DA, Mortada ND (2019) Enhanced allicin cytotoxicity on HEPG-2 cells using glycyrrhetinic acid surface-decorated gelatin nanoparticles. ACS Omega 4:11293–11300. https://doi.org/10.1021/acsomega.9b01580
7. Shah SAA, Firlak M, Berrow SR, Halcovitch NR, Baldock SJ, Yousafzai BM et al (2018) Electrochemically enhanced drug delivery using polypyrrole films. Materials (Basel) 11:1123. https://doi.org/10.3390/ma11071123
8. Beg S, Hasnain MS, Rahman M, Swain S (2019) Chapter 1—Introduction to Quality by Design (QbD): fundamentals, principles, and applications. In: Beg S, Hasnain MS (eds) Pharmaceutical quality by design. Academic, pp 1–17. https://doi.org/10.1016/B978-0-12-815799-2.00001-0
9. Beg S (2021) Screening experimental designs and their applications in pharmaceutical development. In: Beg S (ed) Design of experiments for pharmaceutical product development: volume I: basics and fundamental principles. Springer, Singapore, pp 15–26. https://doi.org/10.1007/978-981-33-4717-5_2
10. Kacker RN, Lagergren ES, Filliben JJ (1991) Taguchi's orthogonal arrays are classical designs of experiments. J Res Natl Inst Stand Technol 96:577–591. https://doi.org/10.6028/jres.096.034
11. Beg S, Swain S, Rahman M, Hasnain MS, Imam SS (2019) Chapter 3—Application of Design of Experiments (DoE) in pharmaceutical product and process optimization. In: Beg S, Hasnain MS (eds) Pharmaceutical quality by design. Academic, pp 43–64. https://doi.org/10.1016/B978-0-12-815799-2.00003-4
12. Mehanny M, Hathout RM, Geneidi AS, Mansour S (2017) Studying the effect of physically-adsorbed coating polymers on the cytotoxic activity of optimized bisdemethoxycurcumin loaded-PLGA nanoparticles. J Biomed Mater Res 105:1433–1445. https://doi.org/10.1002/jbm.a.36028
13. Djuris J, Ibric S, Djuric Z (2013) 3—Experimental design application and interpretation in pharmaceutical technology. In: Djuris J (ed) Computer-aided applications in pharmaceutical technology. Woodhead Publishing Series in Biomedicine. Woodhead Publishing, pp 31–56. https://doi.org/10.1533/9781908818324.31
14. Box GEP, Wilson KB (1951) On the experimental attainment of optimum conditions. J R Stat Soc Ser B (Methodological) 13:1–38. https://doi.org/10.1111/j.2517-6161.1951.tb00067.x
15. Aziz DE, Abdelbary AA, Elassasy AI (2018) Implementing central composite design for developing transdermal diacerein-loaded niosomes: ex vivo permeation and in vivo deposition. Curr Drug Deliv 15:1330–1342. https://doi.org/10.2174/1567201815666180619105419
16. Box GEP, Behnken DW (1960) Some new three level designs for the study of quantitative variables. Technometrics 2:455–475. https://doi.org/10.1080/00401706.1960.10489912

17. Ferreira SLC, dos Santos WNL, Quintella CM, Neto BB, Bosque-Sendra JM (2004) Doehlert matrix: a chemometric tool for analytical chemistry-review. Talanta 63:1061–1067. https://doi.org/10.1016/j.talanta.2004.01.015
18. Abdel-Hafez SM, Hathout RM, Sammour OA (2018) Tracking the transdermal penetration pathways of optimized curcumin-loaded chitosan nanoparticles via confocal laser scanning microscopy. Int J Biol Macromol 108:753–764. https://doi.org/10.1016/j.ijbiomac.2017.10.170
19. Mukherjee R, Halder A, Sansare S, Naik S, Chaudhuri B (2020) A simplex centroid design to quantify triboelectric charging in pharmaceutical mixtures. J Pharm Sci 109:1765–1771. https://doi.org/10.1016/j.xphs.2020.02.001
20. Mandlik SK, Adhikari S, Deshpande AA (2012) Application of simplex lattice design in formulation and development of buoyant matrices of dipyridamole. J Appl Pharm Sci 2:107–111. https://doi.org/10.7324/JAPS.2012.21221
21. Safwat S, Hathout RM, Ishak RA, Mortada ND (2017) Augmented simvastatin cytotoxicity using optimized lipid nanocapsules: a potential for breast cancer treatment. J Liposome Res 27:1–10. https://doi.org/10.3109/08982104.2015.1137313
22. Mohamad Zen NI, Abd Gani SS, Shamsudin R, Fard Masoumi HR (2015) The use of D-optimal mixture design in optimizing development of Okara tablet formulation as a dietary supplement. Sci World J 2015:684319. https://doi.org/10.1155/2015/684319
23. Naguib SS, Hathout RM, Mansour S (2017) Optimizing novel penetration enhancing hybridized vesicles for augmenting the in-vivo effect of an anti-glaucoma drug. Drug Deliv 24:99–108. https://doi.org/10.1080/10717544.2016.1233588
24. Badawi MA, El-Khordagui LK (2014) A quality by design approach to optimization of emulsions for electrospinning using factorial and D-optimal designs. Eur J Pharm Sci 58:44–54. https://doi.org/10.1016/j.ejps.2014.03.004
25. Hathout RM, Woodman TJ (2012) Applications of NMR in the characterization of pharmaceutical microemulsions. J Control Release 161:62–72. https://doi.org/10.1016/j.jconrel.2012.04.032
26. Hathout RM (2014) Using principal component analysis in studying the transdermal delivery of a lipophilic drug from soft nano-colloidal carriers to develop a quantitative composition effect permeability relationship. Pharm Dev Technol 19:598–604. https://doi.org/10.3109/10837450.2013.813544
27. Yehia R, Hathout RM, Attia DA, Elmazar MM, Mortada ND (2017) Anti-tumor efficacy of an integrated methyl dihydrojasmonate transdermal microemulsion system targeting breast cancer cells: in vitro and in vivo studies. Colloids Surf B Biointerfaces 155:512–521. https://doi.org/10.1016/j.colsurfb.2017.04.031
28. Abdelhamid HN, El-Bery HM, Metwally AA, Elshazly M, Hathout RM (2019) Synthesis of CdS-modified chitosan quantum dots for the drug delivery of Sesamol. Carbohydr Polym 214:90–99. https://doi.org/10.1016/j.carbpol.2019.03.024
29. Abdelrady H, Hathout RM, Osman R, Saleem I, Mortada ND (2019) Exploiting gelatin nanocarriers in the pulmonary delivery of methotrexate for lung cancer therapy. Eur J Pharm Sci 133:115–126. https://doi.org/10.1016/j.ejps.2019.03.016
30. Hathout RM, Gad HA, Metwally AA (2017) Gelatinized-core liposomes: toward a more robust carrier for hydrophilic molecules. J Biomed Mater Res A 105:3086–3092. https://doi.org/10.1002/jbm.a.36175
31. Hathout RM, Gad HA, Abdel-Hafez SM, Nasser N, Khalil N, Ateyya T et al (2019) Gelatinized core liposomes: a new Trojan horse for the development of a novel timolol maleate glaucoma medication. Int J Pharm 556:192–199. https://doi.org/10.1016/j.ijpharm.2018.12.015
32. Tang SY, Manickam S, Wei TK, Nashiru B (2012) Formulation development and optimization of a novel Cremophore EL-based nanoemulsion using ultrasound cavitation. Ultrason Sonochem 19:330–345. https://doi.org/10.1016/j.ultsonch.2011.07.001
33. Cho HJ, Lee DW, Marasini N, Poudel BK, Kim JH, Ramasamy T et al (2013) Optimization of self-microemulsifying drug delivery system for telmisartan using Box-Behnken design and desirability function. J Pharm Pharmacol 65:1440–1450. https://doi.org/10.1111/jphp.12115

34. Duangjit S, Mehr LM, Kumpugdee-Vollrath M, Ngawhirunpat T (2014) Role of simplex lattice statistical design in the formulation and optimization of microemulsions for transdermal delivery. Biol Pharm Bull 37:1948–1957. https://doi.org/10.1248/bpb.b14-00549
35. Patel J, Dhingani A, Garala K, Raval M, Sheth N (2014) Quality by design approach for oral bioavailability enhancement of Irbesartan by self-nanoemulsifying tablets. Drug Deliv 21:412–435. https://doi.org/10.3109/10717544.2013.853709
36. Visetvichaporn V, Kim KH, Jung K, Cho YS, Kim DD (2020) Formulation of self-microemulsifying drug delivery system (SMEDDS) by D-optimal mixture design to enhance the oral bioavailability of a new cathepsin K inhibitor (HL235). Int J Pharm 573:118772. https://doi.org/10.1016/j.ijpharm.2019.118772
37. Fasolo D, Pippi B, Meirelles G, Zorzi G, Fuentefria AM, von Poser G et al (2020) Topical delivery of antifungal Brazilian red propolis benzophenones-rich extract by means of cationic lipid nanoemulsions optimized by means of Box-Behnken Design. J Drug Deliv Sci Technol 56:101573. https://doi.org/10.1016/j.jddst.2020.101573
38. Shi J, Xue SJ, Wang B, Wang W, Ye X, Quek SY (2015) Optimization of formulation and influence of environmental stresses on stability of lycopene-microemulsion. LWT Food Sci Technol 60:999–1008. https://doi.org/10.1016/j.lwt.2014.10.066
39. Dhaval M, Devani J, Parmar R, Soniwala MM, Chavda J (2020) Formulation and optimization of microemulsion based sparfloxacin in-situ gel for ocular delivery: in vitro and ex vivo characterization. J Drug Deliv Sci Technol 55:101373. https://doi.org/10.1016/j.jddst.2019.101373

Quality by Design in Pharmaceutical Development

4

Rajesh Kumar, Amit Mittal, and Mangesh Pradeep Kulkarni

Contents

R. Kumar (✉)
Department of Pharmaceutics, School of Pharmaceutical Sciences, Lovely Professional University, Phagwara, Punjab, India

A. Mittal
Department of Pharmaceutical Chemistry, School of Pharmaceutical Sciences, Lovely Professional University, Phagwara, Punjab, India
e-mail: amit.13145@lpu.co.in

M. P. Kulkarni
School of Pharmaceutical Sciences, Lovely Professional University, Phagwara, Punjab, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_4

Abstract

Design and formulation of an ideal pharmaceutical product is a tedious job for a formulator as it comprises multiple objectives. The traditional method followed for years is not only expensive and time-consuming but also needs a lot of efforts in spite of which they are unfavorable and unpredictable at times. "Pharmaceutical Quality by Design" (QbD) is a well-organized approach which focuses an attention on product, process understanding, and process control, based on a thorough knowledge of science and quality risk management. QbD is increasingly gaining confidence with an assured supply of safe and effective pharmaceuticals. Regulatory bodies like International Conference on Harmonization and United States Food and Drug Administration have realized its importance and thus guided the pharmaceutical industries to adopt its principles and applications in product development. The approach to optimize the best product and process characteristics under the given conditions utilizes the "Design of Experiment" that determines the relationship between the critical factors affecting the process and its output. Nowadays, there are various software available for the optimization of pharmaceutical products. This chapter attempts to encompass thorough knowledge about development of QbD along with its application in pharmaceutical product development and suitable software available for it.

Keywords

Quality by design · Optimization · Product development · ICH · Software

Chapter Objectives

- To discuss the need and advantages of QbD and its comparison over traditional technique.

- Understand the ICH guidelines and a brief introduction to optimization and design of experiment.
- Elements of QbD and its role in formulation development with some examples.
- Concept of post-approval lifecycle management plan.
- Perspectives of regulatory bodies and the reflection of QbD over pharmaceutical industries.
- Summary of some useful QbD software and QbD-based case studies.

4.1 Introduction

It is a well-known fact that a product is of best quality if it has been produced from a perfectly well-defined, controlled, and validated process. Generally, issues related to the quality of product arise from an improperly maintained process of manufacture or due to the various ingredients of the dosage form itself. To ensure the same conventionally, researchers had to evaluate each ingredient, parameter of the product and process, respectively, which was done by the trial-and-error technique. This involved the study of the composition and their amount in the dosage form that can influence the final product in either way [1, 2]. This method of development involved the changing of one variable or factor at a time, that is, one variable at a time (OVAT) or one factor at a time (OFAT), and the others were kept constant. This method of optimization had several limitations such as time-consuming, tedious, uneconomical, inefficient, prone to misinterpretations, result is only satisfactory, and lack of interactions and risk assessment. These problems led the way to the development of optimization which itself means finding the best way of utilizing the available resources and considering the various influential factors and variables that influence the formulation characteristics to develop the best possible product [3]. "Quality by Design" (QbD) is something that states the designing and development of a product or process which ensures predefined product characteristics and outcomes [4]. This was first quoted by Joseph M Juran who is known as the father of QbD [2, 5]. He was the one who believed that a quality product can be developed by prior design, and the most problems occur due to improper design in its first place.

Earlier to QbD, Quality by Testing (QbT) was employed by the industries, and the quality was checked by performing several tests of quality control which ultimately led to the conventional approaches itself. In the year 2002, an initiative was taken by the regulatory bodies called the cGMP (current Good Manufacturing Practices) for twenty-first century [1], a risk-based approach which made the drug developer and the manufacturers focus upon the need and importance of QbD, quality risk management, and quality systems. Down the lane, pharmaceutical QbD has evolved with the guidelines and procedures quoted by ICH (i.e., the Q8 (pharmaceutical product development), Q9 (quality risk management), Q10 (pharmaceutical quality system), etc.) [4, 6]. Overall, QbD is an approach that takes into account the various processes and product parameters which can influence the finished product and develop a design to get the best outcome.

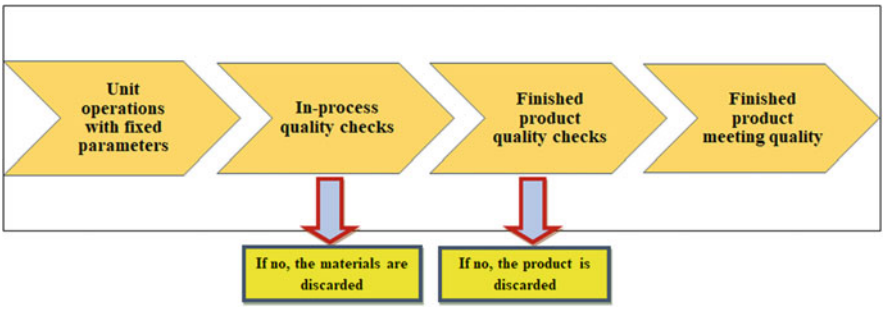


Fig. 4.1 Traditional OVAT technique

Optimization of drug product and process of its manufacturing with the prerequisite quality attributes is a difficult task. Using traditional approach, that is, OVAT, a genuine optimum formulation attainment was never a guarantee since it optimized only a single factor at a time leading to just satisfactory products due to factor–factor interactions. Many inconsistencies prevail in this approach which can be attributed to the inadequate knowledge about variable–response relationship. Moreover, OVAT approach becomes ineffective when all the factors vary simultaneously. Figure 4.1 exhibits traditional approach [7].

The limitations of traditional approach demanded development of some new techniques that can overcome its shortcomings. Though QbD has been a new paradigm, it is not that new in the pharmaceutical industry. The term QbD was coined by Joseph M Juran in the 1970s. From then to the decade of 1980s, various approaches such as molecular mechanics, molecular simulations, quantum chemistry, quantitative structure–activity relationship, and molecular graphics converged into modern computational chemistry. In the 1980s, several technical advances encouraged the improvement of computer usage in pharmaceutical companies. The popularization QbD was done in the 1990s with the help of several publications [8]. In 2002, the United States Food and Drug Administration (US-FDA) made its first attempt toward integrating QbD concept with "current Good Manufacturing Practices" (cGMPs). A final report on "Pharmaceutical cGMPs for the 21st century: A Risk Based Approach" guideline was released by US-FDA with the aim to modernize the regulation of the pharmaceutical product manufacture, quality, and development [9, 10]. In 2004, QbD was first introduced in chemistry, manufacture, and control review process as a result of the pharmaceutical cGMPs for the twenty-first century initiative. Various advantages of implementation of QbD have been summarized in Box 4.1.

This chapter aims to review the role of systematic approach (i.e., QbD in the development of dosage forms). It involves a comparison of QbD with traditionally adopted approach. Various relevant ICH guidelines and the elements of QbD have been discussed in this chapter. In addition, various experimental study designs have been addressed briefly. A few case studies have been discussed here to increase the understanding of implementation of QbD in dosage form development. This will

Table 4.1 Comparison between traditional and QbD approaches according to ICH Q8 guideline [11, 12]

Feature	Traditional approach	QbD approach
Pharmaceutical development	Chiefly empirical and is based on OVAT data only	Systematic understanding of its critical attributes; multivariate experiments for better understanding; establishment of design space and use of PAT tools
Manufacturing process	Fixed; reproducibility and optimization-centric; initial full-scale batch-based validation	Adjustable within design space and continuous process verification; focuses on control strategy and robustness; and utilizes statistical control process methods
Process control	In-process tests are performed to decide the progression of manufacturing	PAT tools are used; process operations are tracked to support continual post-approval development
Product specification	Primary method of control based on batch data available at time of registration	A part of overall control strategy and is based on desired performance of product
Life cycle management	Problem-solving corrective in nature	Preventive in nature with ensured continuous improvement
Control strategy	Control of product quality by intermediates	Risk-based control strategy; real-time testing with lesser end product testing

facilitate better communication between those involved in risk-based drug development and drug application. The perspectives of some important regulatory agencies and viewpoints of industries toward implementation of QbD have been incorporated. A list of software is used in QbD, and a brief summary of a few of them is provided here. Moreover, a few web resources have been given at the end for readers to refer to for detailed content.

Box 4.1 Advantages of QbD Implementation

1. Cost-effective
2. Reduced risk of product failures/rejection
3. Enhanced manufacturing efficiency
4. Provides opportunities for continuous improvement of the product
5. Enhances the opportunities for first cycle approval
6. Minimized/eliminate potential compliance actions
7. Streamline post-approval changes and regulatory processes
8. To better understand the concept of QbD, we need to take an overview over other measures such as design of experiment, model-based approach, etc.

A comparison of traditional and QbD approaches has been discussed in Table 4.1.

4.2 International Council for Harmonization (ICH) of Technical Requirements of Pharmaceuticals for Human Use Guidelines

4.2.1 ICH Q8 (R2)

ICH-Q8 guidelines provide information about the pharmaceutical product development which is of interest when quality by design is considered. ICH-Q8 guideline is sub-divided into two parts: part I deals with pharmaceutical development and part II is the annexure to the guideline stating the principles for QbD [13]. This document entails the details regarding the quality-based product development and the scope of common technical documents [14]. It also says that the development of manufacturing process must be accompanied by its continuous ongoing verification instead of only one-time validation and guides that the critical aspects of the product and process must be determined, and a proper control strategy must be developed. If the product has organized and robust control strategies, the post-approval improvement and, thus, the regulatory compliance will never be an issue. ICH-Q8 also quotes about the application of tools such as experimental designs, process analytical technology (PAT), and quality risk management principles which must be involved in the process of development [15, 16]. It can be defined as system for designing, analyzing, developing through time-based measurements, and controlling the critical process and product attributes which has properties such as a multidimensional combination and analyzation to ensure quality, provides a link between process inputs and outputs, and can be established for the complete process [17, 18]. The elements of pharmaceutical product development such as the quality target profile (QTP), critical parameters, and attributes, and the use of medialization and optimization techniques have been defined in these guidelines.

4.2.2 ICH Q9

As there is a probability of occurring risk in manufacturing a dosage form along with its components, there is a need for quality risk management (QRM) and for obtaining it, this guideline is very helpful. The main aim of this guideline is to provide principles and examples for QRM which can be applied to different pharmaceutical quality aspects. QRM is a systematic process for the assessment, control, communication, and review of risks for obtaining quality products. The risk management methodology includes various risk management tools that can be basic risk management facilitation methods, failure mode and effect analysis (FMEA), failure mode, effect, and criticality analysis, fault-tree analysis, hazard analysis and critical control points, hazard operability analysis, preliminary hazard analysis, risk ranking and filtering, and supporting statistical tools [19].

4.2.3 ICH Q10

This is the guideline that describes a comprehensive model for effective pharmaceutical quality system, which is based on ISO (i.e., International Standards Organization quality concepts including GMP regulations). ICH Q10, when implemented throughout the product's lifecycle, should facilitate the innovations and continuous improvements and strengthen the link between pharmaceutical development and related manufacturing activities. The objectives of ICH Q10 are to achieve product realization, establish and maintain a state of control, and facilitate continuous improvement. The product lifecycle for which ICH Q10 guideline can be implemented includes various technical activities for new as well as existing products like pharmaceutical development, technology transfer, commercial manufacturing, and product discontinuation [20].

4.2.4 ICH Q12

The main aim of this guideline is to provide a framework for facilitating the management of post-approval Chemistry, Manufacturing and Control (CMC) changes in an efficient and predictable manner. This guideline is helpful in applying pharmaceutical drug substances and products which require marketing authorization. ICH Q12 regulatory tools and enablers are categorization of post-approval CMC changes, established conditions, post-approval change management protocol, product lifecycle manage document, pharmaceutical quality system and change management, relationship between regulatory assessment and inspection, structured approaches for frequent CMC post-approval changes, and stability data approaches to support the evaluation of CMC changes [21].

4.3 Design of Experiment (DoE)

In simple terms, DoE can be explained as the systematic study of factors which can affect the output of a product and/or process significantly. Statistically, it is the planning of experiment in such a way to draw essential experimental data and a valid output. Application of this method can help in reducing the number of experiments and also identification of critical factors that are affecting the product [22]. The DoE plays a vital role in various domains; so considering its applications in formulation development, a specific term was assigned to it which is Formulation by Development (FbD). As the name itself indicates, FbD is specifically applicable to use of DoE in the designing of drug formulation. It is a statistical approach which has been found to offer the best possible formulation with lesser time, efforts, and expenditure. The amenability to scale-up and post-approval changes are the best merits of implementing FbD [23].

4.3.1 Design of Experiment Plan

It has been evident that for a better controlled process and then a product/formulation as an outcome of that, there has to be a well-planned sequence of steps that also define the control strategies and validation requirements of the plan. Design of experiment itself means the pre-determination of the formulation design, and the activity done in sequential steps is called the design of experiment plan which is shown in Fig. 4.2 [5, 24].

4.3.2 Methodologies in Optimization

The holistic QbD approach for formulation development (i.e., FbD) is based on the determination of five basic elements which are quality target product profile (QTPP), critical quality attributes, critical formulation attributes, critical process parameters, an apt DoE-led design, and control space. Figure 4.3 demonstrates a five-step ladder for optimization of a drug delivery system [25] (Table 4.2).

In design of experiment, the optimization methodology can be classified into the following types.

4.3.3 Simultaneous Optimization

This is a model-dependent approach and is also generally termed as response surface methodology. A response surface can be defined as the area of space defined within the lower and upper limits of the independent variables to the measured responses

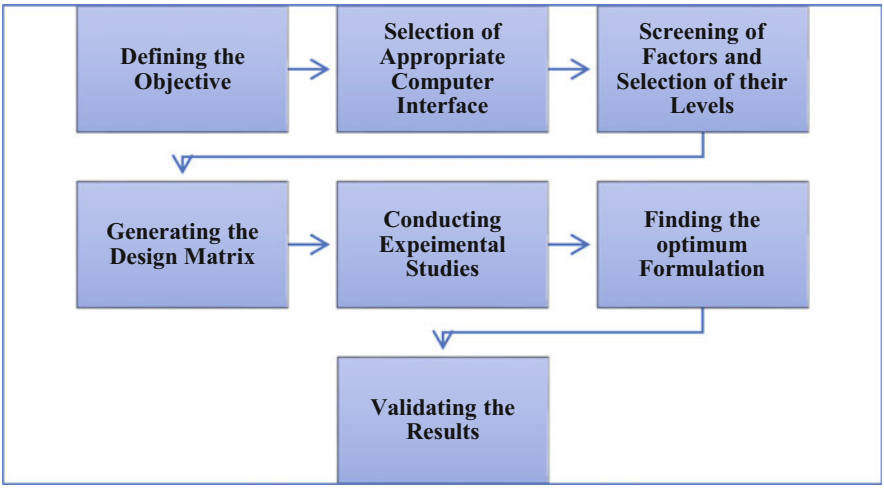


Fig. 4.2 Design of experiment plan

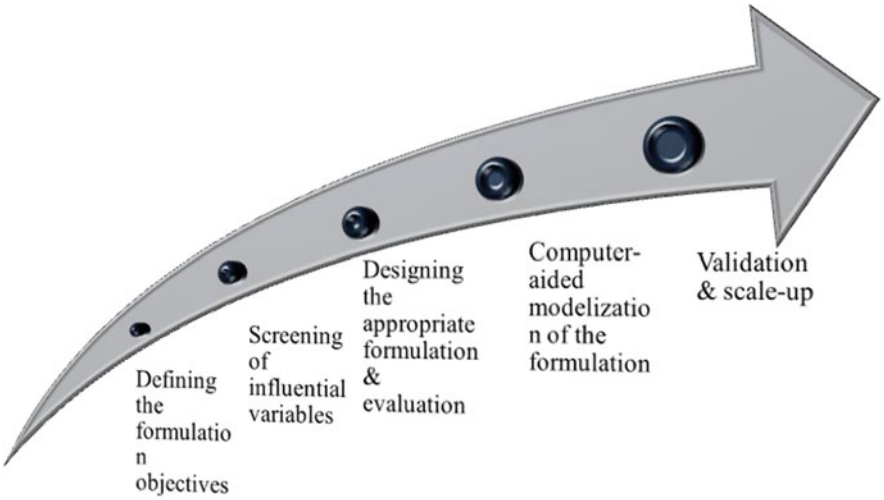


Fig. 4.3 The five-step ladder of optimization

Table 4.2 Classification of model's contribution in assuming product quality (according to ICH Q-IWG document) [26]

Sr. no.	Model class	Description
1	Low-impact models	For supporting the product and process development (e.g., formulation optimization)
2	Medium-impact models	Used for assuring the quality but are not the sole indicators of it (e.g., in-process and design space models)
3	High-impact models	Significant indicator of the product quality (e.g., surrogate dissolution model)

[27, 28]. Figure 4.4 shows various experimental study designs that are employed in pharmaceutical research studies.

4.3.4 Sequential Optimization

The simultaneous optimization is itself used widely and has a lot of demerits but in cases where the exact effects of a factor or a parameter are not known, then to get acceptable data the sequential type of optimization is used [29]. Its major advantage is that it can give data about the unknown factors and their interaction, while its robustness and availability of optimum data is a matter of question.

The experimental models of sequential optimization include the following:

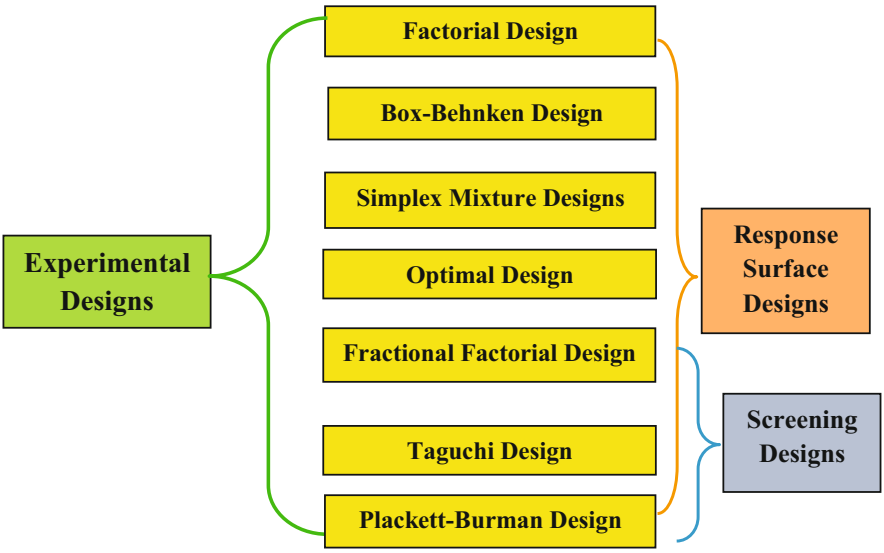


Fig. 4.4 Classification of experimental designs

1. Steepest ascent/descent method: It is mixture of the model-dependent and non-dependent designs that involve the measurement of the steepest fall or rise of the slope followed by the measurement of response till the optimum is reached [30].
2. Optimum path method: This method is similar to the first method but here the optimum is searched by extrapolation outside the domain of experiment.
3. Sequential simplex techniques: In this technique, the $(n + 1)$ data are considered and the decisions are made depending on the predetermined rules till the optimum is reached.
4. Evolutionary operations: This involves the simultaneous generation of data for improvement and also its production using the factorial and simplex designs [31, 32].

4.4 QbD in Formulation Development

In the pharmaceutical QbD, the most basic considerations include the safety and quality perspectives of the patient and the regulatory body that makes the manufacturers consider certain critical product/process elements to develop a desired product which includes the following:

- The route of administration, intended use, dose strength dosage form, delivery system, etc.
- The container/close systems and packaging considerations

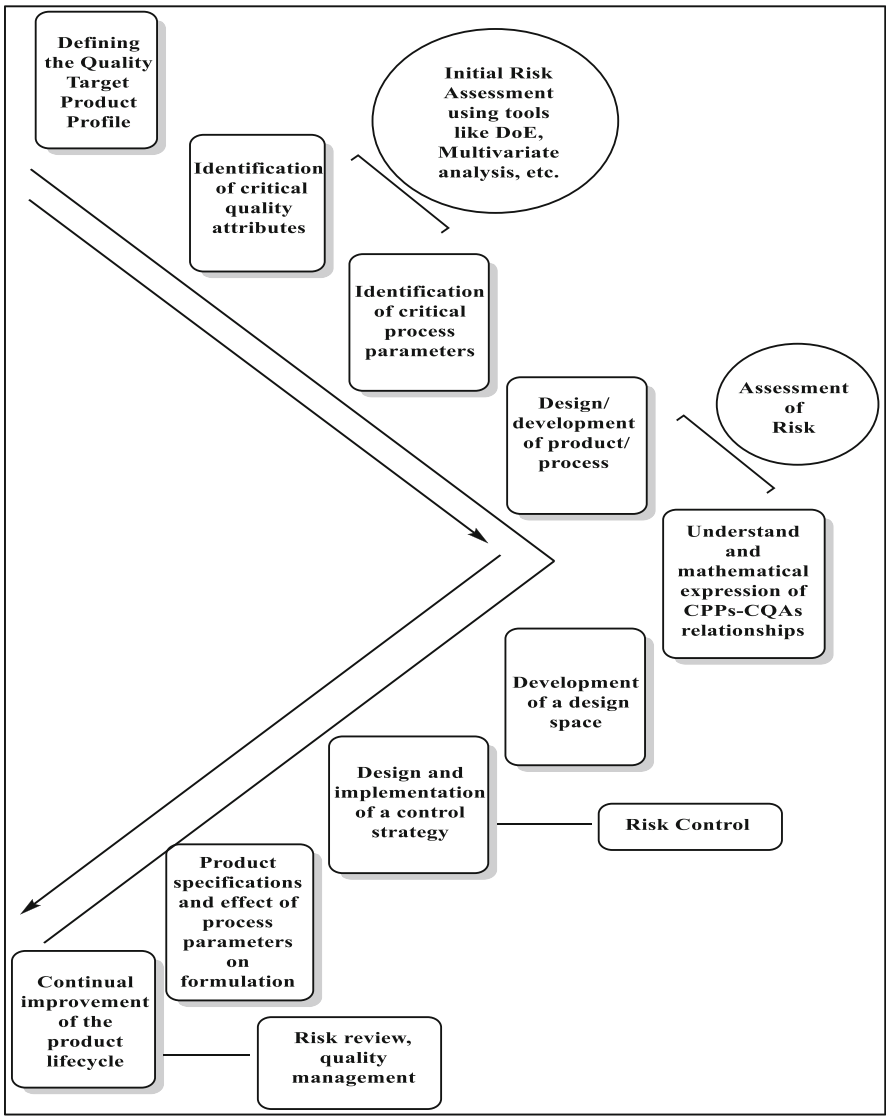


Fig. 4.5 A QbD flowchart combining design space development and risk management tools [11]

- Its pharmacokinetic attributes
- The quality attributes like its purity, stability, etc. [33]

Figure 4.5 demonstrates a QbD approach that combines risk management tools with the development of design space for the experiment [11].

4.4.1 Quality Target Product Profile

It provides the basis of design for the product development. It can be called the heart of QbD as it is the one that relates with the predefined objectives of the work that includes the Critical Quality Attributes (CQAs), Critical Process Parameters (CPPs), and the Critical Material Attributes (CMAs) [13, 34]. Parameters like dosage form/delivery system, dose strength, route of administration, dissolution performance, container-closure system, and criteria for drug product quality (e.g., purity, sterility, stability) can be considered for inclusion under QTPP [4]. Earlier, the importance of predefined target characteristics of a drug product was underestimated, and thus a lack of properly defined QTPP resulted in wastage of valuable time and resources. In view of that, Raw et al. illustrated the importance of a correct and predefining QTPP before conducting any product development [35].

4.4.2 Critical Quality Attributes

CQA is the next step in the development of a drug product. As the name itself explains, CQA includes all the parameters of a product/ingredient/additive considering its physical, chemical, biological, microbiological, or any other parameter that has to be controlled or possess any limits [36, 37]. These quality attributes may include identity, assay, content uniformity, degradation products, drug release study, moisture content, microbial load, and physical characteristics like color, odor, size, shape, and friability. The attributes may or may not be called critical based upon the extent of harm it can cause to the patient in case if the product does not comply with the acceptable range of a particular attribute [4].

4.4.3 Critical Process Parameters

This section of the QTPPs includes the various process attributes that are involved in the manufacture, development, and also packaging of the product or the finished goods thus named as the critical process parameters that can ultimately influence the quality of a dosage form [1, 36]. A CPP is a parameter that can influence the CQA and therefore must be monitored to enable early and accurate detection of deviations from the defined limits which otherwise will affect the quality of product. Not all the process parameters are critical; however, the one with profound impact on CQAs would be rated critical. Consequently, it becomes important to prioritize the CPPs over the others and control them rigorously. Even if the manufacturing process is the same, CPPs may vary depending on the type of product, material attributes, and target profile of the product. Some of the examples of CPPs are temperature, pH, cooling rate, speed of rotation, etc. [38].

4.4.4 Critical Manufacturing Parameters

End product's CQAs influence the product performance within desired quality, efficacy, and safety. These attributes may be those that affect specifications such as impurity, potency, stability, drug release, and microbiological attributes. At the same time, they may be the active ingredient attributes that influence the product performance or reproducibility. They are known as CMAs [39]. The CMAs include the conditions in which the product has been prepared, and any undesirable changes in the manufacturing conditions can hamper the quality of the product [40]. Identification, optimization, and control of these critical attributes are extremely necessary to ensure the desired quality in a dosage form (Fig. 4.6, Table 4.3).

4.4.5 Unclassified Process Parameters

Sometimes, a few parameters that are involved in the processing of the formulation might not be known perfectly to the formulator, but once a small idea is developed on its function, its effects can be reclassified as a critical or a non-critical parameter. For an instance, in the granulation process, the speed of impeller can be clearly identified as unclassified process parameters since the process would not be successful at zero speed; however, it does not imply that impeller speed is always critical. If product performance indicates that the granulation was not significantly affected by changes in impeller speed, it will not be considered critical [46].

4.4.6 Control Strategy

In the development and the manufacturing of any product, once the critical parameters of it are defined, then it is very necessary to also define its control strategy without which a desired output might not be possible [36, 47]. In QbD, it

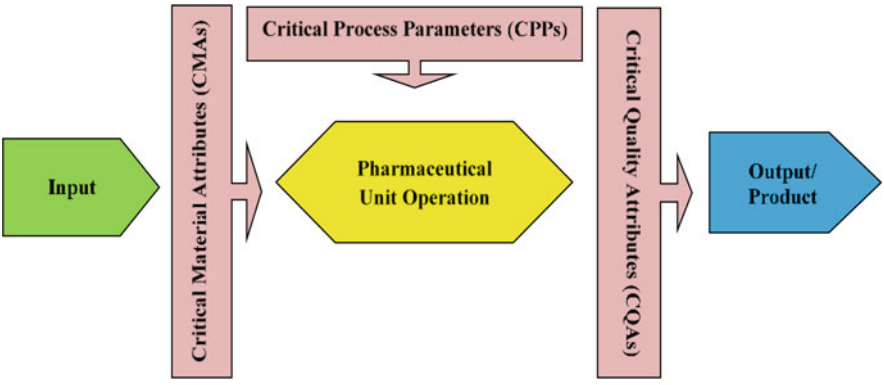


Fig. 4.6 Linking of input with CMAs and CPPs to output CQAs for a unit operation

Table 4.3 Applications of QbD elements in pharmaceutical unit operations (UO) and various dosage forms

UO	Dosage form	Drug	DoE	CMA	CPP	CQA
Film coating	Tablets	Placebo	Face-centered CCD	Solid % in coating dispersion	Inlet air temperature, airflow rate, coating pan speed, spray rate	Tablet appearance, disintegration time of tablets [41]
Hot–melt extrusion	Solid lipid nanoparticles	Fenofibrate	Plackett–Burman design	Concentration of lipid and surfactant	Screw speed, barrel temperature	Particle size, zeta potential, entrapment efficiency [42]
Homogenization	Nanoparticles	Paclitaxel	Box–Behnken design	Concentration of surfactant in aqueous phase	Rate of homogenization	Average particle size, zeta potential, encapsulation efficiency [43]
O/W emulsion–solvent evaporation	Nanoparticles	Cyclosporine A	Plackett–Burman	Organic to aqueous phase ratio, concentration of drug, polymer, surfactant	Stirring rate	Encapsulation efficiency, particle size, zeta potential, dissolution efficiency [44]
Homogenate membrane method	Orodispersible film	Theophylline	Central composite design	Percentage of HPMC and glycerol	Drying temperature	Tensile strength, Young's modulus, disintegration time [45]
Physical mixture, solvent evaporation	Controlled-release tablets	Felodipine	Box–Behnken design	Amount of polymer HPMC, amount of Pluronic F127	Preparation technique	Maximum solubility after 30 min, equilibrium solubility after 24 h, dissolution efficiency [22]

is done through the risk-assessment approach and may involve the approaches such as procedural controls, in-process controls, comparability, and stability testing [48].

4.4.7 Design Space Verification

Design space is "the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality" as defined by ICH Q8 guidelines. It is a scientific concept used in the pharmaceutical industries to assure and support the product quality. The foundation for the design space is provided by the data and knowledge acquired during the development of a product. Fig. 4.7 exhibits the design space [49].

A well-defined design space assures uniformity in product performance across various batches. According to EMA's (European Medicines Agency) definition, design space verification is the verification that material inputs and processes are able to scale to commercial manufacturing levels while maintaining a standard of quality. It is a part of process validation, and its main purpose is to guarantee the quality of product within boundaries of manufacturing. Thus, conducting design space verification is difficult when being operated at target levels; it should be conducted throughout the manufacturing lifecycle. It should be dependent on the results of risk assessment that are involved with the "scale-up activities." As the main aim of design space verification is confirming quality of output within operating ranges, it helps in avoiding re-evaluation of process of production.

4.4.8 Real-Time Release Testing

According to ICH Q8 (R2), real-time release testing is defined as the ability to evaluate and ensure the quality of in-process and/or final product based on process data (combination of measured material attributes and process controls). From the application of real-time release testing benefits from industrial standpoint, economic benefits can be gained from efficient manufacturing like lowered laboratory costs

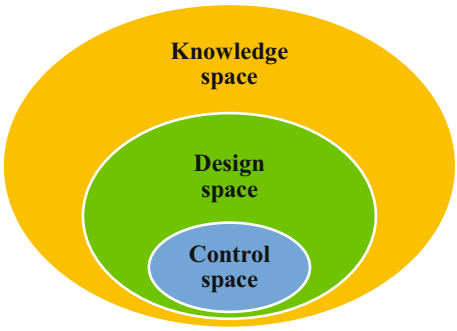


Fig. 4.7 An illustration of design space

and reduced inventory. Before applying real-time release testing to a manufacturing process, which operations are to be used for manufacturing must be determined by pharmaceutical firms. The purpose or main aim of application of real-time release testing is to understand and control one's product and processes in such a manner that ensures quality end product [50, 51].

4.4.9 Normal Operating Range

According to ICH, normal operating range is a region around the target operating conditions that contain common operational variability which cannot always be precisely controlled to a single and specific value. As normal operating range does not represent a deliberate adaption of processes, it can be established for several parameters of process. For better quantification of actual uncontrollable operational variability (of process parameters), normal operating range is used, and it is also intended for introducing flexibility in manufacturing conditions.

4.4.10 Proven Acceptable Range

According to ICH Q8 (R2), proven acceptable range is defined as a defined range of a process parameter for which operation within this range will result in a material complying with the quality criteria, while keeping other parameters within normal operating ranges.

4.4.11 Product/Process Design and Understanding

Once the formulator is aware of the critical quality attributes, a relation between them, and the safety, efficacy has to be ensured with the help of any available data [47]. In QbD, the optimum understanding of this characteristic of the ingredients, their proportions, and also the processes are very important to develop a product with the desired quality and safety [15, 52]. The product design involves everything of a product from its physicochemical properties to the pharmacological properties only then the efficient and useful output is possible [53]. The biopharmaceutical assessment of the drug and the excipients being used in the formulation will help the formulator to select the appropriate dosage form and its classification into the biopharmaceutical classification system [4, 36]. For demonstrating reproducibility and robustness of process, the process capability index has to be studied. Process capability index is a statistical measure of extent of variation in a process relative to the end-user specification which can be determined by using the most popularly used formula as given in Eq. 4.1 [54]:

$$\text{Process capability index} = \text{Upper limit of specification} - \text{Lower limit of specification}/6 \text{ standard deviation} \quad (4.1)$$

4.4.12 Post-approval Lifecycle Management

Even after the approval of a product, it is continuously monitored by the industries on its performance and quality throughout the product lifecycle. When the regulatory bodies are of concern, the changes done to the product within the stipulated space do not require approval, and thus, an extended process design space is more desirable in this case [6, 55]. The process/product robustness is also an important consideration with respect to the parameters like preventive/corrective action, process/product monitoring, quality, and performance [56, 57]. The post-approval lifecycle management (PALM) plan provides a regulatory agreement between the product manufacture and health authorities. It specifies how the product–process attributes can be ensured in a controlled state after approval, how changes in parameters are controlled in the design space, and how to update the control system based on knowledge of product and process [58]. The PALM plan comprises certain elements mentioned in Box 4.2.

Box 4.2 Elements of PALM

Design space definition: This includes sequence of steps, raw material definition, and defined ranges for critical- and non-critical process parameters.

Summary of testing strategy: This comprises control system testing, monitored CQAs considered for stability in comparative practices for both drug substance and the drug product.

Process and product monitoring: This strategy is aimed at continuous verification for ensuring the control state and consistency of commercial process. The monitoring program takes care of routine monitoring of critical and non-critical process parameters, in-process controls, key performance indicators, CQAs assigned to control system testing, and monitoring category. Attribute Testing Strategy Risk Ranking and Filtering (ATS RRF) tool provides assurance regarding control of any inherent residual uncertainty in the application of risk management tools and prediction models [59].

Change management of process parameters: The extent of testing is determined by knowing the level of risk involved (low/medium/high risk) for incorporating a change in a process parameter target within the design space. If needed, assessment testing is performed after the change to verify that the change resulted in the desired result and to verify the design space. Additionally, the testing frequency may also be increased after the change.

(continued)

Box 4.2 (continued)

Change management of the control system: Post-market authorization, the control system should be re-evaluated, if significant new information is available during the product lifecycle. Any changes in the control system require approval from the health authority; however, no such approval is required for changing the frequency of testing [58].

4.5 Regulator's Views on QbD

All the regulatory bodies give an importance to quality of pharmaceutical products. Considering that, FDA realized that more stringent controls were required for pharmaceutical manufacturing operations and that led to implementation of cGMP in 2002 and later in 2005, a more systematic approach (i.e., QbD came into implementation). It was expected to ensure the following:

- Scientific assessment of the product/process and its manufacturing and development
- Evaluation of product quality as per established standards [25]
- Post-approval evaluation of products based on scientific knowledge and risks [60]

4.5.1 FDA ONDQA's CMC (Office of New Drug Quality Assessment; Chemistry, Manufacturing and Control) Pilot Program

This pilot program was initiated in 2005 with objective of offering participating firms an opportunity to submit CMC information demonstrating QbD and enable FDA to implement new QbD concepts. It was able to generate a valuable experience for FDA and industry in the implementation of QbD. All these efforts have led to good understanding of standard QbD approaches (e.g., defining CQAs, quality risk assessment, design space, etc.) among applicants and reviewers and an increased interest in real-time release testing (RTRT), continuous manufacturing, and biopharmaceutics approaches to QbD. However, more experience is still required to standardize the approaches for more complex concepts like RTRT and continuous manufacturing, and hence, FDA has been continuously working toward increased collaboratively across various review offices. Taking a step further in 2011, FDA-EMA Parallel Assessment Pilot program was started with objectives of ensuring consistent implementation of ICH guidelines, establishing a pathway for EU and FDA knowledge sharing on QbD containing application, facilitating existing

collaborations between FDA-EMA on inspections, and providing opportunities for joint training [61].

4.5.2 EMA's Perspective

EMA made guidelines for submission of QbD, its process of review and management, wherein EU-PAT team may offer expert advice and ensure evaluation consistency on request from the Rapporteurs and the completion of review process in a maximum of 30 weeks. EMA believes in the implementation and education of QbD through guidance documents and training, periodic interaction with stakeholders, and activities of EU-PAT team. It also guides to focus on the issues like adequacy of data to support the design space and its validation, control strategy, and life cycle management during the process of review [61].

4.5.3 Pharmaceuticals and Medical Devices Agency (PMDA) Perspective

Apart from the similarities with the FDA/EMA guidelines, PMDA advocates the inclusion of expertise from National Institute of Health and Science, Japan for the review process of QbD and its management. It permits the total review time per application as 1.5 times that of traditional submission so as to evaluate the additional information on QbD. Further, PMDA considers a need for inclusion of statistician/chemometrician to assist in understanding DoE and mathematic models. According to PMDA, the major focus should be on logical explanation of quality risk management and strategy for its control, process of defining QTPP, and identification of CQAs [62].

4.5.4 Health Canada (HC) Perspective

HC focuses on systematic peer review process for any submission. Each review team includes a minimum of three senior reviewers at least one of whom must be having an exposure to QbD-type submissions [62].

4.6 Reflections of QbD on the Industry

Pharmaceutical industry, being governed by authoritative regulatory bodies, is known to be the utmost regulated industry. The safety of efficacy of the drug product is given a prime importance to achieve which QbD supports by offering a thorough understanding of process and related parameters, which is the ultimate aim of incorporating QbD. It does so by providing

- Scientific understanding of critical attributes
- Controls established at the developing stage
- Utilization of knowledge gain for its continuous improvement [63]
- Reduction of costs and time
- Ensured first cycle approval and less chances of rejection [58, 64]

Over the past few decades, the pharmaceutical industries have started using innovative and cutting-edge technology for research and development. The advent of QbD invited the industries to utilize their skills for addressing the issues related to process operation and efficiency. Many pharmaceutical industries have been buzzing about QbD and remained in a phase of experimentation with the concept. Most of the contract manufacturers try to stay out of the discussion on implementation of QbD with an assumption that it has little to do with their role. In reality, applying the principles underlying QbD can help them be superior to their peer competitors. Moreover, QbD helps strengthen the relationship between contractor and client and facilitates the establishment of an effective business strategy for products, processes, and facilities. All of this has made pharmaceutical industries start adopting QbD concepts to facilitate continuous improvement strategies for the enhancement of product quality and as well as productivity. FDA's strong encouragement for QbD elements in Abbreviated New Drug Applications (ANDA) for generic drugs (2013) has made it very clear that the pharmaceutical developers will have to comply with QbD approach in all the license applications ranging from small molecule to biologics [65].

However, there are certain challenges in implementing QbD the most important of which is the training. The regulatory bodies and industries need to conduct periodic training sessions for their personnel [66]. Keeping this in view, FDA initiated various pilot programs like ONDQA CMC (2005), an Office of Biotechnology Products program (2008), and an EMA-FDA joint pilot program of the QbD parallel assessment to convince the pharmaceutical companies to embrace QbD. More such efforts are required to be made to sensitize the stakeholders [66]. Being a relatively new concept in pharma sector, there is a lack of concrete guidance for industry and how the inspection will be conducted once the QbD is made mandatory is required to be addressed since earlier, and the inspections were conducted using FDA's system-based approach [67]. The cost associated with implementing the QbD in product development and associated unit operations is another concern among the stakeholders. Establishing a balance between QbD-based approach and traditional approach of incorporation of quality is in transition phase yet and would be better with awareness of the concept among the masses [68].

4.7 QbD Software

The QbD approach has tremendous potential. The practical application of such a rational approach usually involves a great deal of mathematics and statistics. The availability of robust and economical hardware and comprehensive QbD software

Table 4.4 Software used in Quality by Design (QbD)

Sr. no.	Software	Features
1.	Fusion QbD	Advanced approach toward experimental design and multivariate analysis
2.	Master control QbD	Effective and efficient implementation of principles of QbD
3.	Lean QbD	Knowledge-assisted structured applications that can perform several tasks related to QbD implementation
4.	QbD Vision	
5.	QbD Software	

has greatly simplified the computational technicalities [69]. A list of some software that is used in QbD has been tabulated in Table 4.4, and profile of a few selected software has been given in Box 4.3 (Fusion QbD®), Box 4.4 (JMP QbD), and Box 4.5 (Lean QbD).

Box 4.3 Summary of fusion QbD® software

Sr. no.	Feature	Description
1	Software/App. Name	Fusion QbD®: Quality by Design Software Platform
2	Recent Version	Fusion Pro®
3	Company Name	S-Matrix
4	Location and Website	Northern California, United States; http://www.smatrix.com/index.html
5	Machine Requirements for Installation	Hardware requirement—8 GB RAM; I3 Processor; 500 GB hard disk Software requirement—Windows 7, 8.0, 8.1, 10
6	First Launch Year/ Date	1985
7	Availability (Free/ Paid/Trial Version)	A free demo is available on request on the website
8	Applications in Pharmaceutical R&D	Ultrahigh-performance liquid chromatography methods facilitate the development of glucose-responsive insulin therapeutics (https://link.springer.com/article/10.1007/s00216-019-02249-4)

Box 4.4 Summary of JMP QbD Software

Sr. no.	Feature	Description
1	Software/App. Name	JMP Statistical Discovery
2	Recent Version	JMP® Pro
3	Company Name	SAS, UK
4	Location and Website	United Kingdom https://www.jmp.com/en_in/home.html
5	Machine Requirements for Installation	Hardware requirement—8 GB RAM; I3 Processor; 500 GB hard disk Software requirement—Windows 7, 8.0, 8.1, 10
6	First Launch Year/ Date	1989
7	Availability (Free/ Paid/Trial Version)	A free 30-day trial software is available to download
8	Applications in Pharmaceutical R&D	Applying quality by design (QbD) concept for fabrication of chitosan-coated nanoliposomes (https://www.tandfonline.com/doi/abs/10.3109/08982104.2013.826243)

Box 4.5 Summary of Lean QbD Software

Sr. no.	Feature	Description
1	Software/App. Name	Lean QbD; QbD Software—for the Smart Scientists
2	Recent Version	Knowledge Management Pharma 4.0
3	Company Name	QbD Works (US), LLC402A W Palm Valley Blvd Suite, USA
4	Location and Website	United States; https://www.leanqbd.com/
5	Machine Requirements for Installation	Hardware requirement—8 GB RAM; I3 Processor; 500 GB hard disk Software requirement—Windows 7, 8.0, 8.1, 10
6	First Launch Year/ Date	2013
7	Availability (Free/ Paid/Trial Version)	A free demo is available on the website
8	Applications in Pharmaceutical R&D	Quality by Design for ANDAs: An Example for Immediate-Release Dosage Forms (https://www.fda.gov/media/83664/download)

4.8 Scientific Examples of QbD (Case Studies)

4.8.1 5-Fluorouracil (5-FU)-Loaded Thermosensitive Hydrogel

The main purpose of this study was the implementation of QbD principle in developing 5-fluorouracil (5-FU)-loaded thermosensitive hydrogel. Though 5-FU is an anticancer drug having broad-spectrum activity against solid tumors, it has limitations like getting metabolized rapidly in the body with a half-life of just 8–20 min that demands continuous administration of drug for maintaining the therapeutic activity. Thus, its mode of administration had to be changed. Considering this, 5-fluorouracil (5-FU)-loaded thermosensitive hydrogel was prepared for solid tumors by administration through intra-tumoral injection by using QRM and experimental design tools.

The preparation of 5-FU-loaded thermosensitive hydrogel was done under aseptic conditions using physical crosslinking method. QTPP elements for 5-FU-loaded thermosensitive hydrogel were dosage form, route of administration, drug product quality attributes, container and closure system, and stability. CQAs of 5-FU-loaded thermosensitive hydrogel were physical attributes such as appearance, performance attributes like gelling time, gelling temperature, viscosity at liquid state, swelling index, gel strength, drug release, and microbiological attributes like microbial limit. QRM tools used for the preparation of 5-FU-loaded thermosensitive hydrogel were risk identification, identification of various material, process, and environmental factors that were more likely to affect CQAs of the product by fish-bone diagram. Risk analysis was done by usage of qualitative and quantitative analysis tools. Risk evaluation was done by experimental design.

For evaluating physical appearance, final formulation was evaluated for percent transmittance, pH, particulate matter, and clarity. Gelation temperature and time at which gel did not flow in inverted tube test method were recorded as its gelation temperature and time. Rheological measurement was done for assessing viscosity. For assessing water retention, water absorption capacity of the final formulation was assessed by swelling study. The time required for evaluation of complete hydrogel degradation was observed through biodegradation study. In-vitro drug release study was carried out by using "dialysis tube method." For defining the in-vitro drug release mechanism, various kinetic models like zero order, first order, Higuchi, and Korsmeyer–Peppas were used. Sterility test was performed by Method-B direct inoculation method according to US pharmacopeia 35. Stability study was performed for the determination of temperature that affects the stability of the formulation over a defined period. Thermosensitive hydrogel formulation development was achieved by reducing the gelation temperature of methylcellulose by the addition of different additives like trisodium citrate (TSC). Interaction of TSC with water permitted high intramolecular crosslinking between methylcellulose chains, thus resulting in gelation.

Through this study, it was assessed that hazardous effect, nature of excipient, concentration of methylcellulose, concentration of TSC, mixing speed, and time were considered highly risk factors for which high care must be taken.

Concentrations of MC and TSC were considered critical parameters. Full factorial design with two factors and three levels was used for optimization where dependent variables were gelling temperature and gelling time and viscosity-independent variables are concentrations of MC and TSC. The prepared formulation was found to be clear and transparent. With the increase in temperature, the solution turned from clear to turbid and then converted into gel. This gel was thermo-reversible as the gel could be converted into solution on further cooling. The solution showed pseudoplastic flow in both physiological and non-physiological states. From equilibrium swelling that was attained by the final formulation at 6 h, it can be said that hydrogel can absorb water and saturate rapidly. According to different kinetic models applied, diffusion was found to be the predominant mechanism of release of drug. The formulation had also passed sterility test as there was neither turbidity nor the microbial contamination when incubated for more than 14 days. The formulation was found to be stable even after 6 months at room temperature, accelerated conditions, and refrigerated conditions. As the formulation converted into gel at room temperature and accelerated conditions, the recommended storage condition of the formulation was refrigerated condition [70].

4.8.2 Naproxen Enteric-Coated Pellets

This study aimed at fluid-bed coating preparation of "naproxen enteric-coated pellets" (NAP-ECPs) by using QbD principle. Naproxen is a non-steroidal anti-inflammatory drug, and its solubility has a positive association with pH and also has a disadvantage of local gastric irritation. Thus, there was a need for the preparation of naproxen enteric-coated pellets since they offer advantages like less affected by gastric emptying rhythm, maximum drug absorption due to homogenous distribution in gastrointestinal tract (GIT), and reduced risk of GIT interaction. Plackett–Burman screening design was used for the determination of the significant factors that affected formulation composition. For obtaining the relationship between material attributes or process parameters and CQAs, Box–Behnken optimization design was used in the response surface method. Statistical analyses were done by analyzing results of Plackett–Burman study by statistical software of Minitab. The regression equation that described the linear, interactive, and quadratic effects of variable on the responses was used in case of Box–Behnken analyses.

The quantification of concentrations in dissolution medium of NAP was done by UV–visible spectroscopy (at wavelength 331 nm). Apparatus II (paddle method) was used for the dissolution studies with 50 rpm speed at a temperature of 37 °C. For the assessment of risk factors and corresponding causes, fish-bone diagram was constructed. For the risk analysis of pellet coating parameters, failure mode and effect analysis method (FMEA) was applied. Each variable was seen in terms of probability, detectability, and severity. Risk assessment consisted of two basic components (i.e., risk identification and risk analysis). In risk analysis study, eight high-risk factors were identified which were triethyl citrate (TEC) percentage, glycerol monostearate (GMS) percentage, spray rate, atomizing pressure, batch

size, coating aqueous dispersion solid content, coating weight gain, and curing time. Thus, for the identification of significant factors affecting CQAs, eight variables at two levels—12 runs Plackett–Burman screening design—were employed. From the results of screening study, it was observed that factors of significant impact on responses were found to be TEC, GMS percentage, and coating weight gain. To further examine these parameters for their interactions and effects on quality attributes, Box–Behnken DOE was used. From the results of Box–Behnken DOE, it was noted that spray rate had inverse impact on acid resistance (Y1) and cumulative drug release (Y2). All other remaining factors were found to be less significant. The responses surface was obtained through Box–Behnken DOE for Y1 and Y2.

For the evaluation of model significance, ANOVA was performed. The most suitable suggested model for responses of Y1 and Y2 was quadratic. From regression analysis, it was analyzed that with the increase in Y2, coating weight gain was inversely correlated. With the increase in TEC and GMS percentage, there was a significant increase in Y2. To evaluate the robustness and accuracy of obtained model, conformation tests with three levels (low, medium, and high) of the variables of all eight factors within control space were evaluated. Since there was a good agreement between model prediction and experimental observation, process parameters and formulation variables were found to be robust within control space. The achievement of process model of ECPs preparation was the final aim of this approach. Thus, based on this approach, the design space can be established and control space can be obtained further [71].

4.9 Conclusion

Pharmaceutical optimization is a vast field with a variety of applications in almost all the areas of pharmaceutical science. Optimization techniques have proven their role in reducing the product cost with minimized experimental trials and desirable quality attributes. The content presented in this chapter indicates the need of better understanding of product and process development, wherein QbD seems to be capable of giving solutions to the problems encountered giving assurance to the regulatory agencies as well to face approval-related challenges. The concepts relevant to product development or overall pharmaceutical quality system mentioned under ICH guidelines are essential for formulation development and processes involved therein under QbD which covers all the aspects like ensuring drug product quality profile, prioritizing input variables, and validation of QbD approach to ensure a quality end product. The results obtained so far are very encouraging, and so the regulatory agencies are continuously striving to convince the stakeholders for the effective implementation of QbD and the views of pharmaceutical industries are also encouraging. Except a few challenges in the effective implementation of QbD, it seems to be capable of taking care of all the concerns involved, and hence, QbD, in coming times, is anticipated to become a necessity in sciences. All the stakeholders and their associates should welcome its implementation with a positive intent.

References

1. Chodankar RS, Dev A (2016) Optimisaton techniques: a futuristic approach for formulating and processing of pharmaceuticals. Ind J Pharm Biol Res 4(2):32–40. https://doi.org/10.30750/ijpbr.4.2.5
2. Schwartz JB, Flamholz JR, Press RH (1973) Computer optimization of pharmaceutical formulations. I. General procedure. J Pharm Sci 62(7):1165–1170. https://doi.org/10.1002/jps.2600620722
3. Laínez JM, Schaefer E, Reklaitis GV (2012) Challenges and opportunities in enterprise-wide optimization in the pharmaceutical industry. Comp Chem Eng 47:19–28. https://doi.org/10.1016/j.compchemeng.2012.07.002
4. Lawrence XY, Amidon G, Khan MA, Hoag SW, Polli J, Raju G et al (2014) Understanding pharmaceutical quality by design. AAPS J 16(4):771–783. https://doi.org/10.1208/s12248-014-9598-3
5. Singh B, Kumar R, Ahuja N (2005) Optimizing drug delivery systems using systematic "design of experiments." Part I: fundamental aspects. Crit Rev Ther Drug Carrier Syst 22(1):27–105. https://doi.org/10.1615/CritRevTherDrugCarrierSyst.v22.i1.20
6. Roy S (2012) Quality by design: a holistic concept of building quality in pharmaceuticals. Int J Pharm Biomed Res 3(2):100–108
7. Mishra V, Kesharwani P, Amin MCIM, Iyer A (2017) Nanotechnology-based approaches for targeting and delivery of drugs and genes. Academic, Cambridge
8. Elliott P, Billingham S, Bi J, Zhang H (2013) Quality by design for biopharmaceuticals: a historical review and guide for implementation. Pharm Bioprocess 1(1):105–122. https://doi.org/10.4155/pbp.13.6
9. Rathore AS, Winkle H (2009) Quality by design for biopharmaceuticals. Nat Biotechnol 27(1):26–34. https://doi.org/10.1038/nbt0109-26
10. Kelley B, Cromwell M, Jerkins J (2016) Integration of QbD risk assessment tools and overall risk management. Biologicals 44(5):341–351. https://doi.org/10.1016/j.biologicals.2016.06.001
11. Djuris J, Ibric S, Djuric Z (2013) Quality-by-design in pharmaceutical development. In: Djuris S (ed) Computer-aided applications in pharmaceutical technology. Woodhead Publishing, Cambridge, pp 1–16. https://doi.org/10.1533/9781908818324.1
12. Singh L, Sharma V (2015) Quality by Design (QbD) approach in pharmaceuticals: status, challenges and next steps. Drug Deliv Lett 5(1):2–8. https://doi.org/10.2174/2210303104666141112220253
13. Gandhi A, Roy C (2016) Quality by design (QbD) in pharmaceutical industry: tools, perspectives and challenges. PharmaTutor 4(11):12–20
14. FDA (2004) Final report on pharmaceutical cGMPs for the 21st century—a risk based approach. http://www.fda.gov/cder/gmp/gmp2004/GMP_finalreport2004. Accessed 16 Mar 2021
15. Woodcock J, Woosley R (2008) The FDA critical path initiative and its influence on new drug development. Annu Rev Med 59:1–12. https://doi.org/10.1146/annurev.med.59.090506.155819
16. Adepu AR, Bhogale V (2015) Quality by design: a modern approach to pharmaceutical quality. World J Pharm Sci 4(5):2420–2433
17. Lawrence XY, Lionberger RA, Raw AS, D'Costa R, Wu H, Hussain AS (2004) Applications of process analytical technology to crystallization processes. Adv Drug Deliv Rev 56(3):349–369. https://doi.org/10.1016/j.addr.2003.10.012
18. Aksu B, De Beer T, Folestad S, Ketolainen J, Linden H, Lopes JA et al (2012) Strategic funding priorities in the pharmaceutical sciences allied to Quality by Design (QbD) and Process Analytical Technology (PAT). Eur J Pharm Sci 47(2):402–405. https://doi.org/10.1016/j.ejps.2012.06.009

19. ICH Harmonised Tripartite Guideline (2005) Quality risk management. Q9, current step. https://database.ich.org/sites/default/files/Q9_Guideline.pdf. Accessed 15 Feb 2021
20. ICH Harmonised Tripartite Guideline (2008) Pharmaceutical quality system q10 Current Step. https://database.ich.org/sites/default/files/Q10%20Guideline.pdf. Accessed 15 Feb 2021
21. ICH Harmonised Guideline (2019) Technical and regulatory considerations for pharmaceutical product lifecycle management Q12. https://database.ich.org/sites/default/files/Q12_Guideline_Step4_2019_1119.pdf. Accessed 10 Jan 2021
22. Basalious EB, El-Sebaie W, El-Gazayerly O (2011) Application of pharmaceutical QbD for enhancement of the solubility and dissolution of a class II BCS drug using polymeric surfactants and crystallization inhibitors: development of controlled-release tablets. AAPS PharmSciTech 12(3):799–810. https://doi.org/10.1208/s12249-011-9646-6
23. Singh B, Kapil R, Nandi M, Ahuja N (2011) Developing oral drug delivery systems using formulation by design: vital precepts, retrospect and prospects. Exp Opin Drug Deliv 8(10):1341–1360. https://doi.org/10.1517/17425247.2011.605120
24. Ye C, Liu J, Ren F, Okafo N (2000) Design of experiment and data analysis by JMP®(SAS institute) in analytical method validation. J Pharm Biomed Anal 23(2–3):581–589. https://doi.org/10.1016/S0731-7085(00)00335-6
25. Bhoop BS (2014) Quality by Design (QbD) for holistic pharma excellence and regulatory compliance. Pharm Times 46(8):26–33
26. ICH (2012) Quality IWGPoints to consider for ICH Q8/Q9/Q10 guidelines. European Medicines Agency. https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-use/q9/q10-guidelines_en.pdf. Accessed 10 Jan 2021
27. Shin S, Choi Du H, Truong NK, Kim NA, Chu KR, Jeong SH (2011) Time-oriented experimental design method to optimize hydrophilic matrix formulations with gelation kinetics and drug release profiles. Int J Pharm 407(1–2):53–62. https://doi.org/10.1016/j.ijpharm.2011.01.013
28. Gohel M, Amin AF (1998) Formulation optimization of controlled release diclofenac sodium microspheres using factorial design. J Control Release 51(2–3):115–122. https://doi.org/10.1016/S0168-3659(97)00102-8
29. Ladani R, Patel M, Rakesh P, Bhatt T (2010) Modern optimization techniques in field of pharmacy. Res J Pharm Biol Chem Sci 1(2):148
30. Huisman R, Van Kamp H, Weyland J, Doornbos D, Bolhuis G, Lerk C (1984) Development and optimization of pharmaceutical formulations using a simplex lattice design. Pharm Weekbl 6(5):185–194. https://doi.org/10.1007/BF01999941
31. Singh BS, Kapil R, Ahuja N (2008) Systematic optimisation of drug delivery systems: an insight. Pharm Rev 7:146–186
32. Khanam N, Alam M, Ali Y, Siddiqui A (2018) A review on optimization of drug delivery system with experimental designs. Int J Appl Pharm 10(2):7–12. https://doi.org/10.22159/ijap.2018v10i2.24482
33. Gavan A, Porfire A, Marina C, Tomuta I (2017) Formulation and pharmaceutical development of quetiapine fumarate sustained release matrix tablets using a QbD approach. Acta Pharma 67(1):53–70. https://doi.org/10.1515/acph-2017-0009
34. Chaudhari RA, Pingle AP, Chaudhari CS, Yewale C, Patil KA, Baroda V (2014) Quality by design (QbD): an emerging paradigm for development of pharmaceuticals. Int J Pharm 4(2):138–146
35. Raw AS, Lionberger R, Yu LX (2011) Pharmaceutical equivalence by design for generic drugs: modified-release products. Pharm Res 28(7):1445–1453. https://doi.org/10.1007/s11095-011-0397-6
36. Glodek M, Liebowitz S, McCarthy R, McNally G, Oksanen C, Schultz T et al (2006) Process robustness—a PQRI white paper. Pharm Eng 26(6):1–11

37. Blackburn TD, Mazzuchi TA, Sarkani S (2011) Overcoming inherent limits to pharmaceutical manufacturing quality performance with QbD (Quality by Design). J Pharm Innov 6(2):69–76. https://doi.org/10.1007/s12247-011-9102-x
38. Mesut B, ?zsoy Y, Aksu B (2015) The place of drug product critical quality parameters in quality by design (QBD). Turk J Pharm Sci 12(1):75–92
39. Singh SK, Venkateshwaran TG, Simmons SP (2010) Oral Controlled Drug Delivery: quality by design (QbD) approach to drug development. In: Wan H, Park K (eds) Oral controlled release formulation design and drug delivery. Wiley, New York, pp 279–303. https://doi.org/10.1002/9780470640487
40. Rathore AS (2014) QbD/PAT for bioprocessing: moving from theory to implementation. Curr Opin Chem Eng 6:1–8. https://doi.org/10.1016/j.coche.2014.05.006
41. Teckoe J, Mascaro T, Farrell TP, Rajabi-Siahboomi AR (2013) Process optimization of a novel immediate release film coating system using QbD principles. AAPS PharmSciTech 14(2):531–540. https://doi.org/10.1208/s12249-013-9935-3
42. Patil H, Feng X, Ye X, Majumdar S, Repka MA (2015) Continuous production of fenofibrate solid lipid nanoparticles by hot-melt extrusion technology: a systematic study based on a quality by design approach. AAPS J 17(1):194–205. https://doi.org/10.1208/s12248-014-9674-8
43. Yerlikaya F, Ozgen A, Vural I, Guven O, Karaagaoglu E, Khan MA et al (2013) Development and evaluation of paclitaxel nanoparticles using a quality-by-design approach. J Pharm Sci 102(10):3748–3761. https://doi.org/10.1002/jps.23686
44. Shah B, Khunt D, Bhatt H, Misra M, Padh H (2015) Application of quality by design approach for intranasal delivery of rivastigmine loaded solid lipid nanoparticles: effect on formulation and characterization parameters. Eur J Pharm Sci 78:54–66. https://doi.org/10.1016/j.ejps.2015.07.002
45. Visser JC, Dohmen WM, Hinrichs WL, Breitkreutz J, Frijlink HW, Woerdenbag HJ (2015) Quality by design approach for optimizing the formulation and physical properties of extemporaneously prepared orodispersible films. Int J Pharm 485(1–2):70–76. https://doi.org/10.1016/j.ijpharm.2015.03.005
46. Lionberger RA, Lee SL, Lee L, Raw A, Yu LX (2008) Quality by design: concepts for ANDAs. AAPS J 10(2):268–276. https://doi.org/10.1208/s12248-008-9026-7
47. Lee SL, O'Connor TF, Yang X, Cruz CN, Chatterjee S, Madurawe RD et al (2015) Modernizing pharmaceutical manufacturing: from batch to continuous production. J Pharm Innov 10(3):191–199. https://doi.org/10.1007/s12247-015-9215-8
48. Flåten GR, Ferreira AP, Bellamy L, Frake P (2012) PAT within the QbD framework: real-time end point detection for powder blends in a compliant environment. J Pharm Innov 7(1):38–45. https://doi.org/10.1007/s12247-012-9119-9
49. Beg S, Rahman M, Panda SS (2017) Pharmaceutical QbD: omnipresence in the product development lifecycle. Eur Pharm Rev 22(1):58–64
50. Berridge JC (2004) ICH Q8Pharmaceutical development. Pharmaceutical Quality Forum. http://www.nihs.go.jp/drug/PhForum/documents041122/Berridge041122.pdf. Accessed Mar 2021
51. EMA (2012) Guideline on real time release testing (formerly guideline on parametric release). https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-real-time-release-testing-formerly-guideline-parametric-release-revision-1_en.pdf. Accessed 10 Mar 2021
52. Yu LX, Amidon GL, Polli JE, Zhao H, Mehta MU, Conner DP et al (2002) Biopharmaceutics classification system: the scientific basis for biowaiver extensions. Pharm Res 19(7):921–925. https://doi.org/10.1023/a:1016473601633
53. Vaddemukkala Y, Syed M, Srinivasarao D (2015) Research article on optimization of olmesartan tablet formulation by 2^3 factorial design. Int J Res Pharm Nano Sci 4(4):188–195
54. Rath T, Do S (2006) Rath & Strong's six sigma pocket guide. AON Consulting Worldwide, Lexington, pp 1–180
55. Patil AS, Pethe AM (2013) Quality by Design (QbD): a new concept for development of quality pharmaceuticals. Int J Pharm Qual Ass 4(2):13–19

56. CDER (2007) Draft guidance for industry and review staff. Target product profile—a strategic development process tool. http://www.ncai-cc.ccf.org/skills/documents/U.S.%20FDA%20Target%20Product%20Profile%20Guidance%20Document%20(2007).pdf. Accessed 15 Mar 2021
57. FDA (2004) Guidance for industry, PAT—a framework for innovative pharmaceutical development, manufacturing and quality assurance. https://www.fda.gov/media/71012/download. Accessed 10 Mar 2021
58. Ohage E, Iverson R, Krummen L, Taticek R, Vega M (2016) QbD implementation and post approval lifecycle management (PALM). Biologicals 44(5):332–340. https://doi.org/10.1016/j.biologicals.2016.06.007
59. Kepert JF, Cromwell M, Engler N, Finkler C, Gellermann G, Gennaro L et al (2016) Establishing a control system using QbD principles. Biologicals 44(5):319–331. https://doi.org/10.1016/j.biologicals.2016.06.003
60. Menard F. Quality by design in generic drug development. Presentation to FDA Office of Generic Drugs, 25 Sept 2006
61. Moore C (2012) Quality by Design—FDA lessons learned and challenges for international harmonization. In: International conference on drug development, Austin, 27–29 Feb 2012. http://sites.utexas.edu/cpe-icdd/files/2015/11/12icddagenda.pdf. Accessed Mar 10 2021
62. Chen C-w (2011) Major regulators' perspectives on Quality-by-Design (QbD) implementation. https://www.diaglobal.org/productfiles/25849/20110517/track4/07%20t4-3_chen%20chi-wan.pdf. Accessed 10 Mar 2021
63. Dave VS, Saoji SD, Raut NA, Haware RV (2015) Excipient variability and its impact on dosage form functionality. J Pharm Sci 104(3):906–915. https://doi.org/10.1002/jps.24299
64. Wu H, Khan MA (2011) Quality-by-design: an integrated process analytical technology approach to determine the nucleation and growth mechanisms during a dynamic pharmaceutical coprecipitation process. J Pharm Sci 100(5):1969–1986. https://doi.org/10.1002/jps.22430
65. Rosencrance S (2011) QbD status update generic drugs. GPhA Fall technical conference, 3–5 Oct 2011. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Rosencrance+S.+QbD+status+update+generic+drugs.+GPhA+Fall+technical+conference&btnG=. Accessed 15 Mar 2021
66. Aksu B, Sezer AD, Yegen G, Kusçu L (2016) QbD implementation in biotechnological product development studies. In: Chen T, Chai S (eds) Special topics in drug discovery. Intech Open, Croatia, pp 133–155. https://doi.org/10.5772/62591
67. Sangshetti JN, Deshpande M, Zaheer Z, Shinde DB, Arote R (2017) Quality by design approach: regulatory need. Arab J Chem 10:S3412–S3S25. https://doi.org/10.1016/j.arabjc.2014.01.025
68. Varu RK, Khanna A (2010) Opportunities and challenges to implementing Quality by Design approach in generic drug development. J Generic Med 7(1):60–73. https://doi.org/10.1057/jgm.2009.37
69. Singh B, Beg S (2013) Quality by design in product development life cycle. Chronicle Pharmabiz 22:72–79
70. Dalwadi C, Patel G (2016) Implementation of "quality by design (QbD)" approach for the development of 5-fluorouracil loaded thermosensitive hydrogel. Curr Drug Deliv 13(4):512–527. https://doi.org/10.2174/1567201812666150817122506
71. Kan S, Lu J, Liu J, Wang J, Zhao Y (2014) A quality by design (QbD) case study on enteric-coated pellets: screening of critical variables and establishment of design space at laboratory scale. Asian J Pharm Sci 9(5):268–278. https://doi.org/10.1016/j.ajps.2014.07.005

Teaching Principles of DoE as an Element of QbD for Pharmacy Students

5

Rania M. Hathout

Contents

Abstract

Design of Experiment (DoE) is an emerging field that combines mathematics and statistics in order to establish successful cause–effect correlations between factors and responses. Moreover, with the help of planned experiments, accurate predictions of outcomes based on the constructed models can be achieved. The huge development and advent of the computer software, especially those possessing user-friendly interfaces, could help in relaying the concepts of DoE as a crucial element of quality to the undergraduate students. Implementing this concept in the mind of pharmacy students would have futuristic beneficial

R. M. Hathout (✉)
Department of Pharmaceutics and Industrial Pharmacy, Faculty of Pharmacy, Ain Shams University, Cairo, Egypt
e-mail: rania.hathout@pharma.asu.edu.eg

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_5

impacts on the pharmaceutical industry. This chapter introduces a simple way of conveying the topic to the students using fishbone diagrams and through conducting a step-by-step protocol of an experimental design used for modeling a drug–carrier.

Keywords

Design · Quality · Model · Optimization · Teaching · Software · Drug · Formulation · Delivery

Chapter Objectives

- Introducing the importance of teaching "Quality by Design" to pharmacy students
- Shedding the light on the "Design of Experiments" approach as an important element of "Quality by Design"
- Explaining the use of D-optimal designs as accurate and robust experimental designs
- Gaining information on a basic protocol of an experimental design dealing with a pharmaceutical case utilizing a user-friendly software package, Design Expert®

5.1 Introduction

The pharmaceutical industry and drug manufacturing and research are gaining more grounds every year. Accordingly, the search for new approaches and disciplines in order to optimize the pharmaceutical and engineering processes is emerging at high rates [1, 2]. Additionally, these optimization approaches are highly warranted in order to cut the huge expenses that the pharmaceutical companies exert in providing resources, energy, and labor for experimenting with new drug molecules and formulations and pairing the drug with its optimum carrier (reaching an optimized formulation of the drug–carrier pair) [3].

In this context, the term "Quality by Design" has evolved in order to introduce the concept of building quality and high standards in pharmaceutical products by the means of an initial good designing of the drug formulation and the manufacturing process conditions [4]. It involves a thorough understanding of a process, a goal, or an objective before the actual start of the process [5]. The International Conference on Harmonization includes Quality by Design (QbD) in Q8 pharmaceutical development and Q9 quality risk assessment guidelines. The main aim is that to stress the concept that quality should not only to be tested into products but also to be built in products by design. It can be thought of as a second generation of other quality terms such as the Good Manufacturing Practice as it reduces the risks and expenses through good designing [4].

The Design of Experiments (DoE) is a very important element of Quality by Design. Yet, unfortunately, it is the part that most pharmacists have the most trouble with. Therefore, teaching the basics of this element to undergraduate pharmacy students became inevitable as it would be a very useful update in the profession

and would also help to implement its ideas and concepts in the students' (afterward, graduates) minds. By implementing the approach, a great progress in the pharmaceutical industry could be achieved.

5.2 The Fishbone Diagram as a Means of Delivering the Ideas of DoE

The fishbone diagram (Ishikawa diagram or the cause–effect diagram) was found to be a very simple and clear way to approximate the DoE terms in the students' minds. It helps them to imagine the effect or the problem to be analyzed in a systematic way and hence, determine the main factors (causes) to a certain effect (response or outcome) [6].

An example of a simple form of a fishbone diagram is shown in Fig. 5.1 where generally, a problem or rather an effect is analyzed to five main causes gathered as 4 M + 1 E corresponding to materials, manpower, methods, machines, and environment, respectively.

Projecting this fishbone diagram specifically into pharmaceutical product formulation and manufacturing cases can be summarized in Fig. 5.2, where the analysis led to two main categories of causes or independent factors of the materials and formulation parameters at one hand and the critical process parameters (CPPs) at the other hand. The effects or responses were named critical quality attributes abbreviated as CQAs, and those were considered the dependent factors [7].

The students should comprehend that the ultimate goal of performing design of experiments is to carry few practical experiments and reach a mathematical model that correlates the independent factors (inputs) such as the CPPs and the materials

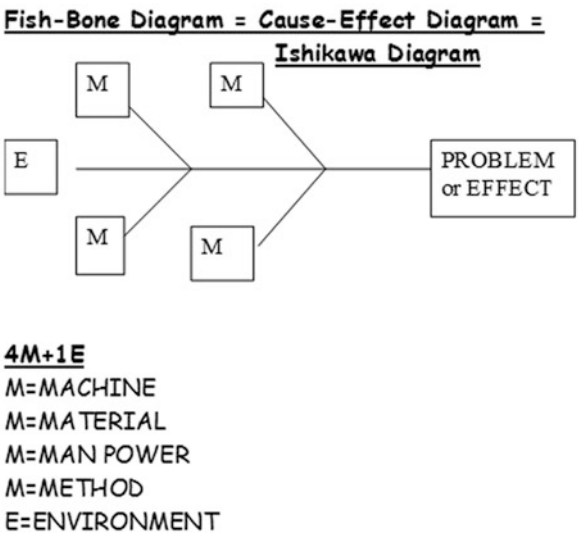


Fig. 5.1 A simple fishbone diagram

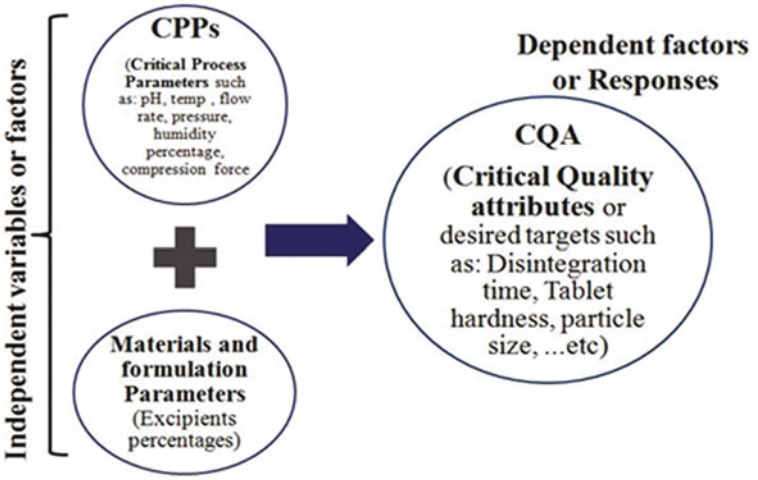


Fig. 5.2 Independent and dependent factors in pharmaceutical production cases

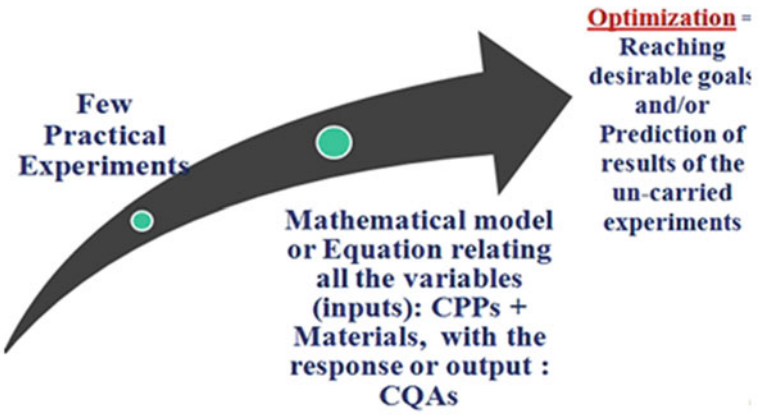


Fig. 5.3 Optimization of a pharmaceutical process

factors with the responses (outputs) aiming for an optimized outcome or process in addition to the ability of the prediction of the outcomes of un-carried experiments (Fig. 5.3).

5.3 Types of Experimental Designs

The undergraduate students could be successfully encountered with several commonly used experimental designs such as the factorial designs; full and fractional, the mixture, designs; D-optimal mixture, simplex lattice and the simplex centroid,

and the response surface designs; and central composite, Box–Behnken, D-optimal, and the distance-based. Table 5.1 summarizes the pros and cons of each of the mentioned designs.

5.4 D-Optimal Designs

The D-optimal designs (DOD) with their two types, response surface and mixture, are considered asymmetrical computer-based designs. They are very flexible designs that can be used in many cases where the conventional DoE protocols do not apply. They are suitable for experiments possessing irregular domains providing results with high accuracy. They are multi-level designs where the response surface can involve from 1 to 30 factors, while the mixture counterpart can accommodate from 2 till 24 factors. They work by selecting the best subset of all the possible experiments that would lead to space-filling sampling. Briefly, this type of design works by maximizing the determinant obtained from the information matrix generated from all the possible combinations of the levels of the involved factors. This explains the use of the letter "D" in the name of the design "D-optimal" designs. Accordingly, the design usually suggests runs that are considered rich-information points. Other optimality designs exist that also depend on the information matrixes such as the A-optimal (aims to minimizing the trace of the inverse of the information matrix which results in minimizing the average variance of the estimates of the regression coefficients) and the V-optimal (aims to minimizing the average prediction variance over a set of specific points). However, the D-optimal remains the most popular and the most robust [18]. DOD was utilized in optimizing many drug delivery systems and formulations including tablets, microparticles, self-nanoemulsifying drug delivery systems, and polymeric and lipidic nanoparticles [22, 23].

5.5 Step-by-Step Protocol to Handle a Drug Delivery Formulation Case Using Design Expert®

During the last decade, immense research has reported the feasibility of lipid nanocarriers as better, prolonged, and efficacious drug delivery systems to deliver the toxic and hydrophobic drugs, anticancer agents, and genetic materials [24]. Hereby, we introduce an optimization case study for rapidly emerging lipid nanocarrier, the lipid nanocapsules [25].

5.5.1 Design

The choice of the components required to prepare lipid nanocapsules together with their upper and lower was performed according to the feasibility domain for the formation of lipid nanocapsules as determined by Hertault et al., 2003 [26] (Fig. 5.4). The lipid nanocapsules to optimize were composed of Labrafac cc® (caprylic–capric

Table 5.1 Utilities, pros, and cons of different experimental designs

Experimental design		Utilities and pros	Cons	References
Screening Designs	Taguchi designs	– Used for screening the main effects rather than determination of the factors' interactions – Limited experimental runs and subsequently limited expenditure and time	– Low-resolution designs	[7]
	Plackett–Burman	– Used for screening the main effects rather than determination of the factors' interactions – Limited experimental runs and subsequently limited expenditure and time	– Low-resolution designs	[7]
Factorial Designs	Full-factorial	– Considers all the possible combinations of factors and levels according to the equation: Number of experiments = (number of levels)$^{\text{number of factors}}$ – Used when the number of factors and levels are reasonable	– Increasing the number of factors and/or levels leads to a large number of experiments that might not be feasible	[8–11]
	Fractional–factorial	– Works by reducing the number of experiments by subtracting number (s) from the power of the full-factorial equation leading to less number of experiments saving resources, time, and efforts	– Subtracting more numbers from the power can cause the design to lose its power to navigate the whole space of the experiments and not to represent the data – Low resolution – Can also be considered a screening design	[7]
Mixture Designs	Simplex lattice	– Used when the factors constitute a mixture and the total is 100% – Optimum in modeling areas or space of experiments representing a simplex (triangle)	– Best option for triangular experimental spaces but not the best in case of other domains	[12, 13]
	Simplex centroid	– Used when the factors constitute a mixture and the total is 100%	– Less accurate than Simplex lattice	[14]

(continued)

Table 5.1 (continued)

Experimental design		Utilities and pros	Cons	References
		– Optimum in modeling areas or space of experiments representing a simplex (triangle)		
	D-optimal mixture design	– Used when the factors constitute a mixture and the total is 100% – Optimum in modeling irregular areas or space of experiments	– Computer-based (needs algorithm so cannot be worked manually)	[15, 16]
Response Surface Designs	Central Composite Deigns (CCD)	– Perfectly navigates the experimental space as it widens the space by adding α and −α levels to each factor – Suitable for stable systems	– Unsuitable for delicate systems that widening the experimental space may cause instability or even loss of their structures	[17]
	Box–Behnken (BBD)	– Contrary to CCD, BBD avoids the corner points or combinations (No points of all high or all low levels of all of the factors) – Suitable for delicate systems or structures where exposing them to harsh conditions of the highest or the lowest levels of the factors at the same time is avoided	– Avoiding the corners may lead to loss of important information – Less accurate than D-optimal	[18–20]
	D-optimal design	– Contrary to CCD, it may be used for any system or structure – More accurate than BBD – Less number of experiments than the full-factorial	– Computer-based (needs algorithm so cannot be worked manually)	[18]
	Distance-based design	– Less experiments than the full-factorial	– Not accurate – Obsolete	[21]

triglyceride), Solutol HS 15® (poly-oxyethylene esters of 12-hydroxystearic acid), and salted water.

Since the space of the experiment (the feasibility domain) represents an irregular shaped area, then the D-optimal mixture design was selected to model and optimize

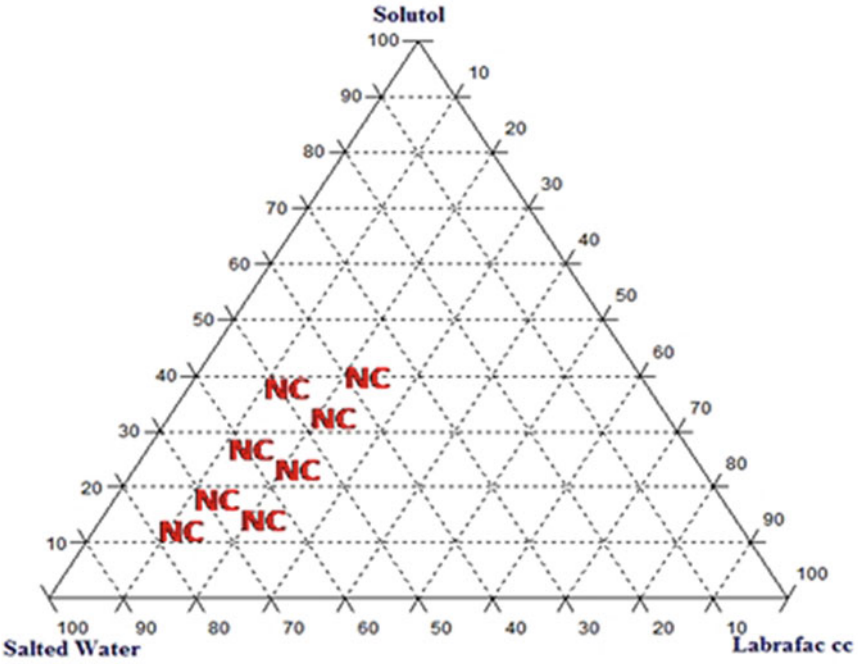


Fig. 5.4 Feasibility domain for lipid nanocapsule formation composed of Labrafac cc, Solutol, and salted water. NC: nanocapsules formation

Table 5.2 Upper and lower limits of the prepared lipid nanocapsules' components

Factor (variable)	Upper limit (%)	Lower limit (%)
Salted water	35	80
Labrafac cc	10	25
Solutol	10	40

the lipid nanocapsules particle size in this domain. Other mixture designs such as the simplex lattice and simplex centroid mixture designs were excluded as the experimental space does not represent a triangular domain. Besides, the simplex centroid usually suffers from decreased accuracy. The particle size was selected as the response or the optimized outcome. Particle size optimization is usually warranted in this kind of system for the better absorption and/or penetration and infiltration of the different body membranes and tissues.

In the upper and lower limits' percentages of the lipid nanocapsules, three main components were provided as follows (Table 5.2) [16]:

A D-optimal mixture design was constructed using point exchange. The candidate points of the model were vertices, centers of edges, axial check blends, interior blends, and overall centroid with an overall number of 12 points (including three replicate points).

The response: Particle size results for the designated runs are shown in Table 5.3:

Table 5.3 Particle size results of 12 D-optimal mixture design-generated runs

Run	Salted water (%)	Solutol (%)	Labrafac cc (%)	Particle size (nm)
1	35	40	25	26.81
2	65	25	10	19.56
3	50	40	10	16.36
4	35	40	25	25.61
5	65	10	25	73.5
6	80	10	10	33.36
7	65	10	25	72.61
8	80	10	10	32.77
9	42.5	40	17.5	21.2
10	65	17.5	17.5	35.76
11	72.5	10	17.5	56.56
12	50	25	25	39.54

These are considered the usually and basically asked questions for any experimental design (modeling case study) that can be easily demonstrated to pharmaceutic students:

What is the suggested model?
Is the model significant?
What is the r^2 of the model?
Provide the equation of the model
Provide a contour plot of the feasibility domain with respect to the provided response.
Predict the particle size of the following lipid nanocapsule formula containing 66% water, 24% surfactant, and 10% oil.

A brief description of the needed steps to construct and analyze the given data using the D-optimal mixture design is provided.

The steps were performed using Design Expert® trial version as follows:

- Start the software and open a new design (Fig. 5.5)
- As a default, the software usually opens on the two-level factorial screen. Choose "Mixture" from the left-hand panel (Fig. 5.6).
- Choose the "D-optimal" and construct the mixture design by choosing the number of the mixture components, naming them, determining their low and high levels, and finally including their total and corresponding units (Fig. 5.7).
- Choose the model points and adjust to the desired total number of runs. As an example, the total number of desired points was adjusted to 12 (Fig. 5.8).
- Enter all the values of the response results corresponding to each of the 12 total model runs (The particle size experimental results of the prepared different formulations) (Fig. 5.9).

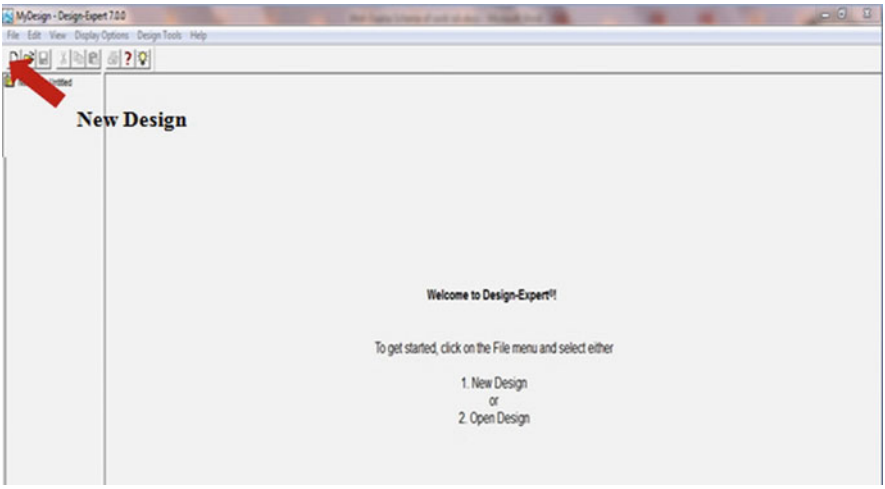


Fig. 5.5 Opening in a new design in Design Expert® software

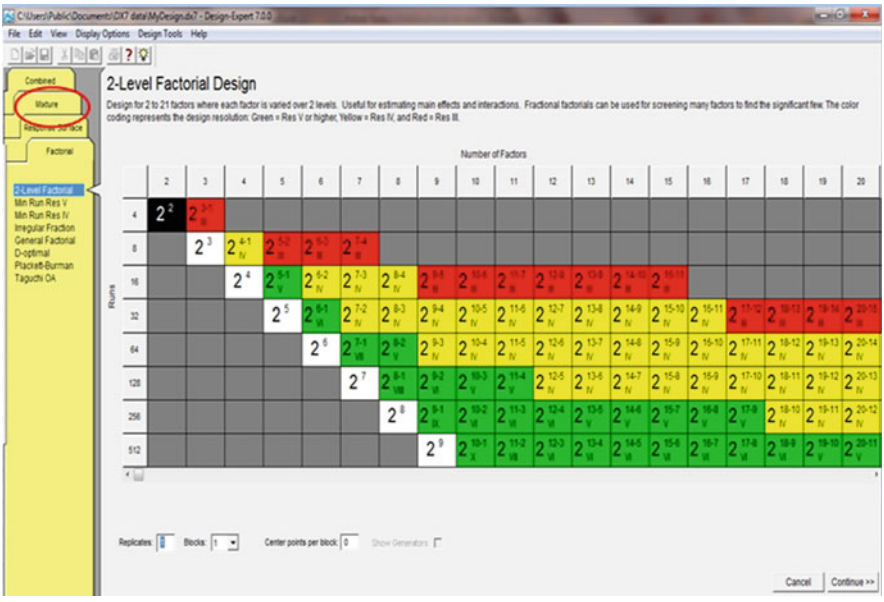


Fig. 5.6 Choosing the category of the mixture designs from the left panel of the Design Expert® software

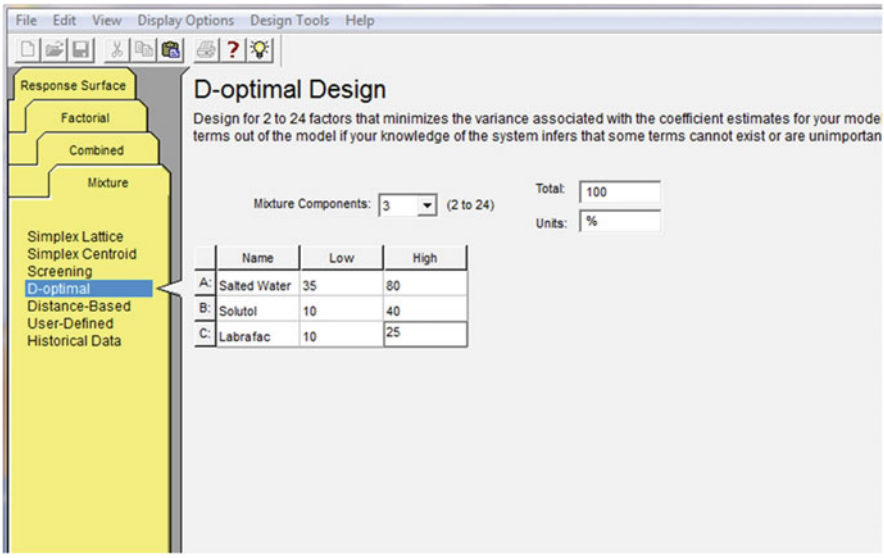


Fig. 5.7 Constructing the parameters of the required D-optimal mixture design

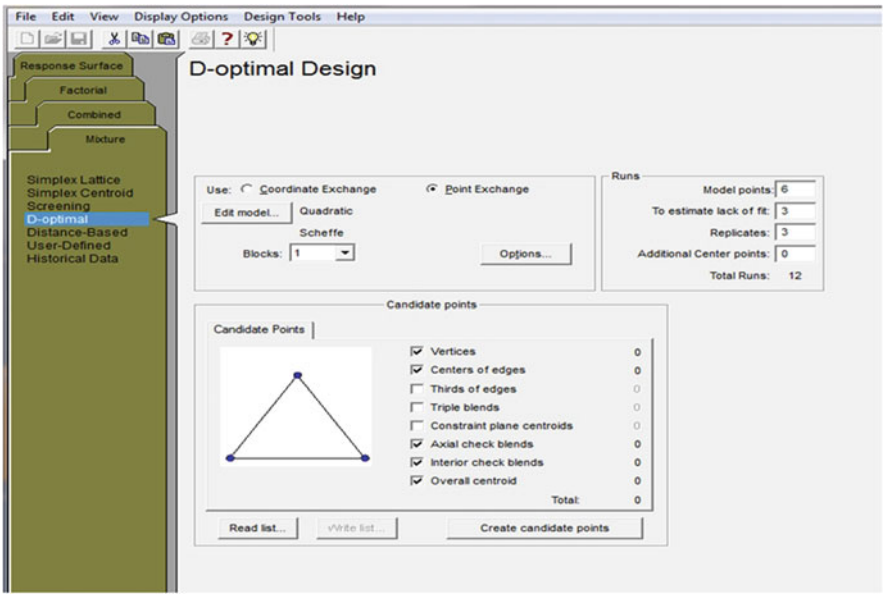
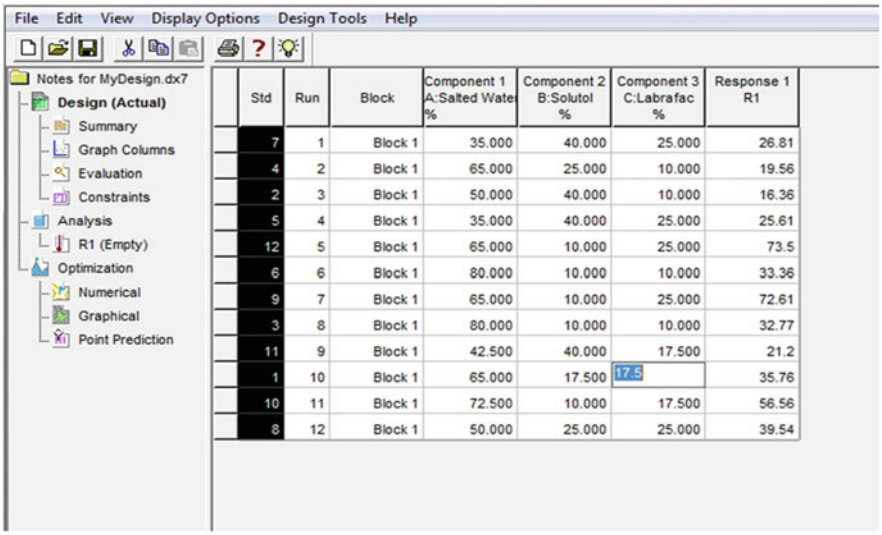


Fig. 5.8 Selecting the D-optimal mixture design model points



Std	Run	Block	Component 1 A:Salted Water %	Component 2 B:Solutol %	Component 3 C:Labrafac %	Response 1 R1
7	1	Block 1	35.000	40.000	25.000	26.81
4	2	Block 1	65.000	25.000	10.000	19.56
2	3	Block 1	50.000	40.000	10.000	16.36
5	4	Block 1	35.000	40.000	25.000	25.61
12	5	Block 1	65.000	10.000	25.000	73.5
6	6	Block 1	80.000	10.000	10.000	33.36
9	7	Block 1	65.000	10.000	25.000	72.61
3	8	Block 1	80.000	10.000	10.000	32.77
11	9	Block 1	42.500	40.000	17.500	21.2
1	10	Block 1	65.000	17.500	17.5	35.76
10	11	Block 1	72.500	10.000	17.500	56.56
8	12	Block 1	50.000	25.000	25.000	39.54

Fig. 5.9 Entering the values of the response results (particle size) that correspond to each of the 12 model runs

5.5.2 Analysis

- Start the analysis by clicking on the "fit Summary" tab where a model function is suggested. In the current case study, a "Quadratic" model was suggested as depicted in Fig. 5.10.
- Move to the next tab; "Model" and adjust the mix order to quadratic as suggested by the previous tab (Fig. 5.11).
- Press the "ANOVA" tab to carry the analysis of variance and determine the significance of the model. The studied model was significant with $P < 0.0001$. However, the lack-of-fit was also found significant which is unfavorable (Fig. 5.12)
- Scroll down to obtain the model parameters (Fig. 5.13) of *r*-squared indicating the goodness of fit of the results to the model equation, adjusted *r*-squared which is the fitting after removal of the insignificant terms of the model, and the predicted *r*-squared which determines the power of the model to predict the results of non-model points (un-carried experiments). Values larger than 0.7 with the values of the last former parameters not differing than a value of 0.2 imply a sufficient model [27–30]. Moreover, the adequate precision is essentially calculated. This parameter indicates the signal-to-noise ratio and the capability of the model to navigate the whole space of the experiment. In Design Expert®, values greater than 4 mean high signal-to-noise ratio [18, 29].
- More scrolling down would lead to the actual equation of the model relating the response (P.S.) to the components and their interactions (Fig. 5.14).

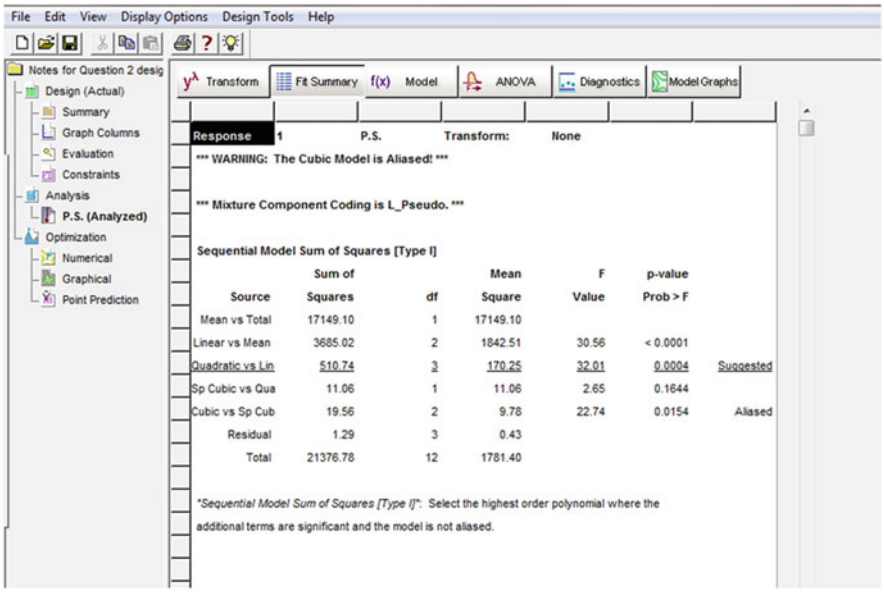


Fig. 5.10 Pressing "Fit Summary" in order to determine the suggested model function; Quadratic (underlined and denoted)

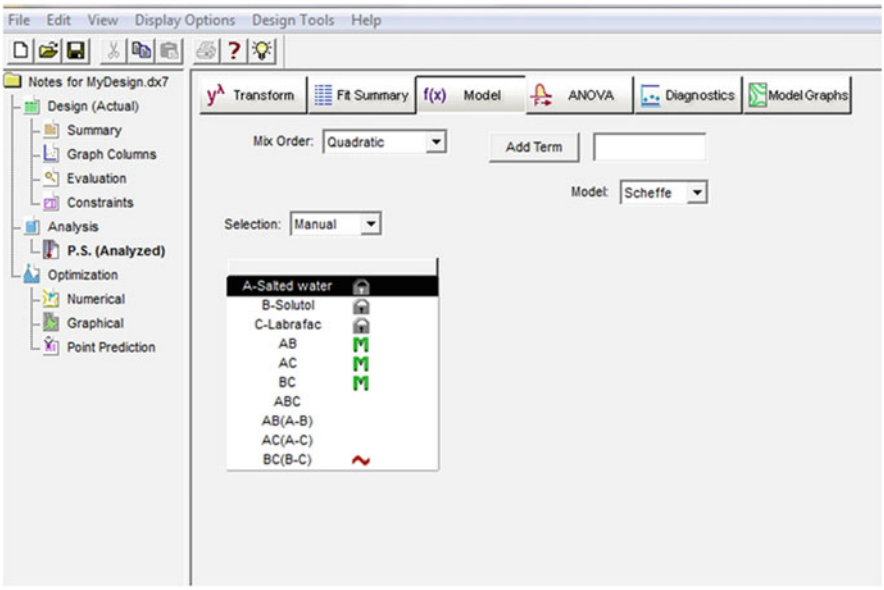


Fig. 5.11 Adjusting the model function

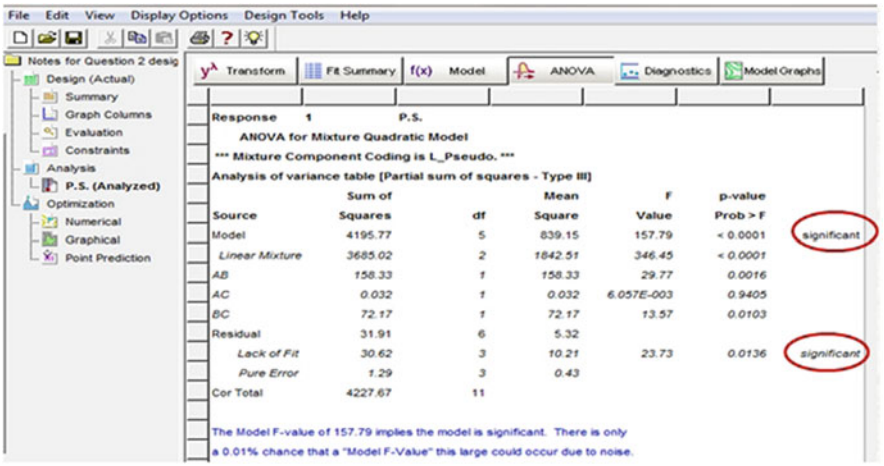


Fig. 5.12 Carrying ANOVA for the developed model to determine its significance

Fig. 5.13 Results of the model parameters

R-Squared	0.9925
Adj R-Squared	0.9862
Pred R-Squared	0.9666
Adeq Precision	34.640

Fig. 5.14 The actual equation of the generated model

P.S. =
+0.23459 * Water
+2.37957 * Surfactant
+3.65778 * Oil
–0.040407 * Water * Surfactant
–2.13668E–003 * Water * Oil
–0.11063 * Surfactant * Oil

5.5.3 Diagnostics

- Move to "Diagnostics" to diagnose the model, determine its flaws, and subsequently improve it (Fig. 5.15).

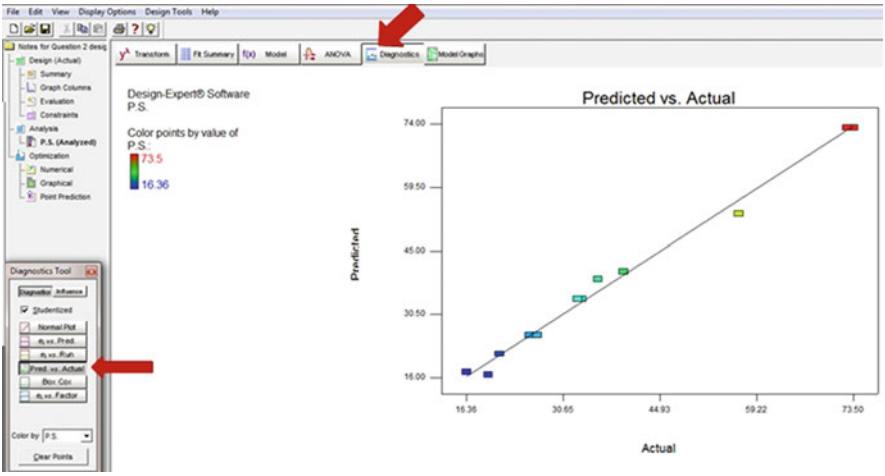


Fig. 5.15 Predicted versus actual diagnostic test of the model

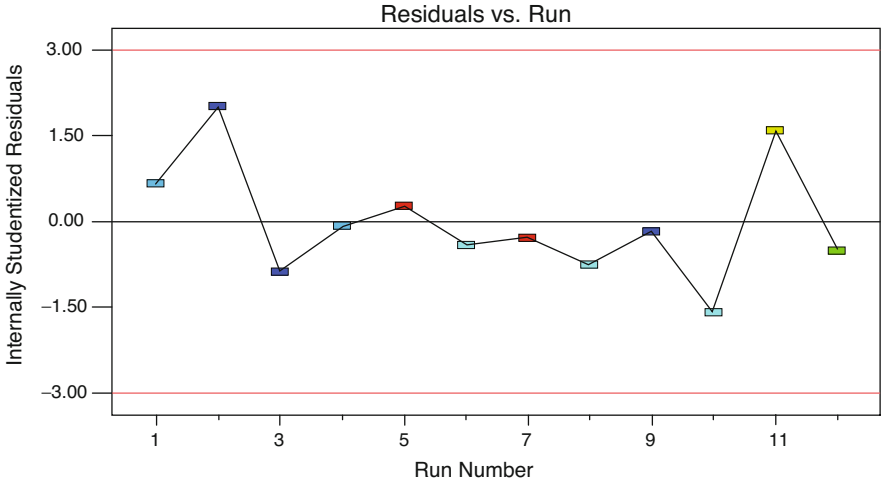


Fig. 5.16 Residuals versus run diagnostic test of the generated model

- First, choose "Pred vs. actual" from the "Diagnostic Tool" dialog box. Proximity of the points to the 45° line means the agreement of the predicted values to the actual counterparts [15].
- Choose "e_i vs. run" indicating the residuals versus run (Fig. 5.16). Scattering of the points (up and down) around the "0" line indicates a homogenous model.
- Select the "Box-Cox" tab, where a plot of ln the residuals sum of squares versus lambda (the power of the response) would be provided (Fig. 5.17). This test provides a recommendation of the best lambda (power) of the response that yields the lowest value for ln of the residuals sum of squares (i.e., least error). Locating

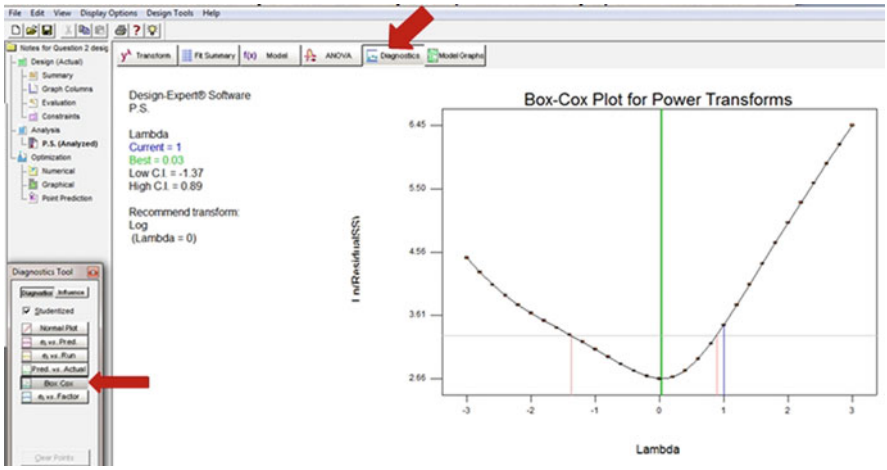


Fig. 5.17 Box-Cox diagnostic test of the generated model

the current lambda outside the best lambda confidence intervals warrants the need for power transformation [18].

5.5.4 Optimization

- Press the tab "y^{λ} Transform" to adjust the power (constant = 0) of the response (Fig. 5.18) as recommended by the Box-Cox diagnostic test at the previous screen (According to the used software package, a value of 0 corresponds to a log function).
- Adjusting the lambda (power) of the response led to a non-significant lack-of-fit (favorable) (Fig. 5.19). All the model parameter values are consequently improved (Fig. 5.20).
- For better optimization, model reduction can also be performed by returning again to the "Model" tab and clicking on the letter "M" beside any non-significant term (as determined by the ANOVA test) leading to its removal from the model (Fig. 5.21).
- The model reduction led to further improvement in the model values (Fig. 5.22).
- Scrolling down the ANOVA screen would lead to the reduced actual equation as follows (Fig. 5.23):

5.5.5 Model Graphs

- Move to the "Model Graphs" tab and through "View" choose "Contour" to obtain the contour plot (Fig. 5.24). Moving from blue to red colors indicates moving from low response values to higher ones [12].

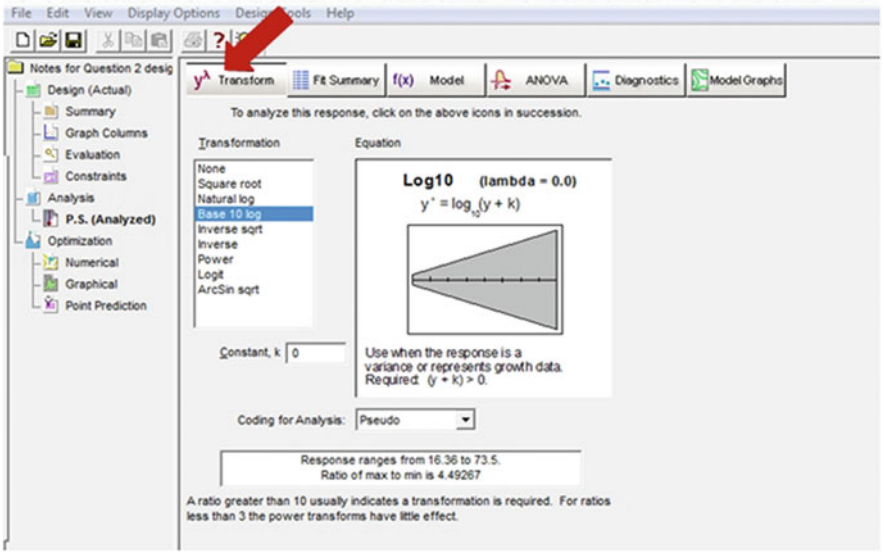


Fig. 5.18 Transforming the lambda (power) of the response as recommended by the Box-Cox diagnostic test

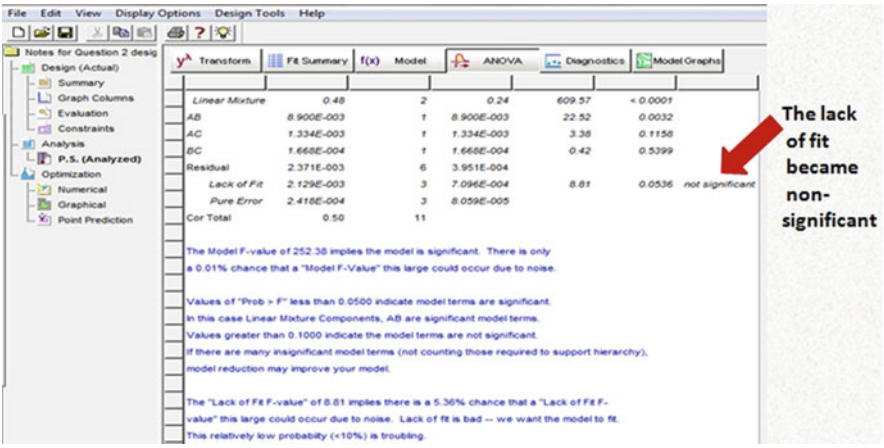


Fig. 5.19 ANOVA results after power transformation

Fig. 5.20 Improvement of the model values after power transformation

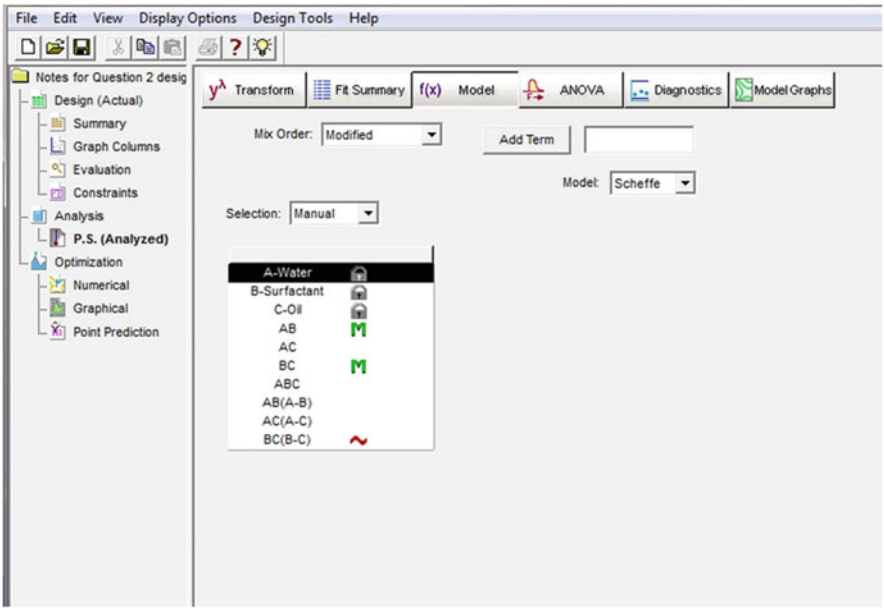


Fig. 5.21 Model reduction by removal of the insignificant terms

Fig. 5.22 Improvement of the model values after model reduction

R-Squared	0.9924
Adj R-Squared	0.9881
Pred R-Squared	0.9779
Adeq Precision	40.970

Fig. 5.23 The reduced actual equation of the generated model

P.S. =

+0.22909 * Water

+2.37400 * Surfactant

+3.51284 * Oil

−0.040420 * Water * Surfactant

−0.10846 * Surfactant * Oil

- From "View" choose "3D surface" and obtain the 3D surface plot (Fig. 5.25) where the response is displayed at a third dimension. From the "Display Options" tab, choose "Response in Original Scale" in order to display the response dimension in its original rather than its transformed scale (log).

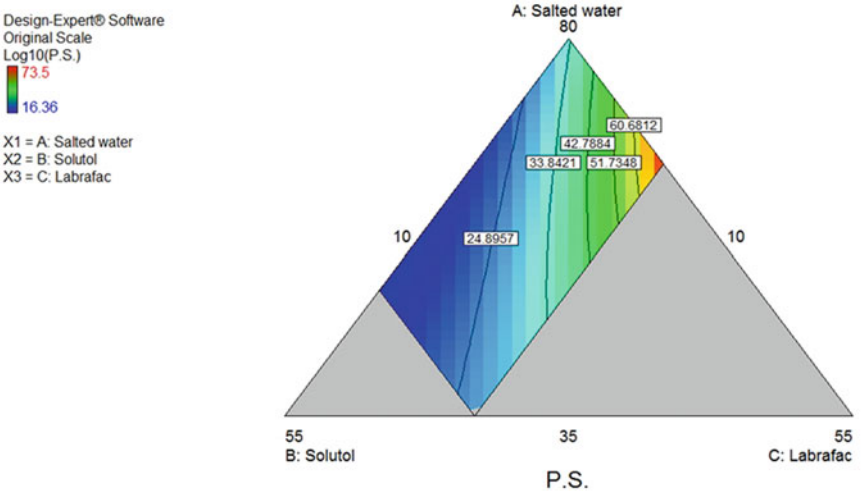


Fig. 5.24 Contour plot of the generated model

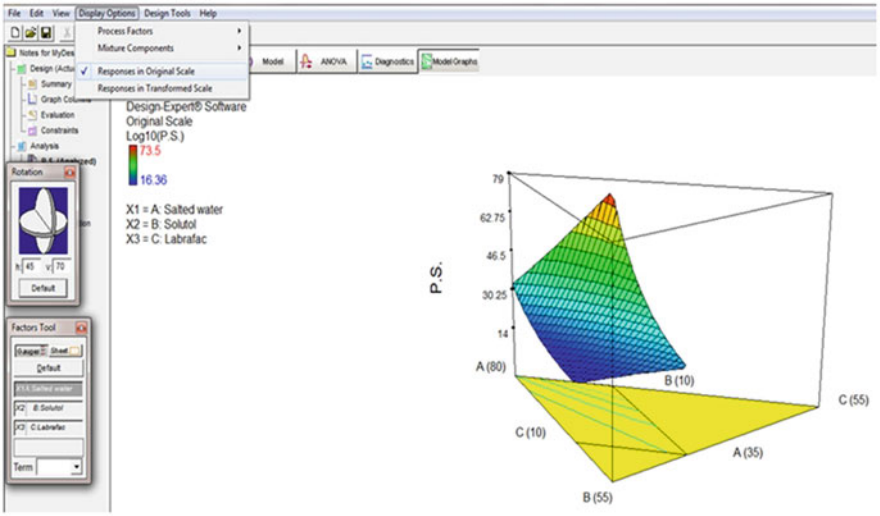


Fig. 5.25 3D surface plot of the generated model

5.5.6 Prediction

- Finally, from the left panel, choose "Point Prediction" (Fig. 5.26) in order to predict the result of the response (P.S.) of an un-carried run (experiment) of the

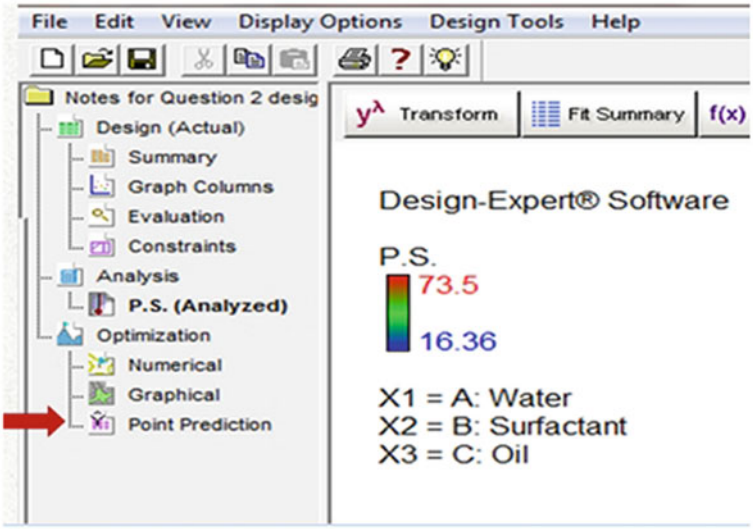


Fig. 5.26 Point prediction of the response result of a non-model formulation

following prepared formulation consisting of 66% salted water, 24% surfactant (Solutol), and 10% oil (Labrafac cc).

- Adjust the values of the factors as desired from the "Factors Tool" dialog box. Consequently, the predicted result will be displayed under the "Prediction" title (Fig. 5.27).

5.6 Conclusions and Future Prospects

Quality by design (QbD) is an important discipline that should be incorporated in pharmaceutics and drug delivery courses. We hereby conclude that there is significant potential for the Design of Experiments (DoE) activities as an important element of QbD to be taught in the undergraduate programs of pharmacy schools. The staff can help to engage the students with real case studies from the industry or from research-based studies via problem-solving and systematic thinking activities. The developed skills would lead to graduates with special abilities to implement quality through good designing of drug formulations and dosage forms that would ultimately and beneficially impact the economy and society by impacting the pharmaceutical industry positively.

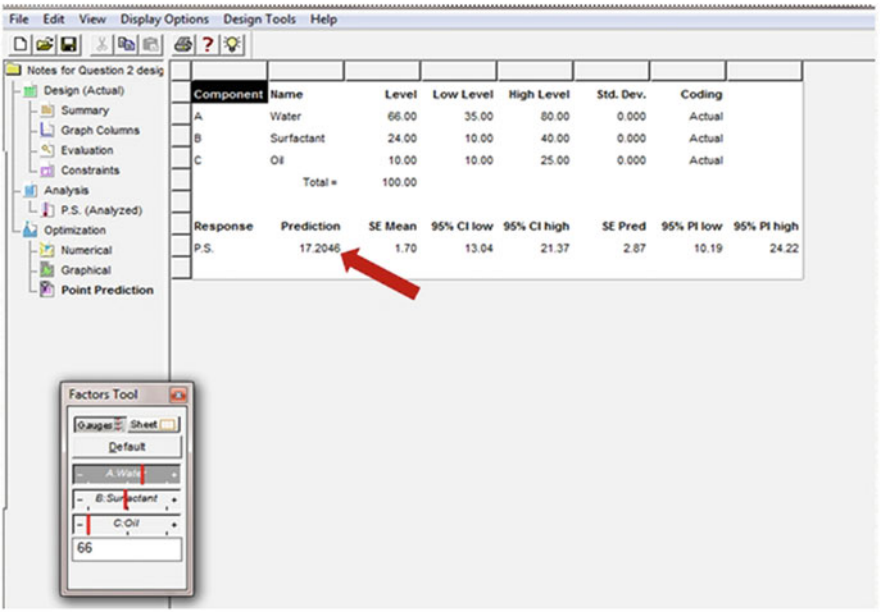


Fig. 5.27 Adjusting the values for a point prediction using the "Factors Tool" dialog box and obtaining the result

References

1. Amit AP, Ravish JP (2018) Nanomedicine for intranasal delivery to improve brain uptake. Curr Drug Deliv 15:461–469. https://doi.org/10.2174/1567201814666171013150534
2. Hathout RM, Metwally AA (2016) Towards better modelling of drug-loading in solid lipid nanoparticles: molecular dynamics, docking experiments and Gaussian Processes machine learning. Eur J Pharm Biopharm 108:262–268. https://doi.org/10.1016/j.ejpb.2016.07.019
3. Metwally AA, Hathout RM (2015) Computer-assisted drug formulation design: novel approach in drug delivery. Mol Pharm 12:2800–2810. https://doi.org/10.1021/mp500740d
4. Swarbrick B (2018) Chapter 6—Quality by design in practice. In: Ferreira AP, Menezes JC, Tobyn M (eds) Multivariate analysis in the pharmaceutical industry. Academic, pp 125–171. https://doi.org/10.1016/B978-0-12-811065-2.00007-2
5. Wang L, Dong J, Chen J, Eastoe J, Li X (2009) Design and optimization of a new self-nanoemulsifying drug delivery system. J Colloid Interface Sci 330:443–448. https://doi.org/10.1016/j.jcis.2008.10.077
6. Beg S, Hasnain MS, Rahman M, Swain S (2019) Chapter 1—Introduction to Quality by Design (QbD): fundamentals, principles, and applications. In: Beg S, Hasnain MS (eds) Pharmaceutical quality by design. Academic, pp 1–17. https://doi.org/10.1016/B978-0-12-815799-2.00001-0
7. Beg S, Swain S, Rahman M, Hasnain MS, Imam SS (2019) Chapter 3—Application of Design of Experiments (DoE) in pharmaceutical product and process optimization. In: Beg S, Hasnain MS (eds) Pharmaceutical quality by design. Academic, pp 43–64. https://doi.org/10.1016/B978-0-12-815799-2.00003-4

8. Abdelhamid HN, El-Bery HM, Metwally AA, Elshazly M, Hathout RM (2019) Synthesis of CdS-modified chitosan quantum dots for the drug delivery of Sesamol. Carbohydr Polym 214:90–99. https://doi.org/10.1016/j.carbpol.2019.03.024
9. Mehanny M, Hathout RM, Geneidi AS, Mansour S (2017) Studying the effect of physically-adsorbed coating polymers on the cytotoxic activity of optimized bisdemethoxycurcumin loaded-PLGA nanoparticles. J Biomed Mater Res 105:1433–1445. https://doi.org/10.1002/jbm.a.36028
10. Hathout RM, Gad HA, Metwally AA (2017) Gelatinized-core liposomes: toward a more robust carrier for hydrophilic molecules. J Biomed Mater Res A 105:3086–3092. https://doi.org/10.1002/jbm.a.36175
11. Hathout RM, Gad HA, Abdel-Hafez SM, Nasser N, Khalil N, Ateyya T et al (2019) Gelatinized core liposomes: a new Trojan horse for the development of a novel timolol maleate glaucoma medication. Int J Pharm 556:192–199. https://doi.org/10.1016/j.ijpharm.2018.12.015
12. Yehia R, Hathout RM, Attia DA, Elmazar MM, Mortada ND (2017) Anti-tumor efficacy of an integrated methyl dihydrojasmonate transdermal microemulsion system targeting breast cancer cells: in vitro and in vivo studies. Colloids Surf B Biointerfaces 155:512–521. https://doi.org/10.1016/j.colsurfb.2017.04.031
13. Mandlik SK, Adhikari S, Deshpande AA (2012) Application of simplex lattice design in formulation and development of buoyant matrices of dipyridamole. J Appl Pharm Sci 2:107–111. https://doi.org/10.7324/JAPS.2012.21221
14. Mukherjee R, Halder A, Sansare S, Naik S, Chaudhuri B (2020) A simplex centroid design to quantify triboelectric charging in pharmaceutical mixtures. J Pharm Sci 109:1765–1771. https://doi.org/10.1016/j.xphs.2020.02.001
15. Naguib SS, Hathout RM, Mansour S (2017) Optimizing novel penetration enhancing hybridized vesicles for augmenting the in-vivo effect of an anti-glaucoma drug. Drug Deliv 24:99–108. https://doi.org/10.1080/10717544.2016.1233588
16. Safwat S, Hathout RM, Ishak RA, Mortada ND (2017) Augmented simvastatin cytotoxicity using optimized lipid nanocapsules: a potential for breast cancer treatment. J Liposome Res 27:1–10. https://doi.org/10.3109/08982104.2015.1137313
17. Aziz DE, Abdelbary AA, Elassasy AI (2018) Implementing central composite design for developing transdermal diacerein-loaded niosomes: ex vivo permeation and in vivo deposition. Curr Drug Deliv 15:1330–1342. https://doi.org/10.2174/1567201815666180619105419
18. Abdel-Hafez SM, Hathout RM, Sammour OA (2014) Towards better modeling of chitosan nanoparticles production: screening different factors and comparing two experimental designs. Int J Biol Macromol 64:334–340. https://doi.org/10.1016/j.ijbiomac.2013.11.041
19. Beg S, Swain S, Singh HP, Patra C, Rao ME (2012) Development, optimization, and characterization of solid self-nanoemulsifying drug delivery systems of valsartan using porous carriers. AAPS PharmSciTech 13:1416–1427. https://doi.org/10.1208/s12249-012-9865-5
20. Honary S, Ebrahimi P, Hadianamrei R (2014) Optimization of particle size and encapsulation efficiency of vancomycin nanoparticles by response surface methodology. Pharm Dev Technol 19:987–998. https://doi.org/10.3109/10837450.2013.846375
21. Marengo E, Todeschini R (1992) A new algorithm for optimal, distance-based experimental design. Chemom Intell Lab Syst 16:37–44. https://doi.org/10.1016/0169-7439(92)80076-G
22. Abdel-Hafez SM, Hathout RM, Sammour OA (2018) Tracking the transdermal penetration pathways of optimized curcumin-loaded chitosan nanoparticles via confocal laser scanning microscopy. Int J Biol Macromol 108:753–764. https://doi.org/10.1016/j.ijbiomac.2017.10.170
23. Mohamad Zen NI, Abd Gani SS, Shamsudin R, Fard Masoumi HR (2015) The use of D-optimal mixture design in optimizing development of Okara tablet formulation as a dietary supplement. Sci World J 2015:684319. https://doi.org/10.1155/2015/684319
24. Safwat S, Ishak RAH, Hathout RM, Mortada ND (2017) Nanostructured lipid carriers loaded with simvastatin: effect of PEG/glycerides on characterization, stability, cellular uptake efficiency and in vitro cytotoxicity. Drug Dev Ind Pharm 43:1112–1125. https://doi.org/10.1080/03639045.2017.1293681

25. Safwat S, Ishak RA, Hathout RM, Mortada ND (2017) Statins anticancer targeted delivery systems: re-purposing an old molecule. J Pharm Pharmacol 69:613–624. https://doi.org/10.1111/jphp.12707
26. Heurtault B, Saulnier P, Pech B, Venier-Julienne MC, Proust JE, Phan-Tan-Luu R et al (2003) The influence of lipid nanocapsule composition on their size distribution. Eur J Pharm Sci 18:55–61. https://doi.org/10.1016/S0928-0987(02)00241-5
27. Abdelrady H, Hathout RM, Osman R, Saleem I, Mortada ND (2019) Exploiting gelatin nanocarriers in the pulmonary delivery of methotrexate for lung cancer therapy. Eur J Pharm Sci 133:115–126. https://doi.org/10.1016/j.ejps.2019.03.016
28. Arai H, Suzuki T, Kaseda C, Takayama K (2009) Effect of an experimental design for evaluating the nonlinear optimal formulation of theophylline tablets using a bootstrap resampling technique. Chem Pharm Bull (Tokyo) 57:572–579. https://doi.org/10.1248/cpb.57.572
29. Shokry M, Hathout RM, Mansour S (2018) Exploring gelatin nanoparticles as novel nanocarriers for Timolol Maleate: augmented in-vivo efficacy and safe histological profile. Int J Pharm 545:229–239. https://doi.org/10.1016/j.ijpharm.2018.04.059
30. Ossama M, Hathout RM, Attia DA, Mortada ND (2019) Enhanced allicin cytotoxicity on HEPG-2 cells using glycyrrhetinic acid surface-decorated gelatin nanoparticles. ACS Omega 4:11293–11300. https://doi.org/10.1021/acsomega.9b01580

Computer-Assisted Manufacturing of Medicines

6

Lalji Baldaniya and Bhumika Patel

Contents

L. Baldaniya
Faculty of Pharmacy, Marwadi University, Rajkot-Morbi Highway, Rajkot, Gujarat, India
e-mail: lalji.baldaniya@marwadieducation.edu.in

B. Patel (✉)
Department of Pharmaceutical Chemistry, Institute of Pharmacy, Nirma University, Ahmedabad, Gujarat, India
e-mail: bhumika.patel@nirmauni.ac.in

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_6

Abstract

Computer-aided manufacturing (CAM), in the twenty-first century, is the automated manufacturing process that works in line with computer-assisted design (CAD) and allows machines to build artefacts directly from the fed designs into the software without any manual intervention to set up machines and processes. CAM machines can manufacture thousands of similar products automatically in very little time. As conventional pharmaceutical manufacturing processes are very complex, the use of CAM technology in the production of tablets, powder, and liquids is far more accurate, precise, cost-effective, timely, and responsive to the goals of manufacturing with higher quality standards. An emerging CAM technique, 3D printing, has even made it possible to manufacture personalized formulations as per the individual patient's need, and inclusive of potent drug having a narrow therapeutic index. In 2016, the very first 3D printlet of Levetiracetam (approved by USFDA) was successfully manufactured by Aprecia Pharmaceuticals, Pennsylvania which is used to treat epilepsy. Despite the exciting recent advances in 3D printing, there are still notable scientific and regulatory obstacles. As a result, the most groundbreaking applications of this technology will take time to get developed.

Keywords

Additive manufacturing · Computer-assisted manufacturing · Computer-aided design · 3D printing · Rapid prototyping · Inkjet printing · Tailor-made drug delivery system · Personalized medicine

Chapter Objectives

- To explore revolutionary CAM technology in the field of pharmaceuticals
- To understand various facets of additive manufacturing to avail dose personalization, multi-drug combinations, tailored release profiles, improved patient acceptability, and on-demand and precise medicine printing
- To obtain detailed information for various 3D printing technologies and software and hardware requirements
- To get an insight into 3D printing application in innovative drug delivery devices
- To discuss the advantages, risks, and challenges of 3D printing in the manufacturing of personalized medicine.

- To understand the importance of solvent in the design of smart polymers for drug encapsulation and release
- To know the differences in the behaviour of linear, hyperbranched, and dendrimers' polymers

6.1 Introduction

Computer-aided manufacturing (CAM), in the twenty-first century, is the automated manufacturing process with the help of software and computer-controlled machinery. Computer-assisted manufacturing typically works in line with computer-assisted design (CAD) and allows machines to build artefacts directly from the fed designs into the software without any manual intervention to set up machines and processes. CAM software converts data and blueprint or sketch of the object into detailed instructions of the manufacturing process which help designers to send specifications to machines directly. Therefore, CAM machines can manufacture thousands of similar products automatically in very little time. CAD software is mainly used to create model blueprints.

The pharmaceutical manufacturing processes are deeply rooted in complexities. The days of manually addressing process problems are long gone and have been replaced by CAM technology that is far more cost-effective, timely, and responsive to the goals of manufacturing facilities. The method may include powder, tablets, and liquids in pharmaceutical manufacturing, each of which relies on quantifying equipment for the processing of products and packaging equipment. CAM software applications include safe, precise drug component measurements, and define packaging forms for pharmaceutical liquid and dry products. Perhaps the greatest advantage of CAM is that it produces a higher quality standard and higher volumes of products with greater accuracy and precision. CAM results in cost savings for many manufacturers by reducing the need for materials used in the processing and also by a reduction in waste. As technology evolves, manufacturing processes that are assisted by computers can result in a wider range of opportunities to expand their output while reducing production costs.

CAD and CAM have been a part of the biomedical industry to provide absolute precision wherever possible in the manufacturing field of clinical medicine, dentistry, personalized medical implants, artificial joints, tissue engineering, biomedical engineering, and robotic surgery using 3D printing (additive manufacturing) technology. Applications of 3D printing are also rapidly expanding in the field of pharmaceutical research for drug discovery, dosage formulation, and delivery. In 2016, the very first 3D printlet of Levetiracetam (approved by the U.S. Food and Drug Administration (USFDA)) was successfully manufactured by Aprecia Pharmaceuticals, Pennsylvania which is used to treat epilepsy. The company is also focusing on developing three other 3D printed products, which is expected to enter the market eventually. It should be noted that despite these exciting recent healthcare advancements involving additive manufacturing, there are still notable

scientific and regulatory obstacles. As a result, the most groundbreaking applications of this technology will take time to get developed.

Currently, the focus of customized or personalized medicines is adaptive drug dose, targeting, and delivery according to the patient's needs, especially with the dosage of active pharmaceutical ingredient (API) with a narrow therapeutic window. Traditionally, the oral tablet manufacturing method includes the use of large batches, various production phases, specified and costly installations, and professional operators. Combined with its static design, the high cost of this method made it less ideal for preparing personalized medicine. Ideally, it should be (1) affordable, (2) safe, (3) highly adjustable, (4) network-controllable, and (5) should occupy minimal space for a production system to meet the new challenges of patient-specific medicine. In contrast to a traditional technique, a newer CAD-based three-dimensional (3D) printing technique consists of laying powder bed initially followed by a multilayer, 3D deposition of a binder solution from the print head. Existing 3D printers operate on a high-degree heart involvement, which in turn is responsible for the degradation of APIs (drugs). Technologies such as Stereolithography (SLA), Fused Deposition Modelling (FDM), and Selective Laser Sintering (SLS) are available for engineering-based spare parts design and development. However, these are inadequate for pharmaceutical ingredients that are thermolabile. Therefore, the binder jetting method would be the most suitable to allow layer-by-layer aggregation of the unit mass of particles in the tablet dosage shape using a benchtop 3D printer. This low-cost production method uses thermally stable pharmaceutically approved polymers and fillers for the manufacturing of individualized dosage forms.

Computer-aided manufacturing (CAM) is a set of techniques that integrate various sub-techniques for the design and manufacture of products using digital technologies in various stages of manufacturing like inventory control, automatic ordering of materials, factory scheduling, predicting material usage and machine changeover, projecting manpower requirements, etc. It was initially developed to ensure the transition from 2D design to 3D manufacturing and is often used after computer-aided design (CAD) [1]. CAM process converts CAD data directly and very fast without any manual intervention in the workflow which enables organizations to take advantage of a fast response. The design of the product itself controls the machine to manufacture the product. Because of this innovation in manufacturing technology, nowadays, mass production based on customized design becomes very easy, efficient, and fast.

6.2 Computer-Assisted Dosage Form Manufacturing

Due to the high patient compliance, solid oral dosage forms of drugs constitute approximately 90% of all dosage forms. Especially, tablets are widely accepted and used by patients all over the world due to their availability in diverse types like conventional compressed, enteric/film/sugar-coated, effervescent, chewable, sublingual, vaginal tablets, etc. [2].

The revolutionary CAD-assisted three-dimensional (3D) printing is an emerging and promising innovative manufacturing technology that rapidly converts digital information of 3D models into physical objects by adding layer upon layer of material. ISO/ASTM (American Society for Testing and Materials) has defined it as "the process of joining materials to make parts from 3D model data, usually layer upon layer" [3].

The required technologies for fabricating a 3D product are 3D modelling software (computer-aided design or CAD), a computer-controlled 3D printing machine printer, and appropriate layering/printing material for the object. The layering materials commonly used are metal, plastics, ceramics, glass, liquids, or even living cells and tissues in bioprinting.

In the engineering and biomedical industry, 3D printing technology has been widely used [4]. Many other widely used terms for 3D printing are rapid prototyping (RP), additive manufacturing (AM), layered manufacturing, or solid freeform fabrication. In contrast to conventional manufacturing technique which is called subtractive manufacturing, 3D printing constructs the objects in a layer-by-layer fashion and hence called "additive manufacturing". Originally, this concept was developed at MIT and involved the printing of a liquid binder onto a thin powder bed. Subsequently, various types of 3D printers with special applications in pharmaceutical products have been developed. At first, CAD software creates a file of the 3D object to be printed/manufactured which is then exported to the 3D printer. According to this file information, the object is printed layer by layer with limited restrictions on its spatial arrangement. The technology prints a single material or a combination of multiple materials with precise regulation of the shape of every individual layer. With the recent advancement of technology, 3D printing of functional tissue constructs (called "bioprinting") using conventional biocompatible materials and viable cells is also possible [5]. Bioprinting is an emerging technique that has the potential to construct desired tissues/organs for various biomedical applications like organ transplantation or cancer drug screening [5, 6].

6.3 History of 3D Printing

The concept of 3D printing (3DP) was introduced long back in 1884 by Charles Hull when he invented the first-ever 3D printing technology, called stereolithography (SLA) [7]. He first ever used a CAD file to interact with a rapid prototyping system, developed computer-modelled objects successfully, and finally, the SLA technology was commercialized into the first 3D printer named SLA-1 in 1987 [8]. Later, other technologies also became available for 3D printing (Fig. 6.1) like selective laser sintering [11], inkjet 3D printing [12], fused deposition modelling [13], etc. The FDM 3D printer was launched in 1991 by Stratasys. In the early 90s, Z-corp released their 3D printer based on the technology called Z-printing (Inkjet printing) [14].

Over the last 15 years, innovations in 3D printing technologies like robotic-assisted printing and laser-assisted bioprinting have been evolved where the idea of RP has been transformed into AM and this made the widespread application of

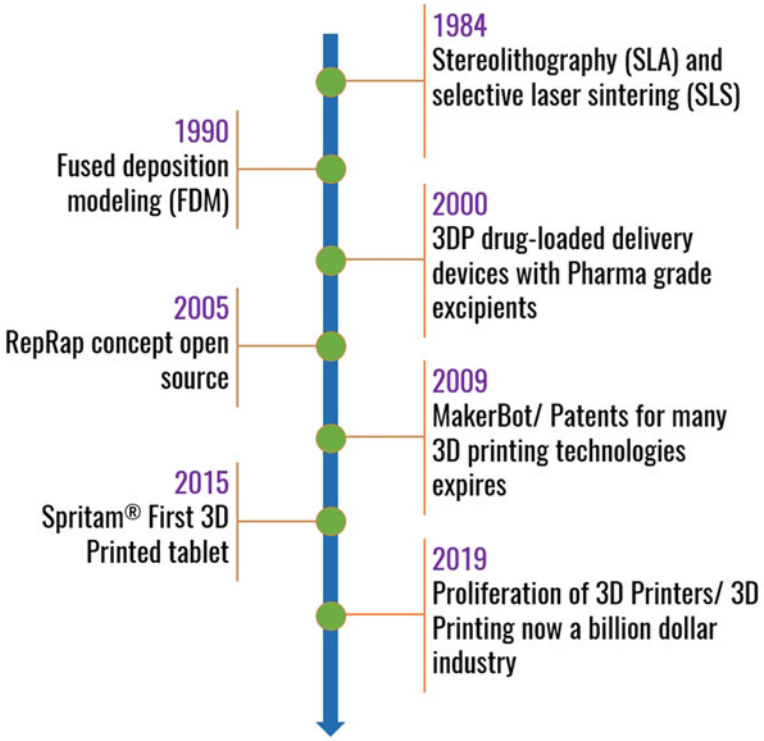


Fig. 6.1 Timeline of 3D printing technologies [9, 10]

them into a biomedical field [15]. The modern era of bioprinting started in the early twenty-first century with inkjet technology when the first-ever human kidney synthetic scaffolds were bio-printed at the Wake Forest Institute for Regenerative Medicine, North Carolina, United States. It was then coated with actual cells taken from the human patient and was successfully implanted into humans. This was a breakthrough discovery that started the era of regenerative medicines which has a strong potential to revolutionize the entire healthcare system [16]. Later, the ASTM was established in 2009 dedicated to the specifications of standards for AM [17].

In 2013, McKinsey Global Institute, Washington, DC, USA, identified 3D printing as one of the twelve potentially disruptive technologies along with mobile Internet, robotics, autonomous vehicles, and others. Disruptive technologies are those characterized by the rapid and high rate of technology change, high potential for social and economic impacts, thereby transforming life, business, and the global economy. 3D printing has this potential. According to the literature, AM/3DP is expected to affect about 12% of the global workforce currently employed in conventional manufacturing by 2025 [18].

Hideo Kodama of Nagoya, Japan created the first 3D printed plastic objects in 1981. In the same year, French General Electric filed a patent application for a process of 3D printing but soon withdrew it as they did not foresee significant business potential in it. The same year US engineer Chuck Hull filed his patent application for stereolithography in which photopolymers were printed in layers and cured by ultraviolet laser creating a 3D object. Chuck Hull is regarded by many as the pioneer of 3D printing. The 1980s and 1990s saw rapid progress in the development of several printing technologies based on inkjet printing, laser sintering, and extrusion processes such as fused filament fabrication and using a variety of materials like metal powders, thermoplastics, and ceramics for creating 3D objects.

3D printing was also being investigated and rapidly developed for 3D models to assist in medical diagnostics, in surgery planning, for making dental implants and medical prosthetics. From being a research and laboratory curiosity, 3D printing had come to its own as a mainstream technology.

In the year 2007, McKinsey filed a patent as Disruptive Technology, Morgan Stanley research Blue Paper published on MedTech: 3D printing reported that medical applications topped the list for patents applied for with 38% of all applications in 2012. In terms of the number of patents granted, medical (28%) ranked second [19].

The applications of AM/3DP in various segments of health care are as below:

(a) In medical diagnosis and care: 3D imaging coupled with 3D printing can help physicians in the diagnosis and planning of complex surgeries.
(b) Personalized medicine: It opened up avenues for mass customization not possible otherwise in areas like dentistry and prosthetics.
(c) It made new solutions hitherto not available; 3D printed organs and tissues have been made possible aided by other medical developments in stem cells.

On an industrial scale, the advantages are faster production of prototypes of new models and parts, on-demand production leading to lower manufacturing inventories, less wastage of materials compared to conventional "subtractive" manufacture, shorter supply chains, and the possibility of manufacturing more complex geometries hitherto very difficult.

More and more uses were being conceived in multifarious fields by individuals, academic research institutes, and industrial organizations. While initially the technology was used in making models and prototypes quickly for new designs, it soon made inroads into manufacturing too, notably using metals and plastics. Precision parts, prototypes, and tooling in the automotive and aerospace industries were the major beneficiaries of this trend.

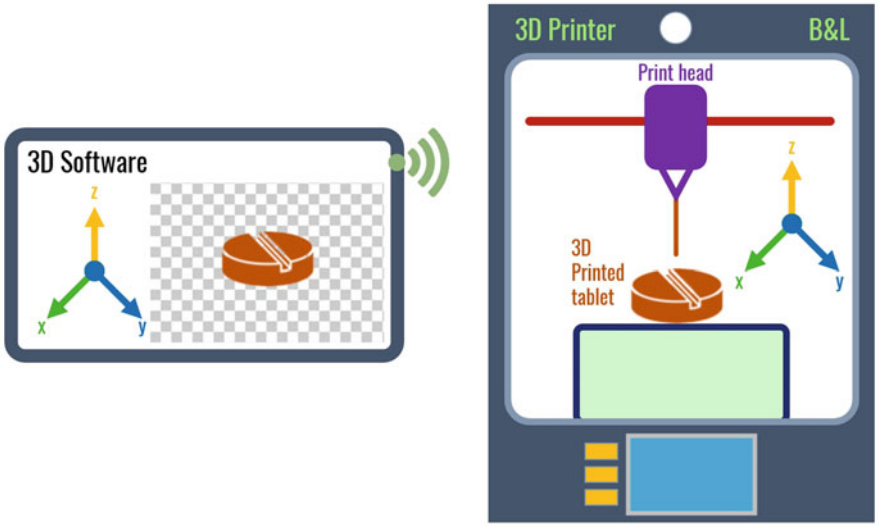


Fig. 6.2 Basic diagram of 3D printer to show the main components (Additional info www.makerbot.com/learn/ [20])

6.4 Basic Components of 3D Printing

The three main components of 3D printing [8] are (a) hardware, a 3D printer itself, (b) CAD software to process the blueprint image so that a printer can recognize the design, and (c) printing material as shown in Fig. 6.2.

6.5 Three-Dimensional Printing Processing Steps

First, the CAD programme or software designs the 3D pattern of the object (tablet) and converts it into a. STL file which is then converted into a G-code file or any other file extension and exported to the printer. Here, the basic information of the. STL file about the design of the object is divided into a series of layers of specific thickness. Then, layer-by-layer approach, materials (drug and excipients) are deposited to construct the object.

The first step in the process of 3D printing is the 3D imaging of the body tissue or part. The three common medical imaging technologies are X-rays, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) scanning. Another technique, cone beam computed tomography using X-rays in a divergent beam, is used for 3D images in dentistry, orthopaedics, and interventional radiography. Digital Imaging and Communications in Medicine (DICOM) is the internationally accepted standard to transmit, store, retrieve, print, process, and display medical information

from 3D imaging. International Standard Organization (ISO) has recognized the DICOM standard as ISO 12052:2017 applicable for information related to the production and management of 3D images, between both medical imaging equipment and systems such as X-ray, MRI, CT, and devices used for their management such as PCs, servers, scanners, and other computer hardware.

The second step is to convert the DICOM data into special software developed for 3D printing. As 3D printers got developed independently by several companies and institutions in several countries, more and more such software came into existence. The de-facto, the most widely used file format, is. STL (Standard Triangle Language), developed by the Albert Consulting Group, a US company of 3D systems, for the stereolithography process. This format is supported by many other packages. As the newer technologies were developed, drawbacks of. STL format such as lack of provisions to represent colour, texture, material, and other attributes were realized. ASTM developed an international standard additive manufacturing file format. Later, ISO and ASTM have joined hands to standardize it into ISO/ASTM 52915: 2013 (revised as 52915: 2016) to ensure compatibility among the various systems used [21].

6.6 3D Printing Software

Digital information of the 3D object is the first and foremost requirement of 3D printing which is generated in a modelling programme or software. The instructions from the software are exported to the 3D printers which then construct the 3D object by joining the materials layer by layer until they become complete.

The process of translating the 3D object through its breaking down into a readable file format for the printer is known as slicing. A 3D printing software, also known as a slicer, effectively performs this job of slicing and generates a readable language, known as G-code, for the 3D printer at the end.

Table 6.1 represents the list of CAD software along with their function, minimum skill level of the user, and system requirement. Most of this software can be tried for free, and there are free tutorial videos available for all of them.

A cartridge of the 3D printer is filled with the printing material and required colours to build the objects. Utmost care must be taken to optimize the design of the model through the software before final printing so that model neither breaks nor deform during printing. Nowadays, many advanced tools are available which are compatible with the printing software to simulate the 3D model efficiently.

6.6.1 MakerBot Print Software

A desktop application, MakerBot Print, allows to preparation, management, and monitor 3D prints. The latest version is MAKERBOT PRINT 4.10.1. Detailed information on this software would be acquired from https://www.makerbot.com/3d-printers/apps/makerbot-print.

Table 6.1 List of software used for 3D printing

Software	Function	Level	System requirement
3D Builder	Design	Beginner	Windows
3D Slash	Design	Beginner	Browser
3DPrinterOS	STL Editor	Beginner	Windows, Mac, Ubuntu, Raspberry Pi
3D-Tool Free Viewer	STL Analysis	Intermediate	Windows
AstroPrint	Slicer, 3D Printer Host	Beginner	Browser
Blender	Design	Professional	Windows, Mac, Linux
FreeCAD	Design	Intermediate	Windows, Mac, Linux
Fusion 360	Design	Intermediate	Windows, Mac
KISSlicer	Slicer	Intermediate	Windows, Mac, Linux
MakerBot Print	Slicer, 3D Printer Design	Beginner	Windows, Mac, Linux
MatterControl 2.0	Slicer, 3D Printer Host, Design	Beginner	Windows, Mac, Linux
MeshLab	STL Editor, STL Repair	Professional	Windows, Mac, Linux
Netfabb	STL Repair, Slicer	Professional	Windows
OctoPrint	Slicer, 3D Printer Host	Intermediate	Windows, Mac, Linux
OnShape	Design	Professional	Browser
OpenSCAD	Design	Intermediate	Windows, Mac, Linux
PrusaSlicer	Slicer	Beginner	Windows, Mac, Linux
Repetier-Host	Slicer, 3D Printer Host	Intermediate	Windows, Mac, Linux
Rhinoceros 3D	Design	Beginner	Windows, Mac, Linux
SketchUp Free	Design	Intermediate	Browser
SliceCrafter	Slicer	Intermediate	Browser
Solidworks	Design	Beginner	Windows, Mac, Linux
TinkerCAD	Design	Beginner	Browser
Ultimaker Cura	Slicer, 3D Printer Host	Beginner	Windows, Mac, Linux
Vectary	Design	Intermediate	Browser
ZBrushCoreMini	Design	Beginner	Windows, Mac

6.7 Various 3D Techniques Used to Fabricate DDS

Some of the methods used for 3D printing are stereolithography, photolithography, magnetic bioprinting, and direct cell extrusion which are described in the next section [22].

3D printing technologies are over 40 3D printing technologies developed to date and in use. There are various groups of 3D printing technologies as per ISO based on their main characteristics. These are as follows [3]:

1. Binder jetting technique in which fine droplets of liquid binder are used to form bonding between powder particles,

2. Guided energetic deposition technique in which molten material is being fused through guided heat energy,
3. Extrusion technique in which material is extruded out in the form of the cylindrical extrudate,
4. Material jetting technique in which fine mist of liquid binder is being deposited selectively on the given powder bed,
5. Powder layer fusion technique in which selective located powder layer is to be fused through heat energy,
6. Sheet layer technique in which layered sheets are fused to built object,
7. Vat photo-activated polymerization technique in which selective curing of liquified photo-sensitive polymer is done using selective radiation.

Some of the technologies have several variants. For example, thermal inkjet (TIJ) and piezoelectric inkjet also called drop-on-demand to come under the classification of material jetting. Stereolithography comes under photopolymerization. Fused Deposition Modelling (FDM) and fabrication of fused filament are extrusion technologies. In the following sections, we will take a brief look at some of the more commonly used technologies especially in medical applications [16].

6.7.1 Stereolithographic 3D Printing (SLA)

The photopolymerization process, called stereolithography, was the first 3D printing process developed by Chuck Hull of the Unite States in 1984 [8, 23]. Figure 6.3 shows a simplified depiction of the setup for stereolithography.

The main parts of this 3D printing setup are (1) a reservoir for liquified polymeric material, (2) a height-adjustable perforated base plate, (3) a UV radiation source, and (4) a control unit for radiation and base plate.

In the beginning, the laser is exposed to the photopolymer just above the perforated platform. The UV laser moves exactly tracing the geometry of the object to be printed, and a 3D digital model becomes available with the controlling computer. The UV curable photopolymer instantly hardens along the path traced by the laser and forms the initial slice of the 3D object.

When the platform drops a fraction of a millimetre, the laser gets exposed to the liquid plastic and moves in the pattern of the next layer. The new layer then bonds and hardens with the first layer. This process gets repeated again and again forming successive layers of the object till the entire solid object gets formed. The perforated platform is raised, and the solid object formed is rinsed with a suitable solvent to remove the photopolymer resin. Finally, the object is baked and cured in a UV oven.

Stereolithography has been used in medical modelling since the 1990s. First, the part to be modelled is scanned in CT or MRI. Here, different tissues are recorded as different shades of grey. The data are translated into digital files suitable for 3D modelling and converted to the actual 3D model by stereolithography. The process has been extensively used in diagnosis, complex surgery planning, for example, brain surgery, and for designing and making implants.

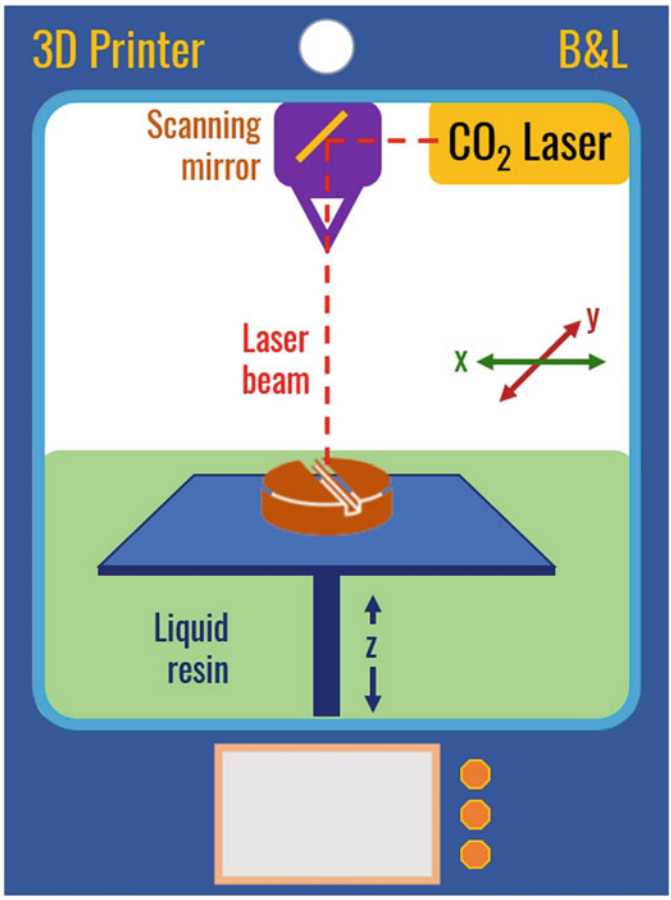


Fig. 6.3 Schematic of the stereolithography printer (Additional info, URL: rb.gy/wnnqt7)

6.7.2 Fused Filament Fabrication/Fused Deposition Modelling (FDM) 3D Printing

Fused deposition modelling is widely being used in extrusion processes on a benchtop 3D printer. In this process, the solid filament of thermoplastic materials like polylactic acid (PLA) or acrylonitrile butadiene styrene (ABS) is heated and melted in the print head (extruder). The print head can move in the horizontal axis in the *X*- and *Y*-directions while the tray, supporting the build platform, moves in the vertical *z*-axis. While moving in the *x*- and *y*-axes, the print head goes on depositing the melted filament till one layer or section of the object gets completed. When the first film formation is complete, the build platform moves lowered down by a fraction of a millimetre, and the second layer is deposited by the printing head on

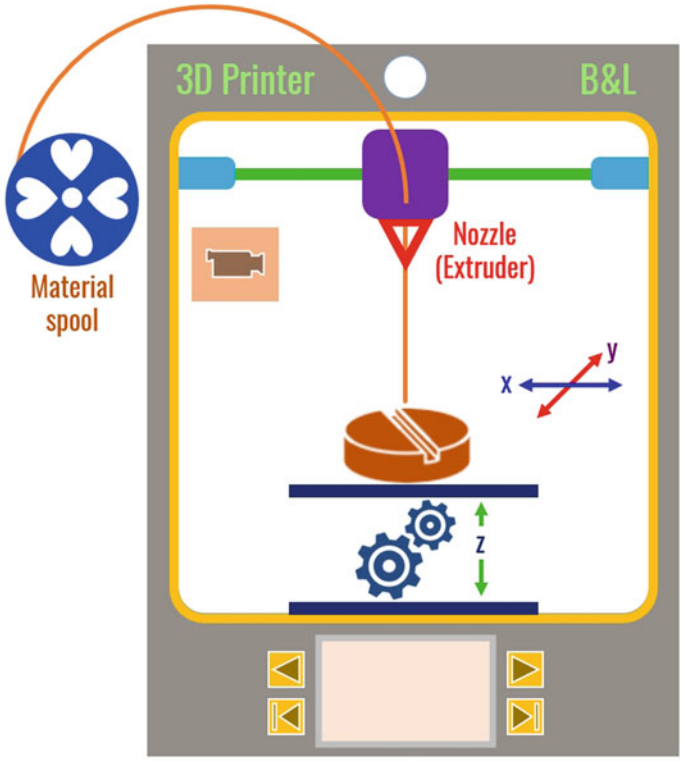


Fig. 6.4 Schematic of the fused deposition modelling printer

top of the first. The layers keep on getting fused till the end of the 3D print as illustrated in Fig. 6.4 (Additional info, URL: rb.gy/tptnmx) [23].

6.7.3 Binder Deposition

A liquid formulation is used as the adhesive material. A fine mist is deposited onto a powder layer, which is composed of APIs and excipients at an optimal rate to fabricate a 3D object [24]. Here, the powder layer is kept on a base plate, accomplished by lifting the bottom of the powder distribution base plate placed next to it. Usually, a scraping roll is used to push the powder through the bed. The movement of the liquid binder laden print head is programmed to move on the *x*-*y* axis and spray the liquid onto the thin layer of powder. The liquid binds dust particles together, causing the layer to solidify. After that base plate moves down along the *z*-axis. The successive steps are repeated to build a 3D object by the distribution of a subsequent thin layer of powder on top.

6.7.4 Selective Laser Sintering (SLS) 3D Printing

Lasers are used as energy sources to sinter metal powders or thermoplastics. As the laser traces the contours of the digital 3D model on the powder surface, the material exposed to the laser selectively gets fused. The product platform gets lowered by a tiny distance once a layer is completed. Intricate shapes, not possible earlier, can now be fabricated by SLS, making it one of the widely employed 3D printing methodologies, despite the high cost of powered laser required. Figure 6.5 depicts the schematic setup.

Few thermoplastic materials used in SLS are polyamides, polyether ether ketone, PLA, and polycaprolactone (PCL). Ceramic powder with polyvinyl alcohol and epoxy resin-based composite is also used. Surgical guides in orthopaedics, dental reconstruction parts as well as 3D models for help in surgical planning are some of the medical applications.

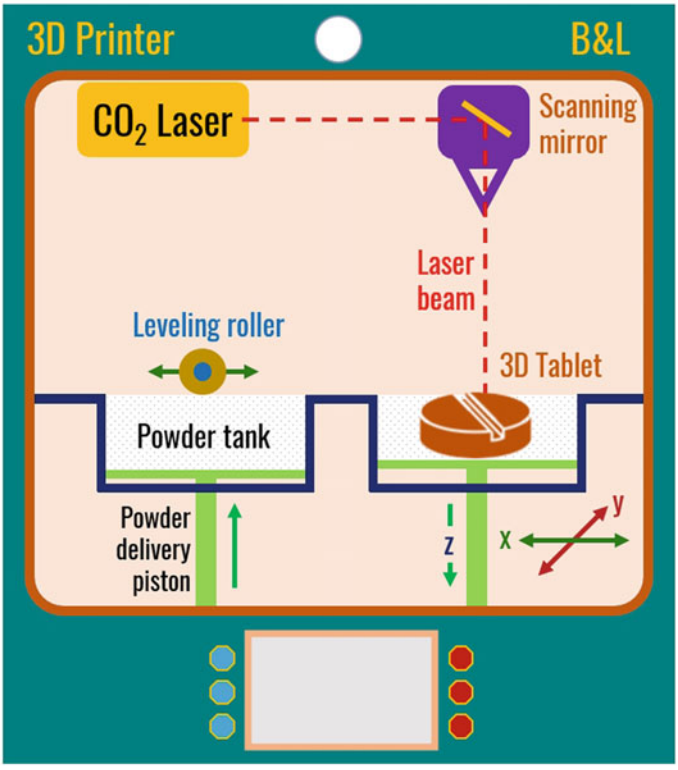


Fig. 6.5 Schematic of the selective laser sintering/melting printer (Additional info, URL: rb.gy/tptnmx)

6.7.5 Digital Light Processing (DLP)

Stereolithography was modified by adding a digital micromirror device (DMD) to the laser's optical path, eliminating the need for scanning, allowing a complete layer to be created in a single exposure [25]. The movable and on and off switchable micromirrors are fixed in DMD. Based on the loaded image, the technology reflects light in a defined pattern, passes through the optical system, illuminates the surface of the resin, and causes the desired area to cure. From inception, this technology was used in the display industry called digital light processing (DLP) technology [26].

6.7.6 Inkjet Printing

Inkjet printing/drop-on-demand printing uses thermal, electromagnetic, or piezoelectric techniques to deposit ink droplets on a substrate according to the digital instructions as shown in Fig. 6.6. In a thermal inkjet (TIJ) printer, heating the print head produces tiny bubbles, which burst and generate pressure pulses, eventually ejecting ink droplets as small as 10–150 pL from the nozzle. In TIJ, piezoelectric pressure will expel droplets. Bioprinting uses "biological ink".

Owing to its precision, controllability, versatility, and positive effects on mammalian cells, it is a prospective technology for tissue engineering and regenerative medicine. This technology has been applied to the 3D bioprinting of tissues and organs.

In TIJ, even though the heating element raises the temperature by about 300 °C, as the contact time of cells is only about 2 μs, the cell temperature rise is only below 10°. So, the cells remain viable. However, in PIJ printing, the frequency of 15–25 kHz, used for the sonication process, generates the pressure which damages the living cell. Hence, PIJ is seldom used in bioprinting applications.

Multijet printing (MJP) is also a process of inkjet printing that uses a piezoelectric print head to deposit light-curing plastic resin layer by layer. MJP is expected to be the key to making the smallest middle ear in the human body a "stapes". Presently, there are no medications, and hearing aids are seldom satisfactory in treating otosclerosis or abnormal hardening of these bones. Human trials with prosthesis printed on a high-resolution 3D multijet printer have been successful [27].

6.7.7 Pressure-Assisted Microsyringe (PAM)

The PAM and FDM methods are identical. Fusion of materials is excluded in the case of PAM method. The mechanism of this technique is the use of a syringe to release viscous semi-liquid materials to design a 3D object [28]. The PAM 3D printer system is used to design and verify the diverse configurations. PAM technology uses a microinjector that is controlled by a computer and provides the required structure and releases the dissolved polymer at constant low pressure [28, 29]. To

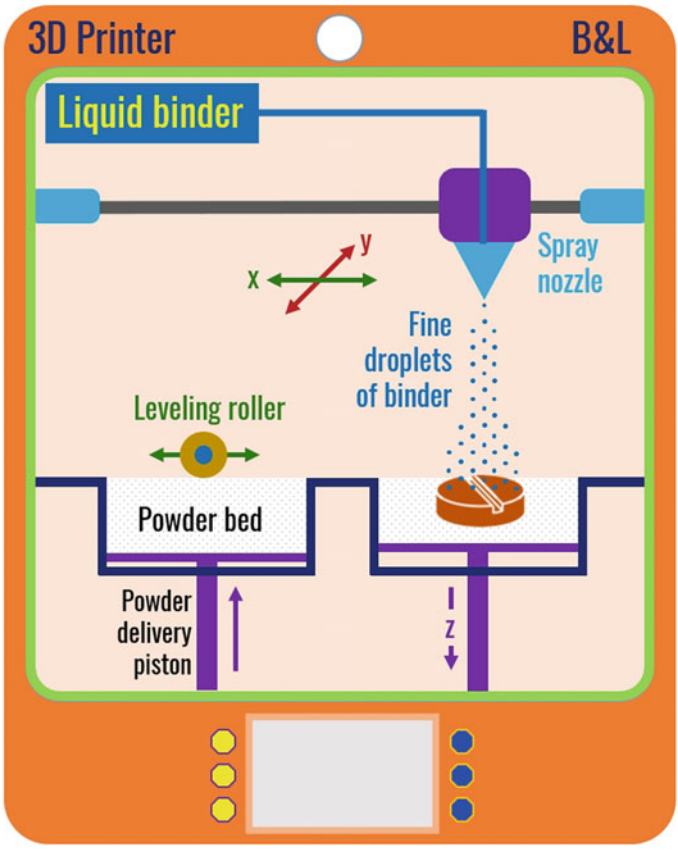


Fig. 6.6 Schematic of the inkjet printer (Additional info, URL: rb.gy/atlhkv)

obtain sufficient physical strength in the dosage form, drying and curing are the subsequent processing steps.

6.7.8 Embedded 3D Printing

This technology can freely produce multi-minerals with difficult designs [30]. The technique includes using a deposition nozzle to extrude viscoelastic ink (embedded phase) into a curing vessel (embedded phase) in a predetermined path. After printing, the deposit is usually cured to built a uniform design. E3DP has been used to produce integrated and highly programmable structures such as portable electronic devices [31, 32]. The application of this method in the pharmaceutical field provides an opportunity to provide a universal matrix of modular systems that contain one or more personalized doses of drugs to fulfil the requirement of one or a few patients.

6.7.9 Stencil Printing

The ink is purged through stencil orifices to form a flatbed or rotary printing process with the help of a knife-edge. This technology has lately been utilized to fabricate wearable electronics [33]. The internal vision system aligns the stencil to the board after its first installation, and then printing work starts with the help of a knife-edge. In the end, the separation of stencil and board is done followed by unloading. The stencil is cleaned frequently after every ten prints to remove excess solder paste.

6.8 Advantages, Limitations, and Challenges of 3DP

3D printing technology can manufacture small batches or different parts according to the needs of specific patients, so it has significant advantages in the fields of pharmaceuticals, tissue engineering, and biomedical equipment. For example, at present, moulds are made for the required parts through machining operations, and then specific surface treatments are performed for the required surfaces, aesthetic effects, and mechanical properties to manufacture surgical implants. It is a costly affair, and hence, patient-specific or specific devices are very costly and rarely fabricated. Other challenges are the difficulty of manufacturing titanium alloys to make them larger compared to 316LSS steel [34, 35]. In addition, these technologies are energy-intensive, generate a lot of material waste, and cannot produce functionally graded implants. Additive manufacturing represents a new hope for the production of various biomechanical parts such as orthopaedic implants.

Particularly for implants, 3D printing can achieve customized complex geometries and on-demand manufacturing of functional implants, which can significantly reduce costs and inventory. Owing to its exclusive application and mechanism of 3D printing, the unit cost of each part remains the same. This economic benefit of 3D printing is considered to be the basis for its application in biomedical orthopaedic implants. Despite some notable successes, the use of 3D printing to develop human tissues or entire organs still faces great challenges [36–39]. From minimally invasive surgery to cancer treatment, from the treatment of congenital defects to functional prostheses for amputees, all fields of medicine and surgery seek advancements in 3D printing technology to make human life easy or help patients extend their lives.

Although 3D printing can use a variety of biological materials (including metals, ceramics, polymers, and composites) to make highly customized and complex designs at low cost, it is having limited application in the medical field due to a lack of diversified properties. The printability of biomaterials, adequate mechanical strength, biodegradability, and biocompatibility are the main concerns.

FDM-based bioprinting requires higher concentrations of polymers to build biological inks to obtain the structural integrity of the final product. Due to its high density, it limits the functional characteristics of biological tissue. To make any medium-sized biological scaffold function, vascularization is the most important, and current 3D printing technology is impossible. The survival diffusion in the small printed stents still needs to explore, but functional organs must have abundant

vascularization. To solve this problem, many researchers have used a combination of sacrificial materials in the stent manufacturing process. These materials provide mechanical support for the impression material by filling the voids, and once the construction is complete, it will be removed using post-processing methods. A carbohydrate glass [40], pluronic glass [41], and gelatin particles [42] are currently being studied as sacrificial/escape materials [43].

Design-induced constraints can cause discontinuities in material because complex CAD designs cannot be well translated into machine instructions. Process-induced limitations include the difference in porosity between CAD objects and 3D printed finished products [44].

Hazard identification is an important step to prevent failure of quality control parameters (such as appearance, content consistency, content, etc.). Analyse processes and process variables to ensure that quality products are manufactured. When the printer cannot print a certain layout, software control should be used. The change in layer thickness should be controlled by monitoring the layer thickness in real time. The environmental factors in the production area must be controlled to cope with improper delamination. By controlling the height of the print head and the speed of the print head, inaccurate positions can be avoided during printing. Non-uniformity in the layers can be avoided by controlling the humidity and particle size distribution of a powder bed. By ensuring the PSD and controlling the flow of the inkjet, the print head can be prevented from clogging. Inconsistent caking or bonding may be due to changes in adhesive viscosity or adhesive surface tension.

6.9 Applications of 3D Printing Technology

Will a teenager buy the same garments as his grandparents? The answer is "Big NO". But when they are sick, despite their many biological differences, they are likely to receive the same treatment. This is a big challenge for the medical practitioner and scientists to understand that how the development of disease response to treatment by all individual persons. The result is a "one size fits all" medical approach based on a broad population average. These traditional ten practices did not meet the standard because each individual has different genetic makeups from others, specifically related to health [45].

The emergence of personalized medicine brings us closer to more accurate, predictable, and powerful medical care, tailored to individual patients. Our increasing understanding of genetics and genomics and how they drive everyone's health, disease, and drug response enable doctors to provide better disease prevention, more accurate diagnosis, safer drug prescriptions, and more treatments. Our increasing understanding of genetics and genomics and how they drive everyone's health, disease, and drug response enable doctors to provide better disease prevention, more accurate diagnosis, safer drug prescriptions, and more treatments for many diseases and conditions that diminish Our increasing understanding of genetics and genomics and how they drive everyone's health, disease, and drug response enable doctors to provide better disease prevention, more accurate diagnosis, safer drug

prescriptions, and more treatments for many diseases and conditions that diminish our health [45].

Personalized medicines are benefitting patients across many different diseases and their therapeutic areas, like haematology, oncology, psychiatry, infectious disease, cardiology, endocrinology, neurology, gastroenterology, rheumatology, etc. [46, 47].

Throughout history, medical practice has been largely passive. Even today, we usually have to wait for diseases to appear before trying to treat or cure them, though all the efforts and available treatments are still lagging complete cure from this disease. The drugs and treatments we designed were tested in a wide range of populations, and statistical averages were used for prescriptions. Therefore, due to genetic differences between populations, they are effective for some patients but not for many others. On average, any prescription drug on the market is only effective for half of the users [45].

Personalized medicine is a medical service to give appropriate and efficacious treatment to all individuals based on their unique genetic makeup and body requirements [48].

Personalized medicine can be equipped with more precise tools, and clinicians can choose therapies or treatment options based on the molecular characteristics of the patient, which not only minimizes harmful side effects but also ensures more successful results. Personalized medicine has the potential to change the way we think about, identify, and manage health problems. The concept of personalized medicine would be more promising to give advanced treatment compared to conventional one without having significant side effects or underdose treatment [48].

Personalized medicine can help improve health care. Patient stratification is another result and requirement of the biobank, which can be transformed from "one size fits all" to more specific treatments, treatment models, and in-silico treatments. A stratified approach can improve personal health care, including age, gender, demographics, and related costs [49].

A patient may respond well to a low dose of a given drug, while other patients may have faster drug metabolism and require higher doses. Response to this treatment may be seen differently due to the difference in body requirements of both the patients. In addition to selecting the appropriate drug type and concentration for the specific indications of a specific patient, due attention must also be paid to the design of the drug product to fully meet the specific needs of the patient (one pill of three drugs). Zolpidem is a narrow therapeutic window, which is prescribed to patients to treat insomnia. The patient took a 5 mg dose per day and no effect was seen, but there were adverse reactions when taking a 10 mg dose. There is no intermediate dose available.

Significant study and efforts to develop patient-centric dosage form will lead to better quality personalized medicines and better treatments. All of these are directly related to healthcare costs and treatment time. Therefore, better medicines and well-trained doctors can result in affordable healthcare treatment [50]. India accounts for 20% of global generic drug exports. In the 20th fiscal year (as of January 2020), India's pharmaceutical exports were USD 13.69 billion [51]. It is estimated that the

Table 6.2 Applications of 3D printing in various fields

Field	Applications
Pharmaceutical	Tailor-made implants and oral solid dosage forms
Medical	• 3D models prepared for perioperative surgery • Dental care fixtures • Implants and prosthetics customized for specific patients • Living tissue engineering and regenerative drugs
Industry	Templates, accessories, and end-use parts for prototypes in the aerospace and automotive industry
Food	3D printed cakes, biscuits, and other food items in complex shapes
Fashion	Ornament, garments, and other decoration accessories
Household	Utensils and related items
Miscellaneous	Prototype items related to space engineering. Designing complex molecules and compounds in the chemical industry

export volume in 2020 will reach 22 billion USD. The Indian pharmaceutical industry is expected to grow at a compound annual growth rate of 22.4% in the short term, and the medical device market is expected to grow by US$25 billion. By 2025, India is the second-largest contributor to the global biotechnology and pharmaceutical labour force. It is estimated that by 2020, the national generic drug market will reach the US $ 27.9 billion. The Indian generic drug market has enormous growth potential. Indian pharmaceutical companies received a total of 415 product approvals and 73 interim approvals in 2018, and the generic drug market is expected to reach the US $ 88 billion by 2021 [52].

Personalized medicine will be a promising treatment to overcome the shortfall of conventional medicine because it is based on the genetic makeup of each individual patient. Owing to this attribute, it will certainly facilitate the healthcare professionals to

- Change the focus of medication from response to prevention
- Predict disease susceptibility
- Improve disease diagnosis
- Prevent the progression of the disease
- Individualize healthcare treatment strategies
- Effective and quality prescription
- Eliminate unwanted effects
- Rapid screening and clinical study of drugs
- Cost-effective healthcare system without compromising patient care

There was limited use of this technology to print anatomical models for teaching purpose. But now, due to the latest developments in the field of new biodegradable materials, the application of 3D printing in the medical and pharmaceutical fields is booming. Table 6.2 depicts various fields where there is the emergence of 3D printing technology.

6.9.1 In Medicine and Dentistry

Orthodontics is one of the branches of dentistry where 3D printing has made an early mark. 3D system technology, one of the early innovators who employed additive technology in orthodontics, joined hands with a start-up company Align Technology in 1997 to develop a system named Invisalign. It uses digital scanning and 3D printing to produce customized orthodontic aligners without using metal wires or ceramics. These near-invisible aligners or braces substantially reduce the discomfort and aesthetic limitations of conventional metal braces. The invention was approved by the Food and Drug Administration (FDA) in 1998, and by 2012, the company was already generating about USD500 million revenue from mass-customized orthodontic treatment devices. Now there are several "clear-aligners" from different sources available in many parts of the world including India.

Dental crowns are another product of 3D printing. They are used to cover or "cap" a damaged tooth as well as used as a dental bridge in place of a missing tooth. Surgical guides made through 3D modelling and printing are also one of the useful innovations made for dental surgeons. The guides can be placed over teeth to guide the surgeon accurately for drilling and so can avoid accidental damage to nerves.

Titanium alloys, bioceramics, and some thermoplastics are the commonly used materials for the manufacturing of 3D printed objects in dentistry. Because a large part of the dental practice is handled by individual dentists and not hospitals, the cost of 3D printing objects is becoming a bottleneck. According to one market research report, the global dental 3D printing market would grow from less than USD 1.6 billion in 2016 to over USD 3.7 billion in 2021 [53]. The market is segmented based on the material used in metal, photopolymer-based, ceramic, and others, with photopolymers accounting for over 58% of the market. Technology in the form of 3D printing is slowly but surely making medical prosthetics affordable to the underprivileged giving hope to over 30 million people all over the world in need of artificial limbs as per the Guardian report [54].

6.9.2 3D Printed Drugs

Rapid disintegrating tablets (Aprecia ZipDose®) were fabricated to release the drug quickly as soon as it immerses into media through the rapid breaking of fused layers made from 3D printing technology. It supports drug dosing up to 1000 mg/1 gm. Spritam (levetiracetam) tablet used to treat epilepsy is the first printlet officially permitted by the USFDA in 2015 [55]. The patented process called ZipDose produces the tablet by binding together layer upon layer of the powdered medication using an aqueous fluid to get a porous, water-soluble matrix that rapidly disintegrates with a sip of water. The technology behind ZipDose is called powder-liquid 3DP, which was originally developed at MIT in 1980 [56]. The technology for pharmaceuticals was acquired and further developed, modified, and patented by Aprecia Pharmaceuticals, an Ohio-based US Company.

The Zip Dose printer deposits a thin layer of powder blend in the disc shape. Then deposition of tiny droplets on the powder blend to bond them together. Repetitive layers are formed till the desired height of the tablet is achieved. The interaction between the powder layers and the intervening aqueous layer creates a unique highly porous matrix which makes it disintegrate easily. ZipDose technology allows the active ingredient only. One small tablet can have a relatively high dose of medication, and so patients need to consume a lesser number of tablets. It also eliminates the need for compression forces, dies, and punches that are presently used in tabletting machines. At present, the 3D printing process technology used by Aprecia is used in a conventional cGMP plant and uses traditional distribution channels.

3D printing may someday enable custom drugs or personalized medicine possible, to be made and sold at the point of sale (POS). It is presumed that in the future where a doctor sends his prescription to the pharmacy that uses a 3D printer to create a custom formulation based on the individual's special needs. It can be visualized that instead of popping several tablets, a patient may be able to take a single patient-specific combination of multiple drugs into a tablet custom printed for him/her made up of different layers of medications bound together in a single tablet. However, it is not a simple matter to do so because several factors such as the release mechanism of the individual drug compound, location, and time in which drug gets released in the gastrointestinal tract and several other factors have to be considered in designing the medication. Researchers are working on finding solutions to such problems. Scientists in the United Kingdom were reportedly experimenting with making 3D tablets of odd shapes like dinosaurs and octopuses to distract and make it easier for kids to take their medicines.

The continuous manufacturing and 3D printing together can change current practices of oral dosages manufacturing in the future, when mass customization and personalized medicine will become a reality. It will also overcome some of the limitations of manufacturing or orphan drugs [57].

The oral dual-compartment dosage unit laden with anti-tuberculosis drugs was magnificently manufactured by 3D printing technology [58]. The restoration of damaged tissues or organs could be done by tissue engineering [59]. The American biotechnology start-up BioBots is at the intersection of computer science and chemistry [60]. Its first product is a desktop 3D printer for biological materials, combining hardware, software, and wet parts.

However, due to the lack of production of stents for individual patients, the same level of effectiveness may not be achieved in treating two different clinical conditions like percutaneous coronary intervention. This study presents the feasibility of a patient-specific stent implantation process constructed from direct 3D segmentation of medical images using direct 3D printing of biodegradable polymer–graphene compounds and the incorporation of two drugs. A biodegradable polymer carbon composite material doped with graphene nanosheets was prepared to achieve the controlled release of combined drugs such as anticoagulant and anti-restenosis agents [61].

A step-by-step process for the development of a custom coronary stent is as follows: (a) The cross section of the artery shown on the surface of the heart shows

plaque accumulation in the lumen due to various factors such as ageing, diet, and genetics; (b) A clot-binding probe, i.e., fibrin-targeted iodinated computed tomography (CT) contrast probe, is administered to locate the blood clot. At the same time, volumetric CT imaging helps to get an accurate measurement of the blockage; (c) Transmit the image information to computer-aided design (CAD) software to design custom stents; (d) PCLGR polymeric compounds are used for fused deposition construction in commercial 3D printers additive manufacturing technology model; (e) The prototype stent is placed in the artery; (f) The two drugs are combined for sequential release; (g) The healing process is further monitored by CT images; and (h) The polymer is in the widened artery [61] and it biodegrades inside.

6.9.3 New Geometries and Designs

It has been proven that additive manufacturing can build precise dosage forms with a small quantity of API [62]. Unlike traditional powder compression, the compact unit tablet dosage form can also be achieved through tabletting technology. This opens the window to a large number of complex designs and geometries that cannot be achieved with traditional manufacturing methods [63]. For example, partially filled or hollow tablet designs [64]. 3D printing technology can also place active materials on manufactured objects [65]. The design of the tablet is made with different additive manufacturing technologies [66]. Unique shaped dosage form and geometries can also result in desired release profiles. Katstra et al. used 3D printing to make tablets with complex release profiles, including immediate prolongation, collapse, double intestinal pulsation, and double pulsation [67]. Sun and Soh published a study where they controlled the release profile of dyes contained in surface-eroding polymers by manufacturing polymers in different shapes. The change in matrix shape results in a constant, pulsating, increasing, or decreasing dye release profile [68]. Przemislaw et al. used 3D printing for rapid prototyping of drug dissolution test equipment for non-standard applications [69]. Basel et al. created a new design for rapid releasing printlet composed of cellulose as a polymer [70]. Okwuosa et al. recommended to manufacture patient-specific liquid capsules on-demand using coordinated 3D printing and liquid dispensing [71]. Sadia et al. proposed an innovative method to accelerate drug release from 3D printed channel tablets [72].

6.10 Regulatory Concerns

AM/3DP has quite a several regulatory and quality system considerations that are still unresolved. For example, bio-printed organs or tissues cannot be considered as "original" body parts by any stretch of the imagination.

The FDA's Guide to 3D Printing Medical Devices: Advancing 3D Printing Technology has brought the perspective of personalized dosing one step closer. However, can regulators and the current legal framework deal with the ambiguity of this boldly advancing technology? The difference between compound drugs and

artificial drugs is the central issue of 3D printing drug regulations. This issue has a significant impact on the regulatory level of 3D printing products. Tragic incidents like the 2012 New England Compounds Center [73] and dozens of other dangerous safety issues at compounding pharmacies [74] have brought drug safety to the fore.

USFDA took the first steps by issuing a guidance note "Technical considerations for additive manufactured medical devices" in December 2017 [75] following a workshop held in 2014, a subsequent consulting process, and a draft guidance note earlier in 2017. The guidance note covers the following:

- Design and manufacturing considerations to be addressed as part of the quality system
- Device testing requirements for premarket notification submissions, commonly called 510(k) process, premarket approvals, De Novo requests for novel devices, and applications for investigational device exemption and humanitarian device exemption applications.

Most of the 3D printed medical devices approved by the FDA were under the 510 (k) pathways, which essentially requires the manufacturer to demonstrate that a new device planned to be marketed is at least as safe and effective as a device already legally marketed without premarket approval. Some devices were permitted under an emergency provision that allows a physician to treat a patient with an unapproved device in an emergency. A few others were permitted under a compassionate use pathway, which allows an investigational device to be used on a patient as there is no other satisfactory therapy available.

FDA guidelines describe the technical aspects of 3D printing equipment, which must be considered during the design and development stages, production process, process verification, and equipment testing of finished and semi-finished products.

The general guidelines are similar to existing drugs and devices. Manufacturers must establish and maintain procedures for control of design identifying the key design parameters; they must have procedures for monitoring and control of process parameters; there must be a validated process to ensure that the device can perform as intended. Specific issues relating to 3D printed devices are discussed in the guidelines in some detail.

Depending on the materials used to manufacture the device and used for cleaning of the component or device becomes important, removal of residual material, which is especially critical if any materials used in the processing steps are not biocompatible. As such cleaning procedure must be demonstrated to be consistently adequate for parts that are of complex shapes with cavities of irregular shapes. Additional finishing such as heat treatment may be required in some cases.

For implantable medical devices, sterility is important and the sterilization process has to be compatible with the properties of materials used for the 3D printed device. The guidance note also deals with starting material control, software workflow, and file conversion between software used for imaging and modelling, process validation, device testing considerations, and labelling aspects of the quality systems.

The FDA has made it clear that this is a "breakthrough document", and new information on the mechanism is available to guide the FDA to share its initial thoughts on emerging technologies.

The FDA has also clarified that it is not applicable for point of care manufactured devices that may be made in hospitals, clinics, or elsewhere, as well as products involving biological, cellular, or tissue-based processes which may necessitate additional manufacturing and regulatory considerations or a different set of regulatory pathways. The guidance also excludes 3D printed drugs like tablets.

The USFDA guidance note is the first such document on 3D printing for medical devices. Earlier in 2015, the Japanese regulatory agency Pharmaceuticals and Medical Devices Agency (PMDA) released a technical document covering only orthopaedic surgical implants involving 3D printing. Regulatory authorities in the EU, Brazil, and Canada have not yet issued any guidance on 3D printed devices. The Therapeutic Goods Administration (TGA) has released a consultation paper on personalized and 3D printed medical devices in November 2017 with the idea of formalizing the Australian guidance notes sometime in 2018 [76].

China FDA is the latest national regulator to issue a draft guidance note on 3D printed medical devices in March 2018. Earlier, in 2015, the agency had the approval of metallic implants manufacturing by AM technology [77]. Table 6.3 highlights research finding from literature in the field of 3D printed drug delivery systems.

These features enable the 3D printer to act as a mini-dispenser, making it possible to make tablet computers closer to the patient. For this reason, it is necessary to mass-produce stable and reproducible starting materials [102]. There is minimal space requirement for AM technology, and make it more promising technology with efficient manufacturing with high precision. They are reasonably priced and can be controlled remotely using virtual reality [85]. In addition, 3D printing technology not only allows small batch production but also allows a single project to be produced in a single build run.

6.11 Future of 3D Printing Technology in the Pharmaceutical Industry

As the pharmaceutical industry shifts from the manufacturing of traditional medicines to personalized medicines, 3D printing is gaining more attention due to its potential ability to produce customized products like oral tablets and dental/medical devices in very small quantities according to the ordered unique configurations in a cost-effective way. 3D printing promises a bright future of drugs that will be printed as per the need of dose, design, and patient age, especially for geriatric and paediatric patients, for example, the printing of paracetamol tablet in "star" shape to make it more kid's friendly.

A major advantage of 3D printing is that both the processes, design and manufacturing, are separate from each other and can be easily outsourced. The software specialists or designers can contract with companies for products based on their designs. Alternatively, a manufacturer or pharmacist can use/download the

Table 6.3 List of research findings done on the 3D printed formulations

3D printing technology	Formulation	Application	Materials	Ref.
3D printing based on powder layering	Absorbable device	Novel controlled releasing drug delivery system	Polycaprolactone (PCL) with methylene blue and polyethylene oxide matrix materials	[78]
	Tablet dosage form	Novel delayed drug release formulation	Drug and fluorescein disodium salt, Eudragit, acetone, polyvinylpyrrolidone, and other excipients	[62]
	Implants	Multi-drug implants for bone tuberculosis	Isoniazid, rifampicin, and other excipients	[79]
	Bioceramic implants	Formulation comprising antibiotics	Vancomycin hydrochloride, tetracycline hydrochloride, and ofloxacin with other excipients	[80]
	Mesoporous biologically active glass	Biologically active glass for restoration of bone	Dexamethasone powder and polyvinyl alcohol	[81]
	Tablet dosage form	Rapid disintegrating soft tablets	Paracetamol and other excipients	[65]
	Tablet dosage form	Controlled drug delivery system	Paracetamol and other excipients	[82]
	Tablet dosage form	Featured complex drug-releasing formulation	Chlorpheniramine maleate and other excipients	[67]
	Cubical architecture	Consistent rate drug-releasing formulation	Pseudoephedrine hydrochloride and related excipients	[83]
	Implantable formulation	Antibiotic laden implant	Antibiotic drug and excipients	[84]
	Multiple-layered drug-eluting device	A multiple-layered drug-eluting device in the form of a doughnut	Paracetamol and excipients	[85]
	Orally dispersible formulation	Quick dispersible dosage form	Levetiracetam and other excipients	[86]
	Implantable formulation	The implantable complex design meant for prophylaxis action	Antibiotic drug and excipients	[87]
	Hydraulic adhesive material	Restoration of bone structure	Tricalcium derivative cement and other excipients	[88]

(continued)

Table 6.3 (continued)

3D printing technology	Formulation	Application	Materials	Ref.
		using cementing material		
Selective Laser Sintering 3D printing	Shell core structure	Controlled drug-releasing device	Polyamide	[89]
	Cubical hollow matrices	Perforated matrix devices	Nylon and other dyes	[90]
	Biopolymeric microstructures	Drug-eluting polymeric discs	PLA	[91]
Fused Deposition Modelling technology	Tablet dosage form	Controlled drug delivery system as personalized medicine	PVA	[64]
	Tablet dosage form	Various geometrical-shaped printlets	Paracetamol	[92]
	Capsule	Pulsatile drug delivery system	Paracetamol and other excipients	[93]
	Tablet dosage form	Modified release drug delivery form	PVA	[94]
	Tablet dosage form	Extended drug-releasing formulation	Steroid and PVA	[95]
	Tablet dosage form	Flexi-dose formulation	Theophylline and Eudragit	[96]
	Discs	Medical devices	Nitrofurantoin and excipients	[97]
	Discs	Drug laden implants	Nitrofurantoin and excipients	[98]
	Structured matrix reservoir	Controlled releasing matrices	Dye and excipients	[99]
	Medical devices	Sustained releasing dosage form containing chemotherapeutic drugs	Chemotherapeutic drugs and excipients	[100]
Combinational 3D printing	Implantable device	3D printed formulation	Antibiotic and excipients	[101]

digital CAD file of the object and print it on a 3D printer [39]. As 3D printers can be operated and controlled remotely through computer software, printing and dispensing of the desired dose of the medicine can also be possible at the locations, accessible by the patients. A new and more dynamic supply chain can be established through 3D printing. This improves patient compliance and shortens the time of clinical response.

Nowadays, inline clinical data monitoring is possible due to the developments in sensor technology. A biosensor is implanted in or worn by patients through which continuous monitoring of body parameters can be done as well as stored by the healthcare network. With the help of 3D printing, it can be made possible to modify the next dose of drug or drug combination according to the continuously generated data and the patient's changing needs.

6.12 Risks and Challenges in 3D Printing

Although 3D printing seems very promising in transforming few industries, its application in the pharmaceutical and medical industry is still facing certain challenges and limitations.

6.12.1 Product Liability Risk

As 3D printing of medicines is new to the healthcare sector, one of the main challenges in its wide application is drug product liability in case of quality and safety issues. The product blueprint is the liability of pharmaceutical companies, but then 3D printing can be done by any healthcare providers or manufactures locally. So, in case of any manufacturing defects, the entire parties across the whole manufacturing like software designers, material suppliers, printer manufacturers, and product manufacturers could be responsible. So, whom to sue in case of any product defect claims and adverse incident? There are no structured regulations and litigation in this area yet. So, pharmaceutical companies must develop a strategy or policy for licensing their blueprints with the involvement of lawyers and insurance brokers to ensure their legal and financial protection.

6.12.2 Cyber Risk

3D printed medicines are most susceptible to counterfeiting. In contrast to the traditional manufacturing process, it is very easy for hackers to hack the digital blueprint or printers used in 3D printing which increases the risk of counterfeiting manifold. Hacking of a single drug blueprint can lead to bulk production of fake/counterfeit pills all over the world which can cause a significant impact on the reputation, financials, and/or IP protection right of the owner company by harming the patients' health.

6.12.3 Safety and Efficacy of 3D Printers

Conventional manufacturing of the medicines is strict as per the cGMP norms and supervised by the concerned regulatory agencies of the state/country which ensures

the product quality and safety. Unlike this, there are no regulations to enforce the quality of a drug product manufactured by a 3D printing process. Also, risk factors like printer defect, contaminated or defective materials, or manufacturing malfunction are the major concerns. Thus, there is a question in the final quality, efficacy, and safety of the 3D printed medicines which can pose a large threat to the future of 3D printing medicines.

6.13 Summary

In nutshell, CAD and CAM technologies are becoming rapidly expanding fields with revolutionary application in the field of pharmaceuticals and medical sciences like the personalization of drug dose and delivery, biosynthetic organs, fabrication of patient-specific implants, etc. With the advancement of 3D modelling software and 3D printers, 3D printing is becoming one of the emerging cost-effective CAM techniques with enhanced speed and productivity [16]. Medical 3D printing is also another budding technology with immense potential in health care.

The biggest contribution towards the ongoing progress of personalized medicines is from the 3D printing technology which has opened up the doors of product design, manufacturing, and distribution for pharmaceutical companies. Simultaneously, companies also need to understand and determine the risk exposures while doing worthwhile investment. For that, they must work closely in consultation with IT and manufacturing colleagues, insurance experts, their broker, etc. and ensure the full coverage of insurance to address the risk factors. Organ printing or bioprinting is still a challenging field that will need some more time to develop and evolve fully.

6.14 Credible Online Resources for Further Reading

URL	Category of source	What to refer?
https://www.usi3dt.org/	Private company	Training and certification course provider
https://make.3dexperience.3ds.com/processes/photopolymerization	Private company	3D printing—additive processes
https://www.autodesk.com/products/fusion-360/overview	Private company	Software and 3D printing service
https://www.aniwaa.com/	Private company	3D printer and technologies
https://www.fabrx.co.uk/technologies/	Institution	Additive manufacturing in medical application
https://pocketdentistry.com/computer-aided-manufacturing-in-medicine/	Personal webpage	Pocket dentistry: Fastest clinical dentistry insight engine

(continued)

URL	Category of source	What to refer?
https://www.stratasys.com/3d-printers	Private company	3D printer manufacturer
https://3dprinting.com/what-is-3d-printing/	Personal webpage	Detailed information for 3D printing technology
https://www.3dhubs.com/guides/3d-printing/	Private company	3D printed parts supply and services
https://www.meity.gov.in	Government webpage	National strategy for additive manufacturing
https://3dprintingindustry.com	Personal news webpage	Educational 3D printing centres
https://www.3dprintingworld.in	Personal webpage	International virtual conference and exhibition

References

1. Wang X, Bi Z (2019) New CAD/CAM course framework in digital manufacturing. Comput Appl Eng Educ 27(1):128–144. https://doi.org/10.1002/cae.22063
2. Jones DS (2016) Solid dosage forms-1: tablets. FASTtrack: pharmaceutics—dosage form and design, 2nd edn. Pharmaceutical Press, pp 251–310
3. ISO/ASTM 52900:2015(en). Additive manufacturing—general principles—terminology. Online Brows Platform. https://www.iso.org/obp/ui#iso:std:iso-astm:52900:ed-1:v1:en. Accessed 29 Mar 2021
4. Bandyopadhyay A, Bose S, Das S (2015) 3D printing of biomaterials. MRS Bull 40(2):108–114. https://doi.org/10.1557/mrs.2015.3
5. Crump SS (1992) Apparatus and method for creating three-dimensional objects. US Patent 5,121,329. 9 Jun 1992
6. Sanders RC, Forsyth JL, Philbrook KF (1998) 3-D model making. US Patent 5,740,051. 14 Apr 1998
7. Munaz A, Vadivelu RK, St. John J, Barton M, Kamble H, Nguyen NT (2016) Three-dimensional printing of biological matters. J Sci Adv Mater Devices 1(1):1–17. https://doi.org/10.1016/j.jsamd.2016.04.001
8. Gupta V, Nesterenko P, Brett P (2019) 3D printing in chemical sciences. Applications across chemistry. Royal Society of Chemistry
9. Simplify3D. Ultimate 3D printing materials guide. https://www.simplify3d.com/support/materials-guide. Accessed 2 July 2020
10. Jamróz W, Szafraniec J, Kurek M, Jachowicz R (2018) 3D printing in pharmaceutical and medical applications-recent achievements and challenges. Pharm Res 35(9):1–22. https://doi.org/10.1007/s11095-018-2454-x
11. Al-maliki JQ, Al-maliki AJQ (2015) The processes and technologies of 3D printing. Int J Adv Comput Sci Technol 4(10):161–165
12. Abboud M, Orentlicher G (2012) Computer-aided manufacturing in medicine. Atlas Oral Maxillofac Surg Clin North Am 20(1):19–36. https://doi.org/10.1016/j.cxom.2012.01.001
13. Melnikova R, Ehrmann A, Finsterbusch K (2014) 3D printing of textile-based structures by fused deposition modelling (FDM) with different polymer materials. IOP Conf Ser Mater Sci Eng 62(1):1–6. https://doi.org/10.1088/1757-899X/62/1/012018

14. Giannatsis J, Dedoussis V (2009) Additive fabrication technologies applied to medicine and health care: a review. Int J Adv Manuf Technol 40(1–2):116–127. https://doi.org/10.1007/s00170-007-1308-1
15. Mathur S, Sutton J (2017) Personalized medicine could transform healthcare. Biomed Rep 7(1):3–5. https://doi.org/10.3892/br.2017.922
16. Lee VC (2014) Medical applications for 3D printing: current and projected uses. P T 39(10):704–711
17. Brown C, Lubell J, Lipman R (2013) Additive manufacturing technical workshop summary report. National Institute of Standards and Technology Technical:1–60. https://doi.org/10.6028/NIST.TN.1823
18. Manyika J, Chui M, Bughin J et al (2013) Disruptive technologies: advances that will transform life, business, and the global economy. McKinsey Global Institute
19. Jungling MK, Wood PA, Koike Y (2013) MedTech: 3D printing—a solution for innovation. Morgan Stanley Blue Paper
20. Makerbot Educators Guidebook—The Definitive Guide to 3D Printing in the Classroom, MakerBot Industries, LLC (2017)
21. The global manufacturing additive leader in standards (2014) ASTM Int 42(6):1–12. https://www.astm.org/ABOUT/OverviewsforWeb2014/Additive-Manufacturing.pdf. Accessed 25 Mar 2021
22. Vanderploeg A, Lee SE, Mamp M (2017) The application of 3D printing technology in the fashion industry. Int J Fash Des Technol Educ 10(2):170–179. https://doi.org/10.1080/17543266.2016.1223355
23. Miller E (2012) PADT. Inc.—The Blog, 3D Printing Technology Animations. http://www.padtinc.com/blog/tag/stereolithography. Accessed 29 Mar 2021
24. Gibson I, Rosen D (2015) Stucker B Additive manufacturing technologies: 3D printing, rapid prototyping, and direct digital manufacturing. Springer, New York
25. Lu Y, Mapili G, Suhali G, Chen S, Roy K (2006) A digital micro-mirror device-based system for the microfabrication of complex, spatially patterned tissue engineering scaffolds. J Biomed Mater Res A 77(2):396–405. https://doi.org/10.1002/jbm.a.30601
26. Kadry H, Wadnap S, Xu C, Ahsan F (2019) Digital light processing (DLP) 3D-printing technology and photoreactive polymers in the fabrication of modified-release tablets. Eur J Pharm Sci 135:60–67. https://doi.org/10.1016/j.ejps.2019.05.008
27. Aimar A, Palermo A, Innocenti B (2019) The role of 3D printing in medical applications: a state of the art. J Healthc Eng 1–10. https://doi.org/10.1155/2019/5340616
28. Goole J, Amighi K (2016) 3D printing in pharmaceutics: a new tool for designing customized drug delivery systems. Int J Pharm 499(1–2):376–394. https://doi.org/10.1016/j.ijpharm.2015.12.071
29. Wen H, He B, Wang H, Chen F, Li P, Cui M (2019) Structure-based gastro-retentive and controlled-release drug delivery with novel 3D printing. AAPS PharmSciTech 20(2):20–68. https://doi.org/10.1208/s12249-018-1237-3
30. Pandey M, Choudhury H, Fern JLC, Kee ATK, Kou J, Jing JLJ (2020) 3D printing for oral drug delivery: a new tool to customize drug delivery. Drug Deliv Transl Res 10(4):986–1001. https://doi.org/10.1007/s13346-020-00737-0
31. Khan S, Ali S, Bermak A (2019) Recent developments in printing flexible and wearable sensing electronics for healthcare applications. Sensors 19(5):1230. https://doi.org/10.3390/s19051230
32. Rycerz K, Stepien KA, Czapiewska M, Arafat BT, Habashy R, Isreb A (2019) Embedded 3D printing of novel bespoke soft dosage form concept for pediatrics. Pharmaceutics 11(12):630. https://doi.org/10.3390/pharmaceutics11120630
33. Wickström H, Koppolu R, Mäkilä E, Toivakka M, Sandler N (2020) Stencil printing-a novel manufacturing platform for orodispersible discs. Pharmaceutics 12(1):33. https://doi.org/10.3390/pharmaceutics12010033

34. Liou FF (2019) Additive manufacturing processes. In: Rapid prototyping and engineering applications—a toolbox for prototype development, 2nd edn. CRC Press, pp 215–286
35. Jackson MJ, Kopac J, Balazic M et al (2016) Titanium and titanium alloy applications in medicine. In: Waqar A, Mark J (eds) Surgical tools and medical devices, 2nd edn, pp 533–576. https://doi.org/10.1007/978-3-319-33489-9_15
36. Woodruff MA, Lange C, Reichert J, Berner A, Chen F, Fratzl P (2012) Bone tissue engineering: from bench to bedside. Mater Today 15(10):430–435. https://doi.org/10.1016/S1369-7021(12)70194-3
37. Hutmacher DW (2013) A road map for a tissue engineering concept for restoring structure and function after limb loss. J Mater Sci Mater Med 24(11):2659–2663. https://doi.org/10.1007/s10856-013-5049-4
38. Reichert JC, Cipitria A, Epari DR, Saifzadeh S, Krishnakanth P, Berner A (2012) A tissue engineering solution for segmental defect regeneration in load-bearing long bones. Sci Transl Med 4(141):141ra93. https://doi.org/10.1126/scitranslmed.3003720
39. Berman B (2012) 3D printing: the new industrial revolution. Bus Horiz 55(2):155–162. https://doi.org/10.1016/j.bushor.2011.11.003
40. Miller JS, Stevens KR, Yang MT et al (2012) Rapid casting of patterned vascular networks for perfusable engineered 3D tissues. Nat Mater 11(9):768–774. https://doi.org/10.1038/nmat3357
41. Kolesky DB, Truby RL, Gladman AS, Busbee TA, Homan KA, Lewis JA (2014) 3D bioprinting of vascularized, heterogeneous cell-laden tissue constructs. Adv Mater 26(19):3124–3130. https://doi.org/10.1002/adma.201305506
42. Hinton TJ, Jallerat Q, Palchesko RN, Park JH, Grodzicki MS, Shue HJ (2015) Three-dimensional printing of complex biological structures by freeform reversible embedding of suspended hydrogels. Sci Adv 1(9):e1500758. https://doi.org/10.1126/sciadv.1500758
43. Tappa K, Jammalamadaka U (2018) Novel biomaterials used in medical 3D printing techniques. J Funct Biomater 9(1):17. https://doi.org/10.3390/jfb9010017
44. Hassana OB, Guessasma S, Belhabib S, Nouri H (2016) Explaining the difference between real part and virtual design of 3D printed porous polymer at the microstructural level. Macromol Mater Eng 301(5):566–576. https://doi.org/10.1002/mame.201500360
45. The Jackson Laboratory (2021) What is personalized medicine? https://www.jax.org/personalized-medicine/precision-medicine-and-you/what-is-precision-medicine#. Accessed 25 July 2020
46. U.S. Department of Health and Human Services, US Food and Drug Administration (2013) Paving the way for personalized medicine-FDA's role in a new era of medical product development. https://www.fdanews.com/ext/resources/files/10/10-28-13-Personalized-Medicine.pdf. Accessed 25 July 2020
47. Vogenberg FR, Barash CI, Pursel M (2010) Personalized medicine Part-1: evolution and development into theranostics. P T 35(10):560–576
48. Personalized Medicine Coalition. The age of personalized medicine. http://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/pmc_age_of_pmc_factsheet.pdf. Accessed 14 Dec 2020
49. Kinkorová J (2016) Biobanks in the era of personalized medicine: objectives, challenges, and innovation. EPMA J 7(1):4. https://doi.org/10.1186/s13167-016-0053-7
50. Misra SC, Bisui S (2018) Modelling vital success factors in adopting personalized medicine system in healthcare technology and management. Eng Sci Technol Int J 21(3):532–545. https://doi.org/10.1016/j.jestch.2018.04.011
51. Pharmaceuticals (2020) India Brand Equity Foundation. https://www.ibef.org/pages/27443. Accessed 10 July 2020
52. KPMG in India, Gujarat pharma industry: striding into the future (2008) Gujarat Pharma Industry. http://www.in.kpmg.com/pdf/gujarat-pharma08.pdf. Accessed 20 June 2020

53. Gasman L (2016) SmarTech analysis, revenues from 3D printing in dentistry will reach $ 3.7 billion by 2021. https://www.smartechanalysis.com/news/3d-printing-dentistry-2021/. Accessed 20 June 2020
54. Birrell Ian (2017) The Guardian, 3D-printed prosthetic limbs: the next revolution in medicine. https://www.theguardian.com/technology/2017/feb/19/3d-printed-prosthetic-limbs-revolution-in-medicine. Accessed 24 July 2020
55. Solanki NG, Tahsin M, Shah AV, Serajuddin ATM (2018) Formulation of 3D printed tablet for rapid drug release by fused deposition modeling: screening polymers for drug release, drug-polymer miscibility and printability. J Pharm Sci 107(1):390–401. https://doi.org/10.1016/j.xphs.2017.10.021
56. Tan DK, Maniruzzaman M, Nokhodchi A (2018) Advanced pharmaceutical applications of hot-melt extrusion coupled with fused deposition modelling (FDM) 3D printing for personalised drug delivery. Pharmaceutics 10(4):203. https://doi.org/10.3390/pharmaceutics10040203
57. CPhI Pharma Insights. Prospects, analysis and trends in global pharma. CPhI Annual Industry Report 2016. https://cupdf.com/document/cphi-annual-industry-report-2016.html. Accessed 15 Oct 2020
58. Genina N, Boetker JP, Colombo S, Harmankaya N, Rantanen J, Bohr A (2017) Anti-tuberculosis drug combination for controlled oral delivery using 3D printed compartmental dosage forms: from drug product design to in vivo testing. J Control Release 268(28):40–48. https://doi.org/10.1016/j.jconrel.2017.10.003
59. Quint C, Niklason L (2009) Tissue engineering and regenerative medicine. In: Simon TL, Snyder EL et al (eds) Rossi's principles of transfusion medicine, 4th edn. Wiley, pp 950–971
60. Ersumo N, Witherel CE, Spiller KL (2016) Differences in time-dependent mechanical properties between extruded and molded hydrogels. Biofabrication 8(3):035012. https://doi.org/10.1088/1758-5090/8/3/035012
61. Misra SK, Ostadhossein F, Babu R, Kus J, Tankasala D, Sutrisno A (2017) 3D-printed multidrug-eluting stent from graphene-nanoplatelet-doped biodegradable polymer composite. Adv Healthc Mater 6(11):1–14. https://doi.org/10.1002/adhm.201700008
62. Katstra WE, Palazzolo RD, Rowe CW, Giritlioglu B, Teung P, Cima MJ (2000) Oral dosage forms fabricated by three-dimensional printing. J Control Release 66(1):1–9. https://doi.org/10.1016/s0168-3659(99)00225-4
63. Alhnan MA, Okwuosa TC, Sadia M, Wan KW, Ahmed W, Arafat B (2016) Emergence of 3D printed dosage forms: opportunities and challenges. Pharm Res 33(8):1817–1832. https://doi.org/10.1007/s11095-016-1933-1
64. Goyanes A, Buanz ABM, Basit AW, Gaisford S (2014) Fused-filament 3D printing (3DP) for fabrication of tablets. Int J Pharm 476(1):88–92. https://doi.org/10.1016/j.ijpharm.2014.09.044
65. Yu DG, Branford-White C, Yang YC, Zhu LM, Welbeck EW, Yang XL (2009) A novel fast disintegrating tablet fabricated by three-dimensional printing. Drug Dev Ind Pharm 35(12):1530–1536. https://doi.org/10.3109/03639040903059359
66. Ursan I, Chiu L, Pierce A (2013) Three-dimensional drug printing: a structured review. J Am Pharm Assoc 53(2):136–144. https://doi.org/10.1331/JAPhA.2013.12217
67. Katstra WE, Palazzolo RD, Rowe CW, Giritlioglu B, Teung P, Cima MJ (2000) Multimechanism oral dosage forms fabricated by three-dimensional printing. J Control Release 66(1):11–17. https://doi.org/10.1016/s0168-3659(99)00224-2
68. Sun Y, Soh S (2015) Printing tablets with fully customizable release profiles for personalized medicine. Adv Mater 27(47):7847–7853. https://doi.org/10.1002/adma.201504122
69. Dorożyński P, Jamróz W, Węglarz WP, Kulinowski W, Zaborowski M, Kulinowski P (2018) 3D printing for fast prototyping of pharmaceutical dissolution testing equipment for nonstandard applications. Dissolut Technol 25(4):48–53. https://doi.org/10.14227/DT250418P48
70. Arafat B, Wojsz M, Isreb A, Forbes RT, Isreb M, Ahmed W (2018) Tablet fragmentation without a disintegrant: a novel design approach for accelerating disintegration and drug release

from 3D printed cellulosic tablets. Eur J Pharm Sci 118:191–199. https://doi.org/10.1016/j.ejps.2018.03.019
71. Okwuosa TC, Soares C, Gollwitzer V, Habashy R, Timmins P, Alhnan MA (2018) On demand manufacturing of patient-specific liquid capsules via co-ordinated 3D printing and liquid dispensing. Eur J Pharm Sci 118:134–143. https://doi.org/10.1016/j.ejps.2018.03.010
72. Sadia M, Arafat B, Ahmed W, Forbes RT, Alhnan MA (2018) Channelled tablets: an innovative approach to accelerating drug release from 3D printed tablets. J Control Release 269:355–363. https://doi.org/10.1016/j.jconrel.2017.11.022
73. Outterson K (2012) Regulating compounding pharmacies after NECC. N Engl J Med 367(21):1969–1972. https://doi.org/10.1056/NEJMp1212667
74. Sun LH (2013) The Washington Post, FDA finds widespread safety issues at compounding pharmacies. https://www.washingtonpost.com/national/health-science/fda-finds-widespread-safety-issues-at-compounding-pharmacies/2013/04/11/5321e17a-a20d-11e2-be47-b44febada3a8_story.html. Accessed 29 July 2020
75. Gottlieb S (2017) US Food and Drug Administration, Statement by FDA Commissioner Scott Gottlieb, M.D., on FDA ushering in new era of 3D printing of medical products; provides guidance to manufacturers of medical devices. https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-fda-ushering-new-era-3d-printing-medical-products. Accessed 29 Sept 2020
76. Australian Government, Department of Health-Therapeutic Goods Administration (2017) Proposed regulatory changes related to personalised and 3D printed medical devices. https://www.tga.gov.au/consultation/consultation-proposed-regulatory-changes-related-personalised-and-3d-printed-medical-devices. Accessed 14 Oct 2020
77. Saunders S (2018) China food and drug administration issues new draft guidance for 3D printed medical devices. China Food and Drug Administration. https://3dprint.com/207575/cfda-3d-printing-draft-guidance. Accessed 12 Nov 2020
78. Wu BM, Borland SW, Giordano RA, Cima LG, Sachs EM, Cima MJ (1996) Solid free-form fabrication of drug delivery devices. J Control Release 40(1–2):77–87. https://doi.org/10.1016/0168-3659(95)00173-5
79. Wu W, Zheng Q, Guo X, Sun J, Liu Y (2009) A programmed release multi-drug implant fabricated by three-dimensional printing technology for bone tuberculosis therapy. Biomed Mater 4(6):065005. https://doi.org/10.1088/1748-6041/4/6/065005
80. Gbureck U, Vorndran E, Müller FA, Barralet JE (2007) Low temperature direct 3D printed bioceramics and biocomposites as drug release matrices. J Control Release 122(2):173–180. https://doi.org/10.1016/j.jconrel.2007.06.022
81. Wu C, Luo Y, Cuniberti G, Xiao Y, Gelinsky M (2011) Three-dimensional printing of hierarchical and tough mesoporous bioactive glass scaffolds with a controllable pore architecture, excellent mechanical strength and mineralization ability. Acta Biomater 7(6):2644–2650. https://doi.org/10.1016/j.actbio.2011.03.009
82. Yu DG, Yang XL, Huang WD, Liu J, Wang YG, Xu H (2007) Tablets with material gradients fabricated by three-dimensional printing. J Pharm Sci 96(9):2446–2456. https://doi.org/10.1002/jps.20864
83. Wang CC, Tejwani MR, Roach WJ, Kay JL, Yoo J, Surprenant HL (2006) Development of near zero-order release dosage forms using three-dimensional printing (3-DPTM) technology. Drug Dev Ind Pharm 32(3):367–376. https://doi.org/10.1080/03639040500519300
84. Huang W, Zheng Q, Sun W, Xu H, Yang X (2007) Levofloxacin implants with predefined microstructure fabricated by three-dimensional printing technique. Int J Pharm 339(1–2):33–38. https://doi.org/10.1016/j.ijpharm.2007.02.021
85. Yu DG, Branford-White C, Ma ZH, Zhu LM, Li XY, Yang XL (2009) Novel drug delivery devices for providing linear release profiles fabricated by 3DP. Int J Pharm 370(1–2):160–166. https://doi.org/10.1016/j.ijpharm.2008.12.008
86. Jacob J, Coyle N, West TG, Monkhouse DC et al (2014) Rapid disperse dosage form containing levetiracetam. US 2014/0271862 A1. 18 Sept 2014

87. Weigang WU, Qixin Z, Xiaodong GUO, Weidong H (2009) The controlled-releasing drug implant based on the three-dimensional printing technology: fabrication and properties of drug releasing in vivo. J Wuhan Univ Technol Mater Sci Ed 24(6):977–981. https://doi.org/10.1007/s11595-009-6977-1
88. Maier AK, Dezmirean L, Will J, Greil P (2011) Three-dimensional printing of flash-setting calcium aluminate cement. J Mater Sci 46(9):2947–2954. https://doi.org/10.1007/s10853-010-5170-4
89. Cheah CM, Leong KF, Chua CK, Low KH, Quek HS (2002) Characterization of microfeatures in selective laser sintered drug delivery devices. Proc Inst Mech Eng Part H J Eng Med 216(6):369–383. https://doi.org/10.1243/095441102321032166
90. Low KH, Leong KF, Chua CK, Du ZH, Cheah CM (2001) Characterization of SLS parts for drug delivery devices. Rapid Prototyp J 7(5):262–267. https://doi.org/10.1108/13552540110410468
91. Leong KF, Chua CK, Gui WS, Verani (2006) Building porous biopolymeric microstructures for controlled drug delivery devices using selective laser sintering. Int J Adv Manuf Technol 31(5–6):483–489. https://doi.org/10.1007/s00170-005-0217-4
92. Goyanes A, Robles Martinez P, Buanz A, Basit AW, Gaisford S (2015) Effect of geometry on drug release from 3D printed tablets. Int J Pharm 494(2):657–663. https://doi.org/10.1016/j.ijpharm.2015.04.069
93. Melocchi A, Parietti F, Loreti G, Maroni A, Gazzaniga A, Zema L (2015) 3D printing by fused deposition modeling (FDM) of a swellable/erodible capsular device for oral pulsatile release of drugs. J Drug Deliv Sci Technol 30(B):360–367. https://doi.org/10.1016/j.jddst.2015.07.016
94. Goyanes A, Buanz ABM, Hatton GB, Gaisford S, Basit AW (2015) 3D printing of modified-release aminosalicylate (4-ASA and 5-ASA) tablets. Eur J Pharm Biopharm 89:157–162. https://doi.org/10.1016/j.ejpb.2014.12.003
95. Skowyra J, Pietrzak K, Alhnan MA (2014) Fabrication of extendedrelease patient-tailored prednisolone tablets via fused deposition modelling (FDM) 3D printing. Eur J Pharm Sci 68:1–7. https://doi.org/10.1016/j.ejps.2014.11.009
96. Pietrzak K, Isreb A, Alhnan MA (2015) A flexible-dose dispenser for immediate and extended-release 3D printed tablets. Eur J Pharm Biopharm 96:380–387. https://doi.org/10.1016/j.ejpb.2015.07.027
97. Sandler N, Salmela I, Fallarero A, Rosling A, Khajeheian M, Kolakovic R (2014) Towards fabrication of 3D printed medical devices to prevent biofilm formation. Int J Pharm 459(1–2):62–64. https://doi.org/10.1016/j.ijpharm.2013.11.001
98. Water JJ, Bohr A, Boetker J, Aho J, Sandler N, Nielsen HM (2015) Three-dimensional printing of drug-eluting implants: preparation of an antimicrobial polylactide feedstock material. J Pharm Sci 104(3):1099–1107. https://doi.org/10.1002/jps.24305
99. Masood SH (2007) Application of fused deposition modelling in controlled drug delivery devices. Assem Autom 27(3):215–221. https://doi.org/10.1108/01445150710763231
100. Weisman JA, Nicholson JC, Tappa K, Jammalamadaka U, Wilson CG, Mills D (2015) Antibiotic and chemotherapeutic enhanced three-dimensional printer filaments and constructs for biomedical applications. Int J Nanomed 10:357–370. https://doi.org/10.2147/IJN.S74811
101. Petrak M, Rogers L (2014) Antimicrobial articles produced by additive manufacturing. WO2014/075185Al. 22 May 2014
102. Khaled SA, Burley JC, Alexander MR, Roberts CJ (2014) Desktop 3D printing of controlled release pharmaceutical bilayer tablets. Int J Pharm 461(1–2):105–111. https://doi.org/10.1016/j.ijpharm.2013.11.021

7 Computer-Aided Biopharmaceutical Characterization: Gastrointestinal Absorption Simulation and In Silico Computational Modeling

Daisy Arora and Bharat Khurana

Contents

Abstract

Biopharmaceutical characterization of drugs is the most important fundamental part of their development and discovery process. This plays a pivotal role in formulating an efficient dosage form with appropriate bioavailability. Absorption of drugs is a multifaceted process affected by several factors including the physicochemical properties of the drug and the pharmaco-technical parameters of the formulation. Several drugs during their development stages fail due to poor

D. Arora (✉)
Department of Pharmacy, Panipat Institute of Engineering and Technology, Panipat, Haryana, India

B. Khurana
Department of Pharmaceutics, Adarsh Vijendra Institute of Pharmaceutical Sciences, Shobhit University, Gangoh, Uttar Pradesh, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_7

biopharmaceutical properties. Thus to decrease the cost and time involved in the drug discovery process and to develop more effective dosage regimens, computer-aided in silico absorption models are required for better characterization of biopharmaceutical properties. One of the major objectives of in silico absorption models is to envisage the drug's physicochemical properties virtually. Computer simulations can be applied to predict the oral absorption of virtual compounds and thus offer the potential to screen the molecules under development that is having a prerequisite absorption profile. The present chapter deals with basics and recent advances along with applications and limitations of commonly used in silico and computational models for biopharmaceutical characterization particularly the ACAT model-based GastroPlus™ software package.

Keywords

GI absorption · ACAT model · In silico modeling · GatroPlus™ · Virtual trials · Parameter sensitivity analysis

Chapter Objectives

- To focus on the importance of biopharmaceutical characterization of drugs for successful product development.
- To understand the concept of gastrointestinal absorption simulation.
- To discuss various conventional and mechanistic approaches for predicting oral drug absorption.
- To demonstrate the application of computer-aided in silico absorption models for better characterization of biopharmaceutical properties of drugs.
- To be acquainted with several computer software packages for biopharmaceutical characterization, for example, GastroPlus™ software package.

7.1 Introduction

Though the oral route for administering drugs is the most accepted route as it is associated with certain advantages like relative safety and patient compliance, it also has certain limitations, like high variability and reduced bioavailability due to first-pass metabolism or poor biopharmaceutical properties of drugs [1]. Biopharmaceutical characterization of drugs plays a significant role in the successful discovery and development of drugs. Also, it is a basic tool for developing clinically successful dosage forms or pharmaceutical products. In conventional words, biopharmaceutics can be defined as the study of all associated influencing factors affecting the absorption of drugs. Orally administered drugs get absorbed through the gastrointestinal tract (GIT) which is a multifaceted process [2]. It includes several steps including the dissolution of the drug, gastric and intestinal transits, intestinal permeation, and intestinal metabolism as shown in Fig. 7.1. The most crucial factors which influence this complex process include physicochemical properties of the drug molecule, pharmaco-technical properties, and physiological factors [3]. Profiling of physicochemical properties of drugs during biopharmaceutical evaluation acts

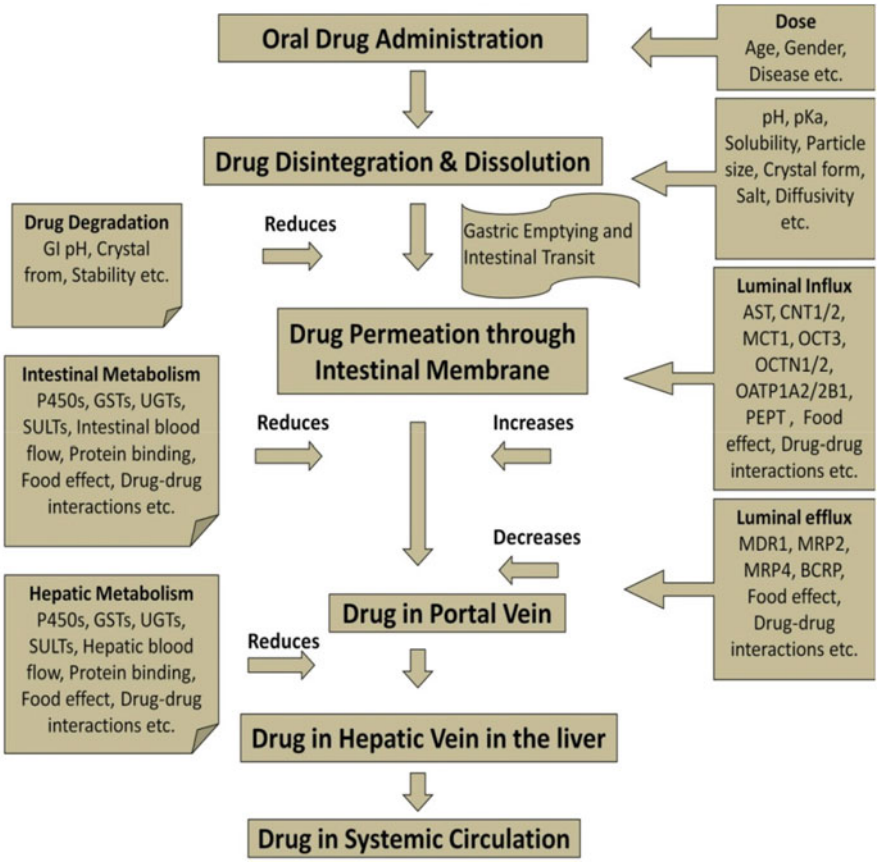


Fig. 7.1 Steps involved in the multifaceted oral absorption process and related factors

as a major tool to steer the formulation strategy and to predict the success of formulation clinically [4]. Biopharmaceutical information assists in designing formulation strategy as well as help in forecasting various factors affecting drug absorption, for example, effect of fasting or feeding, type of food, etc. These factors can be suitably considered for the improvement of the oral bioavailability of drugs during the initial stages of their development [5]. Several drugs during their development stages fail due to poor biopharmaceutical properties. Certain factors such as drug permeability, solubility, dissolution rate, and metabolic stability can be assessed by employing in vitro assays. Thus, the oral bioavailability of several drugs may be evaluated in a much efficient and cost-effective manner. But it is not fully capable of capturing the complexity of these multifarious absorption processes due to its dynamic nature [6]. Several modeling approaches can be utilized to anticipate and predict these properties such as quasi-equilibrium, steady-state, and dynamic models. More sophisticated statistical and mechanistic models such as physiology-based pharmacokinetic modeling can also be utilized to forecast the

biopharmaceutical properties of the drug molecules. However, these are also associated with one or more limitations [7].

Thus to decrease the cost and time implicated in the discovery process of drugs and to develop more effective dosage regimens, computer-aided in silico absorption models are an alternative for better characterization of biopharmaceutical properties [8].

In silico simulation of gastrointestinal absorption utilizes fewer resources with better outcomes and bestows a better perspective for selecting and screening virtual compounds. Since the 1990s, several mechanistic dynamic models have been developed and validated [9], such as Compartmental Absorption and Transit model (CAT); Grass model; GI Transit Absorption model (GITA); Advanced Compartmental Absorption and Transit model (ACAT); and Advanced Dissolution, Absorption, and Metabolism model (ADAM) which are discussed in this chapter in brief. Along with this, an exemplary demonstration of ACAT model integrated GastroPlus™ software is given in detail.

7.2 Biopharmaceutical Characterization: Theoretical Background

Several drugs in the formulation phases of their development were found to have poor biopharmaceutical properties concerning their oral bioavailability [6]. Failure of the development process at later stages results in the increased cost involved during extensive in vitro and preclinical in vivo studies. Thus, it is requisite to assess these properties much earlier during development stages to reduce concomitant costs and time [10]. Several factors and principles have been known as empirical approaches that establish the absorption profile of drugs. Figure 7.2 represents "Lipinski's Rule of Five" developed by Lipinski and the team in 1997. According to the Lipinski rule, certain properties of drug compounds can be utilized in predicting their oral absorption. This method suggests that if a drug follows all the rules, it is expected that drug would show better absorption and permeation profile [11].

Biopharmaceutics Classification System (BCS) is another empirical method widely used for predicting oral absorption based on water solubility and intestinal permeability of drug. This classification was developed by Amidon and co-workers in 1995 [12]. Accordingly, the drug compounds are classified into four categories (i.e., BCS class I, class II, class III, and class IV) (described in Fig. 7.3). This is a widely used basic tool in early drug development as well as in both preclinical and clinical drug development stages and can predict rate-limiting steps in the absorption process.

BCS is associated with three dimensionless numbers, that is, the absorption number (An), the dose number (Do), and the dissolution number (Dn) represented as Eqs. 7.1, 7.2, and 7.3, respectively. These are based on physicochemical and physiological properties affecting drug absorption [13].

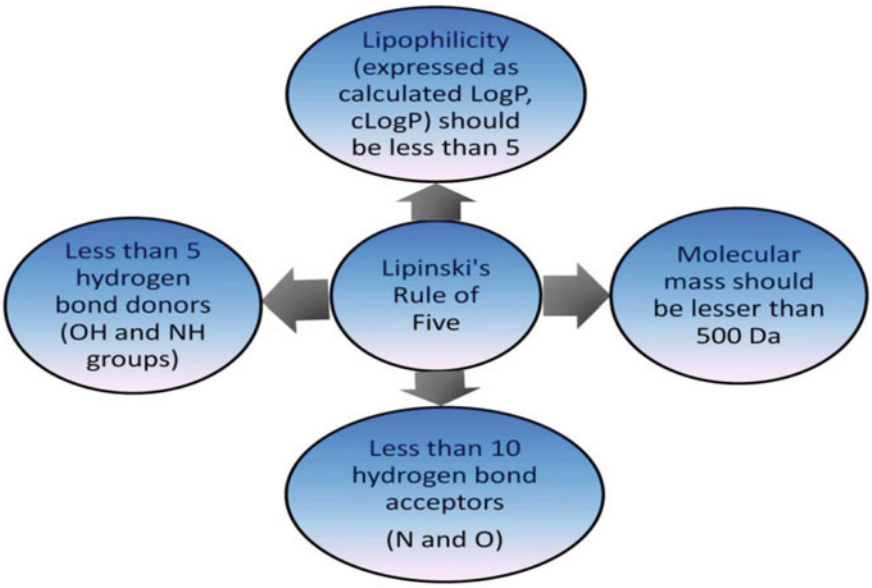


Fig. 7.2 Lipinski's rule of five for predicting oral absorption of drugs

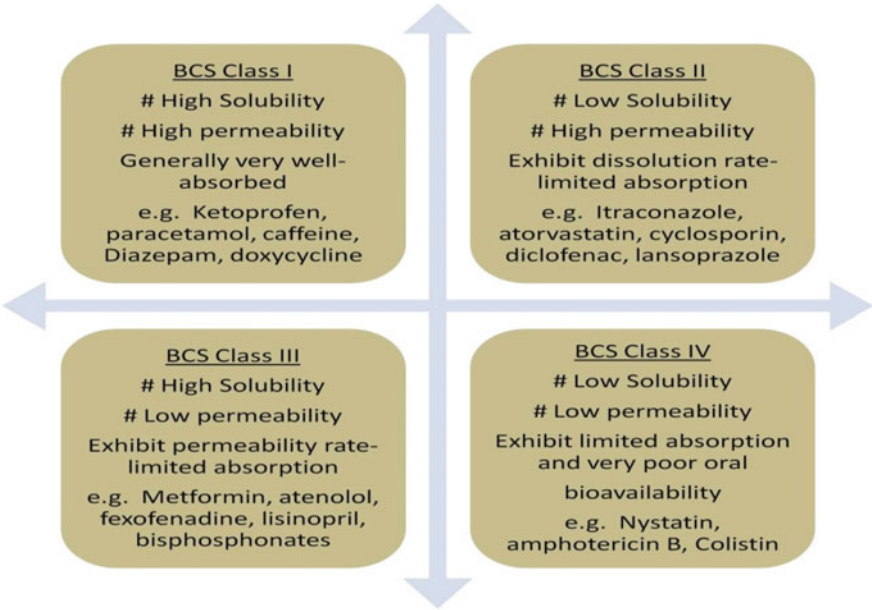


Fig. 7.3 Biopharmaceutical Classification System (BCS) for drugs

$$\text{Absorption number (An)} = \frac{\text{The mean transit time in the small intestine}}{\text{The mean absorption time}} \quad (7.1)$$

$$\text{Dose number (Do)} = \frac{\text{Dose of drug}}{\text{Intrinsic solubility}} \quad (7.2)$$

$$\text{Dissolution number (Dn)} = \frac{\text{Mean transit time}}{\text{Mean dissolution time}} \quad (7.3)$$

The concept of BCS was extended by including the elimination process of drugs. It also considers the effects of drug transporters and efflux phenomenon on drug oral absorption. This modified version is named "Biopharmaceutics Drug Disposition Classification System" [14]. It can also predict the disposition of the drug through transport/absorption/elimination interplay. Besides these, several non-empirical approaches based on absorption modeling and simulations for biopharmaceutical characterization are described below.

7.3 Gastrointestinal Absorption Modeling and Simulation

The static methods of predicting physicochemical parameters of the drug, for example, drug permeability, dissolution rate, and stability, could not completely demonstrate the dynamic and multifaceted process of in vivo absorption [6, 15]. Thus, mathematical and mechanistic models were developed which aided in integrating and extrapolating in vitro data into in vivo information.

The conventional mathematical models were based on two elementary parameters of drug molecule (i.e., solubility and permeability). Absorption potential model and Maximum Absorbable Dose equations are examples of such simple and quick models that can estimate the extent of absorption under "static" conditions [16, 17]. The former can predict the fraction of dose that can be absorbed, while the latter can simulate residence time in small intestinal and thus can predict the maximum amount of a drug that can be absorbed within a period of 6 h [6]. However, these models are also associated with certain limitations such as limited prediction capacity in a dynamic manner like it cannot predict the influence of pH variability in different compartments of GIT, blood perfusion rate, the effect of food, etc. [18, 19].

Further, the applications based on mechanistic models, as well as physiology-based in silico absorption models have become more popular and gained significant interest as prediction tools for gastrointestinal absorption [20]. The mechanistic models are based on physicochemical properties of the drug, pharmaco-technical factors, and physiological factors, and their relationship helps in simulating the gastrointestinal absorption for drugs to be intended for oral administration [21]. Mechanistic and sophisticated absorption models based on physiology can predict the bioavailability and absorption parameters more accurately by improving the understanding related to the factors that govern intestinal absorption [22].

7.4 Mechanistic Approaches for Predicting Oral Drug Absorption

In the last few decades, several strategies like quasi-equilibrium, steady-state, and dynamic models were anticipated with different features and properties as explained in Fig. 7.4. In the earlier 1980s, pharmacokinetic modeling was used to be carried out by treating the entire GIT as a single compartment. This "black box" approach was considered as a simplified model of a very complicated method.

This resulted in the generation of quantitative and mechanistic methods for investigating GI absorption. At that time, general methods which were widely adopted for predicting dynamic oral absorption of drugs were based on physiological approaches, namely, dispersion models and mixing tank (compartment) models [23]. In the simplest mixing tank model, it is assumed that the gastrointestinal tract is a "single well-stirred compartment" associated with linear kinetics for all transfer processes. In this, each compartment is considered to be homogenously mixed having a uniform drug concentration. In the late 1980s, some scientists have simulated dose-dependent absorption using two-compartment models, while a few scientists have also predicted dissolution-controlled drug absorption by treating GIT

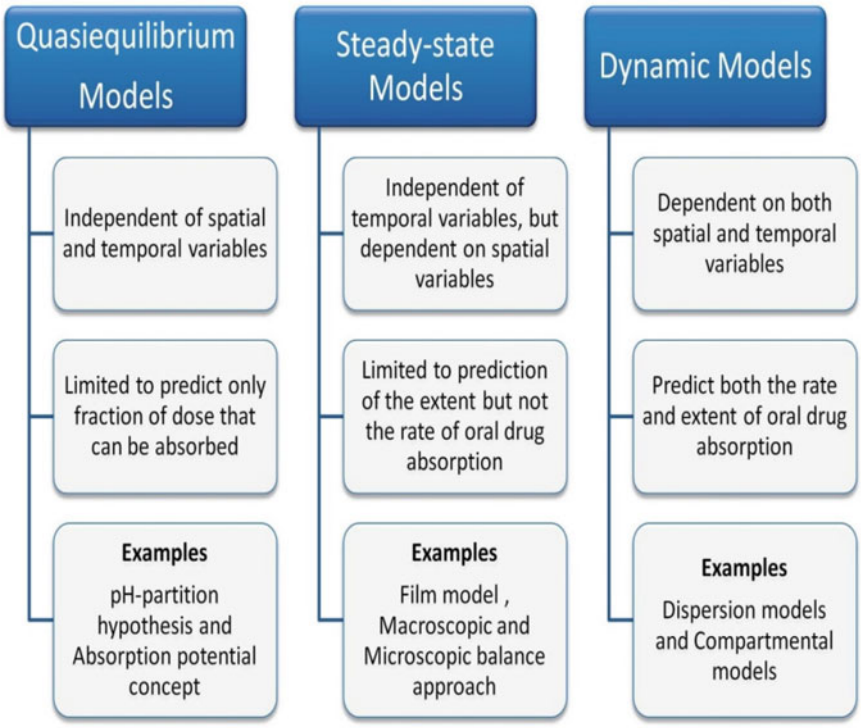


Fig. 7.4 Classification of mechanistic models with their examples

as a single compartment [24]. In a study, the plasma-level double-peak phenomenon was also explained using four compartments [25].

Another conventional absorption model is known as the dispersion model. It portrays the small intestine as a uniform tube with axial velocity, dispersion behavior, and concentration profile across the tube [26].

More advanced dynamic models such as the CAT model, the Advanced CAT (ACAT) model, the Advanced Dissolution, Absorption and Metabolism (ADAM) model, the Grass model, and the GI Transit Absorption (GITA) model represent physiology of the gastrointestinal tracts like drug transit, dissolution, and absorption could be successfully used to demonstrate biopharmaceutical properties more accurately and precisely. Thus, drug plasma concentrations versus time profiles can also be estimated virtually. The biopharmaceutical properties of drugs estimated by dynamic models can be used during various phases of the development process. In silico tools associated with these may also help in assessing critical factors related to the absorption process utilizing certain input information and database virtually. Several industry-oriented and commercial computer software are available which are developed to integrate these dynamic models for computer-aided biopharmaceutical characterization such as NONMEM, PhysioLab®, WinNonLin®, WinNonMix®, SimCYP, PK-Sim®, GastroPlus®, IDEA® (no longer available), Cloe® PK, Cloe®HIA, and INTELLIPHARM® PKCR. Several advanced dynamic models are discussed below in brief.

7.4.1 Compartmental Absorption and Transit (CAT) Model

Following the mixing tank model, Yu et al., in 1996, introduced the Compartmental Absorption Transit (CAT) model [24]. It is constituted of several compartments in series representing the whole intestinal tract with the particular physiological meaning of every compartment. Thus it also represents multiple compartments and describes various effects such as gastric emptying, intestinal transit, etc. According to the CAT model, the small intestine could be best represented by seven major compartments as represented by Fig. 7.5. Among these, the duodenum is the first compartment; the jejunum is represented as the next two compartments, while the last four compartments illustrated the ileum. The movement or transit from one intestinal compartment to another can be represented and controlled by a particular transit rate constant (k_t) and absorption from each compartment can be represented

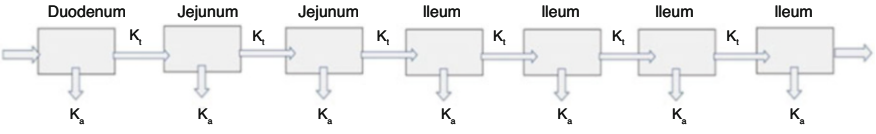


Fig. 7.5 A schematic diagram of the Compartmental Absorption and Transit (CAT) model representing small intestine separated into seven individual compartments

by absorption rate constant (k_a). The basic equation for the CAT model can be represented as follows (Eq. 7.4):

$$\frac{dY_n}{dt} = k_t Y_{n-1} - k_t\, Y_n - k_a\, Y_n \tag{7.4}$$

where n represents the total number of compartments (like $n = 1, 2, 3, ----7$), and Y_n is the amount of drug in the nth compartment. Also, the amount of drug in the previous compartment ($n - 1$) is represented by Y_{n-1}, k_t is the transit rate constant, and k_a is the absorption rate constant [9].

The model is based on certain assumptions such as passive absorption, instantaneous dissolution, linear transfer kinetics for each segment, and negligible absorption from the stomach and colon [23]. Thus, formerly it was developed for predicting the oral drug absorption of stable and highly soluble drugs.

However, the extent or fraction of drug absorption based on effective permeability of drug molecules can also be described particularly for the drugs showing different absorption characteristics [27]. Instead of considering drugs as a single species, the absorption profile can be simulated based on unreleased, undissolved, and dissolved molecules of the drug. Later on, a few modifications were carried out, and the CAT model was extended for several other estimations as follows:

1. Dose-dependent drug absorption or carrier/transporter-mediated absorption can also be computed by applying the fundaments Michaelis–Menten kinetics.
2. The absorption of drugs that can be degraded in GIT could be predicted by considering the gastric emptying rate constant and compartment-dependent degradation rate constant.
3. An additional compartment representing a controlled release dosage form could be used to predict the absorption of part of the dose from controlled release dosage forms.
4. It can also be employed to simulate the fraction of dose absorbed for poorly absorptive drugs as well as can be used to determine the reason for poor oral absorption by considering gastric emptying and dissolution.

7.4.2 Grass Model

Analogous to the CAT model, a scientist named Grass has developed another physiologically based multiple-compartmental model in 1997 [28]. This model predicts absorption from the gastrointestinal compartments based on drug solubility and permeability and tissue surface area. The movement of fluids in the gastrointestinal tract (i.e., transit and emptying) can be described by this model. Additionally, the flux of drug absorption in each gastrointestinal tract can be calculated. But still, it has limited applications because it cannot fully consider pre-systemic metabolism, drug degradation, or the active transport of drugs [29]. It was integrated with the IDEA™ and IDEA pkEXPRESS™ software which are not currently available [1]. STELLA® (isee Systems, Inc.) is one of the software based on this Grass model which has also limited applications.

7.4.3 GI Transit Absorption (GITA) Model

This model was first presented by Sawamoto and co-workers in 1997 to predict oral absorption in rats [30]. The basis of this model is variation in intestinal transit time across different segments of the intestine which can affect the absorption of drugs [29]. Variation in physiology and structures as well as dissimilar expression patterns of metabolizing enzymes and transporters also result in dissimilar absorption patterns in each segment [9]. Thus, the developers have divided the GI tract into eight compartments (stomach, duodenum, upper jejunum, lower jejunum, upper ileum, lower ileum, cecum, and large intestine), with each compartment having different transit and absorption kinetics allowing more flexibility in the transit process [29, 30].

According to this model, a standard in situ closed-loop method was suggested to determine the absorption rate constant for each compartment. Moreover, in vivo studies employing a nonabsorbable marker (i.e., phenyl red) were used to predict GI transit for each compartment. Due to the invasive nature of experimentation, the GITA model was originally proposed to predict oral drug absorption in rats instead of humans. Lately, it was modified by Kimura and Higaki for predicting oral absorption in humans where gamma scintigraphy was used to determine the transit rate constant in humans [31]. While the absorption rate constant in a human was determined by extrapolating the data of rats after normalizing these based on interspecies differences in surface area and luminal volume of the small intestine. The segmental absorption profile and rate of movement of drug from one segment/compartment to another can be described by applying the following equations (Eqs. 7.5 and 7.6):

$$\text{For stomach}: \frac{dX_s}{dt} = -(k_s + k_{as})X_s \tag{7.5}$$

$$\text{For intestine}: \frac{dX_{i+1}}{dt} = X_i k_i - (k_{i+1} + k_{ai+1})X_{i+1} \tag{7.6}$$

where the initial dose administered at the initial time is represented by X_s. The amount, the transit rate constant, and the absorption rate constant are indicated by X, k, and k_a, respectively. The stomach and each site of the intestine are indicated by the subscripts "*s*" and "*i*," respectively.

This is also used for predicting site-specific oral drug absorption owing to flexibility in its transit process [32]. The effect of drug–drug and food–drug interactions on the absorption of drugs can also be studied by this model [31].

7.4.4 Advanced Dissolution, Absorption, and Metabolism (ADAM) Model

The ADAM model is quite similar to the basic CAT and ACAT model. It also represents the small intestine as seven different compartments similar to other

compartment models. Similar to the CAT model, the ADAM model accounts for the processes of dissolution, GI fluid transit, gut wall permeation, drug degradation, intestinal metabolism, and active transport processes [9, 33]. Considerations regarding variability in the GI tract such as the varied distribution of enterocytic blood flow, enzymes in the gut wall, food-induced changes in gastric emptying, splanchnic blood flow, and luminal pH are also included and simulated in the ADAM model. The main difference between the ADAM and CAT model is that in the ADAM model, the dissolution is calculated using the more sophisticated Wang–Flanagan generalized model [34] rather than the conventional Noyes–Whitney equation. Commercially available simulation software Simcyp® (http://www.simCYP.com) was developed to integrate the ADAM model [35]. Formerly, it was introduced as a simulator for the metabolism process, but later on, it was expanded for determining pharmacokinetic properties of drugs including absorption.

7.4.5 Advanced Compartmental Absorption and Transit (ACAT) Model

ACAT model is the extended and advanced version of the previous basic CAT model. It was developed with the added processes such as first-pass metabolism and colon absorption. Biopharmaceutical Classification System (BCS) and previous understanding of gastrointestinal physiology are the main basis of this semi-physiological absorption model. A series of linear and nonlinear rate equations are usually combined to frame this model which imitates the effect of physiological conditions on drug absorption as it passes through succeeding gastrointestinal compartments. It includes both linear transfer kinetics and nonlinear metabolism/transport kinetics. Several factors that are considered in the ACAT model for determining oral drug absorption are enlisted in Table 7.1.

The ACAT model is comprised of serially linked nine compartments. Different parts/segments of the gastrointestinal tract are represented by these compartments as represented in Fig. 7.6 with the below-given details:

- Compartment 1: Stomach
- Compartment 2: Duodenum
- Compartment 3 and 4: Jejunum

Table 7.1 Factors that are considered in the ACAT model for determining oral drug absorption

Physicochemical factors of drug	Physiological factors	Dosage factors
pKa	pH of GIT fluids	Dosage form
Solubility	Gastric emptying	Dose
Particle size	Intestinal transit time and rate	
Effective surface area	First-pass metabolism Luminal transport	
Particle density	Transporter expression	
Permeability		

- Compartment 5, 6, and 7: Ileum
- Compartment 8: Caecum
- Compartment 9: Ascending colon

It is also associated with the following six states of drug component:

1. Unreleased
2. Undissolved
3. Dissolved
4. Degraded
5. Metabolized
6. Absorbed

The following three states of excreted material are considered under the ACAT model:

1. Unreleased
2. Undissolved
3. Dissolved

A sequence of differential equations is used to describe each compartment and sub-compartment of the model. Several processes which determine the dissolved drug concentration change in each gastrointestinal segment have been described in Table 7.2.

Kinetics of every process can be represented by an appropriate "rate constant." For example, lumenal transit can be represented as k_t, which can be evaluated from the mean transit time within each compartment.

Appropriate formulation parameters and the drug properties can be utilized to compute the "dissolution rate constant" (k_d) for the individual compartment. Product of drug effective permeability and individual compartment's absorption scale factor is utilized in the determination of "absorption rate constant" (k_a).

According to it, the ionization of drugs at GI pH leads to decreased effective permeability of drugs. Thus, ACAT model can be used to simulate nonlinear saturable Michaelis–Menten kinetics of metabolism [3, 9].

The half-life or degradation rate of the drug versus pH is used to interpolate the degradation rate constant (k_{degrad}) in different compartments of different pH conditions. The absorption and exsorption rates are primarily based on the difference in drug concentration on the apical and basolateral enterocyte membranes [3].

According to the depicted ACAT model, drug absorption through the basolateral membrane of enterocytes undergoes the first-pass metabolism as it passes through the portal vein and liver and finally reaches into the systemic circulation. After this, the conservative pharmacokinetic model or a PBPK disposition model is associated with the ACAT model. This is a supplementary attribute integrated with the latest edition of ACAT-based software packages.

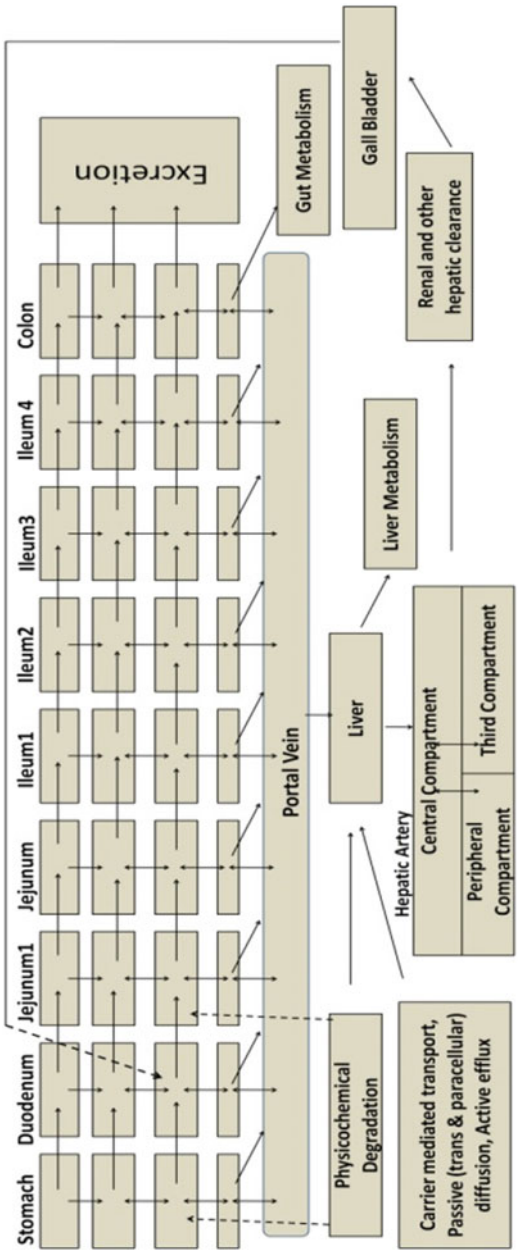


Fig. 7.6 Diagrammatic representation of the ACAT model on which the commercial GastroPlus™ software is based

Table 7.2 Processes responsible for the change of the rate of dissolved drug concentration in each GI compartment as per the ACAT model

Processes
1. Transfer of drug molecule in the compartment
2. Transfer of drug molecule outside the compartment
3. Liberation of the drug molecule into the compartment
4. Dissolution of drug molecules
5. Precipitation of drug molecules
6. Lumenal degradation of the drug molecules
7. Absorption of the drug molecules into the enterocytes
8. Exsorption of the drug molecules from enterocytes to lumen
9. Absorption of the drug molecules into the portal vein via the paracellular pathway
10. Exsorption of the drug from a portal vein via the intercellular pathway

The ACAT model is implemented in very well-known and broadly employed commercially available software, GastroPlus™ (Simulations Plus Inc., Lancaster, CA, USA). This software has incorporated certain improvements into the conventional ACAT model which enhanced the absorption predicting competence of the model. The application and other features of the GastroPlus™ software are discussed in the following sections of the chapter.

7.5 Exemplary Demonstration of ACAT Model Integrated GastroPlus™ Software

Overall, among all physiology-based pharmacokinetic models, the widely prominent models are the ACAT and ADAM models [6]. GastroPlus™ is an advanced technology computer program-based simulation software package that is designed to simulate and predict the pharmacokinetic and pharmacodynamic properties of the drugs in humans as well as animals including their absorption through different routes such as dermal, inhalation, ocular, oral, etc. The project decisions could be achieved much faster and more accurately using this user-friendly interface which in combination with prevailing facts makes it an integrated platform. Currently, in 2020, its 9.8 version has been launched. A few relevant information regarding the software installation, its application, and availability are described in Table 7.3.

Various biopharmaceutical factors which may cause a significant effect on the bioavailability of oral drugs could be evaluated using GastroPlus™ computer simulations. This can be used to achieve the designing of optimized experimental formulations accurately [5].

Population Estimates for Age-Related Physiology (PEAR Physiology™) is an internal module of GastroPlus™ which could be implemented to calculate organ physiology for a wider age group of human models ranging from 0 to 85 years. It is also very much efficient in conducting population simulations using user-defined

Table 7.3 Features of GastroPlus software

Software/App. name	GastroPlus
Recent version	9.8
Company name	Simulations Plus, Inc., USA
Location and website	https://www.simulations-plus.com/gastroplus
Machine requirements for installation	Essential machine/computer requirements include Pentium-class Windows (7 or above) with a network server. For best performance, a minimum of 1 GB RAM is suggested
First launch	Simulations Plus released its first pharmaceutical software product, GastroPlus®, in August 1998
Availability	Paid
Applications in pharmaceutical R&D	• Simulation: To simulate various absorption models in humans as well as animals such as intravenous, and extravascular absorption, biopharmaceutics, pharmacokinetics, and pharmacodynamics [38, 39] • Drug–drug interaction: To predict certain drug–drug/drug–metabolite interactions including mechanistic and static [40] • Physiologically based pharmacokinetic (PBPK) modeling: To predict the concentration of drugs in individual tissue by simulating the drug distribution and drug elimination throughout the body [41] • ADMET prediction: To predict physicochemical characteristics, pharmacokinetics, and Cytochrome P450 induced metabolism parameters from molecule structure which assist with lead selection and optimization activities [7] • Exploration of additional dosage routes: To simulate mechanistic absorption and disposition through dermal (topical and subcutaneous), intraoral, intranasal, ocular, and i.m. routes [38]

coefficients of variations and other statistical parameters, thereby generating a unique new population [20].

As discussed previously, the ACAT model consists of nine compartments. Complex differential equations already defined in user manuals are used to describe several processes by each of the compartments such as the liberation of drug release from the dosage form, rate of dissolution/rate of solubilization/rate of precipitation, chemical stability, permeability, transporter-based influx/efflux, and first-pass metabolism, especially via gut wall enzymes. Maximum processes which are passive in nature are well defined by linear kinetics, while saturable/carrier-mediated transport processes and metabolic processes can be well defined and described using zero-order kinetics (explained by Michaelis–Menten nonlinear kinetics). The kinetics involved is generally based on certain assumptions as of conventional compartment modeling such as passive absorption and transport through the small intestine, fast dissolution, uniform residence time between compartments, though dissimilar fluid volume and fluid flow rate can be considered for different compartments [36, 37].

Modeling of different processes in these compartments such as the release of the drug, dissolution absorption, and metabolism of drugs after administration can be done by applying several differential equations. The fraction of dose absorbed or

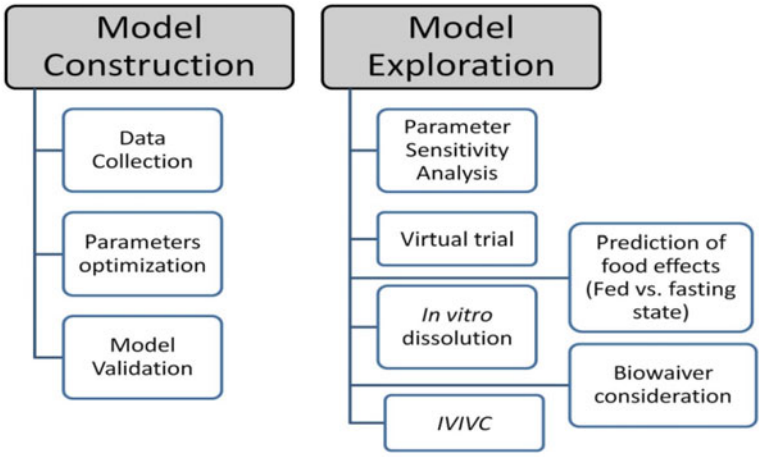


Fig. 7.7 Schematic strategy of GI simulation and modeling with additional features in GastroPlus™ software

unabsorbed can also be determined [42]. Additionally, PBPK modeling can also be achieved as drug distribution in the majority of tissues, including both perfusions limited or permeability limited distribution, can be described. All the compartments are mutually attached to arch other via blood and are represented as individual compartments. Key input factors/parameters related to absorption, distribution, metabolism and excretion (ADME) can be integrated to simulate PK parameters along with plasma and tissue drug concentration versus time profiles.

The requirement of enormous input data is the major challenge in implementing this model. However, several techniques can be used to predict the required data in silico make this approach more favorable [8].

In general, various steps of modeling and simulation via GastroPlus™ include data collection, parameter optimization, and model validation. A schematic strategy of GI simulation and modeling with additional features is represented in Fig. 7.7.

Certain other attributes can also be explored using this developed absorption model specific to the individual drug. A few are mentioned below:

1. Understanding the effect of pharmaco-technical properties or physicochemical properties of the drug on drug PK profile
2. Providing dissolution profile (in vivo)
3. IVIVC
4. Bio-relevant specifications related to dissolution process for the desired formulation
5. Predicting the effect of diverse dosing regimens
6. Predicting food effects on drug pharmacokinetics
7. Achieving stochastic simulations on a group of virtual subjects

7.5.1 Model Construction

The initial step for model construction is data collection. Several input parameters are required in the mechanistic absorption model, which can either be predicted using in silico techniques, literature survey, or can be determined experimentally. Drug or dosage form-related properties such as drug solubility and its relationship with pH of the aqueous solvent, particle size, drug permeability, type of dosage form, etc. can be predicted either experimentally (in vitro and in vivo) or in silico can be used as input data for the model.

Certain physiological factors (such as gastrointestinal transit time, gastric emptying time, gastric pH, the surface area of absorption site, concentration of bile salt and other enzymes, size of compartment, fluid volume, etc.) are by default entered into the model and, however, can be modified also as per the user requirements. Alternatively, the full ADMET Predictor program or the optional ADMET Predictor™ module helps in predicting some input parameters for the GastroPlus model using the structural features of the molecule [20].

There are various unique models reported for several immediate releases as well as controlled release drug formulations constituting drugs of different BCS classes. Based on requirements, several physiological conditions either fasted or fed state for both human and animal physiologies were used for simulation studies. It can also explore their ability to provide a choice between a single simulation and a virtual trial mode (which predicts the possible inter-subject variations in the model). Biopharmaceutical properties of several drugs and dosage forms have been characterized using the same methodologies such as gastrointestinal simulation of "nimesulide oral absorption" which was carried out by Grbic and co-workers in 2012 [43]. In this study, two self-regulating analysts have constructed drug-specific models for absorption utilizing identical data (in vivo); however, in both cases, different assumptions regarding major factors affecting the absorption of nimesulide were considered.

Several physicochemical and pharmacokinetic parameters of nimesulide which were used as input parameters were molecular weight of the drug (g/mol), log *D* value at pH 7.4, pKa, the permeability of human jejunal (cm/s), amount of drug administered (mg), the volume of dose (mL), the solubility of the drug (pH 4.5) (mg/mL), mean precipitation time (s), diffusion coefficient (cm^2/s), the density of drug particles (g/mL), effective particle radius (μm), the weight of the patient (kg), first-pass extraction in the liver (%), the ratio of plasma/blood concentration, plasma protein-bound and free percent (%), clearance (L/h/kg), Vc (L/kg), $t_{1/2}$ (elimination) (h), simulation time (h), and dosage form. Some of these data were adopted from literature, while others were either in silico predicted (ADMET Predictor™ module) or used as default GastroPlus™ values. To select and categorize the best estimation yielding model between both of these, the actual clinical data were compared with the simulation results. The results indicated that both the models have predicted the average blood profile of nimesulide accurately after its oral administration. It is ratified from the results that both the models have predicted the parameters well as the percent prediction errors for C_{max} and area under the curve in both the cases were found to be in specified limits (less than 10%).

Similarly, GastroPlus™-assisted model development for biopharmaceutical characterization was carried out for several other drugs such as gliclazide [44], entrectinib [45], carbamazepine [46], and talinolol [47].

7.5.2 Model Exploration

Following the construction, optimization, and validation of the PK model, GastroPlus™ can be explored for several other unique features for GI modeling and simulation. Such attributes are discussed in the following sections.

7.5.2.1 Parameter Sensitivity Analysis

The intra-model exploration of the absorption model for the specific drug can be used for understanding the effects of formulation parameters and the drug-related properties on the predicted pharmacokinetic profiles. This type of analysis is known as parameter sensitive analysis (PSA) in GastroPlus™. It can be performed by two techniques. In the first technique, one variable is varied within the predetermined range while keeping the others on baseline levels. The second technique is also called three-dimensional PSA because of variation in two variables at a time for assessment of their combined effect [8].

It can be used for generating more bio-relevant data from the roughly estimated input values of physicochemical properties (using in silico predictions). In this case, if a poor correlation exists between predicted values and in vivo results, PSA can also be used as a tool in drug-specific model construction. In the case of drugs with highly varied therapeutic outcomes, PSA can be used to predict the inter-subject variation in pharmacokinetic profiling of absorption of the drug. It may also play an important role in formulation design and development. For example, if the constructed model predicts a poor absorption percentage for a novel drug candidate, PSA can help in the identification of critical parameters and factors limiting drug absorption and bioavailability. After the identification of these possible factors, several approaches can be used to overcome the predicted formulation challenges. Thus, this may help the formulation scientist in saving time and effort during pre-development or development phases, and thus formulation cost can also be reduced.

7.5.2.2 Virtual Trial

After the initial stages of formulation development, a very important part for the formulator is to anticipate the oral bioavailability as well as inter-subject variability of drug oral absorption. These studies can provide a better insight into the outcomes of the formulation. The virtual feature of GastroPlus™ can be used to simulate the effects of formulation in variable population and/or the combined effect of input parameters which are not precise values but are within a predicted range.

The process includes a random sampling of values of certain selected variables from predetermined distributions and the performance of stochastic simulations on several virtual subjects. These values are represented as averages with the absolute or logarithmic form of coefficients of variation (CV %). Coefficients of variation values

are either predicted using in silico prediction models or analyzed based on a literature survey and prior knowledge. The simulation results are represented as average value with CV % for bioavailability, t_{max}, C_{max}, and area under the curve values, the fraction of drug absorbed, etc. Along with these, it can also determine the average plasma concentration–time curve, 90% confidence intervals, and probability contours [8].

7.5.2.3 Prediction of Food Effects (Fed vs. Fasting State)

The food present in the gastrointestinal tract can modify the absorption pattern of the drug due to changes in several physiological mechanisms/factors [34, 48, 49]. Several other physiological changes due to food consumption are represented in Fig. 7.8. In addition to physiological changes, the presence of food can also change the solubility, dissolution, and permeation of drugs. For instance, lipophilic drugs often show high systemic bioavailability with food, because the presence of the food stimulates the bile salts secretions and also high lipid contents which further improve the solubility of lipophilic molecules.

Whereas hydrophilic drugs show low bioavailability, the food present at the absorption site may interfere with the drug permeation process [50]. The frequently used method to determine the effect of food on the absorption of the drug is based on animal studies. However, animal studies conducted on animals for analyzing the

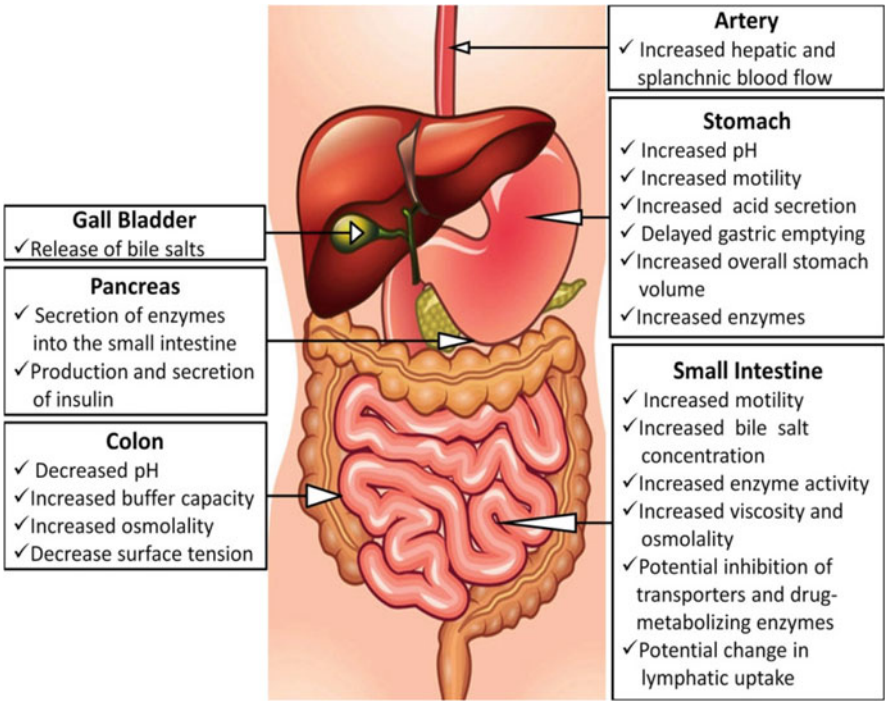


Fig. 7.8 Major physiological changes in GIT due to food consumption (adapted from [34])

effect of food on the drug's bioavailability and absorption cannot predict the accurate food effects in humans due to the different physiological factors across the species.

This further complicates the investigation of food-related factors affecting drug bioavailability in humans. To overcome the disadvantages related to animal experimentation, the best alternative is an in silico simulation using physiological-based absorption models.

These are the mathematical models simulating the different conditions present in the gastrointestinal tract during fed and fasted state; they can depict the drug transit kinetics, kinetics of drug dissolution, and absorption based on drug-related properties like permeability, bio-relevant solubility, dissociation constant values, dose, distribution, and disposition data.

Gastroplus™ utilizes in-built physiologic parameters that vary amid fasted and fed conditions. For example, stomach pH changes from 1.3 (fasted) to 4.9 (fed). Stomach transit time varies from 0.25 h for fasted to 1.0 h for the fed state, while stomach volume varies from 50 mL for fasted and 1000 mL for the fed state. Hepatic blood flow varies from 1.5 L/min for fasted state and 2 L/min for the fed state [42].

In a recent study, physiologically based absorption modeling was conducted using GastroPlus™ to investigate the effect of food and a change in gastric pH on the PK of entrectinib [45]. Entrectinib is a lipophilic base with reasonable permeability and good water solubility at acidic pH. However, a marked reduction in drug solubility is noticeable as pH was increased, and this raises the potential for a pH-dependent drug–drug interaction (DDI) when entrectinib is dosed with gastric acid-reducing agents. Moreover, a significantly higher (~40 times) bio-relevant solubility was observed in Fed State Simulated Intestinal Fluid than in Fasted State Simulated Intestinal Fluid.

Another model was developed for verifying the effect of the acidulant-containing formulations on simulated pharmacokinetics parameters and relative bioavailability studies were carried out.

Non-acidulant formulations were compared with the acidulant formulations with or without lansoprazole (a drug used for gastric acid reduction). The study was based on bio-specific in vitro dissolution measurement studies as well as in silico modeling. Results predicted an insignificant effect of food and very less pH-dependent DDI for the market dosage form which was also established through clinical studies. The process of model construction, verification, and purpose are represented in Fig. 7.9 as adapted from [45].

7.5.2.4 In Vitro Dissolution and In Vitro–In Vivo Correlation

GastroPlus™ is one of the most commonly used software programs of its kind for exploring in vitro–in vivo correlations and widely explore in several publications from the Food and Drug Administration. In vitro–in vivo correlation (IVIVC) along with computational simulation tools can be utilized to simulate distinctive PK properties of a drug. It aids in developing novel generic drug/drug products efficiently with reduced cost and time. It can be described that in vivo performance can be efficiently simulated by in vitro dissolution test conditions if it is demonstrated by level "A" IVIVC [51].

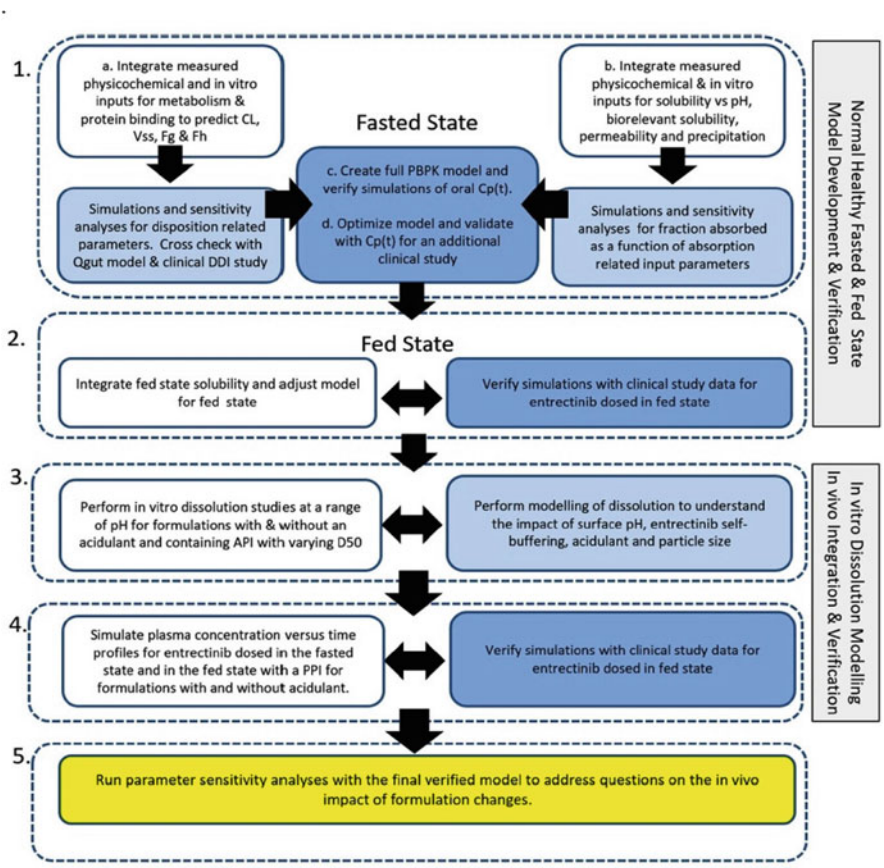


Fig. 7.9 Various steps in model construction, verification, and application for predicting the food effects on absorption of entrectinib using GastroPlus™ (adapted from [45])

Development of new formulations and the selection of bio-relevant dissolution conditions with IVIVC can be achieved. Determination of bio-relevant dissolution conditions with IVIVC can be selected by additional features in GastroPlus™. Convolution and deconvolution are two approaches that can be utilized to establish IVIVC. The plasma drug concentration can be predicted using convolution, while deconvolution approach is used to simulate the in vivo dissolution profile. Accordingly, in vitro dissolution tests can only be employed instead of in vivo studies to predict the changes in the safety and efficacy profile of drugs upon the establishment of IVIVC.

First, anticipated plasma concentration–time profiles can be predicted using a series of in vitro dissolution values by employing the convolution approach is the key input data in GastroPlus™ software. Following this, in vitro–in vivo correlation can be established by correlating in vivo mean drug plasma concentration profile with the obtained dissolution profile. While, in the deconvolution approach, to

identify bio performance' dissolution conditions, a graph can be plotted between software-generated in vivo dissolution profiles versus the obtained in vitro dissolution profiles.

In a study by Honório and co-workers, they successfully established the IVIVC and developed the generic efavirenz tablets using GastroPlus™ [51].

To establish an IVIVC, a set of data was generated by conducting dissolution studies of efavirenz powder and efavirenz tablet. On the basis of physicochemical properties of efavirenz, plasma profile was simulated which was found to be almost similar to the observed in vivo results obtained for both the bio batches. Both Wagner–Nelson ($r^2 = 0.85$) and for Loo–Riegelman models ($r^2 = 0.92$) were used to establish the level "A" IVIVC for the dissolution methods for both generic candidates. A solution of 0.5% sodium Laureth sulfate was considered as a bio-relevant medium for dissolution of efavirenz formulations as per the % of its fraction absorbed values. The simulation of GI bioavailability and in vitro–in vivo correlations achieved from the studied immediate release dosage forms justified that GastroPlus™ is worthful in silico method for establishing in vitro–in vivo correlations and also for developing well-suited formulations for BCS class II drugs.

7.5.2.5 Biowaiver Consideration

Regulatory requirements of conducting the in vivo bioequivalence studies are very much challenging during the development of novel generic medicines. The process of approval and registration of a drug product may become swifter if those in vivo studies could be bypassed or waived off known as biowaiver. In this case, bioavailability of drugs can be estimated by assessment of solubility, permeability, and dissolution parameters instead of in vivo studies. On the basis of BCS, the FDA published a guide for waiving BA and BE studies for the immediate release solid oral dosage forms [52]. According to the guide, waiving these studies is recommended for those drugs which particularly belong to BCS class I, that is, having high solubility as well as high permeability, extending to class II and III drugs under certain situations [51]. Reduced time and cost involved in the development of generic formulations is the major advantage of implementing the biowaiver concept. Several publications are supported by the GastroPlus™ integrated ACAT model for implementing biowaiver.

A study by Okumu and co-workers in 2009 was carried out in which computer simulations were done for justifying biowaiver for etoricoxib solid oral dosage form [53].

Similarly, the relevance of GI simulations for extending biowaivers for the drug having high permeability was demonstrated by Grozdanis et al. [54]. In this study, GI simulation technology was applied and integrated with the physiological properties of the drug with the major aim of predicting the BCS class of the drug. Simulations of absorption profile of several BCS class II drugs which are either weakly acidic or weakly basic were carried out using GastroPlus® employing predetermined physicochemical and PK properties of drugs.

7.6 Conclusion

In summary, here we reviewed various mechanistic models for computer-aided biopharmaceutical characterization of drugs. Each discussed model has certain pros and cons, a combination of in vitro, in vivo, and in silico studies are best suited for accurate predictions. Biopharmaceutical absorption modeling has a wide array of applications and prospective which can rationalize the drug development process with improved regulatory compliance. Thus, for predicting the in vivo performance of drugs, physiological-based pharmacokinetic models are turned up as a much accepted and trustworthy approach. ACAT integrated GastroPlus™ software is the most widely employed tool used by scientists in the last two decades. The strategy, applications, and auxiliary features of GastroPlus™ software have been discussed in detail.

7.7 Credible Online Resources for Further Reading

URL	Category of source	What to read/refer to?
https://www.simulations-plus.com	Commercial software package	Learn mechanistically based simulation Learn to create QSPR/QSAR models for predicting ADMET properties Available learning material at resource center: Blog, posters, webinars, journal articles, etc.
https://www.certara.com/software/simcyp-pbpk/	Commercial software	Development of drugs via modeling and simulation of PK and PD in virtual populations
https://www.iconplc.com/innovation/nonmem/	Commercial software	Press releases, awards, webinars
www.admemodel.com	**PK/ADME simulation modeling software**	Learn the basics of the absorption model in iDEA pkEXPRESS™ which predicts human intestinal absorption and metabolism
https://in.mathworks.com/products/simbiology/pharmacokinetics-software/	Commercial software	Trial software can be downloaded Learn the basics for MATLAB and Simulink using MATLAB tutorials
https://www.iseesystems.com/	Commercial software	Free 30-Day Stella Architect Trial Free Online Webinars, Tutorials, Tips, and Tricks Product News and Updates
https://www.certara.com/software/phoenix-winnonlin/	Commercial software	Resources—Brochure Flexible and powerful compartmental modeling tools

References

1. Dokoumetzidis A, Kalantzi L, Fotaki N (2007) Predictive models for oral drug absorption: from in silico methods to integrated dynamical models. Expert Opin Drug Metab Toxicol 3:491–505. https://doi.org/10.1517/17425225.3.4.491
2. Thomas S (2010) Physiologically-based pharmacokinetic modelling for the reduction of animal use in the discovery of novel pharmaceuticals. Altern Lab Anim 38:81–85. https://doi.org/10.1177/026119291003801S16
3. Agoram B, Woltosz WS, Bolger MB (2001) Predicting the impact of physiological and biochemical processes on oral drug bioavailability. Adv Drug Deliv Rev 50:S41–S67. https://doi.org/10.1016/S0169-409X(01)00179-X
4. Ayad MH (2015) Rational formulation strategy from drug discovery profiling to human proof of concept. Drug Deliv 22:877–884. https://doi.org/10.3109/10717544.2014.898714
5. Kuentz M, Nick S, Parrott N, Röthlisberger D (2006) A strategy for preclinical formulation development using GastroPlus™ as pharmacokinetic simulation tool and a statistical screening design applied to a dog study. Eur J Pharm Sci 27:91–99. https://doi.org/10.1016/j.ejps.2005.08.011
6. Lin L, Wong H (2017) Predicting oral drug absorption: mini review on physiologically-based pharmacokinetic models. Pharmaceutics 9:41. https://doi.org/10.3390/pharmaceutics9040041
7. Daga PR, Bolger MB, Haworth IS, Clark RD, Martin EJ (2018) Physiologically based pharmacokinetic modeling in lead optimization. 1. Evaluation and adaptation of GastroPlus to predict bioavailability of Medchem series. Mol Pharm 15:821–830. https://doi.org/10.1021/acs.molpharmaceut.7b00972
8. Grbic S, Parojcic J, Djuric Z (2013) Computer-aided biopharmaceutical characterization: gastrointestinal absorption simulation. In: Computer-aided applications in pharmaceutical technology. Woodhead Publishing Limited. https://doi.org/10.1533/9781908818324.177
9. Huang W, Lee SL, Yu LX (2009) Mechanistic approaches to predicting oral drug absorption. AAPS J 11:217–224. https://doi.org/10.1208/s12248-009-9098-z
10. Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM et al (2015) An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat Rev Drug Discov 14:475–486. https://doi.org/10.1038/nrd4609
11. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25. https://doi.org/10.1016/S0169-409X(96)00423-1
12. Amidon GL, Lennernäs H, Shah VP, Crison JR (1995) A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. Pharm Res 12:413–420. https://doi.org/10.1023/A:1016212804288
13. Oh DM, Curl RL, Amidon GL (1993) Estimating the fraction dose absorbed from suspensions of poorly soluble compounds in humans: a mathematical model. Pharm Res 10:264–270. https://doi.org/10.1023/A:1018947113238
14. Wu C-Y, Benet LZ (2005) Predicting drug disposition via application of BCS: transport/absorption/elimination interplay and development of a biopharmaceutics drug disposition classification system. Pharm Res 22:11–23. https://doi.org/10.1007/s11095-004-9004-4
15. Chu X, Bleasby K, Evers R (2013) Species differences in drug transporters and implications for translating preclinical findings to humans. Expert Opin Drug Metab Toxicol 9:237–252. https://doi.org/10.1517/17425255.2013.741589
16. Dressman JB, Amidon GL, Fleisher D (1985) Absorption potential: estimating the fraction absorbed for orally administered compounds. J Pharm Sci 74:588–589. https://doi.org/10.1002/jps.2600740523
17. Johnson KC, Swindell AC (1996) Guidance in the setting of drug particle size specifications to minimize variability in absorption. Pharm Res 13:1795–1798. https://doi.org/10.1023/A:1016068705255

18. Ding X, Rose JP, Van Gelder J (2012) Developability assessment of clinical drug products with maximum absorbable doses. Int J Pharm 427:260–269. https://doi.org/10.1016/j.ijpharm.2012.02.003
19. Sun D, Yu LX, Hussain MA, Wall DA, Smith RL, Amidon GL (2004) In vitro testing of drug absorption for drug "developability" assessment: forming an interface between in vitro preclinical data and clinical outcome. Curr Opin Drug Discov Devel 7:75–85
20. Kostewicz ES, Aarons L, Bergstrand M, Bolger MB, Galetin A, Hatley O et al (2014) PBPK models for the prediction of in vivo performance of oral dosage forms. Eur J Pharm Sci 57:300–321. https://doi.org/10.1016/j.ejps.2013.09.008. Elsevier B.V.
21. Sugano K (2009) Introduction to computational oral absorption simulation. Expert Opin Drug Metab Toxicol 5:259–293. https://doi.org/10.1517/17425250902835506
22. Sjögren E, Thörn H, Tannergren C (2016) In silico modeling of gastrointestinal drug absorption: predictive performance of three physiologically based absorption models. Mol Pharm 13:1763–1778. https://doi.org/10.1021/acs.molpharmaceut.5b00861
23. Yu LX, Lipka E, Crison JR, Amidon GL (1996) Transport approaches to the biopharmaceutical design of oral drug delivery systems: prediction of intestinal absorption. Adv Drug Deliv Rev 19:359–376. https://doi.org/10.1016/0169-409X(96)00009-9
24. Yu LX, Crison JR, Amidon GL (1996) Compartmental transit and dispersion model analysis of small intestinal transit flow in humans. Int J Pharm 140:111–118. https://doi.org/10.1016/0378-5173(96)04592-9
25. Oberle RL, Amidon GL (1987) The influence of variable gastric emptying and intestinal transit rates on the plasma level curve of cimetidine; an explanation for the double peak phenomenon. J Pharmacokinet Biopharm 15:529–544. https://doi.org/10.1007/BF01061761
26. Dressman JB, Fleisher D (1986) Mixing-tank model for predicting dissolution rate control of oral absorption. J Pharm Sci 75:109–116. https://doi.org/10.1002/jps.2600750202
27. Yu LX, Amidon GL (1998) Saturable small intestinal drug absorption in humans: modeling and interpretation of cefatrizine data. Eur J Pharm Biopharm 45:199–203. https://doi.org/10.1016/S0939-6411(97)00088-X
28. Grass GM (1997) Simulation models to predict oral drug absorption from in vitro data. Adv Drug Deliv Rev 23:199–219. https://doi.org/10.1016/S0169-409X(96)00436-X
29. Abuhelwa AY, Williams DB, Upton RN, Foster DJR (2017) Food, gastrointestinal pH, and models of oral drug absorption. Eur J Pharm Biopharm 112:234–248. https://doi.org/10.1016/j.ejpb.2016.11.034. Elsevier B.V.
30. Sawamoto T, Haruta S, Kurosaki Y, Higaki K, Kimura T (1997) Prediction of the plasma concentration profiles of orally administered drugs in rats on the basis of gastrointestinal transit kinetics and absorbability. J Pharm Pharmacol 49:450–457. https://doi.org/10.1111/j.2042-7158.1997.tb06823.x
31. Kimura T, Higaki K (2002) Gastrointestinal transit and drug absorption. Biol Pharm Bull 25:149–164. https://doi.org/10.1248/bpb.25.149
32. Yokoe J, Iwasaki N, Haruta S, Kadono K, Ogawara K, Higaki K et al (2003) Analysis and prediction of absorption behavior of colon-targeted prodrug in rats by GI-transit-absorption model. J Control Release 86:305–313. https://doi.org/10.1016/S0168-3659(02)00424-8
33. Jamei M, Turner D, Yang J, Neuhoff S, Polak S, Rostami-Hodjegan A et al (2009) Population-based mechanistic prediction of oral drug absorption. AAPS J 11:225–237. https://doi.org/10.1208/s12248-009-9099-y
34. Cheng L, Wong H (2020) Food effects on oral drug absorption: application of physiologically-based pharmacokinetic modeling as a predictive tool. Pharmaceutics 12:672. https://doi.org/10.3390/pharmaceutics12070672
35. Jamei M, Marciniak S, Feng K, Barnett A, Tucker G, Rostami-Hodjegan A (2009) The Simcyp ® population-based ADME simulator. Expert Opin Drug Metab Toxicol 5:211–223. https://doi.org/10.1517/17425250802691074
36. Zhang T, Wells E (2020) A review of current methods for food effect prediction during drug development. Curr Pharmacol Rep 6:267–279. https://doi.org/10.1007/s40495-020-00230-9

37. Yu LX, Amidon GL (1999) A compartmental absorption and transit model for estimating oral drug absorption. Int J Pharm 186:119–125. https://doi.org/10.1016/S0378-5173(99)00147-7
38. Xia B, Yang Z, Zhou H, Lukacova V, Zhu W, Milewski M et al (2015) Development of a novel oral cavity compartmental absorption and transit model for sublingual administration: illustration with zolpidem. AAPS J 17:631–642. https://doi.org/10.1208/s12248-015-9727-7
39. Le Merdy M, Spires J, Lukacova V, Tan M-L, Babiskin A, Xu X et al (2020) Ocular physiologically based pharmacokinetic modeling for ointment formulations. Pharm Res 37:245. https://doi.org/10.1007/s11095-020-02965-y
40. Miller NA, Reddy MB, Heikkinen AT, Lukacova V, Parrott N (2019) Physiologically based pharmacokinetic modelling for first-in-human predictions: an updated model building strategy illustrated with challenging industry case studies. Clin Pharmacokinet 58:727–746. https://doi.org/10.1007/s40262-019-00741-9
41. Ahmad A, Pepin X, Aarons L, Wang Y, Darwich AS, Wood JM et al (2020) IMI—oral biopharmaceutics tools project—evaluation of bottom-up PBPK prediction success part 4: prediction accuracy and software comparisons with improved data and modelling strategies. Eur J Pharm Biopharm 156:50–63. https://doi.org/10.1016/j.ejpb.2020.08.006
42. Jones HM, Parrott N, Ohlenbusch G, Lavé T (2006) Predicting pharmacokinetic food effects using biorelevant solubility media and physiologically based modelling. Clin Pharmacokinet 45:1213–1226. https://doi.org/10.2165/00003088-200645120-00006
43. Grbic S, Lukic V, Kovacevic I, Parojcic J, Djuric Z (2012) An investigation into the possibilities and limitations of *in silico* absorption modeling: GastroPlus™ simulation of nimesulide oral absorption. In: Proc 2nd Electron Conf Pharm Sci. MDPI, Basel, Switzerland, p 816. https://doi.org/10.3390/ecps2012-00816
44. Grbic S, Parojcic J, Ibric S, Djuric Z (2011) In vitro–in vivo correlation for gliclazide immediate-release tablets based on mechanistic absorption simulation. AAPS PharmSciTech 12:165–171. https://doi.org/10.1208/s12249-010-9573-y
45. Parrott N, Stillhart C, Lindenberg M, Wagner B, Kowalski K, Guerini E et al (2020) Physiologically based absorption modelling to explore the impact of food and gastric pH changes on the pharmacokinetics of entrectinib. AAPS J 22:78. https://doi.org/10.1208/s12248-020-00463-y
46. Zhang X, Lionberger RA, Davit BM, Yu LX (2011) Utility of physiologically based absorption modeling in implementing quality by design in drug development. AAPS J 13:59–71. https://doi.org/10.1208/s12248-010-9250-9
47. Tubic M, Wagner D, Spahn-Langguth H, Bolger MB, Langguth P (2006) In silico modeling of non-linear drug absorption for the P-gp substrate talinolol and of consequences for the resulting pharmacodynamic effect. Pharm Res 23:1712–1720. https://doi.org/10.1007/s11095-006-9020-7
48. Koziolek M, Alcaro S, Augustijns P, Basit AW, Grimm M, Hens B et al (2019) The mechanisms of pharmacokinetic food-drug interactions—a perspective from the UNGAP group. Eur J Pharm Sci 134:31–59. https://doi.org/10.1016/j.ejps.2019.04.003
49. Deng J, Zhu X, Chen Z, Fan CH, Kwan HS, Wong CH et al (2017) A review of food–drug interactions on oral drug absorption. Drugs 77:1833–1855. https://doi.org/10.1007/s40265-017-0832-z
50. Gu C-H, Li H, Levons J, Lentz K, Gandhi RB, Raghavan K et al (2007) Predicting effect of food on extent of drug absorption based on physicochemical properties. Pharm Res 24:1118–1130. https://doi.org/10.1007/s11095-007-9236-1

51. da Silva Honório T, Pinto EC, Rocha HVA, Esteves VSD, dos Santos TC, Castro HCR et al (2013) In vitro–in vivo correlation of efavirenz tablets using GastroPlus®. AAPS PharmSciTech 14:1244–1254. https://doi.org/10.1208/s12249-013-0016-4
52. CDER/FDA (2015) Guidance for Industry, Waiver of in vivo bioavailability and bioequivalence studies for immediate release solid oral dosage forms based on a biopharmaceutics classification system. Cent Drug Eval Res
53. Okumu A, DiMaso M, Löbenberg R (2009) Computer simulations using GastroPlus™ to justify a biowaiver for etoricoxib solid oral drug products. Eur J Pharm Biopharm 72:91–98. https://doi.org/10.1016/j.ejpb.2008.10.019. Elsevier B.V.
54. Tubic-Grozdanis M, Bolger MB, Langguth P (2008) Application of gastrointestinal simulation for extensions for biowaivers of highly permeable compounds. AAPS J 10:213–226. https://doi.org/10.1208/s12248-008-9023-x

Computer Simulation and Modeling in Pharmacokinetics and Pharmacodynamics

8

Ruchi Chawla, Varsha Rani, Mohini Mishra, and Krishan Kumar

Contents

R. Chawla (✉) · V. Rani · M. Mishra · K. Kumar
Department of Pharmaceutical Engineering and Technology, Indian Institute of Technology (Banaras Hindu University), Varanasi, Uttar Pradesh, India
e-mail: rchawla.phe@iitbhu.ac.in

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_8

Abstract

Computer-based modeling and simulation is emerging as a useful tool to complement the analysis and interpretation of biological data. The large volume, scale, and complexity of data generated from in vitro, in vivo, and ex vivo data cannot be analyzed and interpreted by conventional data analysis tools. So, various in silico computational e-resources, databases, and simulation softwares are being used for the determination of pharmacokinetic (PK) and pharmacodynamic (PD) parameters for the management of diseases. These tools help in providing multiscale representation of the biological processes in the order of increasing complexity from that of protein and genes, cells, isolated tissues and organs, and the whole organism. The United States Food and Drug Administration (USFDA) has also directed the use of PK/PD simulation for the evaluation of drugs during the clinical phase, in which the primary focus is on the establishment of relationship between therapeutic drug concentration and patient response. The aim of this chapter is to discuss the role, advancement, and development of biocomputational tools used in research and development in the pharmaceutical industry, wherein the number of experimental studies is exponentially growing. Furthermore, application of these studies to optimize the dosing regimens, dose-response relationship, etc. will be discussed.

Keywords

Biomolecular simulation · In silico modeling · Pharmacokinetic · Pharmacodynamic simulation

Chapter Objectives

The chapter titled "Computer Simulation and Modeling in Pharmacokinetics and Pharmacodynamics" serves the following objectives:

- Importance of computer simulation and modeling in the current scenario of drug discovery and development research
- Conceptualization of the principle of model development, validation, and simulation
- Understanding of the importance of integration of pharmacokinetic and pharmacodynamic simulation for the management of disease
- Application of multiscale modeling and simulation using various computational tools
- Scope of application of the computational tools in understanding the drug-drug interaction (DDI), drug-receptor interaction, and modification of dosing regimens for altering the dose-response relationship with respect to time

8.1 Introduction

In the current era, computers and computational tools have become an integral part of our research and provide a dynamic platform for application in any field of science [1]. Especially, in the biological and medical fields, this science has shown unparalleled results allowing the researchers to simulate in vivo processes for predictive modeling [2]. Computational science also has specific applications in pharmaceutical science and technology, wherein computer simulation offers a valuable approach for the determination of pharmacokinetic (PK) and pharmacodynamic (PD) parameters, prediction of disease progression, therapeutic drug monitoring, and response to therapies [3]. Furthermore, it facilitates the process of drug discovery and development and has shown progressive results by abridging the time required for lead identification and optimization. Use of computational modeling and simulation in the biomedical field has comprehensively played a significant role from the development of simple medical devices, biomaterials, diagnostic equipments to artificial intelligence-based platforms. The interdisciplinary applications of this technology are guided by experts from fields of pharmaceutical sciences, biological sciences, computer engineering, statistics, bioengineering, etc. Furthermore, incorporation of interdisciplinary concepts adds to the complexity of understanding, and integration of the same in the virtual platform, thus making it a time-taking process [4–6]. The steps between administration of drug to the initiation of body response can be divided as: pharmacokinetic and pharmacodynamic [7]. Pharmaco originates from the Greek term Pharmakon for "drug," and kinetics originates from the Greek term kinetikos for "moving." Pharmacokinetics signifies the changes in drug concentration within body with time after administration of drug, that is, it is concerned with drug absorption (A), distribution (D), metabolism (M), and elimination (E) (ADME), as shown in (Fig. 8.1) [8]. Thus, the biodistribution pattern of a drug inside the body is studied in pharmacokinetics, wherein the distribution is compartmentalized and in order to produce a therapeutic response, a sufficiently safe concentration of the drug should be available at the site of action. Though the measurement of plasma drug concentration at the site of action is important, it is not routinely feasible. The relationship between the plasma concentration and the drug concentration at the site of action is generally linear. Any change in plasma concentration proportionally alters the concentration of drug at the site of action [9].

Pharmaco-originates from the Greek term, that is, Pharmakon for "drug" and dynamics means "of or relating to variation of intensity." The response of body to the drug is analyzed in pharmacodynamics. Particularly, PD prognostically determines the onset, duration, and intensity of drug action in relation to the concentration of drug. The interaction of drug with special binding site in the tissue, that is, receptor ensues drug response. The interaction of the drug with the receptor leads to a conformational modification in the receptor, generating the stimulus that causes a physiological and biochemical response.

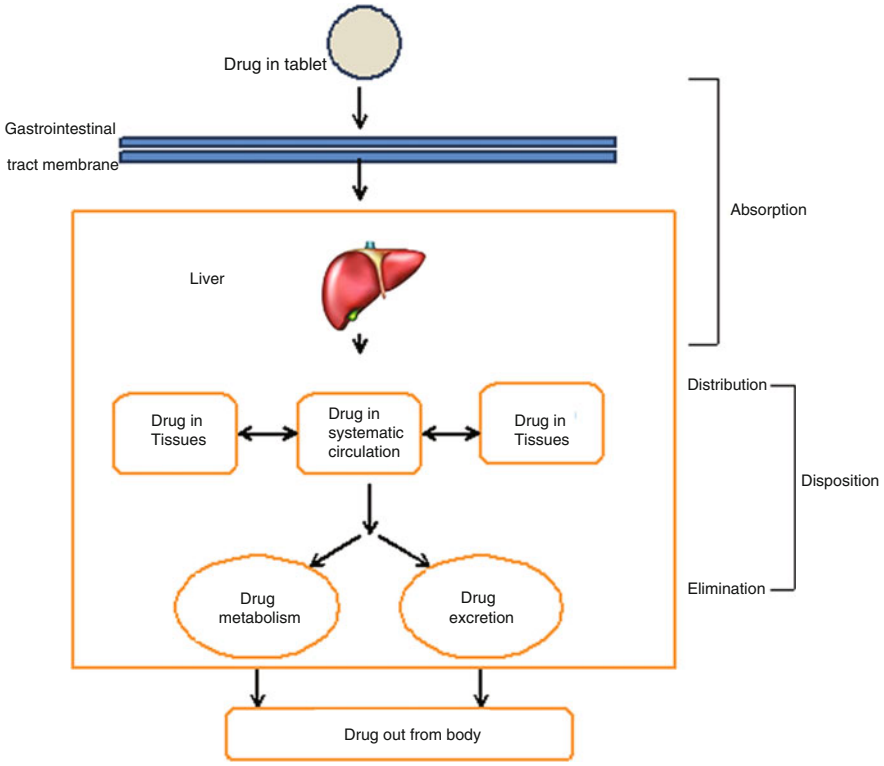


Fig. 8.1 Diagrammatic representation of the fate of drug (ADME) in the body

More than 95% of receptors are proteins; however, other receptor types also exist like the DNA receptors, which are used in chemotherapy. The binding of drug to the receptor is mostly reversible and results in the establishment of an equilibrium between the bound and unbound drug; and when the elimination of drug starts, the response subsides and finally disappears due to the detachment of drug from the receptor. On the other hand, some drugs are attached via irreversible covalent bonds with their receptors, for example, aspirin inhibits the aggregation of platelets by suppressing the generation of thromboxane in platelets. The drug-receptor binding is extremely reliant on the chemical structures of the receptor and the drug. All active endogenous moieties, that is, natural ligands such as hormones and neurotransmitters regulate biochemical and physiological processes in the body by binding to their receptors [10–12]. Optimal dosing regimens can be designed based on the understanding of the correlation of pharmacokinetic and pharmacodynamic phases. This correlation can then be progressively transformed into one or more mathematical models depicting the relation of dose with the mode of drug response [7].

Simulation and modeling are based on experimental datasets generated through previous studies. Higher is the accuracy and precision of the experimental data,

better is the predicted output. Similarly, in the case of pharmacokinetic and pharmacodynamic modeling of a selected drug, data from previous studies are taken as reference to develop the model. The developed model can be used to predict atomic details and simulate mechanisms which are however not possible experimentally or through manual methods [13–16]. Simulation and modeling are being made an integral part of the drug development and assessment processes to evaluate the safety and potency of drugs and drug products before being sent to the market. The safety of patients is of prime concern especially in the case of drugs, so scrutiny and rigorous testing of products before commercialization are highly essential. Some substances like pesticides, insecticides, disinfectants, etc. fall into the category of potentially harmful substances, and preclinical and clinical studies cannot be conducted for them; herein, computer-based studies can be used as an alternative to obtain approximate results. The computer simulation models are thus being developed to cater to these issues, however, still a lot needs to be done in this area [17, 18].

8.2 Computational Modeling and Simulation

Models are used to predict a correlation between the independent factors (inputs) and the dependent variable (output). Linear models describe the interaction with only one factor at a time called as the OVAT (one variable at a time) technique. But, generally, there is an interaction between more than one factor and requires the development of complex mathematical models. These multilevel correlation models induce confidence for optimal parameter determination through the use of appropriate mathematical equations and graphical models. Once a model is finalized, then simulation can be performed to make predictions about the system [5]. Simulation is the use of a model to testify assumptions that can provide an approximate representation of some aspect of the real world. It is a tool to practically investigate the system's transformation under observation. Computer simulation is the field of designing an actual or theoretical model of a physical system, implementing the digital computer model, and evaluating the performance of different strategies (within the limits of set criteria) [19]. Modeling and simulation have distinct points of view. Modeling is based on previously generated data and simulation progressively predicts data by using the model. Simulation can be categorized as: deterministic, Monte Carlo, quantum mechanical, and molecular dynamics (MD). In a deterministic simulation, the data variability is ignored and these simulations are based on the mean values and long-term results. Nomenclature of Monte Carlo (MC) simulation originates from the gambling capital of Monaco and takes into consideration stochastic variability in the data, that is, both random and systematic variability which further encompasses long-term expectation and expected variability around the expectation. The deterministic simulation is easy to perform as compared to Monte Carlo simulation, but the Monte Carlo expectation helps in a better understanding of the resulting data and variability in the result. Simulations of molecular dynamics using computational technology are used to analyze the

physical movements of atoms and molecules and to assess the system's macroscopic thermodynamic properties [5, 20, 21]. Quantum simulation, by quantum mechanical means, can be loosely described as simulating a quantum system. Quantum mechanical simulation is a helpful strategy for unraveling the reaction process. Quantum mechanical simulation also operates with the classical molecular dynamic simulation throughout the field of drug design involving biological macromolecules. In conjunction with classical molecular dynamics, quantum mechanics is also beneficial for evaluating the reliability of interaction energy and sampling of conformation space [22]. Simulation of direct molecular dynamics (MD) using quantum mechanical and molecular mechanical (QM/MM) techniques is very important for understanding the process of chemical reactions in a dynamic situation. It is possible to substantially reduce the calculation cost of QM/MM calculations during MD simulations. In addition to measuring thermodynamic properties, MD simulations have a great potential for characterizing reaction dynamics, which offer a valuable tool for studying chemical or biochemical processes in solutions or enzymes [23]. Modeling and simulation thus help to address a number of issues related to clinical testing of drug molecules like anticipated human PK profile, PK/PD profile, optimal dosing regimen, effect of food, variability in data due to other reasons, etc. [10].

Using computational modeling and simulation, pharmacokinetic and pharmacodynamic-based drug design can be used for the prediction of drug behavior and its properties. This involves the interpretation of data related to disease and its mechanism, characteristics of drug, and interrelationship between each of these parameters using mathematical equations based on well-defined assumptions [24].

Development of a model involves sequential steps, as described below:

- Generation of experimental data and use of these data in the development of a model, which is accurate enough to explain the data obtained
- Translation of structured model into mathematical equations with the help of a suitable modeling software
- Identification of theoretical indicative parameters from the observed data
- Validation of structured model by comparing the data predicted by the model (experimental data) against the observed data

If the experimental data as predicted by proposed model are not in accordance with the observed data, modifications in the model are made till the model is successfully validated. The model can be cross-validated by evaluating the predicted data obtained from simulation by comparing those data with data (validation set) which were not used in the development of the model. Once the model is validated, it can be used for simulation and prediction without real-time experimentation [24].

Furthermore, the huge and sophisticated datasets (or the "Big Data") produced from various clinical studies, chemical libraries, screening studies, biological activities, in vitro and in vivo assays, and various other sources require screening, interpretation, and correlation. Besides this, volume and velocity of generation, diversity and veracity of data put forth challenges and opportunities for the research

Table 8.1 List of tools and databases used in computational modeling

S. no.	Tools	URL address	Category
1.	PubChem	https://pubchem.ncbi.nlm.nih.gov	Ligand databases
2.	ChEMBL	https://www.ebi.ac.uk/chembl/	
3.	DrugBank	https://www.drugbank.ca	
4.	DrugMatrix	https://ntp.niehs.nih.gov/results/drugmatrix/index.html	
5.	Binding database	https://www.bindingdb.org/bind/index.jsp	
6.	Protein information resource (PIR)	https://proteininformationresource.org/	Protein databases
7.	SWISS-PROT	https://www.uniprot.org/statistics/Swiss-Prot	
8.	TrEMBL	https://www.ebi.ac.uk/training/online/glossary/uniprotkbtrembl	
9.	GenBank	https://www.ncbi.nlm.nih.gov/genbank/	
10.	RefSeq	https://www.ncbi.nlm.nih.gov/refseq/	
11.	Protein Data Bank (PDB)	https://www.rcsb.org/	
12.	Simcyp™	http://www.simcyp.com	PBPK model-based software
13.	GastroPlus™	http://www.simulations-plus.com	
14.	PK-Sim™	http://www.pksim.com/	
15.	MEDICI-PK™	http://www.cit-wulkow.de/	
16.	Cloe PK™	http://www.cyprotex.com	

community to develop novel computational techniques that would facilitate mining, validation, retrieval, and organization of data [25].

PubChem (public repository for chemical structures and their biological properties), ChEMBL (database containing binding, functional, ADME, and toxicity data for numerous compounds), DrugBank (database containing all approved drugs with their mechanisms, interactions, and relevant targets), DrugMatrix (contains large-scale gene expression data from tissues of rats administered over 600 drugs, mostly targeting several major organs), Binding Database (resource of drug-target binding data), Protein information resource (PIR), SWISS-PROT, TrEMBL, GenBank, GenPept, RefSeq, and Protein Data Bank (PDB) are some of the open-source platforms for data mining and application that are mentioned in Table 8.1. Fostering of big data and their handling followed by coupling with biological systems and analytical tools help in the development of flexible simulation models [26].

8.2.1 Methods of Biomolecular Simulation

Identification of drug binding sites and explication of mechanism of drug action are possible by the biomolecular simulation of target proteins at different levels. Furthermore, by virtual screening of chemical libraries, potential drug candidates can be identified and their efficacy can be evaluated by in vitro experiments. De novo drug

design and artificial intelligence are being used as sophisticated methods in this area. Chemical reactions (like enzyme catalysis) and the calculation of spectra and electronic properties can be done by a combination of *quantum mechanics* and *molecular mechanics*. These studies pave the way for the elucidation of the mechanism of drug action. Another biomolecular method used in simulation studies is *molecular dynamics,* which can be used in the identification of a drug-binding site on the target protein and the calculation of the binding free energy between the target protein and drug molecule [27, 28].

With the knowledge of the structure of the binding site (binding pocket) of the target protein, it is easier and justifiable either to search (from the chemical database) or to develop an optimized drug candidate that fits into the binding pocket of the receptor site with a high binding affinity [29–31].

Virtual screening is the technique of screening big databases of chemicals for the identification of lead compounds. *Molecular docking* or "induced fit" model of virtual screening is based on the matching of spatial shape and energy of small molecules and proteins [32]. Molecular docking evaluates the interaction between small molecules and proteins as well as between different proteins; and facilitates the prediction of binding potential between them. The determination of correct spatial conformation of ligand binding with the protein receptor helps to study the mechanism of action of lead compounds [31, 33]. In the year 1909, Ehrlich gave the idea of pharmacophores [34], which act as descriptors present in the ligand (with active molecular/structural features) and provide the structural framework for the binding of ligand with target protein receptors. The interaction of pharmacophore with target protein results in a geometrically and energetically matched conformation of the latter [35, 36]. *Quantitative Structure Activity Relationship (QSAR)* helps in correlating the estimated structural properties of a drug molecule like log P, electrostatic constant, acid dissociation constant, pKa, etc. with its biological activity [32].

8.3 Multiscale Modeling and Simulation Studies

Multiscale modeling and simulation is a predictable substitute model that ensures the accuracy of input and output at various levels of modeling and simulation. This approach helps bridge the responses received at different stages of drug discovery and development, starting from in silico up to ADME predictions. Multiscale models simulate the physiological processes that span across various lengths and timescales. The multiscale models are based on empirical data, mathematical approximations, and justified assumptions. Pharmacometrics widely employs multiscale modeling, in which biology, pharmacology, disease, and physiology-based models simulate interactions between drugs and patients, along with an indication of positive and negative effects resulting from these interactions [37].

Model equations are needed for reliable estimation through unknown pre-exponential coefficients or kinetic rate of existing catalytic systems. Pharmacokinetic/pharmacodynamic modeling is the mathematical model that predicts various essential parameters based on a time-dependent infrequent sampling of the drug in

various body compartments. Besides, it also highlights in vitro and in vivo conditions based on experimentation and animal/human studies. Furthermore, drug-binding kinetic modeling aims to develop efficacious drugs with minimal failed clinical trials [38]. The emerging computational and mathematical models have been strategically classified into three levels: molecular-level modeling, cellular-level modeling, and organ-system-level modeling. Molecular-level modeling includes RNA, DNA, proteins, and metabolites. In contrast, cellular- and organ-level modeling consist of genetic and cellular pathways and the pharmacokinetic and pharmacodynamic properties of modeled drug candidates, respectively [39]. Multiscale modeling integrates the inputs from molecular-, cellular-, and organ-levels to form the best-fit model explaining the drug's pharmacological and toxicological effects. Hypothetical elucidation of biophysical and biochemical mechanisms at the cellular-, subcellular-, and tissue-levels is intricate, but these phenomena can be thoroughly studied using computational tools.

8.3.1 Pharmacokinetic and Pharmacodynamic Modeling

The time-course of drug concentration is described by PK models. In PK models, the most widely used method is to represent the body as a compartment framework. The compartment is a collection of tissues or organs in which the drug is well mixed and kinetically homogeneous and acts as a building block for several PK models. The compartments are conceptual and have no physiological or anatomical segregation [8]. Physiologically based pharmacokinetic (PBPK) models typically consist of a greater number of compartments compared to other models, representing various organs or tissues in the body linked by blood flow. Each compartment is characterized by the volume of tissue (or weight) and the rate of tissue blood flow that varies with the tissue as well as the organisms [9]. The association between drug concentration and pharmacological effect is described in pharmacodynamics (PD). The integration of time-course of drug concentration (PK) and concentration-effect relationship (PD) results in the generation of a PK/PD model and helps in the prediction of the relationship of dose and response with respect to time [40].

PK/PD modeling is especially used during drug development studies for the determination of safety and efficacy of drugs. In this computational approach, absorption, distribution, metabolism, and excretion parameters are modeled in silico. Anatomical and physiological parameters are simulated by considering the rate of blood flow, tissue/organ volume, and activity of drug transporters/disposition enzymes present in human body. For PK/PD modeling, two types of input parameters are required: physiological parameters (viz. organ volume, volume and rate of blood flow, etc.) and drug-specific parameters (e.g., molecular weight, solubility, particle size, pKa, log *P*, permeability, metabolic stability, and plasma protein binding). Scientists use an integrated model approach to combine different PK/PD parameters in one integrated system, as shown in Fig. 8.2. By using an integrated model, the data can be linked together and a logical correlation between

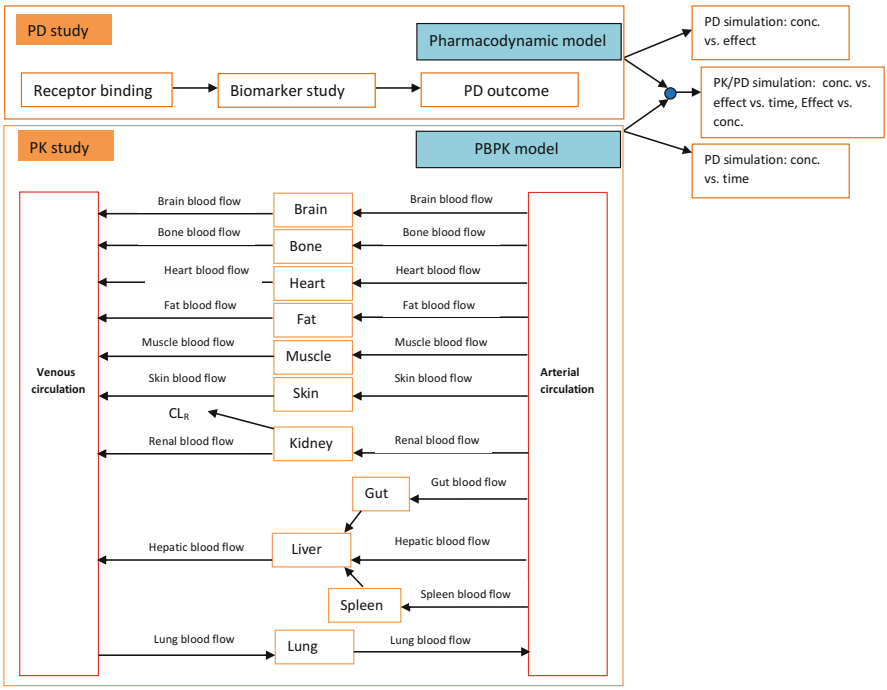


Fig. 8.2 Diagrammatic representation of modeling and simulation for integration of different experiments [41]

each of the experiments can be established in most of the cases. Such models also lead to the discovery of novel pathways or new hypothesis [41].

8.3.2 Physiologically Based Pharmacokinetic (PBPK) Model

The physiologically based pharmacokinetic model in combination with pharmacokinetic-pharmacodynamic (PK/PD) model can be used to predict the pharmacokinetic profile of a drug along with the effect profile and dose of drug to obtain the desired response. The concentration predicted using PBPK modeling can be used as input for PK/PD studies and further, the safety and efficacy of drug can be validated using preclinical studies [42]. PBPK models are represented by distinct compartments based on anatomy and physiology like gastrointestinal tract (GIT), kidney and liver (organs involved in elimination), and other tissues such as muscles, brain, fat, etc. which are connected by the circulatory system. Species-specific physiological parameters like pH, blood perfusion rates, permeability rate, presence of enzymes, organ volume, etc. can be integrated with physicochemical properties of the drug to predict the PK and PD of the drug.

Various softwares have been developed based on the PBPK model for simulation and modeling of PK/PD parameters like:

- Simcyp® developed by Simcyp is a population-based simulator capable of PBPK modeling and simulates in vitro data for pharmacokinetic studies. Thai et al. carried out a bridging study where the research team used PBPK modeling to optimize the dose and sampling times in pediatric oncology. Simcyp® was used to develop a PBPK model for docetaxel in adult cancer patients, which was then used to predict the PK profile of docetaxel in pediatric patients (age group of 0–18 years) taking into consideration age-dependent physiology differences [43].
- GastroPlus™ developed by Simulations Plus, Inc. is a mechanistically based simulation program used for the prediction of pharmacokinetics, biopharmaceutics, and pharmacodynamics in humans and preclinical species. It is basically used to predict intestinal absorption.
- PK-Sim™ software developed by Bayer Technology Services is used for whole-body PBPK modeling in humans and commonly used preclinical species like mouse, rat, minipig, dog, and monkey. It is one of the most functional softwares, as it fully integrates the four basic processes of pharmacokinetics, that is, absorption, distribution, metabolism, and excretion, into one model and hence is of greater utility in terms of practical use [44].
- MEDICI-PK™ developed by Computing in Technology is a virtual laboratory for studying the effect of drug/metabolites on and possible interactions in humans and preclinical species.
- Cloe PK™ software developed by Cyprotex is a multispecies PK simulation platform based on the integration of ADME and physicochemical data used during early stages of drug discovery.

(GastroPlus™ and Simcyp® are single process-based simulation softwares, the former being used for intestinal absorption and the latter for studying metabolism and interactions. However, Cleo PK™ and PK-Sim™ are whole-body simulation softwares) [44].

The integration of pharmacokinetic with pharmacodynamics helps to predict maximum (C_{max}) and minimum concentrations (C_{min}), half-life ($T_{1/2}$), clearance (Cl), and other pharmacokinetic variables. The PBPK modeling along with in vitro-in vivo extrapolation (IVIVE) technique can also be used to predict plasma/tissue drug-concentration-time profiles in diverse species, populations, and diseased conditions. The limitations of PBPK models are the mathematical complexity and requirement of a large number of input parameters. However, with growing attention of regulatory authorities toward the applicability of PBPK modeling, user-friendly computational platforms are being designed for simulation and modeling, especially for use during the early stages of drug discovery [45]. The limitation of incorporating physiological and biological parameters in the model like the presence of a certain concentration of enzymes/transporters, diseased state, variation due to ethnicity, etc. puts constraint in mimicking the in vivo conditions. Drug discovery, drug development, clinical trials, selection of drug candidate, and assessment of human dose, drug interactions profile, etc. can be explored by this model.

8.3.3 Pharmacodynamic Simulation

The understanding of ligand-receptor interaction and the associated kinetics require sophisticated statistical and computational analysis. A generic coarse-grained Springs, Sites, and Langevin Dynamics (SpringSaLaD) model has been developed to capture ligand-receptor kinetics by estimating diffusion dependence affinity and dissociation rates to study chimeric compounds and receptor phenotype overexpression to selectively target unhealthy cells. It effectively models the proteins; binding events and state changes among multivalent molecules; and ligand-receptor interactions. SpringSaLaD has a user-friendly graphical interface that simulates the binding reactions based on diffusion coefficients of the sites, the site radii, and rate kinetics. BINANA (BINdingANAlyzer), a novel Python-implemented computer algorithm, is being used for characterizing the hydrophobic sites that facilitate the binding of ligand to its receptor (protein). BINANA can be used to search through a large ligand-receptor database to identify complexes with superior binding features and identify lead candidates from a virtual screen with specific, desirable binding characteristics [46].

In a typical coarse-grained model, complex C formed as a result of chemical interaction between a freely diffusing ligand L and its corresponding receptor R present on the cell membrane can be represented as:

$$L + R \underset{K_{\text{off}}}{\overset{K_{\text{on}}}{\rightleftarrows}} C \tag{8.1}$$

where the rates of association and dissociation are represented by K_{on} and K_{off}, respectively. A simple ordinary differential equation describing ligand-receptor dynamics based on mass-action kinetics can be represented as:

$$dC/dt = K_{\text{on}}[L] - K_{\text{off}}C \tag{8.2}$$

Coarse-grained model can also be used to study the dynamics of chimeric ligands that profoundly influence the efficiency and selectivity of drug-based therapies. The chimeras basically consist of two subunits: a targeting element (*TE*) that binds to a receptor and an activity element (*AE*) that interacts with a specific receptor. *TE* is connected to *AE* via a polymeric linker and the binding of *TE* to the receptor facilitates the binding of AE to its complementary site on the cell surface. The force of interaction based on molecular dynamics simulations can be described by the following equation:

$$F(r) = \{[K_{\text{B}}T/2lp][2r/r_{\max}] + 1/2\{r_{\max}/[r_{\max} - r]\} - ½\} \tag{8.3}$$

$$r_{\max} = Nl\cos(\theta/2) \tag{8.4}$$

where r represents the distance between *AE* and *TE*, and $r_{\max}$ is the size of the completely stretched polymer chain. N is the number of monomers in the chain, l is the size of each monomer, and angles between the monomers are denoted by θ. The persistence length is given by $lp = 2l/\theta^2$. *AER* and *TER* are, respectively, the active

element receptor and target element receptors that interact with *AE* and *TE* chimera subunits.

$$< r^2 >= 2lpr_{\max} - 2lp^2[1 - \exp(-r_{\max}/lp)] \tag{8.5}$$

To study the dependence of the chimeric efficiency on the number of *TER*s on the cell surface, we calculate the number of *AE-AER* complexes in chimeric versus monomer configuration. The specificity of chimeric ligands toward cells overexpressing *TER*s is enhanced when the affinity of the *AE*s toward their receptors is reduced. The number of *AE-AER* complexes formed is equivalent to a nonchimeric ligand in the absence of *TER*. Increasing the geometric factor γ further reduces the affinity rate of the interaction. Accordingly, raising the value of γ_{AE} enhances the specificity of chimeric ligands for high and low interaction strengths. After performing simulations, the effective affinity rate of *AEs* for different linker lengths and different average distances between *TERs* and *AERs* can be computed. The average distance (d) between receptors for a distributed number of *TER*s on surface is calculated as follows: A is an average area with an associated radius ρTER:

$$d = 1/A_{\mathrm{TER}} \int_0^{2\pi} \int_0^{\rho \mathrm{TER}} \rho^2 d\rho d\theta = 2/3\sqrt{A/[\pi.\mathrm{TER}_0]} \tag{8.6}$$

Thus, the coarse-grained model helps to understand complex ligand-receptor systems, multi-interactions between species with unique geometries, receptor dimerization, and multivalent binding and receptor cross-linking [47].

8.3.3.1 Quantitative Structure-Pharmacokinetic Parameters (QSPkRs)

Quantitative structure-pharmacokinetic relationship (QSPkR) modeling is a reliable approach for high-throughput prediction of pharmacokinetic parameters of new drug candidates during the drug development process. This modeling system correlates with the principles of the quantitative structure-activity relationship (QSAR) which describe the dependence of the structure of the compound on the biological or physicochemical activity of the compound. However, due to complex and partially understood knowledge of the underlying PK mechanisms, and the lack of standardized procedures and acceptance criteria, high-quality QSPkRs still need to be developed. In a study, QSPkR was performed for the determination of the volume of distribution at steady state (*V*ss) and clearance (CL) of 44 antimicrobial agents in humans using the k-nearest-neighbor (k-NN) and partial least-square (PLS) methods with cross-validation of the coefficient of determination $q(2)$ for the training set and external predictive $r(2)$ for the test set. The best simulated annealing (SA)-kNN model predicted values with high accuracy and can thus be used in the identification of compounds with excellent PK properties, along with the selection of a starting dose for Phase I clinical trials [48]. Another experiment was conducted for an early prediction of plasma protein binding (PPB) of new drug candidates using QSPkR. The study involved the development of QSPkR for a negative logarithm for the calculation of free fraction of drug in plasma (pfu) of basic drugs. It was found that the presence of quaternary C-atoms contributes negatively to PPB [49]. In another

study, contemporary three-dimensional (3D) molecular characterization and multivariate statistical analysis were used to predict the behavior of adenosine A_1 receptor agonists. QSPkR was performed using programs like SYBYL/CoMFA, GRID, and Pallas [50].

During submissions for regulatory filing and review, PBPK modeling and simulation is receiving significant attention. PBPK models can be used as a specific method to predict a compound's pharmacokinetic (PK) description on the basis of its preclinical ADME data which can be used to determine sensitivity to the specific organ after the delivery of drugs, by taking into account the rate of absorption and disposition in that organ and, where appropriate, the metabolism within that organ. Based on the PK data of a single dose, the PBPK model can be used to predict the PK profile of multidose and/or variable dosing routes. The PBPK model can be used to estimate the PK profile in different ethnic groups, as well as of different age and disease groups based on the PK data from one ethnic group [51]. Pharmacokinetic models are used to explain the duration of drug exposure in patients and to analyze the cause of patient variability. They can be used to simulate alternative dosage regimens, facilitating an informed evaluation of dosage regimen prior to conduct of research [52]. Via simulation, a deeper understanding of the interrelationship between pharmacokinetics and pharmacodynamics (integrative pharmacology or PK/PD) of a compound may provide predictions of the effects of new, untested scenarios (such as different dose levels and/or regimens) to maximize the likelihood of technical success. The measurement and assessment of drug-concentration-time or drug-effect-concentration data are therefore of critical importance in quantitative pharmacological research.

The various phases of simulation performed in a simulator have been shown schematically in Fig. 8.2. From the main window, the users can pick a module of interest, PK or PD, and then collaboratively set the options. For PK, PD, and PK/PD modeling, the simulator offers several customizable options. In general, the PD module can be used to investigate an exposure-response profile that provides concentrations in a particular limit or to perform an integrated PK/PD simulation using the time-concentration profile generated using the PK module [53]. In combination with PK/PD modeling, the mechanism-based PD drug-drug interaction findings may provide clear mechanistic reasons for potentially conflicting results obtained from various designs of temporal drug regimens implemented in different biological systems. Mathematical modeling and simulation provides a theoretical basis for assessing the design of drug combinations or dosing regimens in PD drug-drug interaction studies [54]. Docking can be used to analyze ligand-protein interaction in the PD study; however, docking protocols are approximate and lack protein versatility, which may impact the reliability of the generated matrix of ligand-protein. Therefore, a more accurate method of molecular dynamic simulation will provide a better complement to docking. In particular, molecular dynamics simulations have been used to examine the behavior of macromolecules [55].

Critical Path Initiative of drug development led by FDA emphasizes on new tools and approaches to reengineer the drug development and discovery enterprise. Of these, one such tool is the utilization of a model-based system in drug development.

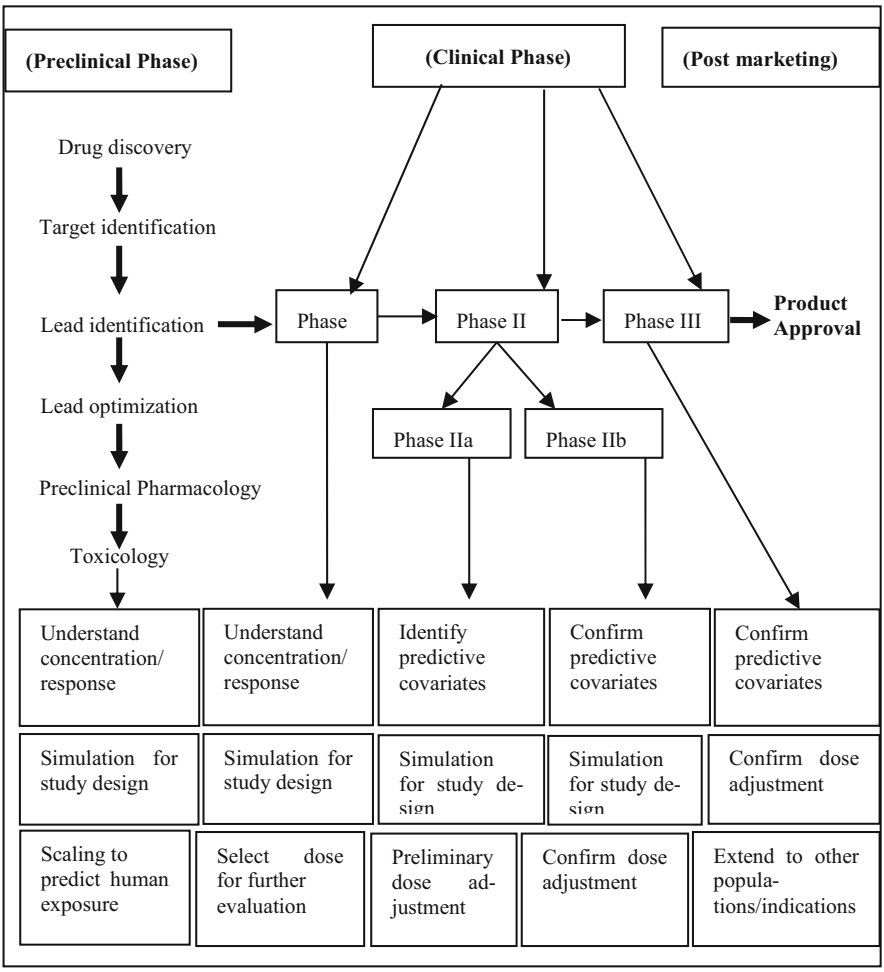


Fig. 8.3 Modeling and simulation during model-based drug development showing the integration of all aspects of drug development from drug discovery to postmarketing

The model-based drug development (MBDD) approach integrates relationships among diseases, drug properties, and variability induced across each study and development stage, so that scientifically informed decisions can be made at each stage of the process [56]. MBDD is a new paradigm based on a combined mathematical and statistical approach to integrating knowledge from all phases of drug development to the postmarketing phase. The model development can be considered as a continuum of iterative steps of learning and confirming during which the knowledge about a new compound is continuously updated (Fig. 8.3). The primary focus of MBDD is on the construction of mathematical models to characterize the causal relationship among prognostic factors, disease state, drug characteristics, safety, and efficacy outcomes and summarizing of new data, previous knowledge,

and assumptions. Leveraging previous information on the target disease, the new chemical entity, and integrating historical information on related compounds can improve model efficiency. Simulations based on such models provide a tool for the selection of effective and safe doses, optimization of study sample sizes, evaluating various trial designs, and creating logical go/no go decisions depending on the possibility of getting predefined study targets. Under the MBDD framework, different model components are used together to facilitate informed judgmental designing of the drug development process [57].

8.4 Structural Components of the Physiologically Based Pharmacokinetic (PBPK) Model

The body can be divided into various compartments connected by the circulatory system (blood). These compartments could be adipose tissue, bone, brain, gut, heart, kidney, liver, lung, muscle, skin, spleen, etc. Each compartment has a specific permeability and perfusion rate kinetics. This model assumes that tissue drug concentration is in equilibrium with the total circulated concentration of the drug denoted by K_p. The time required to arrive at steady state is faster if the tissue is highly perfused. In the case of permeability-limited kinetics mainly for polar molecules, the tissue is divided into intracellular and extracellular space. The drug-specific parameters that are used for developing oral plasma concentration-time profiles using PBPK are intrinsic clearance of a compound (Cl_{int}) and K_p [58]. An overall structural demonstration of the PBPK model illustrating two-compartment (gastrointestinal) or three-compartment (with kidney consisting of proximal tubule and rest of kidney) model of intravenous (IV) and oral routes of administration have been shown in Fig. 8.4 [59].

In pharmacokinetic and toxicokinetic studies, the PBPK tissue compartment models are either flow/perfusion-limited or permeability/diffusion-limited. The mass balance differential equation describing the flow-limited tissue compartment model is:

$$dC/dt = F/V(C_{in} - C/\rho) \tag{8.7}$$

where C denotes the drug concentration in the compartment, C_{in} indicates the inflow concentration of drug, F and V, respectively, represent the flow and total volume of the organ or tissue, and ρ is the partition coefficient of the drug. In whole-body PBPK modeling, vascular connectivity of tissue and organs is accounted by inflow (C_{in}) and outflux (C_{out}) of the drug [60]:

$$F(C_{in} - C_{out}) = V_1(dC_1/dt) + V_2(dC_2/dt) \tag{8.8}$$

Single-compartment PBPK model possessing constant volume with homogeneous concentration throughout the whole tissue and the rate of mass change in a tissue compartment is given by:

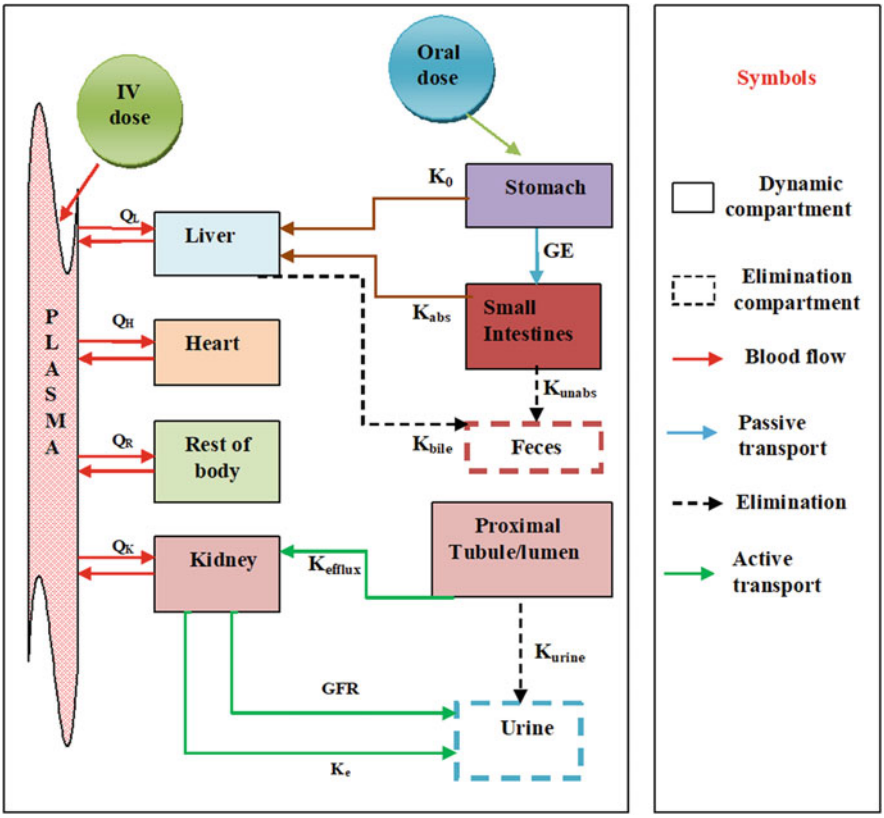


Fig. 8.4 A schematic representation of the PBPK model where Q_L, Q_H, Q_R, and Q_K represent fractional blood flows, respectively, through liver, heart, body, and kidney. K_0, K_{abs}, K_{unabs}, K_{bile}, K_{efflux}, K_{urine}, K_e and GFR represent stomach to liver uptake, small intestine to liver uptake, unabsorbed dose, biliary elimination rate, rate of efflux, urinary elimination rate, elimination rate constant, and glomerular filtration rate constant, respectively

$$V(dC/dt) = F_{in}C_{in} - F_{out}C/\rho \tag{8.9}$$

The vascular and extravascular spaces are well mixed with complete balance and there is no limitation of mass transfer between tissue and blood. When $V_2 \gg V_1$, V_2 expresses the total volume of the tissue compartment, thus leading to an approximation that the whole organ/tissue can be treated as a single compartment.

Multicompartment model has input (I) and excretion (E) related variables and the mass balance equation for these models is represented as [60]:

$$F(C_{in} - C_{out}) + I - E = V_1(dC_1/dt) + V_2(dC_2/dt) + V_3(dC_3/dt) \tag{8.10}$$

8.4.1 PBPK Modeling Simulation Software

Numerous commercial platforms are available for PBPK modeling based on mathematical equations or coding such as MATLAB, ModelMaker, Berkeley Madonna, and acsIX which provide tools for the development of code, numerical equations, and graphical output. These tools are capable of simplifying mathematical complexity, constructing and validating a generic model in a short span of time. Various studies have been conducted using these platforms for the prediction of physiological parameters. A mechanistic oral absorption PBPK model of metformin has been developed using GastroPlus Advanced Compartmental and Transit (ACAT) model (Simulations Plus, Inc., Lancaster, CA) (ACAT model divides the gastrointestinal tract into nine compartments and for each of them, the rate and extent of absorption is calculated as a function of time. The input parameters in correspondence to compound are solubility at different pH values and permeation coefficient) [44]. In the model, plasma membrane monoamine transporter (PMAT) influx and multidrug and toxic compound extrusion (MATE) efflux transporters are incorporated on the apical membrane of the intestinal lumen, in addition to paracellular permeability as the only mode of absorption into the portal vein. For simulation of systemic clearance, permeability-limited liver and kidney tissues have been incorporated with transporters like human organic cation transporter 1 (OCT1) (basolateral intestine and liver inflow), organic cation transporter 2 (OCT2) (basolateral kidney inflow), and MATE2K (apical kidney inflow) [61, 62]. Furthermore, the PBPK model to study the drug-drug interaction (DDI) involving metformin has been developed, wherein substantially lower inhibition constant values of MATE for cimetidine were incorporated to establish a correlation between the in vitro data and in vivo clinical data. Also, the model suggested that the DDI between metformin and cimetidine is mediated through the inhibitory effect of MATEs by cimetidine, instead of OCT2 inhibition [63].

Similarly, PBPK models have also been applied to study the toxic effect of drugs in children. The pediatric PBPK model for the prediction of pharmacokinetic parameters of compounds metabolized by cytochrome P450 (CYP) enzymes using PK-Sim® software has been developed. Increased regulatory demand for such models is propelling the need for development and evaluation of these models [64]. Furthermore, a SIMCYP™ population-based simulator has been used to study enzyme inhibition and induction potential of crizotinib. When administered along with midazolam, the model showed the role of CYP3A enzymes present in hepatocytes in time-dependent inhibition of enzymes [65]. In order to facilitate the process of drug discovery, ADMETlab, a free web platform, has been developed to provide impetus to rapid screening of drug-likeness and drug characteristics, biopharmaceutical properties, mainly the ADMET (absorption, distribution, metabolism, elimination, and toxicity) prediction [66, 67].

8.5 Physiome Project

The International Union of Physiological Sciences (IUPS) Council introduced the Physiome Project in 1993 to provide a quantitative description of physiological dynamics of intact organisms along with their functional behavior. Virtual physiome human (VPH) is a personalized, 3D multilevel modeling of an individual's unique physiology and anatomy that focuses on different biological levels: molecular and cellular levels, tissue and organ levels, and system and human (organism) levels and models, simulates and visualizes them for application in personalized health care (development of personalized medicine and performing virtual "surgery" for validation of in silico versus in vivo data) [68]. Other objectives of the initiative include the identification of relevant biomarkers for diagnosis and understanding of the disease; take individualized medical decisions for the designing of drug regimen, implants, and artificial organs with the application of artificial intelligence and robotics. The initiative comprises more than 20 projects, including euHeart, personalized heart models for diagnosing and treating heart disease; and HAMAM, highly accurate breast cancer diagnosis technology that combines novel imaging methods with modeling.

8.5.1 Cardiovascular Modeling

Cardiac diseases such as coronary artery disease (CAD), congestive heart failure (HF), and cardiac arrhythmias account for the loss of cardiac function that increases mortality and morbidity. Thus, early detection and prediction of cardiovascular disease (CVD) progression are essential for improved treatment and reducing the mortality and morbidity rates. Multiscale numerical scheme (MSNS) is a method developed by Whiteley which helps to solve bidomain equations by exploiting the multiscale nature of the problems modeled by these equations. Bidomain models are generally used to model cardiac electrophysiology by the application of elliptic and parabolic partial differential equations, coupled at each point in space with a huge array of stiff, nonlinear ordinary differential equations [69].

euHeart is an integrated patient-specific personalized cardiovascular modeling platform which has a library of computer-based heart models and functions via web-enabled euHeart database (euHeartDB) or Anatomical Model Database (AMDB) [70]. Being a dynamic sharing platform, this database manages data access and is also linked with Physiome Model Repository via CellML geometrical models and interoperable FieldML data format. euHeart helps in cardiac resynchronization therapy (CRT) by implantation of pacemaker, thereby reducing heart dyssynchrony. Cardiac electromechanical simulations can be performed by developing a computer-based cardiac model representing the subject's heart morphology depending on the clinical data. A realistic model is designed based on the inputs such as patient-specific geometry, tissue properties like elasticity or conductivity, deformation/motion information, and microstructure properties. The simulation is performed using the software framework GIMIAS (Graphical Interface for Medical Image

Analysis and Simulation), integrated with open-source coding platforms like OpenCMISS and Chaste for computational imaging and computational physiology [71].

8.5.2 Breast Cancer Modeling (HAMAM)

The early, rapid, and accurate breast cancer diagnosis is still an unresolved challenge, despite advances in modern imaging technology. Highly Accurate Breast Cancer Diagnosis through Integration of Biological Knowledge, Novel Imaging Modalities, and Modeling (HAMAM) is one of the key points of VHP that integrates the available multimodal images and patient information on a single clinical platform. The breast tissue can be characterized based on the information collected from knowledge databases of tumor models, genetic data, genotype, phenotype, and standardized imaging. The suspicious breast tissue can be diagnosed and resolved by applying statistical knowledge obtained from these relevant databases. Furthermore, novel mammographic image descriptor (magnetic resonance imaging (MRI)) has been combined with kinetic and morphological descriptors to characterize the lesions as benign or malignant. Moreover, a new tool called computer-aided diagnosis (CAD) has been developed to characterize suspicious lesions along with application of automated 3D breast ultrasound (ABUS) [72, 73].

8.6 Isolated Tissues and Organs

Functions and structure of isolated tissues and organs are also being simulated to explore the organ- and tissue-specific role. For example, in the case of cardiovascular diseases (CVDs), understanding of the structure-function relationship of cardiovascular system can help in better diagnosis and treatment. This understanding can be further strengthened by interfaces which model biology, therapeutic targets, and biomechanism of the cardiovascular system. Integration of these multiscale and multiphysical models helps to understand the development and progression of CVD and designing of clinical intervention [74].

To efficiently understand the physiology and function of heart, the electromechanical operation of the heart can be simulated through whole heart modeling. Partial differential equations (PDEs) have been applied to simulate membrane potential of the heart, along with a series of ordinary differential equations (ODEs) and algebraic equations to predict current flow in ion channels, pumps, and exchangers [75]. Besides, the rabbit ventricular geometry model (RVGM) is an anatomical model, consisting of 36 linear cubic Hermitian hexahedral finite elements and 99 nodes, to study fiber orientation in cardiac defibrillation and ventricular electrophysiology [76, 77]. Electroanatomical mapping (EAM) system is a preoperative characterization approach to plan radiofrequency ablation (RFA) protocol by using 3D computational models for accurate simulation of cardiac electrophysiology to understand cardiac disorders, plan therapy, and its management [78]. Multiscale

models of cardiac electrophysiology and biomechanics can be combined to model the electromechanical heart function, which can be of help in health and disease management [79].

Similarly, the epithelial cell rearrangements and morphogenesis have been studied using cellular Potts modeling at the multiscale level. Cellular Potts modeling and subcellular elemental model (SEM), also known as the Glazier-Graner-Hogeweg (GGH) model, are cell-centric modeling frameworks that represent complex multiscale behavior [80, 81]. A visual and realistic 4D simulation model for organogenesis based on in silico experimentation of mouse pancreas has helped to study morphological development. It was deduced that the extracellular matrix (ECM) triggers the development of the pancreas, in the absence of which pancreas fails to develop in mice [82].

xPULM, an electromechanical respiratory lung simulator, actively replicates human breathing frequencies and tidal volumes by combining in silico, ex vivo, and mechanical respiratory data [83]. Similarly, the epithelial cell model and the quantitative kidney knowledge database populated with standard kinetic descriptors of membrane transport can be used to simulate renal epithelial cells with cell-specific transporters. The serial analysis of gene expression (SAGE) technique developed by Chabardes et al. has primarily been employed for high-resolution mapping of the gene expression pattern of nephron segments of the human kidney [84]. Recently, a web-based application known as "*Bodylight.js.Simulator*," consisting of a mathematical model and a graphical user interface, has been developed to understand the physiology of nephrons and anatomy of the kidney [85]. Probability-based bifurcation algorithm, along with Murray's law, is used to study renal microcirculation phenomena such as asymmetric bifurcating tree (ABT) and kidney-specific asymmetric bifurcating tree (KSABT) [86]. Executable simulation model (EXSIMO) composed of data and codes, testing, continuous integration, and computational analysis has been used to replicate hepatic glucose metabolism [87]. Drug-induced liver injury (DILI)-sym is an initiative by the USFDA. It utilizes the quantitative systems toxicology to restructure the DILIsym model to predict liver safety liabilities of new drug candidates and interpretation of liver biosafety biomarkers used during clinical studies [88].

8.7 Cells

"cBioPortal" is an open-access resource that provides interactive modeling of cancer genomics datasets for use in clinical evaluation and real-time application. CompuCell3D is a three-dimensional C++ platform used to simulate and model cellular and multicellular processes, integrated with morphogenesis modules [89, 90]. Glioblastoma multiforme (GBMs) have also been studied with finite-element model (FEM) by simulating the 3D growth of GBMs in brain parenchyma along with comparison of in silico GBM growth with real-time growth as visualized by magnetic resonance images [91].

Embryogenesis is a complex mechanism and requires an integrated understanding of mechanical, molecular, and genetic phenomena. Gene regulatory network (GRN) focuses on multigene interaction and helps to understand genetic signaling involved in a biological process. Various GRN models have been proposed, such as Boolean cis control module function or binary networks containing rate differential equations, which help predict RNA polymerase's probability of initiating or blocking gene transcription [92]. The computer-based mitochondrial simulation model "MITOsym" is commonly used to investigate the glycolytic pathway and cellular parameters such as oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) [93]. Mutant mitochondrial DNA (mtDNA) accumulation and its fusion-fission dynamics have been evaluated through in silico modeling, and it has been observed that an uneven distribution of mutant mtDNA is due to the lower rate of mitochondrial fusion-fission resulting in a higher burden of the mutant mtDNA in the tissues which tend to expand clonally [94]. Several multilevel computational modules have been used to study mitochondrial dysfunction and aging in the nematode *Caenorhabditis elegans*. They focus on the mitochondrial free radical theory of aging, mitochondrial unfolded protein response (UPR^{mt}), mitochondrial biogenesis and autophagy dynamics, and mitochondrial stress response pathways' nicotinamide adenosine dinucleotide (NAD^+)-dependent deacetylase activity [95]. Simulation of the 3D model of cardiomyocytes has helped to study cardiac electrophysiology, Ca^{2+} dynamics, adenosine triphosphate (ATP) metabolism, diffusion of Ca^{2+} and energy metabolites, etc. as a result of sarcomeric forces [96–98]. This 3D cardiomyocyte simulation model has been improvised to assess subsarcolemmal mitochondrial (SSM) and interfibrillar mitochondrial (IFM) dynamics, oxygen diffusion, and oxidative phosphorylation [99]. Monte Carlo (MC) simulation is a fractal chromatin-based human cell nuclear model that is used to investigate radiation-induced cell damage, primarily single-strand breaks (SSBs) and double-strand breaks (DSBs) [100]. To understand Ca^{2+} dynamics (influx/efflux) across mitochondria-associated endoplasmic reticulum membrane, also termed as μd compartment (alongside cytosol, endoplasmic reticulum, and mitochondria), a mathematical compartment closed-cell model has been developed. The model describes the cytosolic Ca^{2+}oscillations upon stimulation of inositol triphosphate (IP_3) in nonexcitable cells [101]. In this line, a Markov model for studying the kinetics of type I and II IP_3 receptors has been developed based on the concentration of intracellular Ca^{2+}, IP_3, and adenosine triphosphate. The channel activity and transition of channels between active and inactive states have been studied [102].

Multicellular organisms with complex structures metaphorically develop from single- or multiple-cell fertilized egg. Their forms, structures, and gene expression patterns emerge through the organization of a myriad of cells. Steinberg has introduced the theoretical concept of differential adhesion hypothesis (DAH), further developed as 2D and 3D DAH-based computer simulation models to study the morphogenetic pattern of the multicellular organisms [103]. In addition, MecaGen is a computational simulation modeling platform that primarily links the genetic regulatory network (GRN) with cell response to analyze morphogenesis during early

embryogenesis. MecaGen is mainly composed of two modules "Meca" and "Gen.": Meca defines the mechanical interaction and cell behavioral properties. Gen deals with the modeling of genetic regulation and molecular signaling [92]. Morphogenetic behavior, structural complexity, and life cycle of the whole organism can be studied by the cellular Potts model (CPM) [104]. A web-accessible database known as a cyber cell database (CCDB) has been designed for in silico simulation and modeling of *Escherichia coli* (strain K12). It describes all characteristics of the organism, including its entire genomic and protein sequences. CCDB primarily consists of four browsable databases, namely: gene and protein information CCDB; structural proteomics 3D database, RNA database (tRNA and rRNA information) (CCRD), and metabolite database (CCMD) [105]. Likewise, a hypothetical E-CELL simulation model of self-replicating parasite "*Mycoplasma genitalium*" has been constructed based on 127 genomic sets. The virtual E-CELL interface generated based on in silico experimentation of metabolic pathways of *Mycoplasma genitalium* describes glycolysis, lactate fermentation, glycerol, and fatty acid uptake, gene transcription, protein, and RNA degradation [106]. The physical and kinetic network-based models are two major approaches for whole-cell modeling. The network-based models are extensively based on experimental kinetic data and complete genomic information of the organism like those of *Mycoplasma genitalium* and *Escherichia coli*. However, physical models have limited applications due to structural information and high computational costs [107]. Some simulation model softwares have been mentioned below (Table 8.2).

8.7.1 CellModeller

CellModeller is physical-genetic 3D simulation software of bacterial colonies and determines the adhesive force required for intercellular interaction and shows viscous drag forces across the cells responsible for adhesion. It can simulate up to 10^6 cells in a day, allowing for a fast and significant system [114]. The statistical analysis helps to determine actual cell architectures and visualization and the CellModeller facilitates the development of morphogenetic models of genetic, hormonal, and physical systems [115].

8.7.2 The Cellular Potts Model

Glazier and Graner's cellular Potts simulation model provides aggregate information on cell dynamics, cell migration, and cell interaction primarily at tissue level [115]. It has become an essential technique for studying differentiation, growth, death, deformation, cell migration of morphogenesis, and the secretion and absorption of extracellular materials. This model has been successfully used to describe cell behavior and morphological development of *Dictyostelium discoideum* [80, 116, 117]. The CPM helps to analyze irregular and fluctuating cell shapes, cell dynamics, and cell deformation with high resolution. This simulation modeling is based on

Table 8.2 Isolated organ and cell simulation modeling software

S. no.	Isolated organ modeling softwares	Application	Reference
1.	Cellular Potts modeling and subcellular elemental model (SEM)	Analyze cell movement, cell-cell interactions, and response to an external chemotactic gradient	[108]
2.	xPULM	An electromechanical respiratory in silico, ex vivo lung simulator	[109]
3.	"*Bodylight.js.Simulator*"	A visual editor for creating in-browser dynamic applications and visualizations	[110]
4.	EXSIMO	An executable simulation model combines data and code with the execution environment to run the computational analysis in an automated manner using tools from software engineering	[111]
5.	DILIsym model	Quantitative systems toxicology software for modeling drug-induced injury	[112]
	Cell modeling software	Application	Reference
6.	cBioPortal	An interactive open-source platform designed for visualizing and analyzing genomics data and selecting better treatments for patients	[113]
7.	CompuCell3D	A multicell computer simulation method for finite-element mechanistic modeling, in silico study of multicell phenomena at the tissue scale based on biologically observed cell behavior and interactions such as movement, adhesion, growth, death, mitosis, secretion of chemicals, chemotaxis, etc.	[90]
8.	"MITOsym"	MITOsym includes the primary, essential biochemical pathways associated with hepatocellular respiration and bioenergetics, including mitochondrial oxidative phosphorylation, electron transport chain activity, mitochondrial membrane potential, and glycolysis	[93]
9.	MecaGen	MecaGen enables the specification and control of complex collective movements in 3D space through a biologically relevant gene regulatory network and parameter space exploration	[92]
10.	E-CELL simulation model	A simulator of cellular system	[106]

phenomenological energy known as the Hamiltonian function described by the equation:

$$H(\acute{o}) = H_A + H_P + H_J, \tag{8.11}$$

where $\acute{o}$ is the cell configuration ($H_A = \lambda a\,(A - a)^2$; $H_P = \lambda_P(P - p)^2$; and $H_J = J$ $(0,1)P$). Here, H_A denotes the energy cost for expansion or contraction of the area "A." Compared to the rest area "a" of the cell, H_P is an energy cost for deviation of

the cell perimeter "P" from its rest perimeter "p." H_J is an energy associated with the cell-medium interface (including cell-cell or cell-medium adhesive energies.) The factors λ_a, λ_p, and $J(0,1)$ set the relative energetic costs of area changes, perimeter changes, and changes in contact with the medium [118].

8.7.3 Chaste

The cell-based chaste calculation framework connects spatial and temporal scales in a single general modeling framework. It is based on a highly nonlinear differential equation approach that describes cell interactions; nutrients' transport pathways, signaling pathways, cell growth, and cell division. Chaste has been developed for analyzing cardiac electrophysiology and cancer biology [116].

8.7.4 Agent-Based Modeling (ABM)

Agent-based modeling is a common technique to explore biological systems mainly in the field of oncology and offer a detailed understanding of cancer initiation, progression, and invasion and metastatic mechanism [119, 120]. According to this model, each cancerous cell appears as an agent with unique characteristics and spatial coordinates. These agents are individually approved to provide details at the cellular level, such as cell-cell interactions, signaling pathways, replication, necrosis, apoptosis, mitosis, tumor heterogeneity, and transient cancer phenotype. This technique involves various computational techniques, neural network approaches, and even Monte Carlo technique to describe the distribution pattern and random phenotypic transition in two- and three-dimensions (2D or 3D). This modeling helps to determine genetic mutation pathways, epithelial and mesenchymal transitions, cell-cell signaling (including heterogeneous cell interactions), and individual susceptibility to drugs.

8.7.5 Steered Molecular Dynamics (SMD) Simulation

SMD investigates several biological phenomena based on molecular dynamics mimicking the principle of atomic force microscopy (AFM). The conformational changes of ligands and biomolecules can be studied with SMD along with drug-receptor interactions and their binding energies [3]. SMD has been used for ligand-residue critical analysis and interaction of M1 aminopeptidase inhibitors in antimalarial drug development. SMD simulations are mainly based on the application of time-dependent external forces on the ligand to facilitate the dissociation process and explore the receptor's structure-function relationship with drug complex. Similarly, the energy pathways and conformational changes of HIV-1 RT/α-APA (human immunodeficiency virus type 1 reverse transcriptase/α-anilinophenylacetamide) and avidin-biotin complex have been analyzed with SMD modeling [121, 122].

8.8 Proteins and Genes

Various simulation models have been developed for structural analysis, characterization, biological activity, and pathway elucidation of several proteins and genes. In silico molecular modeling has been used for 3D structural analysis, functional characterization, and molecular docking of regucalcin against 1,5-anhydrosorbitol [123]. Researchers have developed a simplified version of the off-lattice model to investigate thermodynamics and kinetics of folding pattern (α, β, α/β) of small, large, and mixed proteins [124]. Several methods have been developed to predict the structure (protein structure prediction (PSP)) and complexity of protein, such as template-based modeling (TBM), de novo PSP template-free modeling to predict new protein configurations. Monte Carlo and molecular dynamics simulation approaches have also been applied to the atomistic and coarse-grained models to investigate the folding pattern of peptides and proteins [125]. Furthermore, molecular dynamics has helped in realistic simulation to understand protein folding, RNA folding, protein function, and bimolecular pathways at the atomic level [126, 127]. Structure-based modeling based on the energy landscape theory and minimal frustration principles has been primarily established to study folding pathways, folding kinetics, posttranslational modifications, allosteric conformation, and ligand binding [126].

The Bayesian induction modeling has been applied to analyze the expression and regulation of 800 genes with varied expressions over the cell cycle, including that in *S. Cerevisiae* [126]. A multidimensional simulation toolbox "SELANSI" (SEmi-LAgrangian SImulation of GRNs) provides detailed information about GRNs, genetic expression, and self-regulation with the same or another gene network [128]. Also, BioDynaMo simulator has been developed by a research group to determine heritable genetic information like DNA and base-pair sequences, RNA, and protein by using Euler and Runge-Kutta differential methods [127]. Moreover, 3D simulation model "MODELLER" can validate RV1258c (tetracycline/aminoglycoside efflux pump that implicates multidrug resistance for *Mycobacterium tuberculosis*) structural and molecular dynamics using GROMACS software [129].

3D in silico modeling of La Crosse virus-Orthobunyavirus (LACV) using SWISS-MODEL was used to assess the nucleoprotein of groundnut ringspot virus (GRSV). The aligned sequences were analyzed with the MODELLER simulation platform to develop tetrameric models with or without RNA [130]. The assessment of glycosylation of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) spike protein (S protein) of viral membrane and their model structure has been performed in GALAXY modeling suite present in CHARMM-GUI Membrane Builder [128]. Similarly, 3D structure, dynamics, and molecular docking of the histamine H1 receptor using cloperastine have been evaluated using MODELLER 9 V7 software. It was found that the hydrogen-binding interaction between the H1 receptor and cloperastine derivatives can help develop potential anti-allergy therapy [131]. The computational simulation like NiDelta (de novo predictor) for predicting the tertiary protein structure and its dynamics is crucial for modern biology research

[132]. Other computational simulation approaches used for structural and molecular dynamics estimation have been mentioned in Table 8.3.

8.9 Future Scope and Direction

Pharmacokinetic and pharmacodynamic modeling and simulation have provided tools that have speeded up drug discovery and development. This has also brought down the cost and risks associated with the process. Integration of computational approach with the biological data has been used to predict properties and system pathways to explore process-structure-property relationships. Moreover, these mathematical/computational models based on in vitro, in silico, and in vivo experiments have to lead to design, optimization, and validation of biological hypothesis which can be extrapolated to preclinical and clinical settings. A clear picture of therapeutic regimen can be created with optimal treatment schedules. Especially in research areas like cancer, drug discovery, and life cycle management, issues can be addressed reliably, and failure can be minimized [134].

Furthermore, with the help of multiscale modeling and simulation data integration can be done at the scale of tissue, organs, protein, and genes. As a result of this technology, data related to the detailed description of diseases and response to therapies have been generated. The potential application of computer simulation and modeling is enhancement in the knowledge of pharmacokinetics and pharmacodynamic drug interactions which can help optimize dose and combination drug administration. Reproducible dose titration and clear end points can be determined, which can be clinically applied. With the use of these platforms, validation of interaction models can be done and made available to physicians for real-time application in clinical settings [135].

The simulation and modeling studies hold an enormous scope in the analysis of pediatric drug delivery and clinical trials. Regulatory binding, along with ethical and financial considerations, adds to the complexity of pediatric studies. PBPK/PD helps develop optimal study design and data collection techniques complemented with rationalized sample size and number. With computer simulation and modeling drug development, researchers, clinicians, technicians, and the like, can confidently design clinical trials in pediatric patients [136].

Table 8.3 List of computational simulation software and their applications

S. no.	Modeling and simulation software	Application	URL
1.	Simcyp™ Simcyp Pediatric, Simcyp Cardiac Safety Simulator (CSS), Simcyp Lactation, PBPK simulator	Simulates drug-drug interaction, absorption modeling, dosing for special populations, and PK prediction	https://www.certara.com/software/simcyp-pbpk/
2.	GastroPlus®	Simulates intravenous, oral, oral cavity, ocular, inhalation, dermal, subcutaneous, intramuscular absorption, biopharmaceutics, pharmacokinetics, and pharmacodynamic parameters in humans and animals	https://www.simulation-plus.com/software/gastroplus/
3.	PKPlus™	Estimates noncompartmental pharmacokinetic (PK) parameters, along with one-, two-, and three-compartment PK models from pharmacokinetic studies (IV and/or oral) without the need to run full simulations	https://www.simulation-plus.com/software/gastroplus/pk-models/
4.	DDDPlus™	Simulates in vitro dissolution for formulation and analytical of active pharmaceutical ingredients (API) and formulation excipients under various experimental conditions in seconds	https://www.simulation-plus.com/software/dddplus/
5.	PDPlus™	To predict the PD effect due to alteration in dose, dosage form, and dosing regimens and determine the action's kinetics. Multiple PD models (therapeutic and adverse) can be accommodated for each drug record	https://www.simulation-plus.com//software/gastroplus/pkpd-modeling/
6.	MembranePlus™	A software that determines in vitro permeability, in vivo absorption, and systemic clearance/distribution with the advanced compartmental absorption and transit (ACAT™) and PBPK models. It also provides in vitro permeability (human colon adenocarcinoma (Caco-2), parallel artificial membrane permeability assay	https://www.simulation-plus.com/software/membraneplus/

(continued)

Table 8.3 (continued)

S. no.	Modeling and simulation software	Application	URL
		(PAMPA), or Madin-Derby canine kidney (MDCK)) for optimization	
7.	DILIsym® and NAFLDsym®	DILIsym is quantitative systems toxicology (QST) software for potential drug-induced liver injury (DILI) responses at various development stages	https://www.simulation-plus.com/software/dilisym/
8.	GROMACS	The most widely used open-source free software codes in chemistry, used primarily for dynamical simulations of biomolecules, proteins, lipids, and nucleic acids	https:www.gromacs.org [133]
9.	MODELLER	A program for automated protein homology modeling	https://salilab.org/modeller/ https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4186674/
10.	Macromoltek	A molecular simulations software for antibody modeling, side-chain packaging, renumbering, and other web-based computational tools for antibody development	https://www.macromoltek.com

8.10 Credible Online Resources for Further Reading

URL address	Category of source (Government/institution/private company/personal webpage)	What to read/refer
https://www.certara.com	University of Sheffield, UK, Food and Drug Administration, and Swedish Medical Products Agency	Simcyp™ Pharmacokinetic modeling and simulation, Phoenix™ PK/PD platform, Phoenix WinNonlin, Phoenix NLME, Phoenix in vitro-in vivo correlation (IVIVC) toolkit, SIMCYP™ PBPK simulator, Simcyp Animal, Simcyp Pediatric, quantitative systems pharmacology, quantitative systems toxicology and safety
http://www.simulations-plus.com	Los Angeles based company, California	GastroPlus™, physiologically based biopharmaceutics modelling (PBBM) Physiologically based pharmacokinetic (PBPK) QSP QST modeling, ADMET Predictor®, DILIsym®, KIWI™, Monolix®, PKPlus™, and NAFLDsym®
http://www.open-systems-pharmacology.org/	Bayer AG, a German multinational pharmaceutical and life sciences company	PK-Sim™, PBPK modeling tools, and Mobi
http://www.cit-wulkow.de/	Northern Germany based company	MEDICI-PK™, **simulation and modeling**
http://www.cyprotex.com	Cyprotex Limited was acquired by Evotec AG	Cloe PK™, ADME PK, physicochemical profiling, in silico modeling, and toxicology
http://vcell.org	The Center for Cell Analysis & Modeling. National Institute of General Medical Sciences (NIGMS)	Comprehensive platform for modeling cell biological systems
http://www.cellml.org	Auckland Bioengineering Institute at the University of Auckland	Database and model building softwares

References

1. Gallopoulos E, Houstis E, Rice JR (1994) Computer as thinker/doer: problem-solving environments for computational science. IEEE Comput Sci Eng 1(2):11–23. https://doi.org/10.1109/99.326669
2. Shiflet AB, Shiflet GW (2014) Introduction to computational science: modeling and simulation for the sciences, 2nd edn. pp 1–816

3. Patel JS, Berteotti A, Ronsisvalle S, Rocchia W, Cavalli A (2014) Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5. J Chem Inf Model 54(2):470–480. https://doi.org/10.1021/ci4003574
4. Bonate PL (2000) Clinical trial simulation in drug development. Pharm Res 17(3):252–256. https://doi.org/10.1023/A:1007548719885
5. Bonate PL (2001) A brief introduction to Monte Carlo simulation. Clin Pharmacokinet 40(1):15–22. https://doi.org/10.2165/00003088-200140010-00002
6. Johnson SCD (1998) The role of simulation in the management of research: what can the pharmaceutical industry learn from the aerospace industry? Ther Innov Regul Sci 32(4):961–969. https://doi.org/10.1177/009286159803200414
7. Reigner BG, Williams PEO, Patel IH, Steimer JL, Peck C, Van Brummelen P (1997) An evaluation of the integration of pharmacokinetic and pharmacodynamic principles in clinical drug development. In: Experience with Hoffmann La Roche, vol 33. Clinical pharmacokinetics. Springer International Publishing, pp 142–152. https://doi.org/10.2165/00003088-199733020-00005
8. Mould DR, Upton RN (2012) Basic concepts in population modeling, simulation, and model-based drug development. CPT Pharmacometrics Syst Pharmacol 1(1):1–14. https://doi.org/10.1038/psp.2012.4
9. Abbiati RA, Savoca A, Manca D (2018) An engineering oriented approach to physiologically based pharmacokinetic and pharmacodynamic modeling. In: Computer aided chemical engineering, vol 42, 1st edn. Elsevier B.V., pp 37–63. http://hdl.handle.net/11311/1128778
10. Abdel-Rahman SM, Kauffman RE (2004) The integration of pharmacokinetics and pharmacodynamics: understanding dose-response. Annu Rev Pharmacol Toxicol 44(1):111–136. https://doi.org/10.1146/annurev.pharmtox.44.101802.121347
11. Salahudeen MS, Nishtala PS (2017) An overview of pharmacodynamic modelling, ligand-binding approach and its application in clinical practice. Saudi Pharm J 25:165–175. https://doi.org/10.1097/00000542-199101000-00010. King Saud University
12. Duffull S (2012) Basic pharmacokinetics and pharmacodynamics, an integrated textbook and computer simulations. Br J Clin Pharmacol 73(2):312–313. https://doi.org/10.1111/j.1365-2125.2011.04077.x
13. Anderson BJ, Holford NHG (1997) Rectal paracetamol dosing regimens: determination by computer simulation. Paediatr Anaesth 7(6):451–455. https://doi.org/10.1046/j.1460-9592.1997.d01-125.x
14. Kuentz M, Nick S, Parrott N, Röthlisberger D (2006) A strategy for preclinical formulation development using GastroPlus™ as pharmacokinetic simulation tool and a statistical screening design applied to a dog study. Eur J Pharm Sci 27(1):91–99. https://doi.org/10.1016/j.ejps.2005.08.011
15. Shafer SL, Varvel JR (1991) Pharmacokinetics, pharmacodynamics, and rational opioid selection. Anesthesiology 74(1):53–63. https://doi.org/10.1097/00000542-199101000-00010
16. Bemporad D, Essex JW, Luttmann C (2004) Permeation of small molecules through a lipid bilayer: a computer simulation study. J Phys Chem B 108(15):4875–4884. https://doi.org/10.1021/jp035260s
17. Viceconti M, Henney A, Morley-Fletcher E (2016) In silico clinical trials: how computer simulation will transform the biomedical industry. Int J Clin Trials 3(2):37. https://doi.org/10.18203/2349-3259.ijct20161408
18. Fuchs A, Csajka C, Thoma Y, Buclin T, Widmer N (2013) Benchmarking therapeutic drug monitoring software: a review of available computer tools. Clin Pharmacokinet 52(1):9–22. https://doi.org/10.1007/s40262-012-0020-y
19. Ingalls RG (2011) Introduction to simulation. In: Proceedings of the 2011 winter simulation conference (WSC), pp 1374–1388. https://doi.org/10.1109/WSC.2011.6147858
20. Hossain S, Kabedev A, Parrow A, Bergström CAS, Larsson P (2019) Molecular simulation as a computational pharmaceutics tool to predict drug solubility, solubilization processes and partitioning. Eur J Pharm Biopharm 137:46–55. https://doi.org/10.1016/j.ejpb.2019.02.007

21. Georgescu IM, Ashhab S, Nori F (2014) Quantum simulation. Rev Mod Phys 86(1):153. https://doi.org/10.1103/RevModPhys.86.153
22. Zhou T, Huang DCA (2010) Quantum mechanical methods for drug design. Curr Top Med Chem 1(10):33–45. https://doi.org/10.2174/156802610790232242
23. Shen L, Yang W (2018) Molecular dynamics simulations with quantum mechanics/molecular mechanics and adaptive neural networks. J Chem Theory Comput 14(3):1442–1455. https://doi.org/10.1021/acs.jctc.7b01195
24. Kim TH, Shin S, Shin BS (2018) Model-based drug development: application of modeling and simulation in drug development. J Pharm Investig 48:431–441. https://doi.org/10.1007/s40005-017-0371-3. Springer Netherlands
25. Zhu H (2020) Big data and artificial intelligence modeling for drug discovery. Annu Rev Pharmacol Toxicol 60:573–589. https://doi.org/10.1146/annurev-pharmtox-010919-
26. Fathima AJ, Murugaboopathi G (2018) A novel customized big data analytics framework for drug discovery. J Cyber Secur Mobil 7(1):145–160. http://www.riverpublishers.com/journal_read_html_article.php?j=JCSM/7/1/11
27. Fan J, Lin L, Wang CK (2017) Excited state properties of non-doped thermally activated delayed fluorescence emitters with aggregation-induced emission: a QM/MM study. J Mater Chem C 5(33):8390–8399. https://pubs.rsc.org/en/content/articlehtml/2017/tc/c7tc02541f
28. Jindal G, Warshel A (2016) Exploring the dependence of QM/MM calculations of enzyme catalysis on the size of the QM region. J Phys Chem B 120(37):9913–9921. https://pubs.acs.org/sharingguidelines
29. Levoin N, Calmels T, Krief S, Danvy D, Berrebi-Bertrand I, Lecomte JM et al (2011) Homology model versus X-ray structure in receptor-based drug design: a retrospective analysis with the dopamine D3 receptor. ACS Med Chem Lett 2(4):293–297. https://pubs.acs.org/doi/abs/10.1021/ml100288q
30. Jacobson KA, Costanzi S (2012) New insights for drug design from the X-ray crystallographic structures of G-protein-coupled receptors. Mol Pharmacol 82:361–371. http://molpharm.aspetjournals.org/content/82/3/361. American Society for Pharmacology and Experimental Therapeutics
31. Ferreira L, dos Santos R, Oliva G, Andricopulo A (2015) Molecular docking and structure-based drug design strategies. Molecules 20(7):13384–13421. http://www.mdpi.com/1420-3049/20/7/13384
32. Vucicevic J, Nikolic K, Mitchell JBO (2019) Rational drug design of antineoplastic agents using 3D-QSAR, cheminformatic, and virtual screening approaches. Curr Med Chem 26(21):3874–3889. https://pubmed.ncbi.nlm.nih.gov/28707592/
33. De Ruyck J, Brysbaert G, Blossey R, Lensink M (2016) Molecular docking as a popular tool in drug design, an in silico travel. Adv Appl Bioinforma Chem 9(1):1–11. https://www.dovepress.com/molecular-docking-as-a-popular-tool-in-drug-design-an-in-silico-travel-peer-reviewed-article-AABC
34. Güner OF, Bowen JP (2014) Setting the record straight: the origin of the pharmacophore concept. J Chem Inf Model 54:1269–1283. https://pubmed.ncbi.nlm.nih.gov/24745881/. American Chemical Society
35. Bhuvaneshwari S, Sankaranarayanan K (2019) Identification of potential CRAC channel inhibitors: pharmacophore mapping, 3D-QSAR modelling, and molecular docking approach. SAR QSAR Environ Res 30(2):81–108. https://www.tandfonline.com/doi/abs/10.1080/1062936X.2019.1566172
36. Yang SY (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug Discov Today 15:444–450. https://doi.org/10.1016/j.drudis.2010.03.013. Elsevier Current Trends
37. Van Der Graaf PH (2012) CPT: pharmacometrics and systems pharmacology, vol 1. Wiley-Blackwell, p e8. https://doi.org/10.1038/psp.2012.8

38. Clarelli F, Liang J, Martinecz A, Heiland I, Abel Zur Wiesch P (2020) Multi-scale modeling of drug binding kinetics to predict drug efficacy. Cell Mol Life Sci 77:381–394. https://doi.org/10.1007/s00018-019-03376-y. Springer
39. Zhang JD, Sach-Peltason L, Kramer C, Wang K, Ebeling M (2020) Multiscale modelling of drug mechanism and safety. Drug Discov Today 25:519–534. https://doi.org/10.1016/j.drudis.2019.12.009. Elsevier Ltd
40. Hwan T, Soyoung K, Beom S, Shin S (2017) Model-based drug development: application of modeling and simulation in drug development. J Pharm Investig 0:1–11. https://doi.org/10.1007/s40005-017-0371-3
41. de Lange EC, Bonate P (2019) Modeling and simulation in the translational pharmacology of CNS drugs. In: Handbook of behavioral neuroscience, 1st edn. Elsevier B.V., pp 25–37
42. Lavé T, Parrott N, Grimm HP, Fleury A, Reddy M (2007) Challenges and opportunities with modelling and simulation in drug discovery and drug development. Xenobiotica [Internet] 37(10–11):1295–1310. https://www.tandfonline.com/doi/abs/10.1080/00498250701534885
43. Thai H-T, Mazuir F, Cartot-Cotton S, Veyrat-Follet C (2015) Optimizing pharmacokinetic bridging studies in paediatric oncology using physiologically-based pharmacokinetic modelling: application to docetaxel. Br J Clin Pharmacol [Internet] 80(3):534–547. http://doi.wiley.com/10.1111/bcp.12702
44. Schmitt W, Willmann S (2004) Physiology-based pharmacokinetic modeling: ready to be used, vol 1. Drug discovery today: technologies. Elsevier Ltd, pp 449–456. https://doi.org/10.1016/j.ddtec.2004.09.006
45. Reddy MB, Clewell III HJ, Lave T, Melvin E (2013) Physiologically based pharmacokinetic modeling: a tool for understanding ADMET properties and extrapolating to human. In: New insights into toxicity and drug testing. InTech. https://doi.org/10.5772/54965
46. Durrant JD, McCammon JA (2011) BINANA: a novel algorithm for ligand-binding characterization. J Mol Graph Model 29(6):888–893. https://doi.org/10.1016/j.jmgm.2011.01.004
47. Ruiz-Herrero T, Estrada J, Guantes R, Miguez DG (2013) A tunable coarse-grained model for ligand-receptor interaction. PLoS Comput Biol 9(11):e1003274. https://doi.org/10.1371/journal.pcbi.1003274
48. Ng C, Xiao Y, Putnam W, Lum B, Tropsha A (2004) Quantitative structure-pharmacokinetic parameters relationships (QSPKR) analysis of antimicrobial agents in humans using simulated annealing k-nearest-neighbor and partial least-square analysis methods. J Pharm Sci 93(10):2535–2544. https://doi.org/10.1002/jps.20117
49. Zhivkova ZD (2017) Quantitative structure—pharmacokinetics relationships for plasma protein binding of basic drugs. J Pharm Pharm Sci 20(1):349–359. https://doi.org/10.18433/J33633
50. Van Der Graaf PHD, Nilsson J, Van Schaick EA, Danhof M (1999) Multivariate quantitative structure-pharmacokinetic relationships (QSPKR) analysis of adenosine A1 receptor agonists in rat. J Pharm Sci 88(3):306–312. https://doi.org/10.1021/js980294a
51. Zhuang X, Lu C (2016) PBPK modeling and simulation in drug research and development. Acta Pharm Sin B 6(5):430–440. https://doi.org/10.1016/j.apsb.2016.04.004
52. Mould DR, Upton RN (2013) Basic concepts in population modeling, simulation, and model-based drug development—part 2: introduction to pharmacokinetic modeling methods. CPT Pharmacometrics Syst Pharmacol 2(4):e38. https://doi.org/10.1038/psp.2013.14
53. Germani M, Del Bene F, Rocchetti M, Van Der Graaf PH (2013) A4S: a user-friendly graphical tool for pharmacokinetic and pharmacodynamic (PK/PD) simulation. Comput Methods Prog Biomed 110(2):203–214. https://doi.org/10.1016/j.cmpb.2012.10.006
54. Niu J, DE Straubinger RMM (2019) Pharmacodynamic drug–drug interactions. Clin Pharmacol Ther 6(106):1395–1406. https://doi.org/10.1002/cpt.1434
55. Alnajjar R, Mohamed N, Kawafi N (2020) Bicyclo[1.1.1]Pentane as phenyl substituent in atorvastatin drug to improve physicochemical properties: drug-likeness, DFT, pharmacokinetics, docking, and molecular dynamic simulation. J Mol Struct 1230:129628. https://doi.org/10.1016/j.molstruc.2020.129628

56. Suryawanshi S, Zhang L, Pfister M, Meibohm B (2010) The current role of model-based drug development. Expert Opin Drug Discov 5(4):311–321. https://doi.org/10.1517/17460441003713470
57. Zhang L, Pfister M, Meibohm B (2008) Concepts and challenges in quantitative pharmacology and model-based drug development. AAPS J 10(4):552–559. https://doi.org/10.1208/s12248-008-9062-3
58. Jones HM, Rowland-Yeo K (2013) Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. CPT Pharmacometrics Syst Pharmacol 2(8): e63. https://doi.org/10.1038/psp.2013.41
59. Chou WC, Lin Z (2019) Bayesian evaluation of a physiologically based pharmacokinetic (PBPK) model for perfluorooctane sulfonate (PFOS) to characterize the interspecies uncertainty between mice, rats, monkeys, and humans: development and performance verification. Environ Int 129:408–422. https://doi.org/10.1016/j.envint.2019.03.058
60. Thompson MD, Beard DA (2011) Development of appropriate equations for physiologically based pharmacokinetic modeling of permeability-limited and flow-limited transport. J Pharmacokinet Pharmacodyn 38(4):405–421. https://doi.org/10.1007/s10928-011-9200-x
61. Bolger MB, Macwan J, Lukacova V (2018) The influence of dissolution, PMAT influx, and MATE efflux rates on paracellular absorption of metformin using a mechanistic oral absorption / PBPK model. 2018 AAPS Workshop Drug Transporters in ADME: from the Bench to the Bedside, April 16–18, 2018, Herndon, VA. https://www.simulations-plus.com/resource/the-influence-of-dissolution-pmat-influx-and-mate-efflux-rates-onparacellular-absorption-of-metformin-using-a-mechanistic-oral-absorption-pbpk-model
62. Lobenberg R (2014) Modelling the absorption of metformin with patients post gastric bypass surgery. J Diabetes Metab 5:3. https://doi.org/10.4172/2155-6156.1000353
63. Nishiyama K, Toshimoto K, Lee W, Ishiguro N, Bister B, Sugiyama Y (2019) Physiologically-based pharmacokinetic modeling analysis for quantitative prediction of renal transporter–mediated interactions between metformin and cimetidine. CPT Pharmacometrics Syst Pharmacol 8(6):396–406. https://doi.org/10.1002/psp4.12398
64. Yun YE, Edginton AN (2019) Model qualification of the PK-Sim® pediatric module for pediatric exposure assessment of CYP450 metabolized compounds. J Toxicol Environ Health A 82(14):789–814. https://doi.org/10.1080/15287394.2019.1652215
65. Mao J, Johnson TR, Shen Z, Yamazaki S (2013) Prediction of crizotinib-midazolam interaction using the simcyp population-based simulator: comparison of CYP3A time-dependent inhibition between human liver microsomes versus hepatocytess. Drug Metab Dispos 41(2):343–352. https://doi.org/10.1124/dmd.112.049114
66. Dong J, Wang NN, Yao ZJ, Zhang L, Cheng Y, Ouyang D et al (2018) Admetlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. J Cheminform 10(1):29. https://doi.org/10.1186/s13321-018-0283-x
67. Elmokadem A, Riggs MM, Baron KT (2019) Quantitative systems pharmacology and physiologically-based pharmacokinetic modeling with mrgsolve: a hands-on tutorial. CPT Pharmacometrics Syst Pharmacol 8(12):883–893. https://doi.org/10.1002/psp4.12467
68. Cooper J, Cervenansky F, De Fabritiis G, Fenner J, Friboulet D, Giorgino T et al (2010) The virtual physiological human toolkit. Philos Trans R Soc A Math Phys Eng Sci 368(1925):3925–3936. https://doi.org/10.1098/rsta.2010.0144
69. Whiteley JP (2008) An efficient technique for the numerical solution of the bidomain equations. Ann Biomed Eng 36(8):1398–1408. https://doi.org/10.1007/s10439-008-9513-0
70. Kerfoot E, Lamata P, Niederer S, Hose R, Spaan J, Smith N (2013) Share and enjoy: anatomical models database—generating and sharing cardiovascular model data using web services. Med Biol Eng Comput 51(11):1181–1190. https://doi.org/10.1007/s11517-012-1023-4
71. Smith N, de Vecchi A, McCormick M, Nordsletten D, Camara O, Frangi AF et al (2011) euHeart: personalized and integrated cardiac care using patient-specific cardiovascular modelling. Interface Focus 1(3):349–364. https://doi.org/10.1098/rsfs.2010.0048

72. Highly accurate breast cancer diagnosis through integration of biological knowledge, novel imaging modalities, and modelling | HAMAM Project | FP7 | CORDIS | European Commission
73. HAMAM: successful outcome for project to improve breast cancer diagnosis through integrated clinical workstation. https://www.vph-institute.org/news/hamam-successful-outcome-for-project-to-improve-breast-cancer-diagnosis-through-integrated-clinical-.html
74. Zhang Y, Barocas VH, Berceli SA, Clancy CE, Eckmann DM, Garbey M et al (2016) Multi-scale modeling of the cardiovascular system: disease development, progression, and clinical intervention HHS public access author manuscript. Ann Biomed Eng 44(9):2642–2660. https://doi.org/10.1007/s10439-016-1628-0
75. Trayanova NA (2011) Whole-heart modeling: applications to cardiac electrophysiology and electromechanics. Circ Res 108:113–128. https://doi.org/10.1161/circresaha.110.223610. NIH Public Access
76. Trayanova N, Eason J, Aguel F (2002) Computer simulations of cardiac defibrillation: a look inside the heart. Comput Vis Sci 4(4):259–270. https://doi.org/10.1007/s00791-002-0082-8
77. Krishnamoorthi S, Perotti LE, Borgstrom NP, Ajijola OA, Frid A, Ponnaluri AV et al (2014) Simulation methods and validation criteria for modeling cardiac ventricular electrophysiology. PLoS One 9(12):e114494. https://doi.org/10.1371/JOURNAL.PONE.0114494
78. Lopez-Perez A, Sebastian R, Izquierdo M, Ruiz R, Bishop M, Ferrero JM (2019) Personalized cardiac computational models: from clinical data to simulation of infarct-related ventricular tachycardia. Front Physiol 10:580. https://doi.org/10.3389/fphys.2019.00580
79. Mcculloch AD (2016) Systems biophysics: multiscale biophysical modeling of organ systems. Biophys J 110:1023–1027. https://doi.org/10.1016/j.bpj.2016.02.007
80. Hirashima T, Rens EG, Merks RMH (2017) Cellular Potts modeling of complex multicellular behaviors in tissue morphogenesis. Develop Growth Differ 59(5):329–339. https://doi.org/10.1111/dgd.12358
81. Nematbakhsh A, Sun W, Brodskiy P, Amiri A, Narciso C, Xu Z et al (2016) Multi-scale computational study of the mechanical regulation of cell mitotic rounding in epithelia. bioRxiv 037820. https://doi.org/10.1101/037820
82. Setty Y, Cohen IR, Dor Y, Harel D (2008) Four-dimensional realistic modeling of pancreatic organogenesis. Proc Natl Acad Sci U S A 105(51):20374–20379. https://doi.org/10.1073/pnas.0808725105
83. Pasteka R, Forjan M, Sauermann S, Drauschke A (2019) Electro-mechanical lung simulator using polymer and organic human lung equivalents for realistic breathing simulation. Sci Rep 9(1):1–12. https://doi.org/10.1038/s41598-019-56176-6
84. (PDF) Modelling and simulation of the kidney. https://hal.archives-ouvertes.fr/hal-00339612
85. Šilar J, Ježek F, Mládek A, Polák D, Kofránek J. Model visualization for e-learning, Kidney simulator for medical students. https://doi.org/10.3384/ecp19157393
86. Postnov DD, Marsh DJ, Postnov DE, Braunstein TH, Holstein-Rathlou N-H, Martens EA et al (2016) Modeling of kidney hemodynamics: probability-based topology of an arterial network. Beard DA, editor. PLoS Comput Biol 12(7):e1004922. https://doi.org/10.1371/journal.pcbi.1004922
87. König M. Executable simulation model of the liver running title: executable simulation model of the liver. bioRxiv. https://doi.org/10.1101/2020.01.04.894873
88. Watkins PB (2019) The DILI-sim initiative: insights into hepatotoxicity mechanisms and biomarker interpretation. Clin Transl Sci 12(2):122–129. https://doi.org/10.1111/cts.12629
89. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA et al (2012) The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2(5):401–404. https://doi.org/10.1158/2159-8290.CD-12-0095
90. Swat MH, Thomas GL, Belmonte JM, Shirinifard A, Hmeljak D, Glazier JA (2012) Multi-scale modeling of tissues using CompuCell3D. In: Methods in cell biology. Academic, pp 325–366. https://doi.org/10.1016/B978-0-12-388403-9.00013-8

91. Clatz O, Sermesant M, Bondiau PY, Delingette H, Warfield SK, Malandain G et al (2005) Realistic simulation of the 3-D growth of brain tumors in MR images coupling diffusion with biomechanical deformation. IEEE Trans Med Imaging 24(10):1334–1346. https://doi.org/10.1109/TMI.2005.857217
92. Delile J, Herrmann M, Peyriéras N, Doursat R (2017) A cell-based computational model of early embryogenesis coupling mechanical behaviour and gene regulation. Nat Commun 8(1):1–10. https://doi.org/10.1038/ncomms13929
93. Yang Y, Nadanaciva S, Will Y, Woodhead JL, Howell BA, Watkins PB et al (2015) MITOsym®: a mechanistic, mathematical model of hepatocellular respiration and bioenergetics. Pharm Res 32(6):1975–1992. https://doi.org/10.1007/s11095-014-1591-0
94. Tam ZY, Gruber J, Halliwell B, Gunawan R (2013) Mathematical modeling of the role of mitochondrial fusion and fission in mitochondrial DNA maintenance. Stekel DJ, editor. PLoS One 8(10):e76230. https://doi.org/10.1371/journal.pone.0076230
95. Hoffman TE, Barnett KJ, Wallis L, Hanneman WH (2017) A multimethod computational simulation approach for investigating mitochondrial dynamics and dysfunction in degenerative aging. Aging Cell 16(6):1244–1255. https://doi.org/10.1111/acel.12644
96. Hatano A, Okada JI, Washio T, Hisada T, Sugiura S (2011) A three-dimensional simulation model of cardiomyocyte integrating excitation-contraction coupling and metabolism. Biophys J 101(11):2601–2610. https://doi.org/10.1016/j.bpj.2011.10.020
97. Hatano A, Okada J, Hisada T, Sugiura S (2012) Critical role of cardiac t-tubule system for the maintenance of contractile function revealed by a 3D integrated model of cardiomyocytes. J Biomech 45(5):815–823. https://doi.org/10.1016/j.jbiomech.2011.11.022
98. Hatano A, Okada JI, Washio T, Hisada T, Sugiura S (2013) Mitochondrial colocalization with Ca2+ release sites is crucial to cardiac metabolism. Biophys J 104(2):496–504. https://doi.org/10.1016/j.bpj.2012.12.004
99. Hatano A, Okada J-I, Washio T, Hisada T, Sugiura S (2015) Article distinct functional roles of cardiac mitochondrial subpopulations revealed by a 3D simulation model. https://doi.org/10.1016/j.bpj.2015.04.031
100. Sakata D, Lampe N, Karamitros M, Kyriakou I, Belov O, Bernal MA et al (2019) Evaluation of early radiation DNA damage in a fractal cell nucleus model using Geant4-DNA. Phys Med 62:152–157. https://doi.org/10.1016/j.ejmp.2019.04.010
101. Moshkforoush A, Ashenagar B, Tsoukias NM, Alevriadou BR (2019) Modeling the role of endoplasmic reticulum-mitochondria microdomains in calcium dynamics. Sci Rep 9(1):1–16. https://doi.org/10.1038/s41598-019-53440-7
102. Siekmann I, Wagner LE, Yule D, Crampin EJ, Sneyd J (2012) A kinetic model for type I and II IP3R accounting for mode changes. Biophys J 103(4):658–668. https://doi.org/10.1016/j.bpj.2012.07.016
103. Computational systems biology: Chapter 16. Computational modeling and ...—Julien Delile, René Doursat, Nadine Peyriéras—Google Books
104. Marée AFM, Grieneisen VA, Hogeweg P (2007) The cellular Potts model and biophysical properties of cells, tissues and morphogenesis. In: Single-cell-based models in biology and medicine. Birkhäuser Basel, pp 107–136. https://doi.org/10.1007/978-3-7643-8123-3_5
105. Sundararaj S, Guo A, Habibi-Nazhad B, Rouani M, Stothard P, Ellison M et al The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of Escherichia coli. https://doi.org/10.1093/nar/gkh108
106. Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, Miyoshi F et al (1999) E-CELL: software environment for whole-cell simulation. Bioinformatics 15(1):72–84. https://doi.org/10.1093/bioinformatics/15.1.72
107. Feig M, Sugita Y (2019) Annual review of cell and developmental biology whole-cell models and simulations in molecular detail. Annu Rev Cell Dev Biol 35:191–211. https://doi.org/10.1146/annurev-cellbio-100617
108. Guisoni N, Mazzitello KI, Diambra L (2018) Modeling active cell movement with the Potts model. Front Phys 6:61. https://doi.org/10.3389/fphy.2018.00061

109. Pasteka R, Forjan M, Sauermann S, Drauschke A (2019) Electro-mechanical lung simulator using polymer and organic human lung equivalents for realistic breathing simulation. Sci Rep 9(1):19778. https://doi.org/10.1038/s41598-019-56176-6
110. Šilar J, Ježek F, Mládek A, Polák D, Kofránek J (2019) Model visualization for e-learning, Kidney simulator for medical students. In: Proceedings of the 13th international modelica conference, Regensburg, Germany, 4–6 Mar 2019, Linköing University Electronic Press, pp 393–402. https://doi.org/10.3384/ecp19157393
111. König M (2020) Executable simulation model of the liver. bioRxiv. https://doi.org/10.1101/2020.01.04.894873
112. Woodhead JL, Pellegrini L, Shoda LKM, Howell BA. Comparison of the hepatotoxic potential of two treatments for autosomal-dominant polycystic kidney disease using quantitative systems toxicology modeling. https://doi.org/10.1007/s11095-019-2726-0
113. cBioPortal: a web platform of gene-based data exploration | Office of Cancer Genomics
114. Dupuy L, MacKenzie J, Rudge T, Haseloff J (2008) A system for modelling cell-cell interactions during plant morphogenesis. Ann Bot 101(8):1255–1265. https://doi.org/10.1093/aob/mcm235
115. Gradeci D, Bove A, Charras G, Lowe AR, Banerjee S. Single-cell approaches to cell competition: high-throughput imaging, machine learning and simulations. https://doi.org/10.1016/j.semcancer.2019.05.007
116. Marée AF, Hogeweg P (2002) Modelling *Dictyostelium discoideum* morphogenesis: the culmination. Bull Math Biol 64(2):327–353. https://doi.org/10.1006/bulm.2001.0277
117. Popławski NJ, Swat M, Scott Gens J, Glazier JA (2007) Adhesion between cells, diffusion of growth factors, and elasticity of the AER produce the paddle shape of the chick limb. Phys A Stat Mech Appl 373:521–532. https://doi.org/10.1016/j.physa.2006.05.028
118. Rens EG, Edelstein-Keshet L (2019) From energy to cellular forces in the Cellular Potts Model: an algorithmic approach. PLoS Comput Biol 15(12):e1007459. https://doi.org/10.1371/journal.pcbi.1007459
119. Mirams GR, Arthurs CJ, Bernabeu MO, Bordas R, Cooper J, Corrias A et al (2013) Chaste: an open source C++ library for computational physiology and biology. Prlic A, editor. PLoS Comput Biol 9(3):e1002970. https://doi.org/10.1371/journal.pcbi.1002970
120. Chaste—cancer, heart and soft tissue environment
121. Shen L, Shen J, Luo X, Cheng F, Xu Y, Chen K et al (2003) Steered molecular dynamics simulation on the binding of NNRTI to HIV-1 RT. Biophys J 84(6):3547–3563. https://doi.org/10.1016/S0006-3495(03)75088-7
122. Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K (1997) Molecular dynamics study of unbinding of the avidin-biotin complex. Biophys J 72:1568–1581. https://doi.org/10.1016/S0006-3495(97)78804-0
123. Protein modeling and molecular dynamics simulation of cloned regucalcin (RGN) gene from Bubalus bubalis | Bentham Science. https://doi.org/10.2174/1386207319666161220124532
124. Sorenson JM, Head-Gordon T (2000) Matching simulation and experiment: a new simplified model for simulating protein folding. J Comput Biol 7(3–4):469–481. https://doi.org/10.1089/106652700750050899
125. Pillai H, Yadav BS, Chaturvedi N, Jan AT, Gupta GK, Baig MH et al (2017) Protein modeling and molecular dynamics simulation of cloned regucalcin (RGN) gene from *Bubalus bubalis*. Comb Chem High Throughput Screen 20(3). https://doi.org/10.2174/1386207319666161220124532
126. Lutz B, Sinner C, Bozic S, Kondov I, Schug A (2014) Native structure-based modeling and simulation of biomolecular systems per mouse click. BMC Bioinformatics 15(1):292. https://doi.org/10.1186/1471-2105-15-292
127. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP et al (2010) Atomic-level characterization of the structural dynamics of proteins. Science 330(6002):341–346. https://doi.org/10.1126/science.1187409

128. Pá Jaro M, Otero-Muras I, Vá Zquez C, Alonso AA, Stegle O. SELANSI: a toolbox for simulation of stochastic gene regulatory networks. https://doi.org/10.1093/bioinformatics/btx645
129. Cloete R, Kapp E, Joubert J, Christoffels A, Malan SF (2018) Molecular modelling and simulation studies of the *Mycobacterium tuberculosis* multidrug efflux pump protein Rv1258c. Soares CM, editor. PLoS One 13(11):e0207605. https://doi.org/10.1371/journal.pone.0207605
130. Nunes Lima R, Faheem M, Alexandre J, Gonçalves Barbosa R, Polêto MD, Verli H et al (2016) Homology modeling and molecular dynamics provide structural insights into tospovirus nucleoprotein. BMC Bioinformatics. https://doi.org/10.1186/s12859-016-1339-4
131. Daddam JR, Sreenivasulu B, Peddanna K, Umamahesh K (2020) Designing, docking and molecular dynamics simulation studies of novel cloperastine analogues as anti-allergic agents: homology modeling and active site prediction for the human histamine H1 receptor. RSC Adv 10(8):4745–4754. https://doi.org/10.1039/c9ra09245e
132. Cheung NJ, Yu W (2018) De novo protein structure prediction using ultra-fast molecular dynamics simulation. bioRxiv 262188. https://doi.org/10.1101/262188
133. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B et al (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001
134. Block M (2015) Physiologically based pharmacokinetic and pharmacodynamic modeling in cancer drug development: status, potential and gaps. Expert Opin Drug Metab Toxicol 11:743–756. https://doi.org/10.1517/17425255.2015.1037276. Informa Healthcare
135. Van Den Berg JP, Vereecke HEM, Proost JH, Eleveld DJ, Wietasch JKG, Absalom AR et al (2017) Pharmacokinetic and pharmacodynamic interactions in anaesthesia. A review of current knowledge and how it can be used to optimize anaesthetic drug administration. Br J Anaesth 118:44–57. https://doi.org/10.1093/bja/aew312. Oxford University Press
136. Roberts JK, Stockmann C, Balch A, Yu T, Ward RM, Spigarelli MG et al (2015) Optimal design in pediatric pharmacokinetic and pharmacodynamic clinical studies. Anderson B, editor. Pediatr Anesth 25(3):222–230. https://doi.org/10.1111/pan.12575

Physiologically Based Pharmacokinetic (PBPK) Modelling

9

Ankit Balhara, Sumeet Kale, and Saranjit Singh

Contents

A. Balhara · S. Kale · S. Singh (✉)
National Institute of Pharmaceutical Education and Research (NIPER), S.A.S. Nagar, Punjab, India
e-mail: ssingh@niper.ac.in

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_9

Abstract

Since its introduction in 1937 by Teorell, physiologically based pharmacokinetic (PBPK) modelling has come to a point, that apart from becoming an integral part of drug discovery and development process, it has also gained worldwide acceptance by drug regulatory bodies. PBPK models correlate drug properties with information of the physiology and biology of a species, in order to achieve a mechanistic representation of the drug in biological systems. The prediction involves consideration of different physiological organs of the body (referred to as compartments) that are interconnected via blood circulation. The characteristic volume and blood flow for each compartment is considered, and further, mass balance differential equations are used to describe the fate of the drug in different compartments. The model generates data in the form of concentration-time curves, which are then utilized to generate PK parameters (e.g. clearance, half-life, area under the curve, bioavailability, etc.) for the drug. The results of simulation can be then implemented for the intended purpose. This book chapter delves into the history, regulatory aspects and various components of PBPK, along with general workflow and approaches for model development, along with variety of salient applications of PBPK modelling.

Chapter Objectives

- To discuss the concept of PBPK modelling and its historical aspects
- To provide overview of regulatory considerations of PBPK modelling
- To review the major applications of PBPK modelling
- To elaborate the various components and process of building of a PBPK model
- To introduce the various softwares employed for PBPK modelling

9.1 Introduction

Prediction of human pharmacokinetics (PK) at early stages of the drug development can help in decision making related to dose selection, clinical trial design and development progression. At the same time, liabilities related to PK such as high

clearance, low bioavailability, drug-drug interaction (DDI) potential or the requirement of dose adjustments in special populations can also be identified [1, 2]. For this purpose, various methods have been devised, which range in complexity from simple static mechanistic models for predicting specific PK parameters to the dynamic physiologically based PK (PBPK) models that help in prediction of plasma concentration-time profiles [3].

Static models work on the steady-state assumptions and hence are unable to predict time-dependent changes in inhibition of enzymes and transporters, tissue distribution kinetics and overall plasma concentration-time profiles. These models are quite simple and capable of predicting the human metabolic and transporter-mediated clearance and DDIs [4–6]. On the other hand, dynamic PBPK models allow the prediction of kinetics of drug absorption, distribution and elimination processes along with simulation of time-dependent changes in the concentration of a drug and/or its metabolite(s) in plasma or an organ of interest [7, 8]. Structurally, PBPK model is a mathematical framework comprising of a large number of compartments corresponding to different organs or tissues in the body, which are parameterized using known physiology and biology. Various compartments are connected by flow rates that parallel the circulating blood system. This structure of model, in combination with information on the drug, helps to achieve a mechanistic representation of the drug in biological system, allowing *a priori* simulation of drug concentration-time profiles [9, 10].

But owing to its complex structure and requirement of detailed drug and system physiology data, PBPK modelling is a labour-intensive exercise. Yet, its advantages are numerous. The PBPK modelling has become an indispensable tool for the: (1) selection of optimal sampling times or dosing strategies in different populations; (2) designing of DDI studies where concentrations of perpetrator [drug which affects the PK of interacting drug (victim drug)] are fluctuating over sampling and dosing intervals; (3) prediction of drug bioavailability; (4) mechanistic evaluation of inter-individual variabilities in the PK of a given drug; (5) study of the impact of drug properties on absorption kinetics and intestinal interactions; (6) simulation of the food effects; (7) simulation of the drug disposition in special populations including hepatic and renal impaired patients, children, pregnant women, etc. and (8) pharmacokinetic-pharmacodynamic (PK/PD) simulations through linking of the simulated PK with the PD endpoints [11–15].

Last few years have seen a significant increase in the number of publications and regulatory submissions involving PBPK modelling [3, 16]. Also, regulatory agencies such as United States Food and Drug Administration (US FDA) and European Medicines Agency (EMA), have included PBPK modelling in their multiple guidance documents like those involving hepatic impairment [17], paediatric [18], DDIs [19–21] and pharmacogenetics [22, 23] to guide clinical study design and labelling decisions. PBPK models are also being used in investigating new drugs and hence in new drug applications (NDAs) [16]. Recent boost to PBPK modelling has been provided by the availability of various user friendly commercial softwares such as Simcyp, GastroPlus and PK-Sim, which has made this predictive science accessible even to those without extensive modelling and/or programming experience [7].

The history, regulatory considerations, applications and components of PBPK models are discussed below. Also covered are general approaches and workflow of PBPK model development. A separate section covers features of the commercial softwares.

9.2 History of PBPK Modelling

The concept of PBPK modelling is not entirely new. In 1937, Teorell introduced for the very first time a multi-compartmental model, having biological and physiological components for the simulation of PK data. The model contained five compartments representing the circulatory system, a drug depot, fluid volume, kidney elimination and tissue inactivation and it considered physiological volumes of each [24]. The next historically significant contribution was made by Bellman et al. in 1960, who segregated the normal tissue into intercellular, intracellular and capillary regions. They proposed that diffusional processes were responsible for the spatial concentration differences observed across various regions of the tissue. Partial differential equations were employed to define the mass transport between various tissue regions [25].

The realization based on above studies that it was possible to develop models portraying biological systems realistically, resulted in advancement to a comprehensive model for drug distribution, which was proposed by Bischoff and Brown in 1966. They arranged various body parts and organs in a flow network and their model incorporated physiologic parameters such as volume of the tissues or organs and the blood flow rates through each. Some of the body regions such as head and upper extremities were lumped together, while organs such as liver and kidney were taken separately owing to their unique roles. A series of ordinary differential equations, instead of partial differential equations, was employed to describe concentrations in various organs, including the mass transfer processes [26].

There was slow and steady progress thereafter in PBPK modelling and most of the applications were limited to environmental toxicants or pollutants. The reasons were insurmountable ethical constraints to generate data on these in the human population. So the only approach was to predict the outcome of the exposure of toxicants to humans, followed at the best by preclinical studies. The major reason for minimal use of PBPK modelling in the pharmaceutical domain was the mathematical complexity of the models and the perceived demand for a large number of parameters as inputs. Many of these parameters were not even established and hence many gaps existed in the knowledge space [27].

However, in the last two decades, one finds a tremendous growth in PBPK modelling science and its application in the academia and industry. The three primary reasons for the exponential rise seemingly are: (1) advancements in computer sciences, as modern computers are able to handle the vast number of calculations quite efficiently and rapidly, which these models require; (2) advancements in the field of in vitro assays of human enzymes, cells and tissues, wherein results allow prediction of the behaviour of drugs at the organ level and

(3) the deeper and mechanistic understanding of physiological, biochemical and pharmaceutical parameters that can affect the PK of drugs. Also, the combination of in vitro-in vivo extrapolation (IVIVE) of absorption, distribution, metabolism and excretion (ADME) with PBPK models has greatly widened the scope of latter [7, 28].

9.3 Regulatory Considerations

The late-stage failures in drug discovery and development programmes cause high attrition rates, which is a major challenge faced by the pharmaceutical industry. Less than 10% of the new compounds that enter clinical trials ultimately make it to the market. So to cut short the ever-increasing attrition rates, introduction of innovative methods in drug development and evaluation is a must. Currently, both regulators and industry believe that modelling and simulation (M&S) can increase the robustness of regulatory as well as industrial decision making [9, 16].

The first known case of a review dossier submitted to the US FDA involving PBPK can be traced back to 1990s. In this dossier, application of PBPK simulation was made to establish the potential risk of foetal exposure and birth defects from tretinoin, a highly teratogenic active ingredient of a topical anti-wrinkle cream. PBPK study results suggested that overall risk was minimal [29]. The faith of regulatory agencies in PBPK modelling increased slowly, and in the last decade there has been significant acceptance and realization that predictive results through PBPK modelling can be really helpful in certain situations, especially where no other means exist to generate the required data. Examples of such situations include labelling instructions and dose selection in renal and hepatic impaired patients and other categories of special populations, including pregnancy, paediatrics, etc., [30–34]. Several guidelines have been issued by the US FDA and EMA, which suggest PBPK modelling as a means to guide clinical study design and labelling decisions. The list includes regulatory guidance documents on hepatic impairment [17], paediatrics [18], DDIs [19–21] and pharmacogenetics [22, 23]. Accordingly, PBPK modelling is increasingly being used in investigation of new drugs and hence in NDAs by the industry.

As per the FDA 2019 report, a total of 254 submissions including 94 NDAs involving applications of PBPK modelling were reviewed by the Office of Clinical Pharmacology (OCP) from 2008 to 2017 [35]. The distribution of application areas of PBPK is shown in Fig. 9.1.

The per year count of NDAs, containing predictions from PBPK modelling is separately shown in Fig. 9.2.

Apart from its role in new drug discovery and development, PBPK is also emerging as an important regulatory tool for the development of generic drug policies and standards. The primary areas, where the usage of PBPK models is found in the generic drugs review section of the FDA, include: (1) assessment of locally acting gastrointestinal drugs and those administered by topical and pulmonary routes; (2) assessment of the equivalence of products containing drugs that

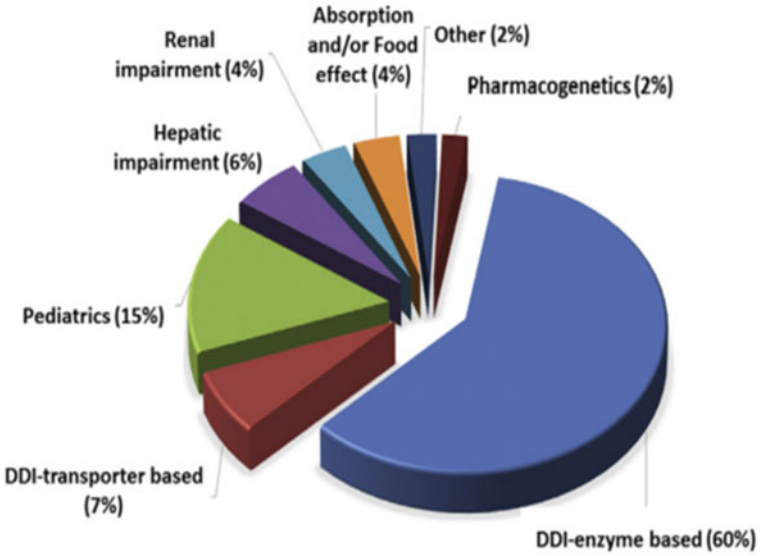


Fig. 9.1 Distribution of application areas of PBPK modelling and simulation in IND/NDA submissions reviewed by the US FDA's Office of Clinical Pharmacology from 2008 to 2017. Adapted from Grimstein et al. [35] with permission

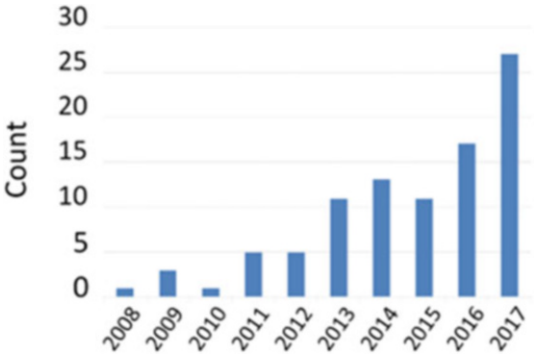


Fig. 9.2 Number of NDAs per year involving PBPK modelling. Adapted from FDA website [36]

exhibit multiple concentration peaks; (3) biowaivers for Biopharmaceutical Classification System Class II and III drugs; (4) improved in vitro-in vivo correlation (IVIVC) and (5) evaluation of formulations involving liposomes, nanotechnology and multiple-component mixtures [7].

Both US FDA [37] and EMA [38] have also issued guidelines on how to detail protocols and results of PBPK modelling in registration dossiers. A white paper was also published in 2018 to guide the PBPK model qualification and reporting procedures for regulatory submissions [39].

9.4 Applications of PBPK Modelling

The spectrum of application of PBPK modelling is very broad. But only the most prominent ones are being discussed below, initially by stages of drug development and later by specific application categories.

9.4.1 Early Stage Drug Development

At early stages of drug development, the target is to generate minimal data on physicochemical and biological properties of a candidate using in silico and in vitro tools. These data are then employed to run simulations for the purpose to prioritize compounds for in vivo experimentation, as the process of determining in vivo characteristics for new compounds is costly and time consuming. Further, the PK data generated in a preclinical species help to develop a model that can be used to predict the PK in humans. Even the developed model can be used to guide further experimentation to arrive at the desired information more rapidly. For example, if a PBPK model is not able to adequately describe animal PK data, this means that a biological phenomenon affecting PK has not been included in the model and is not represented by the assays used to screen the compounds. So, PBPK modelling allows one to determine the possible mechanisms that are consistent with the observed data [40, 41]. This utility of PBPK modelling is well highlighted by Peters [42], who proposed a 'line-shape analysis' method, in which a mismatch between simulated and observed oral profiles was assessed to gain mechanistic insights into processes impacting absorption and PK (e.g. saturable metabolism, enterohepatic cycling, transporter involvement in absorption from the gut and regional variation in gut absorption). PBPK modelling is quite useful in the selection of a clinical candidate in the scenarios, where multiple factors have to be considered, and PK and PD data for the compound need to be combined and compared in a rational way. Parrott et al. [41] used the PBPK model to select the best among the five candidates as the clinical lead. They integrated the preclinical data of each to estimate the efficacious human doses and associated plasma exposures. Thereafter, PBPK models were linked to E_{max} PD model so as to identify the dose that would yield 90% effect.

9.4.2 Cross-Species Extrapolation and First-In-Human (FIH) Dose Predictions

The uncertain comparability of physiological processes among the various laboratory animals makes cross-species extrapolation a challenging task. The provision of incorporating inter-species physiological and biochemical differences in PBPK modelling allows translation of mechanistic knowledge from one species to another. Specifically, the consideration of species-specific physiology, plasma protein binding, enzyme and transport kinetics, as well as tissue-specific gene expression profiles in PBPK modelling greatly improves the accuracy of cross-species extrapolations,

and thus supports first-in-human (FIH) trials based on prior preclinical knowledge [43, 44].

9.4.3 Formulation Development and Optimization

Once the detailed information about the pharmaceutical and biological properties (e.g. solubility, dissolution, precipitation, membrane permeation, transport and metabolic kinetics) of a drug and its PK data are available, PBPK modelling can be used for the identification and verification of the factors impacting oral absorption and to guide formulation development. However, it is important to emphasize that typical physiological conditions should be considered appropriately for a given absorption simulation. At later stages of drug development, absorption modelling can be very useful, for example, for predicting the food effect, effect of formulation variables such as particle size, release profiles, dosage forms, etc. [45, 46].

9.4.4 Prediction of Drug-Drug Interactions

Serious and unmanageable DDIs are one of the most common causes of drug withdrawal from the market. So their prediction at the drug development stage can save a lot of time, money and resources [7]. The widest application (almost 67%) of PBPK modelling in submissions to the US FDA (Sect. 9.3, Fig. 9.1) is the prediction of DDI potential of a particular drug [35]. PBPK models allow 'dynamic' evaluation of the impact of DDI with respect to full concentration-time profiles of the interacting compounds. The time-dependent changes in enzyme abundance and aspects of experimental design such as dosage, choice of dosage form and the timing of dosing of interacting compounds can be easily evaluated. Also, PBPK modelling can help in assessment of the more complex scenarios involving: (1) simultaneous dose-dependent inhibition and induction, (2) competition for plasma binding, (3) inhibitory effects of both the parent drug and the metabolites, (4) lack of adherence and (5) multiple DDIs [5, 47, 48]. Additionally, the possible extremes in a population such as the genetic absence of a functional enzyme (other than the one being inhibited) or hepatic/renal impairment can also be encompassed in model development to study their influence on the outcome of a DDI study [49].

9.4.5 Prediction of the Effect of Age

Physiologically, children are not small adults. Even amongst them, there are heterogeneous groups with marked physiological differences starting from birth up to the age of 14. The variation comes from developing enzyme systems and clearance mechanisms. As a result, PK of xenobiotics differs widely between children and adults. Also, this makes adult-based dosimetric extrapolation uncertain, especially at early postnatal ages [50, 51]. The regulatory bodies still need the information [18],

but the conduct of clinical studies on children are bound by strong ethical concerns. In such a scenario, IVIVE and PBPK modelling are proving as indispensable tools to anticipate the PK differences in paediatric population relative to adults. Also, modelling can assist in the selection and optimal design of in vivo investigations [52]. The same can also be used for precise paediatric dose selection owing to possibility to account for physiological and biochemical changes associated with age [53]. The, coupling with IVIVE enables prediction of drug clearance and its variability in neonates, infants and children with acceptable accuracy, along with the prediction of full concentration-time PK profiles [52, 54].

On similar lines, the changes in physiological and biochemical parameters in geriatrics such as reduction in liver mass, microsomal protein, hepatocellularity, number of functional nephrons, etc., result in decreased renal and hepatic clearance of the drugs [53]. The variability in age-dependent parameters even in this case can be incorporated in existing PBPK models, thus allowing prediction of PK of drugs and hence appropriate dosing in ageing population [55].

9.4.6 Prediction of Genetic Effects

Genetic polymorphisms in drug-metabolizing enzymes and their frequencies vary across different ethnic groups, leading to possibility of differential drug exposure and hence the response. PBPK modelling allows assessment of the impact of polymorphism in drug metabolizing enzymes and transporters on the PK of drugs across different populations [56]. For example, the influence of polymorphism in CYP2D6 on the PK of paroxetine was studied by Jornil et al. [57] using a population-based simulator.

9.4.7 Prediction of the Effects of Disease

The prediction of PK of a drug in a particular diseased condition is possible with PBPK modelling, which is done by incorporating pathological features, along with etiology and severity of the disease. The major diseased states considered within PBPK modelling are hepatic and renal impairment [7]. Reports in literature highlight that modelling was successful in yielding accurate prediction of intravenous (i.v.) clearances in hepatic impairment patients for multiple drugs viz., midazolam, theophylline, metoprolol and omeprazole [58], and the oral clearances of midazolam, caffeine, theophylline, metoprolol, nifedipine, quinidine, diclofenac, sildenafil and omeprazole [59]. The changes incorporated were with respect to hepatic blood flow, Cytochromes P450 (CYPs) enzyme abundance, liver volume, hematocrit and renal functions [58, 59]. The prediction of PK through PBPK models in subjects with impaired hepatic function is also recommended in current regulatory guidelines [17]. Similarly, PBPK modelling has been shown to be particularly helpful in PK and dosing predictions in renal impaired population [60]. The other diseased

conditions that are currently being investigated by PBPK models include inflammatory disease [61], obesity [62], tuberculosis (TB) [63] and Parkinsonism [64].

9.4.8 Assessment of the Food Effect

According to the US FDA regulatory guideline entitled 'Food-Effect Bioavailability and Bioequivalence Studies', the food effect bioavailability study should be conducted for all new chemical entities during the investigational new drug (IND) application period. Also, the abbreviated new drug applications (ANDAs) are recommended to have a fed bioequivalence study data [65]. There are multiple ways to study the food effect, but in recent years PBPK modelling has been used widely to evaluate oral drug absorption, including the food effect. There are several associated advantages behind the popularity of modelling for the purpose, which include simulation of dynamic PK profiles under fasted and fed stages (to investigate variability induced by the intake of food), and potential to integrate the changes prompted by food on human physiology as well as drug and its delivery system [66]. According to a survey of the literature reports and NDA filings, it was observed that PBPK models were able to predict the food effect within acceptance criterion in 75% cases [67].

9.4.9 Other Applications

Innovative applications of PBPK appearing in the literature include introducing time-varying physiology into paediatric PBPK models [68]; modelling drug disposition in lung [69], brain [70] and kidney [71]; virtual bioequivalence studies [72]; modelling of antibody drug conjugates (ADC) [73], etc. The connection and interaction of other tools/platforms such as quantitative system pharmacology (QSP) models of various disease progressions can further expand the applications of PBPK models [3]. In near future, it may be possible that PBPK models will be used in implementing precision dosing at the point of care by incorporating characteristics of patients. Also, integration of information relevant to the biomarkers for enzyme/transporters activities within PBPK models enables determination and optimization of doses [29].

Major applications of PBPK modelling are summarized in Fig. 9.3.

9.5 Components of a PBPK Model

Generally, a whole body PBPK model comprises of the mainly four components viz., PBPK model basic structure, system-related input parameters, drug-specific input parameters and other miscellaneous properties, each one of them are discussed in detail below.

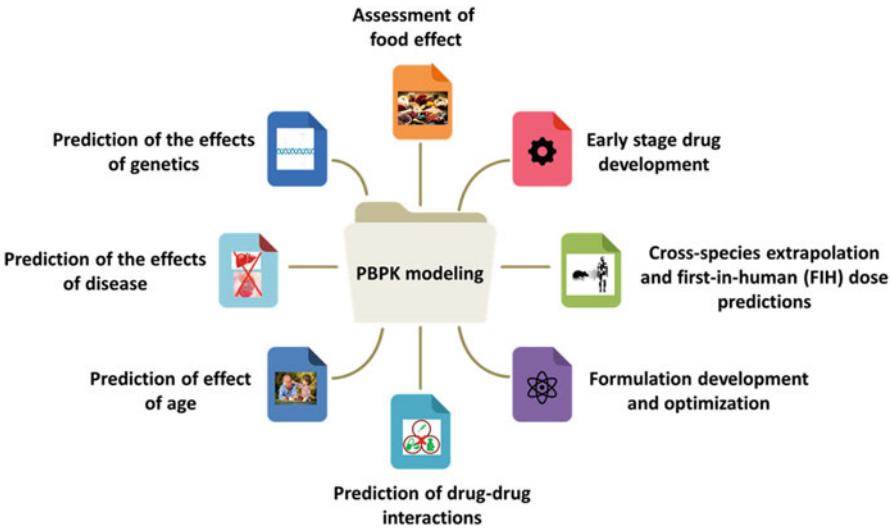


Fig. 9.3 Major applications of PBPK modelling

9.5.1 PBPK Model Basic Structure

Structurally, PBPK models consist of compartments corresponding to different tissues in the body connected by circulating blood system (Fig 9.4). Each compartment is defined by species-specific tissue volume (or weight) and tissue blood flow rate. Main tissues/organs of the body such as adipose, bone, brain, gut, heart, kidney, liver, lung, muscle, skin and spleen are considered as separate compartments [1, 9]. However, complexity in the structure of PBPK model depends on the intended purpose for which it is developed. In some cases, tissues with similar blood flow rate properties can be collated together as a single compartment to generate 'reduced' models [74].

Depending on the characteristics of drug molecule such as its passive permeability and active transport, each tissue is considered as either perfusion rate limited or permeability-rate limited. The permeability-limited tissues allow incorporation of active transport processes. Mostly, generic PBPK models assume perfusion rate-limited kinetics, with liver and kidney considered as the main sites of clearance [7].

The transport of a drug across different compartments is handled by mass balance differential equations. The simulation of plasma or/and tissue concentration-time profiles following intravenous administration is quite simple as compared to oral administration, which is comparatively more complex. In literature, multiple absorption models have been described, but generally, each model separates the gut compartment into a number of sub-compartments corresponding to different regions of the gastrointestinal tract, viz., stomach, duodenum, jejunum, ileum, caecum and colon. Each sub-compartment is divided into the lumen (unabsorbed drug) and the enterocyte (absorbed drug). Further, each of these sub-compartments is defined by a

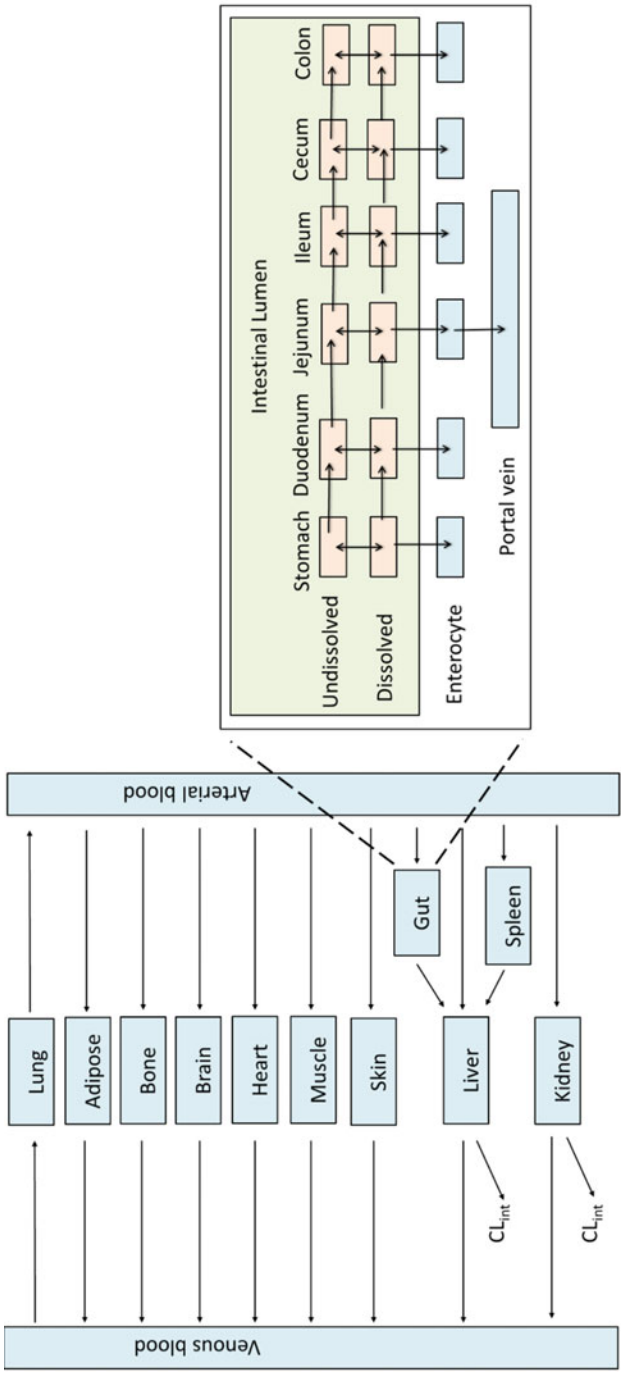


Fig. 9.4 Schematic representation of a PBPK model. A detailed representation of the gut is shown in the insert, as a typical example. Key: CL_{int} intrinsic clearance. Adapted from Jones et al. [9]

sub-tissue volume, transit time and pH [13, 46]. The mass balance equations are used to define various absorption-related processes such as dissolution and precipitation of the drug, with proper consideration of drug-specific parameters, for example, ionization coefficient (pK_a), octanol to water partition coefficient (logP) and solubility. Both passive diffusion as well as active transport can be unified in these models [9, 10]. Further, the absorption as well as disposition models are integrated to predict PK profiles, which are then compared with those obtained experimentally. Commercial platforms like Simcyp, GastroPlus and PK-Sim are employed for the purpose. Alternatively, PBPK models can be coded up with the commonly used modelling softwares like NONMEM, ADAPT, Berkeley Madonna, SAAM, WinNonlin, etc. [7].

9.5.2 System-Related Input Parameters

The benefit of PBPK lies in the fact that it can be developed for any species, provided the physiological and biochemical parameters are available or determined. The system-dependent parameters to support the model development for common species such as rat, dog and human are reported in the literature. Also, all commercial PBPK softwares have embedded parameters for most common species. For humans, the softwares allow inclusion of various physiological and biochemical features such as hepatic blood flow, CYP enzymes abundance and their polymorphism, liver volume, hematocrit and liver/renal function associated with age or specific disease states [15, 75]. This allows prediction of PK of drug in different population groups, including elderly, paediatric (including ontogeny), pregnancy, obesity and comorbid diseases (e.g. cirrhosis, chronic kidney failure, etc.). Also, the influence of environmental factors such as smoking can be added in a model [9, 10]. A correlated Monte Carlo approach can be used to generate a virtual population by using the values and formulae describing demographic, anatomical and physiological variables. This virtual population can be further employed to assess variability among populations before initiating clinical studies [7].

9.5.3 Drug-Specific Input Parameters

Along with the system-related input parameters, the simulation of intravenous and oral plasma concentration-time PK profiles using generic PBPK models also require multiple drug-specific inputs. This is because drug-related properties directly influence ADME mechanisms of a particular compound. For example, physicochemical properties of a drug [lipophilicity, solubility, molecular weight (MW) and pK_a value] and its biological properties [blood to plasma partition coefficient (B/P), plasma protein binding (fu_p), permeability, etc.] directly affect its plasma-to-tissue partition coefficient (K_p), intrinsic clearance, bioavailability, etc. [7, 10]. Another important parameter of any PBPK model is K_p value, which defines the distribution characteristics of a compound and hence ultimately affects the volume of

distribution (V_d). Various methods are available to calculate K_p and hence the V_d, therefore, selection of the most appropriate method is critical for achieving an accurate simulation [9]. Another key parameter, viz., intrinsic clearance can be determined by in vitro experiments and can be further scaled to in vivo clearance using appropriate set of IVIVE equations [11]. The in silico and in vitro determined values of solubility, dissolution, precipitation, uptake, permeability, particle size, etc., can be used to predict the in vivo bioavailability (rate and extent of absorption), which is an important component of any oral PK simulation [13]. The physiological framework of PBPK model also facilitates incorporation of active transport processes in a number of different tissues, for example, liver and intestine, provided relevant input data are available. The number of input parameters that a model may require depends directly on the intended purpose, yet the success of PBPK modelling depends highly on the quality of input data, as 'better data means better models' [39].

9.5.4 Other Miscellaneous Properties

Other properties that can help build a robust PBPK model are: (1) formulation properties, viz., particle size, surface area, mechanism of dissolution, etc.; (2) administration protocol, viz., route of administration, dose and dosing regimen and (3) time-related special events such as gallbladder emptying time, food intake, exercise, enterohepatic recycling (EHC), etc. [10] (Fig. 9.5).

9.6 Approaches for PBPK Model Development

There are three established pathways to PBPK model development, which include bottom-up, top-down and middle-out approaches.

9.6.1 Bottom-Up Approach

A 'bottom-up' approach involves modelling of the mechanisms that define ADME processes and related concentration profiles. The two pre-requisite to this approach are the availability of high-quality in vitro and preclinical data, and the verified (IVIVE) factors and scalars. But at early stages of drug development, the scalars used for extrapolation of the in vitro data to in vivo settings are associated with a high degree of uncertainty, especially when transporters are involved. So they need to be verified later while correlating drug development data with in vivo clinical data. The general application of bottom-up approach is that it provides mechanistic understanding of various ADME processes. Examples of modelling scenarios, involving this approach, include projection of human drug distribution using physiochemical properties and blood binding properties (e.g. logP, pKa, fup and B/P); projection of human PK parameters and FIH dose using in vitro permeability and pharmaceutics information (e.g. solubility, dissolution, particle size, etc.),

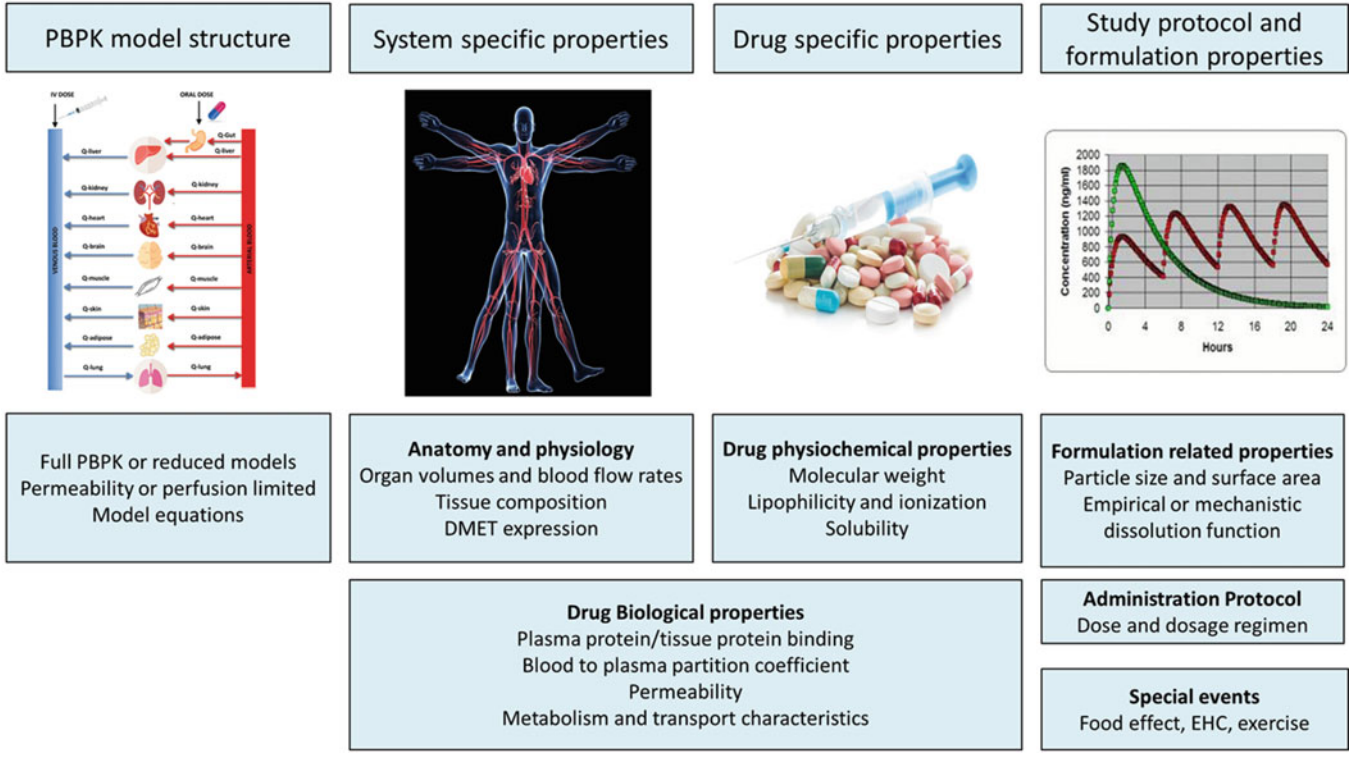


Fig. 9.5 Components of a PBPK model, including system and drug/formulation-related properties required to build an efficient PBPK model. Adapted from Keufer et al. [10] with few modifications

projection of enzyme/transporter DDI (victim and perpetrator) using in vitro metabolism and substrate and perpetrator data of the transporters [8, 9, 39].

9.6.2 Top-Down Approach

The main objective of 'top-down' approach is to build a model that describes the observed PK/clinical data. Other roles are to estimate means of various parameters, their inter-subject variability and to identify significant covariates of the PK parameters. More often, this approach involves fitting of the model parameters to clinically observed plasma concentration-time profile and/or urine data of the drug, established following administration of the single intravenous dose; single or multiple oral ascending doses; DDI scenarios or exposure across multiple formulations. The goodness of fit of the model is assessed by various statistical and visual approaches. While interpolation of data can be carried out easily using the optimized models, it is bit challenging to extrapolate outside the data space used to fit the model. Yet, this approach is commonly used in population PK (PopPK) data analysis. Its major application is towards providing support to clinical trial decisions [8, 9, 39].

9.6.3 Middle-Out Approach

The 'middle-out' approach is a combination of bottom-up and top-down approaches. In this case, initial model development utilizes the high-quality physicochemical, preclinical, in vitro metabolism/transport and mass-balance data in combination with other in silico or built-in PBPK predictions of drug distribution parameters. The observed DDI with inhibitors can then be employed to assess the model predictive performance and estimate some of the key parameters such as the fraction metabolized (f_m) by a particular enzyme. The sparse data methods such as nonlinear mixed effects and Bayesian maximum-like procedures, can be utilized along with parameters sensitivity analysis and parameter-estimation algorithms, to refine model parameters using the clinical data. Also, independently reported clinical studies can be applied for the external verification of the model, whereafter, the refined model can be employed to address clinically relevant questions such as requesting a waiver for the dedicated clinical trials or to extrapolate PK from healthy adults to special populations such as paediatric, organ impaired, etc. [8, 9, 39].

9.7 Best Practices for PBPK Model Building

The increasing applicability of PBPK modelling in drug development and regulatory submissions has propelled the industry, regulatory agencies and academia to discuss and develop best practices for the same [10]. The general recommendations and

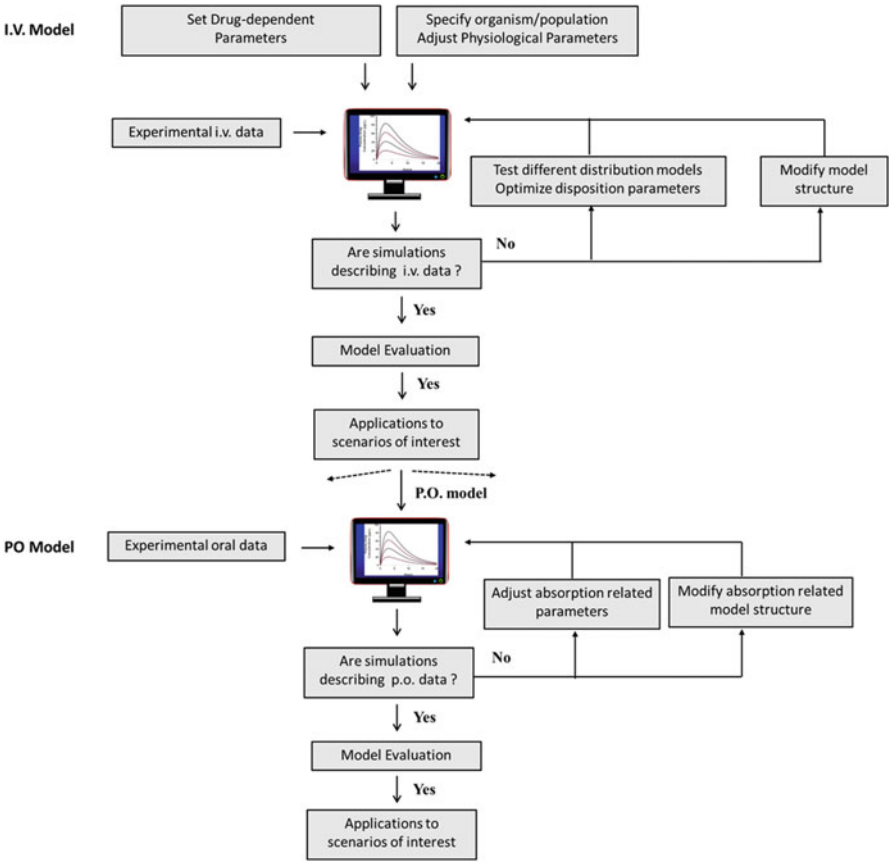


Fig. 9.6 Steps usually involved in PBPK model building. Adapted from Keufer et al. [10]. Key: *i. v.* intravenous, *p.o.* peroral

widely used practices for PBPK model building are compiled in Fig. 9.6 and are discussed below.

9.7.1 Compilation of Available Data and Information

Before initiating model building, it is absolutely necessary to collate all the available information related to ADME of the drug. Usually, the data needed include physiochemical, biological and pharmacokinetic properties, along with the characteristics of the organism or population [76]. Default physiological parameters can be changed in case of special populations on the basis of mechanistic rationale [10]. The ADME information such as clearance processes, transporters, or specific binding partners (e.g. plasma/tissue proteins, carrier proteins, etc.) can help in building physiologically plausible models. The availability of experimental PK

data is a must to validate the model and identify unknown or uncertain parameters. The most commonly used data include concentration-time profiles of the drug in the plasma (e.g. for different preclinical species and/or humans, different routes of administration and different dosage regimen), concentration-time profiles of the drug in relevant tissues, the percentages of drug excreted via the urine and faeces, mass balance data and information on drug metabolism (e.g. enzymes involved, identification of metabolites, rate of metabolism and excretion, etc.) [9, 10, 77].

9.7.2 Establishment of the Intravenous Disposition Model

Once all the available necessary information has been gathered, the next key step involves development of an intravenous (i.v.) disposition model that critically simulates the in vivo i.v. PK profile. Here, one must ensure that V_d and C_L values are as close to the experimental value as possible. Usually in their distribution modules, modelling softwares contain different methods of calculating organ/plasma partition coefficients, therefore, a suitable method shall be chosen, which is able to consistently describe the drug distribution behaviour. While the clearance parameters can be estimated from the experimental data, the role of different elimination pathways is assessed considering the mass balance approach [10]. Also, IVIVE methods are utilized to estimate the rates of metabolism. If required, surrogate compound parameters like lipophilicity can be slightly adapted to obtain good agreement with the experimental data. If insufficiency still exists, one needs to relook at all the relevant processes, if any critical one is missing in the model. At times, incorporation of active transport or binding parameters like K_m and K_d, enzyme abundance, etc., helps in building of simulation profile that nearly matches with the clinically observed i.v. data [8, 10, 42].

9.7.3 Establishment of Oral Absorption Model

Once disposition parameters (distribution or metabolism/excretion) get fixed through establishment of an i.v. model, the same for peroral (p.o.) administration (or for other extravascular routes) can be established by varying the parameters that influence the oral absorption [10]. Typical parameters that need to be considered during development of a p.o. model include intestinal permeability and parameters related to meal events. Additionally, formulation-related parameters such as drug release and solubility may be adjusted, especially if aqueous solubility rather than biorelevant solubility are used as initial values. In further steps, additional absorption-related processes such as enterohepatic recycling (EHC) may be incorporated, if pertinent for the drug under consideration [13, 14]. The sequential development of i.v. and p.o. (or any extravascular administration) models help in identification of the parameters relevant for absorption, as the information regarding distribution and clearance processes is already finalized from the i.v. data. However, in certain circumstances where i.v. model incorporates processes that are also

relevant for oral absorption, for example, uptake transporters, it is advisable to perform parameter identification using the i.v. and p.o. data simultaneously. Also, it is not mandatory to develop an i.v. model, sometimes PBPK model can be developed using p.o. data only, depending on the modelling purpose and the needed model precision. In these cases, IVIVE approaches can be used to predict the information related to absorption or clearance, with proper consideration of drug absorption, metabolism and mass balance data. Sometimes, in the absence of human i.v. data, a human p.o. model may be scaled from an animal model that is established using both i.v. and p.o. data. But it must be noted that the lack of i.v. data may increase the model uncertainty many folds, if there is role of transporters or gut-wall metabolism [10].

9.7.4 Overall Model Evaluation

After the model development, the quality of model should be evaluated. The stringency of model evaluation depends on the goal of modelling project. Overall, the evaluation step decides whether the model fits the purpose or not [39]. The various criteria or tests used to evaluate a model necessarily include the following:

9.7.4.1 Concurrence of Modelling Outcome with the Experimental Data

Usually, a visual comparison of simulated versus experimental concentration-time profiles (for plasma and/or other tissues, if available) with focus on the absolute concentrations and the dynamic shape of the PK profile can be used for model evaluation [42]. Additionally, error functions such as root mean-square deviation (RMSD), the area under the curve (AUC) error or the concordance correlation coefficient can also be used [43].

9.7.4.2 Comparison of Typical PK Parameters like C_{max}, T_{max}, AUC and $T_{½}$ Between Simulated Results Obtained by the Model and the Experimental Data

It is necessary to critically re-evaluate the uncertainty related to both model simulations and experimental data. Apart from the data used for model building, additional data may also be considered for external validation to assess consistency of the model prediction. The consideration of different doses may be an important cross-check for model validation and any deviations in estimations for dose levels, which had not been considered during model development, point to structural shortcomings of the model. Consistency of the model across different species in terms of drug-dependent parameters and calculation methods for distribution and cellular permeability can also be used as a check for model evaluation. Any changes in these properties should be supported by a plausible physiological explanation. The consistency of the model in special populations (hepatic or renal impaired patients, etc.) further aids in its evaluation, provided PK data are available and the physiological parameters can be changed accordingly. Furthermore, the model evaluation can be carried out by monitoring the consistency of the model across

compounds, for which the same processes are relevant in terms of ADME properties. For example, while implementing saturable receptor binding into a PBPK model, it can be checked if the receptor concentration is the same in a PBPK model of a reference compound, which binds to the identical receptor [10, 39, 42].

9.7.5 Sensitivity Analysis or Best/Worst Case Scenarios

It is always recommended to assess uncertainties in the model output through sensitivity analysis of the relevant parameters. Sensitivity analysis can be performed on all the parameters in the PBPK model to identify the most sensitive parameters for a specified model output, for example, plasma concentration or PK parameters or on certain specified parameters like uncertain parameters for active processes included in the model. Additionally, the effect of extreme experimental and physiological alterations in uncertain parameters can be evaluated to simulate the best/worst case scenarios. The results of the sensitivity analysis or best/worst case scenarios can be used to assess if the model development conclusions are robust [10, 39].

9.8 Software Employed for the Modelling

As discussed above, there are quite a few commercial software packages that are dedicated for use in PBPK modelling and simulations. These are classified into three categories on the basis of their intended utility, as discussed below.

9.8.1 Custom Physiologically Based Pharmacokinetics (PBPK) Software

These softwares are specifically developed for PBPK modelling by users involved in drug discovery, development and regulation [7]. These are the ones that are currently preferred by the PBPK scientific community.

9.8.2 General Purpose High-Level Scientific Computing Software

These packages provide general tools for scientific computing in the form of high-level programming or matrix languages. These tools are used generally in cases where the investigators need capabilities, which otherwise are not provided by user friendly softwares [7].

9.8.3 Bio-Mathematical Modelling Software

These tools are designed explicitly for mathematical modelling of biological systems. Some have a user friendly (graphical) interface and the extent of their usage for PBPK modelling is governed by the limitations imposed by the graphical interface, speed of computation and flexibility of the modelling language. Sparse data sets can be analyzed by mixed-effects (population) capabilities provided in some of the tools [7].

The various softwares in each of three named categories, along with their associated developers and links, are listed in Table 9.1.

9.9 Understanding Limitations of PBPK Modelling

There are many limitations that may be encountered in the whole process of PBPK modelling. These are enumerated below:

9.9.1 Inaccurate and Inconsistent Core Information

The data related to physiological, biochemical and physicochemical processes of biological systems, and properties of the drug which are required during PBPK modelling, usually are not available from a single source. When the data are taken from different texts and reports, they may have inherent variability due to the use of different test techniques (practically each method and analytical technique employed to generate the relevant data has its own advantages and limitations in terms of accuracy and dynamic range); variable sample preparation protocols; nature of in vitro or in vivo systems employed; differential style of presentation (whether figures, tables, bar diagrams, etc.); different units, etc. Therefore, very often one lands into state of confusion as to what data to accept and which to reject, so that only reliable, accurate and consistent information is collated and used in the model development [3, 78].

9.9.2 Absence of Key Data

Another limitation that may hinder the use of PBPK modelling is with respect to the absence of reported in vivo data and limited knowledge of tissue-specific changes in enzyme and transporter expression in special populations of interest. The model predictions can be confounded by lack of confidence in individual parameters due to limited availability of in vivo data to verify them. For instance, lack of i.v. data of a drug in humans can introduce uncertainty into model parameters and output, as the distribution and absorption parameters cannot be validated experimentally [79, 80].

Table 9.1 Commercially available PBPK platforms along with their associated links

Software	Developer/distributor	Associated links
Custom physiologically based pharmacokinetics (PBPK) software		
Simcyp Simulator	Simcyp Ltd	https://www.certara.com/software-old/physiologically-based-pharmacokinetic-modeling-and-simulation/simcyp-simulator/?ap=PKPD
GastroPlus	Simulations Plus Inc.	https://www.simulations-plus.com/software/gastroplus/
PK-Sim	Bayer Technology Services	http://www.systems-biology.com/products/pk-sim/
Cloe Predict	Cyprotex Ltd	https://www.cyprotex.com/insilico/physiological_modelling/cloe-pk/
General purpose high-level scientific computing software		
Berkeley Madonna	University of California at Berkeley	http://www.berkeleymadonna.com/
MATLAB and Simulink product families	The MathWorks Inc.	https://in.mathworks.com/products/matlab.html?s_tid=hp_products_matlab
MLAB	Civilized Software Inc.	http://www.civilized.com/mlabdesc.html
GNU Octave	GNU	https://www.gnu.org/software/octave/
Ecolego	AFRY	https://www.ecolego.se/ecolego/
GNU MCSim	GNU	https://www.gnu.org/software/mcsim/
Bio-mathematical modeling software		
ADAPT 5	Biomedical Simulations Resource, University of Southern California	https://bmsr.usc.edu/software/adapt/
ModelMaker	ModelKinetix	https://www.apbenson.com/modelkinetix-downloads
NONMEM	ICON	https://www.iconplc.com/innovation/nonmem/
STELLA	Isee systems Inc.	https://www.iseesystems.com/store/products/stella-architect.aspx
WinNonlin	Pharsight, a Certara company	https://www.certara.com/pkpd-modeling-and-simulation-2/phoenix-winnonlin-2/?ap=PKPD
SAAM II	The Epsilon Group	https://tegvirginia.com/software/saam-ii/
acslX	The AEgis Technologies Group Inc.	http://www.acslx.com/products/toolkits.shtml
PhysioLab	Entelos Inc.	http://www.entelos.com/
PKQuest	University of Minnesota	http://www.pkquest.com/
gCOAS	Process Systems Enterprise	https://www.psenterprise.com/products/gcoas/

(continued)

Table 9.1 (continued)

Software	Developer/distributor	Associated links
COPASI	Biocomplexity Institute of Virginia Tech, University of Heidelberg, University of Connecticut, UConn Health	http://copasi.org/
Maxsim2	Fraunhofer-Chalmers Research Centre	http://www.maxsim2.com/
AIMT7: RVIS	Health and Safety Laboratory	http://cefic-lri.org/projects/aimt7-rvis-open-access-pbpk-modelling-platform/
ADME WorkBench	AEgis Technologies	http://www.admewb.com

9.9.3 Non-availability of Selective Substrates and Inhibitors

The validation of the model for drugs cleared by non-CYP enzymes and transporters, against in vivo data, is prevented by lack of selective substrates and inhibitors for some of these enzymes [79]. Currently, the models describing disposition of transporter substrates rely on the incorporation of empirical scaling factors. Though these scaling factors allow prediction of the kinetics of multiple uptake transporter substrates, it is impossible to experimentally verify the predicted unbound tissue exposures. This tends to have significant implications for IVIVE of efflux clearance, metabolism-transporter interplay and prediction of pharmacological effects [81–83].

9.9.4 Absence of Knowledge of Basic Mechanisms in Certain Instances

Some of the processes as such are not yet well characterized, for instance, transporter abundance or the absorption process in newborns and infants; various processes involved in oral drug absorption; in vivo behaviour of non-conventional drug delivery systems, etc. These knowledge gaps may result in failure of the model to optimally describe PK behaviour in mentioned scenarios [80].

9.9.5 Limited Utility in the Prediction of Disposition of Therapeutic Proteins

Another issue is limited utility of PBPK in the prediction of disposition of therapeutic proteins, as the model structures employed till now to predict the kinetics of monoclonal antibodies are of inconsistent nature [84].

9.10 Conclusion

PBPK modelling has become an indispensable tool in drug development process. The enhanced understanding of processes involved in drug absorption and disposition has greatly widened the spectrum of applications of PBPK modelling. But still there are knowledge gaps, which propel the modeller to make certain assumptions. The latter need careful consideration to avoid uncertainties in simulation and extrapolation results. However, besides all the limitations, the numerous applications and advantages of PBPK make the modelling science an attractive proposition to be pursued by the researchers, whether in academia, industry or regulatory science [85]. It is also a high time that serious efforts are made to fulfil the knowledge gaps, so that PBPK modelling is free of uncertainties, and the predictive research becomes more and more reliable and acceptable.

9.11 Credible Online Resources for Further Reading

URL	Description of source	Recommendation
www.certara.com	Developed and maintained by Certara Ltd. Company, the parent company of PBPK software tool, Simcyp simulator	This website has list of all articles and posters published using Simcyp simulator. Additionally, Certara blogs are also quite useful to the readers for deeper understanding of the concepts of PBPK modelling
www.simulations-plus.com	Developed and maintained by Simulation Plus Inc., the parent company of Gastroplus PBPK tool	This website provides updates related to advanced applications of Gastroplus and its allied tools. Additionally, articles and posters citing the usage of these tools are also available on this website
https://in.mathworks.com/help/simbio/	Developed and maintained by The MathWorks Inc.	This website contains the information related to Matworks tools such as Matlab, simbiology etc., which can be used to develop the custom PBPK models
http://www.open-systems-pharmacology.org/	Developed and maintained by Bayer Technology Services, the parent company of PBPK tool, PK-Sim	This website provides detailed instructions related to usage of PK-Sim. Additionally, posters and papers published using the software are also available at this website

References

1. Jones HM, Gardner IB, Watson KJ (2009) Modelling and PBPK simulation in drug discovery. AAPS J 11(1):155–166. https://doi.org/10.1208/s12248-009-9088-1
2. Shardlow CE, Generaux GT, Patel AH, Tai G, Tran T, Bloomer JC (2013) Impact of physiologically based pharmacokinetic modeling and simulation in drug development. Drug Metab Dispos 41(12):1994–2003. https://doi.org/10.1124/dmd.113.052803
3. Sager JE, Yu J, Ragueneau-Majlessi I, Isoherranen N (2015) Physiologically based pharmacokinetic (PBPK) modeling and simulation approaches: a systematic review of published models, applications, and model verification. Drug Metab Dispos 43(11):1823–1837. https://doi.org/10.1124/dmd.115.065920
4. Obach RS, Baxter JG, Liston TE, Silber BM, Jones BC, Macintyre F et al (1997) The prediction of human pharmacokinetic parameters from preclinical and *in vitro* metabolism data. J Pharmacol Exp 283(1):46–58. PMID: 9336307
5. Fahmi OA, Maurer TS, Kish M, Cardenas E, Boldt S, Nettleton D (2008) A combined model for predicting CYP3A4 clinical net drug-drug interaction based on CYP3A4 inhibition, inactivation, and induction determined *in vitro*. Drug Metab Dispos 36(8):1698–1708. https://doi.org/10.1124/dmd.107.018663
6. Varma MV, Lin J, Bi YA, Rotter CJ, Fahmi OA, Lam JL et al (2013) Quantitative prediction of repaglinide-rifampicin complex drug interactions using dynamic and static mechanistic models: delineating differential CYP3A4 induction and OATP1B1 inhibition potential of rifampicin. Drug Metab Dispos 41(5):966–974. https://doi.org/10.1124/dmd.112.050583
7. Rowland M, Peck C, Tucker G (2011) Physiologically-based pharmacokinetics in drug development and regulatory science. Annu Rev Pharmacol Toxicol 51:45–73. https://doi.org/10.1146/annurev-pharmtox-010510-100540
8. Tsamandouras N, Rostami-Hodjegan A, Aarons L (2015) Combining the 'bottom up' and 'top down' approaches in pharmacokinetic modelling: fitting PBPK models to observed clinical data. Br J Clin Pharmacol 79(1):48–55. https://doi.org/10.1111/bcp.12234
9. Jones HM, Rowland-Yeo K (2013) Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. CPT Pharmacometrics Syst Pharmacol 2(8):1–2. https://doi.org/10.1038/psp.2013.41
10. Kuepfer L, Niederalt C, Wendl T, Schlender JF, Willmann S, Lippert J et al (2016) Applied concepts in PBPK modeling: how to build a PBPK/PD model. CPT Pharmacometrics Syst Pharmacol 5(10):516–531. https://doi.org/10.1002/psp4.12134
11. Rostami-Hodjegan A, Tucker GT (2007) Simulation and prediction of *in vivo* drug metabolism in human populations from *in vitro* data. Nat Rev Drug Discov 6(2):140–148. https://doi.org/10.1038/nrd2173
12. Di L, Feng B, Goosen TC, Lai Y, Steyn SJ, Varma MV et al (2013) A perspective on the prediction of drug pharmacokinetics and disposition in drug research and development. Drug Metab Dispos 41(12):1975–1993. https://doi.org/10.1124/dmd.113.054031
13. Jamei M, Turner D, Yang J, Neuhoff S, Polak S, Rostami-Hodjegan A et al (2009) Population-based mechanistic prediction of oral drug absorption. AAPS J 11(2):225–237. https://doi.org/10.1208/s12248-009-9099-y
14. Chow ECY, Sandy Pang K (2013) Why we need proper PBPK models to examine intestine and liver oral drug absorption. Curr Drug Metab 14(1):57–79. PMID: 22935069
15. Jamei M, Dickinson GL, Rostami-Hodjegan A (2009) A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: a tale of 'bottom-up' vs 'top-down' recognition of covariates. Drug Metab Pharmacokinet 24(1):53–75. https://doi.org/10.2133/dmpk.24.53
16. Huang SM, Abernethy DR, Wang Y, Zhao P, Zineh I (2013) The utility of modeling and simulation in drug development and regulatory review. J Pharm Sci 102(9):2912–2923. https://doi.org/10.1002/jps.23570

17. Committee for Medicinal Products for Human Use (CHMP) (2005) Guideline on the evaluation of the pharmacokinetics of medicinal products in patients with impaired hepatic function. CPMP/EWP/2339/02, European Medicines Agency, London
18. Center for Drug Evaluation and Research (CDER) (2014) General clinical pharmacology considerations for pediatric studies for drugs and biological products. U.S. Food and Drug Administration, Silver Spring
19. Committee for Medicinal Products for Human Use (CHMP) (2012) Guideline on the investigation of drug-drug interactions. European Medicines Agency, London
20. Center for Drug Evaluation and Research (CDER) (2012) Guidance for industry: drug interaction studies study design, data analysis, implications for dosing, and labeling recommendations. U.S. Food and Drug Administration, Silver Spring
21. Ministry of Health, Labour and Welfare Research Group (2014) Drug interaction guideline for drug development and labeling recommendations. Japanese Ministry of Health, Labour and Welfare, Tokyo
22. Committee for Medicinal Products for Human Use (CHMP) (2011) Guideline on the use of pharmacogenetic methodologies in the pharmacokinetic evaluation of medicinal products. EMA/CHMP/37646/2009, European Medicines Agency, London
23. Center for Drug Evaluation and Research (CDER) (2013) Guidance for industry: clinical pharmacogenomics: premarket evaluation in early-phase clinical studies and recommendations for labeling. U.S. Food and Drug Administration, Silver Spring
24. Teorell T (1937) Kinetics of distribution of substances administered to the body, I: the extravascular modes of administration. Arch Int Pharmacodyn 57:205–225
25. Bellman R, Jacquez JA, Kalaba R (1960) Some mathematical aspects of chemotherapy: I. One-organ models. Bull Math Biol 22(2):181–198. https://doi.org/10.1007/BF02478005
26. Bischoff KB (1967) Applications of a mathematical model for drug distribution in mammals. In: Hershey D (ed) Chemical engineering in medicine and biology. Springer, Boston. https://doi.org/10.1007/978-1-4757-4748-5_13
27. Andersen ME, Yang RS, Clewell HJ III, Reddy MB (2005) Introduction: a historical perspective of the development and applications of PBPK models. In: Anderson ME (ed) Physiologically based pharmacokinetic modeling: science and applications, vol 10. Wiley, Hoboken, pp 1–8. https://doi.org/10.1002/0471478768.ch1
28. Clewell HJ III, Reddy MB, Lave T, Andersen ME (2010) Physiologically based pharmacokinetic modeling. In: Clewell HJ (ed) Pharmaceutical sciences encyclopaedia: drug discovery, development, and manufacturing, vol 15. Wiley, Hoboken, pp 1–62. https://doi.org/10.1002/9780470571224.pse065
29. Clewell HJ III, Andersen ME, Wills RJ, Latriano L (1997) A physiologically based pharmacokinetic model for retinoic acid and its metabolites. J Am Acad Dermatol 36(3):S77–S85. https://doi.org/10.1016/s0190-9622(97)70063-x
30. Zhao P, Rowland M, Huang SM (2012) Best practice in the use of physiologically based pharmacokinetic modeling and simulation to address clinical pharmacology regulatory questions. Clin Pharmacol Ther 92(1):17–20. https://doi.org/10.1038/clpt.2012.68
31. Leong R, Vieira ML, Zhao P, Mulugeta Y, Lee CS, Huang SM et al (2012) Regulatory experience with physiologically based pharmacokinetic modeling for pediatric drug trials. Clin Pharmacol Ther 91(5):926–931. https://doi.org/10.1038/clpt.2012.19
32. Huang SM, Rowland M (2012) The role of physiologically based pharmacokinetic modeling in regulatory review. Clin Pharmacol Ther 91(3):542–549. https://doi.org/10.1038/clpt.2011.320
33. Pan Y, Grillo J, Hsu V, Zhang L, Sinha V, Huang SM et al (2014) Application of the FDA PBPK knowledgebase in evaluating model predictability for drug-drug interactions. Clin Pharmacol Ther 95(1)
34. Sinha V, Zhao P, Huang SM, Zineh I (2014) Physiologically based pharmacokinetic modeling: from regulatory science to regulatory policy. Clin Pharmacol Ther 95(5):478–480. https://doi.org/10.1038/clpt.2014.46

35. Grimstein M, Yang Y, Zhang X, Grillo J, Huang SM, Zineh I et al (2019) Physiologically based pharmacokinetic modeling in regulatory science: an update from the US Food and Drug Administration's Office of Clinical Pharmacology. J Pharm Sci 108(1):21–25. https://doi.org/10.1016/j.xphs.2018.10.033
36. US Food and Drug Administration (2019) Impact story: supporting drug development through physiologically based pharmacokinetic modeling. https://www.fda.gov/drugs/regulatory-science-action/impact-story-supporting-drug-development-through-physiologically-based-pharmacokinetic-modeling. Accessed 5 Apr 2020
37. Center for Drug Evaluation and Research (CDER) (2018) Guidance for industry: physiologically based pharmacokinetic analyses—format and content. U.S. Food and Drug Administration, Silver Spring
38. Committee for Medicinal Products for Human Use (CHMP) (2018) Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation. EMA/CHMP/458101/2016. European Medicines Agency, London
39. Shebley M, Sandhu P, Emami Riedmaier A, Jamei M, Narayanan R, Patel A et al (2018) Physiologically based pharmacokinetic model qualification and reporting procedures for regulatory submissions: a consortium perspective. Clin Pharmacol Ther 104(1):88–110. https://doi.org/10.1002/cpt.1013
40. Peters SA, Ungell AL, Dolgos H (2009) Physiologically based pharmacokinetic (PBPK) modeling and simulation: applications in lead optimization. Curr Opin Drug Discov Devel 12(4):509–518. PMID: 19562647
41. Parrott N, Jones H, Paquereau N, Lavé T (2005) Application of full physiological models for pharmaceutical drug candidate selection and extrapolation of pharmacokinetics to man. Basic Clin Pharmacol Toxicol 96(3):193–199. https://doi.org/10.1111/j.1742-7843.2005.pto960308.x
42. Peters SA (2008) Evaluation of a generic physiologically based pharmacokinetic model for lineshape analysis. Clin Pharmacokinet 47(4):261–275. https://doi.org/10.2165/00003088-200847040-00004
43. Thiel C, Schneckener S, Krauss M, Ghallab A, Hofmann U, Kanacher T et al (2015) A systematic evaluation of the use of physiologically based pharmacokinetic modeling for cross-species extrapolation. J Pharm Sci 104(1):191–206. https://doi.org/10.1002/jps.24214
44. Miller NA, Reddy MB, Heikkinen AT, Lukacova V, Parrott N (2019) Physiologically based pharmacokinetic modelling for first-in-human predictions: an updated model building strategy illustrated with challenging industry case studies. Clin Pharmacokinet 58(6):727–746. https://doi.org/10.1007/s40262-019-00741-9
45. Yang J, Jamei M, Yeo KR, Tucker GT, Rostami-Hodjegan A (2007) Prediction of intestinal first-pass drug metabolism. Curr Drug Metab 8(7):676–684. https://doi.org/10.2174/138920007782109733
46. Sugano K (2009) Introduction to computational oral absorption simulation. Expert Opin Drug Metab Toxicol 5(3):259–293. https://doi.org/10.1517/17425250902835506
47. Yeo KR, Jamei M, Yang J, Tucker GT, Rostami-Hodjegan A (2010) Physiologically based mechanistic modelling to predict complex drug-drug interactions involving simultaneous competitive and time-dependent enzyme inhibition by parent compound and its metabolite in both liver and gut—the effect of diltiazem on the time-course of exposure to triazolam. Eur J Pharm Sci 39(5):298–309. https://doi.org/10.1016/j.ejps.2009.12.002
48. Chien JY, Mohutsky MA, Wrighton SA (2003) Physiological approaches to the prediction of drug-drug interactions in study populations. Curr Drug Metab 4(5):347–356. https://doi.org/10.2174/1389200033489307
49. Rostami-Hodjegan A, Tucker G (2004) '*In silico*' simulations to assess the '*in vivo*' consequences of '*in vitro*' metabolic drug-drug interactions. Drug Discov Today Technol 1(4):441–448. https://doi.org/10.1016/j.ddtec.2004.10.002

50. Anderson BJ, Holford NH (2013) Understanding dosing: children are small adults, neonates are immature children. Arch Dis Child 98(9):737–744. https://doi.org/10.1136/archdischild-2013-303720
51. Jong GT (2014) Pediatric development: physiology. Enzymes, drug metabolism, pharmacokinetics and pharmacodynamics. In: Shalom DB (ed) Pediatric formulations: a roadmap. Springer, Boston, pp 9–23. https://doi.org/10.1007/978-1-4899-8011-3
52. Johnson TN, Rostami-Hodjegan A, Tucker GT (2006) Prediction of the clearance of eleven drugs and associated variability in neonates, infants and children. Clin Pharmacokinet 45(9):931–956. https://doi.org/10.2165/00003088-200645090-00005
53. Johnson TN, Tucker GT, Tanner MS, Rostami-Hodjegan A (2005) Changes in liver volume from birth to adulthood: a meta-analysis. Liver Transpl 11(12):1481–1493. https://doi.org/10.2165/00003088-200645090-00005
54. Edginton AN, Schmitt W, Voith B, Willmann S (2006) A mechanistic approach for the scaling of clearance in children. Clin Pharmacokinet 45(7):683–704. https://doi.org/10.2165/00003088-200645070-00004
55. Butler JM, Begg EJ (2008) Free drug metabolic clearance in elderly people. Clin Pharmacokinet 47(5):297–321. https://doi.org/10.2165/00003088-200847050-00002
56. Kusama M, Maeda K, Chiba K, Aoyama A, Sugiyama Y (2009) Prediction of the effects of genetic polymorphism on the pharmacokinetics of CYP2C9 substrates from *in vitro* data. Pharm Res 26(4):822. https://doi.org/10.1007/s11095-008-9781-2
57. Jornil J, Jensen KG, Larsen F, Linnet K (2010) Identification of cytochrome P450 isoforms involved in the metabolism of paroxetine and estimation of their importance for human paroxetine metabolism using a population-based simulator. Drug Metab Dispos 38(3):376–385. https://doi.org/10.1124/dmd.109.030551
58. Edginton AN, Willmann S (2008) Physiology-based simulations of a pathological condition: prediction of pharmacokinetics in patients with liver cirrhosis. Clin Pharmacokinet 47(11):743–752. https://doi.org/10.2165/00003088-200847110-00005
59. Johnson TN, Boussery K, Rowland-Yeo K, Tucker GT, Rostami-Hodjegan A (2010) A semi-mechanistic model to predict the effects of liver cirrhosis on drug clearance. Clin Pharmacokinet 49(3):189–206. https://doi.org/10.2165/11318160-000000000-00000
60. Nolin TD, Naud J, Leblond FA, Pichette V (2008) Emerging evidence of the impact of kidney disease on drug metabolism and transport. Clin Pharmacol Ther 83(6):898–903. https://doi.org/10.1038/clpt.2008.59
61. Schmith VD, Foss JF (2010) Inflammation: planning for a source of pharmacokinetic/pharmacodynamic variability in translational studies. Clin Pharmacol Ther 87(4):488–491. https://doi.org/10.1038/clpt.2009.258
62. Hanley MJ, Abernethy DR, Greenblatt DJ (2010) Effect of obesity on the pharmacokinetics of drugs in humans. Clin Pharmacokinet 49(2):71–87. https://doi.org/10.2165/11318100-000000000-00000
63. Rasool MF, Khalid S, Majeed A, Saeed H, Imran I, Mohany M, Al-Rejaie SS et al (2019) Development and evaluation of physiologically based pharmacokinetic drug-disease models for predicting rifampicin exposure in tuberculosis and cirrhosis populations. Pharmaceutics 11(11):578. https://doi.org/10.3390/pharmaceutics11110578
64. You X, Wu W, Xu J, Jiao Z, Ke M, Huang P, Lin C (2020) Development of a physiologically based pharmacokinetic model for prediction of pramipexole pharmacokinetics in Parkinson's disease patients with renal impairment. J Clin Pharmacol 60(8):999–1010. https://doi.org/10.1002/jcph.1593
65. Center for Drug Evaluation and Research (CDER) (2002) Guidance for industry: food-effect bioavailability and fed bioequivalence studies. U.S. Food and Drug Administration, Silver Spring
66. Tistaert C, Heimbach T, Xia B, Parrott N, Samant TS, Kesisoglou F (2019) Food effect projections via physiologically based pharmacokinetic modeling: predictive case studies. J Pharm Sci 108(1):592–602. https://doi.org/10.1016/j.xphs.2018.05.024

67. Li M, Zhao P, Pan Y, Wagner C (2018) Predictive performance of physiologically based pharmacokinetic models for the effect of food on oral drug absorption: current status. CPT Pharmacometrics Syst Pharmacol 7(2):82–89. https://doi.org/10.1002/psp4.12260
68. Abduljalil K, Jamei M, Rostami-Hodjegan A, Johnson TN (2014) Changes in individual drug-independent system parameters during virtual paediatric pharmacokinetic trials: introducing time-varying physiology into a paediatric PBPK model. AAPS J 16(3):568–576. https://doi.org/10.1208/s12248-014-9592-9
69. Gaohua L, Wedagedera J, Small BG, Almond L, Romero K, Hermann D, Hanna D et al (2015) Development of a multi-compartment permeability-limited lung PBPK model and its application in predicting pulmonary pharmacokinetics of anti-tuberculosis drugs. CPT Pharmacometrics Syst Pharmacol 4(10):605–613. https://doi.org/10.1002/psp4.12034
70. Verscheijden LF, Koenderink JB, de Wildt SN, Russel FG (2019) Development of a physiologically-based pharmacokinetic pediatric brain model for prediction of cerebrospinal fluid drug concentrations and the influence of meningitis. PLoS Comput Biol 15(6):e1007117. https://doi.org/10.1371/journal.pcbi.1007117
71. Zhou L, Tong X, Sharma P, Xu H, Al-Huniti N, Zhou D (2019) Physiologically based pharmacokinetic modelling to predict exposure differences in healthy volunteers and subjects with renal impairment: ceftazidime case study. Basic Clin Pharmacol Toxicol 125(2):100–107. https://doi.org/10.1111/bcpt.13209
72. Loisios-Konstantinidis I, Cristofoletti R, Fotaki N, Turner DB, Dressman J (2020) Establishing virtual bioequivalence and clinically relevant specifications using *in vitro* biorelevant dissolution testing and physiologically-based population pharmacokinetic modeling. Case example: naproxen. Eur J Pharm Sci 143:105170. https://doi.org/10.1016/j.ejps.2019.105170
73. Khot A, Tibbitts J, Rock D, Shah DK (2017) Development of a translational physiologically based pharmacokinetic model for antibody-drug conjugates: a case study with T-DM1. AAPS J 19(6):1715–1734. https://doi.org/10.1208/s12248-017-0131-3
74. Nestorov IA, Aarons LJ, Arundel PA, Rowland M (1998) Lumping of whole-body physiologically based pharmacokinetic models. J Pharmacokinet Biopharm 26(1):21–46. https://doi.org/10.1023/a:1023272707390
75. Jones HM, Parrott N, Jorga K, Lavé T (2006) A novel strategy for physiologically based predictions of human pharmacokinetics. Clin Pharmacokinet 45(5):511–542. https://doi.org/10.2165/00003088-200645050-00006
76. Nestorov I (2007) Whole-body physiologically based pharmacokinetic models. Expert Opin Drug Metab Toxicol 3(2):235–249. https://doi.org/10.1517/17425255.3.2.235
77. Nestorov I (2003) Whole body pharmacokinetic models. Clin Pharmacokinet 42(10):883–908. https://doi.org/10.2165/00003088-200342100-00002
78. Harwood MD, Neuhoff S, Carlson GL, Warhurst G, Rostami-Hodjegan A (2013) Absolute abundance and function of intestinal drug transporters: a prerequisite for fully mechanistic *in vitro-in vivo* extrapolation of oral drug absorption. Biopharm Drug Dispos 34(1):2–8. https://doi.org/10.1002/bdd.1810
79. Jones HM, Chen Y, Gibson C, Heimbach T, Parrott N, Peters SA et al (2015) Physiologically based pharmacokinetic modeling in drug discovery and development: a pharmaceutical industry perspective. Clin Pharmacol Ther 97(3):247–262. https://doi.org/10.1002/cpt.37
80. Edginton AN, Joshi G (2011) Have physiologically-based pharmacokinetic models delivered? Expert Opin Drug Metab Toxicol 7(8):929–934. https://doi.org/10.1517/17425255.2011.585968
81. Varma MV, Lai Y, Feng B, Litchfield J, Goosen TC, Bergman A (2012) Physiologically based modeling of pravastatin transporter-mediated hepatobiliary disposition and drug-drug interactions. Pharm Res 29(10):2860–2873. https://doi.org/10.1007/s11095-012-0792-7

82. Jones HM, Barton HA, Lai Y, Bi YA, Kimoto E, Kempshall S (2012) Mechanistic pharmacokinetic modeling for the prediction of transporter-mediated disposition in humans from sandwich culture human hepatocyte data. Drug Metab Dispos 40(5):1007–1017. https://doi.org/10.1124/dmd.111.042994
83. Chu X, Korzekwa K, Elsby R, Fenner K, Galetin A, Lai Y (2013) Intracellular drug concentrations and transporters: measurement, modeling, and implications for the liver. Clin Pharmacol Ther 94(1):126–141. https://doi.org/10.1038/clpt.2013.78
84. Chetty M, Li L, Rose R, Machavaram K, Jamei M, Rostami-Hodjegan A (2015) Prediction of the pharmacokinetics, pharmacodynamics, and efficacy of a monoclonal antibody, using a physiologically based pharmacokinetic FcRn model. Front Immunol 5:670. https://doi.org/10.3389/fimmu.2014.00670
85. Rostami-Hodjegan A, Tamai I, Pang KS (2012) Physiologically based pharmacokinetic (PBPK) modeling: it is here to stay! Biopharm Drug Dispos 33(2):47–50. https://doi.org/10.1002/bdd.1776

Computers in Clinical Development

10

Swati Changdeo Jagdale and Asawaree Anand Hable

Contents

Abstract

Clinical development of a pharmaceutical product plays an important role as it regulates the future of investigational new drug application approval (IND) and approval of new drug application (NDA). This stage comes next to drug

S. C. Jagdale (✉)
School of Pharmacy, Dr Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India

A. A. Hable
MAEER's Maharashtra Institute of Pharmacy, Pune, Maharashtra, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_10

discovery and development. Clinical development aims for evaluation of drug molecule for its safety, dose range, effectiveness, side effects, and comparison to current treatment. Computers are rapid, reliable, and efficient electronic devices. Utilization of computer systems has tremendously affected our society, working style, and thereby it has gained immense popularity in almost every sector. The utilization of computers in clinical development helps in data collection as to maintain patient records, adequate quality and quantity of data, analysis and statistical treatment of data, interpretation of graphs and other numerical data, etc. There are many softwares and databases which are used in clinical development of a drug product. Studies like drug interaction, dose calculations, stability studies, statistical evaluation, clinical trial study evaluation, pharmacovigilance safety extensively rely on use of softwares, databases, and computers. This chapter highlights the role of computers in clinical development of pharmaceutical products, importance of computer in clinical development, softwares, databases, and electronic data capture tools.

Keywords

Clinical trial · Investigational new drug · New drug application · Computer · Softwares · Database · Artificial intelligence

Chapter Objective

- Review the clinical trial management system
- Know and understand the documents and records involved in clinical trial management system with their electronic form
- Know the softwares used in clinical trials for data management
- Overview of digital wearable medical health devices for record keeping.
- Understand the role of artificial intelligence in clinical trials
- Realize the advantages, limitations, and challenges of e-technologies in clinical research

10.1 Introduction

Clinical research is the approach to evaluate the safety and efficacy of the new treatments. The aim of clinical research is to provide more effective treatment with lesser side effects. The treatments in clinical research can be for newly developed drug, dosage form, dietary supplements, medical devices, vaccines, or combination of the successful therapies. If the treatments are given without conducting a clinical trial, there might be a risk of side effects or no effects or harmful effects of the drug [1]. Clinical research is an important segment in the drug development process because it provides various ways of prevention and treatment of the disease.

A clinical research consists of various kinds of research and has four phases involved in conducting a clinical trial. Before beginning the actual clinical trials with humans, the therapy under research is investigated and tested on laboratory animals.

Table 10.1 Summary of clinical trial phases

Phase	Objective	Number of participants	References
Phase 0: Pre-clinical study	• Study of safety and efficacy in laboratory animals • Determine safety, dose range of drug in experimental laboratory animals	As per study requirement	[1, 2]
Phase I: Clinical pharmacology	• Determine safety of drug dose in human • Evaluate therapeutic dose range • Identify side effects of the treatment	20–80	[1, 2]
Phase II: Drug efficacy/ safety, dose ranging	• Determine effectiveness of the treatment • Evaluate safety to all pharmacological parameters	100–300	[1, 2]
Phase III: Long-term, large scale, confirmatory	• Evaluate therapeutic effectiveness • Monitoring the side effects • Comparison with available treatments	1000–5000	[1, 2]
Phase IV: Post-market monitoring	• Determine adverse drug reaction after marketing of drug in wide population • Evaluate long-term safety	1000+	[1, 2]

This phase in a clinical trial is called as a preclinical trial study. This phase of the study ascertains the safety of research drugs in experimental animals [1]. If the treatment showed beneficial results against specific diseases, then only new therapy will go through the clinical trial. After successful completion of the pre-clinical study, an investigational new drug (IND) application must be filed and the actual clinical testing of a drug in humans can be started. The summary of clinical trial phases is represented in Table 10.1.

To cure diseases with some advanced treatments, many countries are developing novel drugs or modified treatments. To check the effectiveness of the said therapy, a therapy should undergo thorough quality checks in the clinical trials. Thus, the countries are conducting clinical trials in huge numbers. The number of registered studies is 3,70,014 as of March 08, 2021 as per ClinicalTrials.gov data.

A clinical development plan (CDP) is a well-structured, comprehensive document and a foundation of good clinical development. The CDP includes goals, protocol, time, and type of documentation. It should also contain all the documents for regulatory requirements like informed consent forms (ICF), informed assent forms, investigator's brochures (IB), study reports, subject narratives, risk management plans (RMP), and periodic safety update reports (PSUR) [2].

The clinical development branches can broadly be categorized as clinical operations and clinical data management (CDM). The teams of healthcare professionals like doctors, nurses, and pharmacists are involved in all the operations that are actually to be carried out in hospitals during a particular clinical trial. They are also called as clinical operations team. The operational responsibilities of this team include the development of documents like protocols, case report files (CRF),

informed consent, questionnaires, etc. The team carries out all the processes carefully including record management in conducting a clinical trial. The team has to observe and analyze every step minutely for the outcome of the study. The team has to collect all the necessary data and handover it to the clinical data management team [1, 2]. The clinical data management team manages the data collected during the clinical trial. The CDM team segregates analyses and stores the data after receiving it. The CDM team also applies statistical applications to the received data to get the conclusion of the study and the significance of the study treatment.

In the process of drug development, clinical trials have prime importance including the preclinical trials to evaluate therapeutic effect of the new treatment. The clinical and preclinical trials also assess the pharmacological risks and benefits involved with the treatment. Computer softwares can help in predicting the safety, efficacy, and effectiveness of the newly developed treatment.

The utilization of computers in clinical development has been progressing rapidly and advancements in it will be more beneficial. The utilization of computers can be made more practical and effective with its applications in the data collection, data storage, and data analysis for the successful run of a clinical trial [3]. Computers upgrade the team functioning in more accurate and faster ways. It provides more effective results from the clinical trial. Considering the future benefits of the use of computers in clinical trials, some software companies developed the clinical data management software, applications, and tools [3, 4]. Thus, the use of software and databases facilitates the operations involved in clinical trial and data management.

Clinical trial researchers are getting adapted to the new e-technology. Researchers are developing and using electronic technology to assist healthcare professionals in different aspects of a clinical trial. The researchers use the technology for recruitment of the patients in the trial and for the creation of computer-based interventions. The researchers are using computers to develop protocols, to communicate with the study personal involved in the trial, to randomize participants, to collect data, and to analyze the results [5, 6]. The websites can be used to distribute the data or information related to the clinical trial. The website development and updating the information in a timely manner guide the volunteers about the information related to the eligibility criteria or recruitment or retention or consent or any other information related to the trial. Also, websites can be utilized to collect data from online surveys by providing online questionnaires. Most recently, clinical trials have started using social media (e.g., Facebook, Twitter, etc.) for easy and regular communication. Also, there are regulatory guidelines available to practice the technology in clinical research.

Broadly, computer technologies can be categorized as per the operations they use for as clinical operations, data management, and pharmacovigilance (PV), that is, post-marketing surveillance of phase IV of clinical trials. The categories of computer technology in clinical development can be categorized as clinical trial management system (CTMS), clinical data management systems (CDM), pharmacovigilance systems/drug safety database/postmarketing adverse events (PMAE), and electronic data capture tools (EDC). The data entry should be done as per the guidelines provided by the sponsors [5, 6]. Also, the data should be validated according to

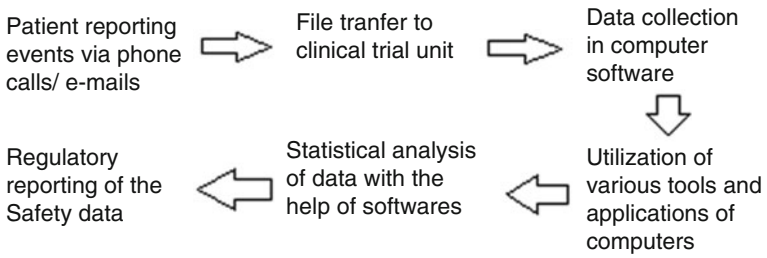


Fig. 10.1 Utilization of computers in various steps in clinical trial management system

the protocol of the trial to avoid any errors or discrepancies in the analysis. If any discrepancies are found, they are investigated and resolved.

The communication between investigators and participants is of prime importance in clinical trial for data collection and data management. If physical meeting is not possible then, the mode of communication can be Email, web sites, File Transfer Protocol (FTP), and video-conferencing via computer or smart phone by using internet. These are used as communication tools for collecting clinical trial data [6]. The tele-health and virtual trials are new emerging concept with computer communication (Fig. 10.1).

There is more use of computers and softwares in clinical trial management system than hardware and other services. Thus, there is an increasing demand of IT-based companies to work in clinical trial sector in a more advanced manner. This will result in accurate and effective data management.

10.2 Clinical Trial Management System

The computerization of all the clinical operations involved in the clinical trials can act as assistance to clinical operations teams. This will help the team in their duties like the storage, management, and analysis of the data from the clinical trial. The information in clinical operations is sponsor information, inclusion criteria for the participants, enrollment of participants, number of participants involved in the clinical trial, information related to the site of the trial and investigators conducting a clinical trial, drug accountability, and financial information related to the trial.

Clinical operations include all the steps involved during the clinical study trial. It consists of receiving of study drug, protocol from the sponsor, recruitment of the patients or volunteers, treatment to volunteers, and record-keeping of all the results including side effects [7, 8]. The recruitment of volunteers is done based on the selection criteria provided by the sponsor. The computers can be used during the recruitment of volunteers in processes like taking the history of the volunteers, their counseling, and informed consent. After the start of the clinical trial, all the records of volunteers including the details of patient like height, weight, allergies, pre-existing diseases, details of prescription of the drug, monitoring of drug administration, administration of medications other than study drug, medication error,

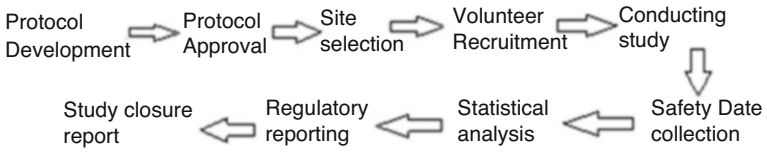


Fig. 10.2 Steps involved in clinical trial management

drug-drug interactions, laboratory test reports and records, blood profile for drug level monitoring, side effects, all the pharmacokinetics parameter of the drug, any adverse reactions, etc., are stored in electronic form in computers [8]. The inventory control and drug information are also managed in the computer database (Fig. 10.2).

The patient record database should be updated on a regular basis as the trial continues. Also, the patient transfer or exclusion from the trial should also be reported. In drug details, name of drug, strength, dose, dosage form, route of administration, schedule of dosage, etc., should be recorded [9]. Software or databases like drug interaction, inventory management, toxicity dose calculation, and statistical evaluation are being used in patient record maintenance during the clinical trial.

The effective way for the improvement in the accuracy of the data used in a clinical trial is the elimination of the data with discrepancies. It is necessary to establish an integrated system which will be capable of running the workflows smoothly with easy-to-use database. The new electronic-based computerized systems used in a clinical trial can be termed generally as 'eClinical'. This will minimize errors and avoid duplicate data collection [9].

Computer-based technology like Health Evaluation through Logical Processing (HELP) is used for defining the criteria and further in medical decision making. These criteria are set by experts from the medical profession. This system provides more accurate monitoring of patient drug treatment during the clinical trial. This improves the efficiency in the work of the team [10]. If similar type of data are entered in the database of the computer, it warns to avoid the duplicity of the results (Table 10.2).

10.2.1 Registries

For the recruitment of patients or volunteers for a specific clinical trial, the use of the registry is done by the investigators. The internet bases registries are used for recruitment, randomization of the trial, and data collection [13]. The registries can also provide information regarding the follow-up schedule and visits of the patient to the study site. The existing registries can be utilized as reference material for the screening of patients, recruiting of patients, randomizing the trial, and collecting the data to draw outcome [14]. This approach improved the efficiency of clinical trials as the updation and maintenance of the registries reduces the time and cost involved. Some of the clinical trial registries are enlisted in Table 10.3.

Table 10.2 Use of computer technology in clinical development stages

Sr. no.	Clinical development stages	Softwares	Applications	References
1	Protocol design	Verified Clinical Trials (VCT), TriNetX	Analyze protocols	[11, 12]
2	Patient management	PatientsLikeMe, TrialSpark, TrialX, SubjectWell, StudyKIK, Seeker Health, mProve Health, Langland, Clinical Connection, Comprehend Clinically, and FindMeCure	Patient recruitment and retention	[11, 12]
3	Clinical data management	Electronic data capturing (EDC) like Oracle Clinical, Medidata Rave and Inform by Phase Forward Electronic health records (EHR)	Capture, store and manage all safety data	[11, 12]
4	Study monitoring	Oracle Clinical, Phase Forward, NetRegulus, Aris Global	Adverse event reporting	[11, 12]
5	Regulatory reporting	SyTech, Wimmer systems	Data analysis and reporting, regulatory submission	[11, 12]

Table 10.3 List of some clinical trial registries with URL

Sr. no.	Clinical development stages	Softwares	Applications	References
1	Protocol design	Verified Clinical Trials (VCT), TriNetX	Analyze protocols	[11, 12]
2	Patient management	PatientsLikeMe, TrialSpark, TrialX, SubjectWell, StudyKIK, Seeker Health, mProve Health, Langland, Clinical Connection, Comprehend Clinically, and FindMeCure	Patient recruitment and retention	[11, 12]
3	Clinical data management	Electronic data capturing (EDC) like Oracle Clinical, Medidata Rave and Inform by Phase Forward Electronic Health Records (EHR)	Capture, store and manage all safety data	[11, 12]
4	Study monitoring	Oracle Clinical, Phase Forward, NetRegulus, Aris Global	Adverse event reporting	[11, 12]
5	Regulatory reporting	SyTech, Wimmer systems	Data analysis and reporting, regulatory submission	[11, 12]

10.2.2 Electronic Health Records (EHR)

The use of electronic health records (EHR) is increasing in the clinical research sector. This provides easy data entry and retrieval of the data related to a specific clinical trial. It also saves the time and cost involved. Clinical investigators use various systems to record the results and outcomes found during the clinical trial [15]. Combinations of systems avoid the duplication of data. The electronic form of case report files is called as electronic case report files (eCRF). The data entered in eCRF are acquired from sponsors, investigators; staff involved in a clinical study, clinical reports, patient laboratory tests, imaging facilities, and electronic health record systems, etc. Electronic health records are essential in clinical trials to recruit and to collect the data. EHRs can identify the potential subjects in the trial [15]. EHRs can recognize the differences between clinical and research procedures. EHRs can withdraw the data to bring into the clinical trial database and can directly collect the study outcome. The investigators also use EHRs to recruit the patients, to acquire the consent, and to collect the data [16].

The health records include number of records like patient registration or recruitment or transfer or exclusion. It also includes patient clinical databases like history or physical examination or laboratory or progress or follow-up reports. Health records also include diagnosis, treatment, and support to the participants involved in the trial [17, 18]. The laboratory or diagnostic reports, food and nutrition intake records, other medication records, and any drug-drug or food-drug interactions are also reported in health records [19]. The reports of all the records of the participants of the trial are case report files (CRF).

10.2.3 Informed Consent

Informed consent has a lot of importance in a clinical trial. The consent of participant is mandatory in a clinical trial. Online consent is prepared with surveys or interviews or questioner. This document should also explain in detail, procedures involved during the clinical trial study, treatment details of the study, and risk and benefits associated with the study treatment. The document should also ask about the volunteer participation of the patients in the clinical trial [19]. The Institutional Review Board (IRB) is also comfortable with the online informed consent form as it has considerable confidentiality than face-to-face consent. In case of face-to-face consent during the recruitment of participants in a clinical trial, the recordings should be kept with the investigator in an electronic format. Only disadvantage of online consent is as there is no face-to-face contact, the investigator is ambiguous and has no assurance about the actual participation of the volunteer [19]. This challenge can be overcome with the more advanced utilization of the e-technologies for effective communication between investigators and participants (Fig. 10.3).

In December 2016, USFDA released guidance document for the use of electronic formats in informed consent. As per this guidelines, electronic informed consent (eIC) can utilize the electronic formats, such as photos, graphics, movie, videos,

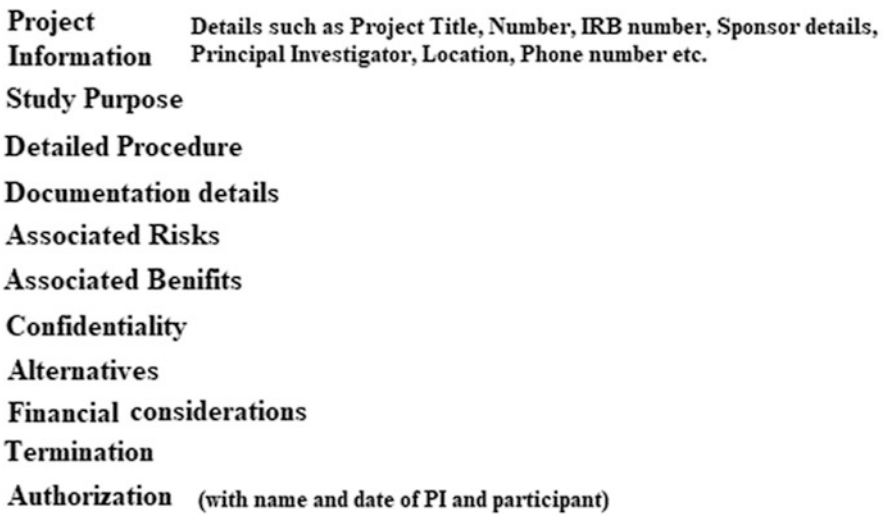


Fig. 10.3 General template and points to be consider for informed consent preparation for clinical trial

audios, etc., for communication with the volunteers and obtain the informed consent for a particular clinical trial. There are some advantages of eIC to the volunteers like more convenient, easy to understand, less anxiety about participation in trial, etc. [20, 21]. Also, there are benefits of eIC to the investigators, such as cost effectiveness as no travel expenses and paperwork are involved, easy and more patient recruitment, more ethical consent, etc.

The documentation and information presented to the subject to obtain the consent from volunteer should meet the applicable regulatory requirements. FDA's requirements for electronic records/electronic signatures, informed consent, and IRBs are set forth in 21 CFR parts 11, 50, and 56, respectively. Health and Human Services (HHS) requirements regarding the protection of human subjects are set forth in 45 CFR part 46. IRB requirements are same for both the paper and electronic form of the informed consent. IRB needs format in approved language, all elements in IC required by FDA/HHS, signature, participation of staff, and copy to volunteers [20, 21].

The guidance for use of electronic informed consent has been prepared jointly by the Department of Health and Human Services (HHS) Office for Human Research Protections (OHRP) and the Food and Drug Administration (FDA). This guidance is intended for Institutional Review Boards (IRBs), investigators, and sponsors engaged in or responsible for oversight of human subject research under HHS and/or FDA regulations [20, 21]. This guidance provides recommendations on procedures that may be followed when using an eIC are as follows:

- Ensure protection of the rights, safety, and welfare of human subjects
- Facilitate the subject's comprehension of the information presented during the eIC process
- Ensure that appropriate documentation of consent is obtained when electronic systems and processes that may employ multiple electronic media are used to obtain informed consent
- Ensure the quality and integrity of eIC data included in FDA applications and made available to FDA during inspections

10.2.4 Clinical Data Management Systems (CDMS)

The responsibilities of the clinical data management team members are to collect, manage, store, track, and analyze the clinical data acquired during the clinical trial. The data are collected through direct face-to-face meetings, video conferencing, telephonic conversations, fax, e-mail, reporting on websites, etc. This collected data are managed for their confidentiality, security, validity, and assessment for a particular outcome. The manpower involved in clinical data management is head of the CDM department, project managers, data managers, programming managers, clinical data associates, medical order, and biostatisticians [22]. The computer technology systems assist the team for accurate and effective management of data. The computerized technology is called as clinical data management system (CDMS).

The software programs or databases are designed in such a way that one can easily track and retrieve the data for analysis at any point in time. The database can examine blank or incomplete data to avoid any deviation from the rules. The database provides accurate information and helps to site investigators about the follow-up check-ups of the participants in a clinical trial. The date and time are also saved with the name of personnel from the clinical data management team who enters data in the database during a clinical trial. If there are changes like addition, deletion of data, or any modification or updating in data, the database stores the previous version of data with date and time and updates changes. As the database saves all the previous versions of data entered in software, it is effortless data tracking which helps during the periodic checks, clinical audits, and regulatory checks. All the data entered into CDMS are converted into electronic datasets. This electronic form can be analyzed further with the use of the statistical analysis system (SAS). These software programs can be well utilized by statisticians to report a trend in treatment and response and also to draw the final conclusion of the data driven from a clinical trial. The submission of data analysis is finally done to the regulatory authorities. CDMS software programs are updated from time-to-time to capture the data automatically from electronic documents. CDMS works more efficiently in data storage, management, and analysis. Some examples of CDMS include Oracle Clinical (Oracle Corp.), ClinTrial (Phase-forward), and the very new SAS Drug development (SAS Inc).

Clinical data management constructs a fundamental part in the clinical trial studies. The accuracy of the data has vital importance during the conclusive

statements and authentication of the clinical study. The quality of data entered in the CDMS database is crucial and must be valid. Thus, it is obligatory that the software programs or database used during the data entry in clinical data management (CDM) must be validated. Clinical data management certifies the collection, integration, validation, and availability of the clinical trial data. The standard operating procedures (SOPs) and good clinical practices (GCPs) carried out during the clinical trial ensure the quality of the data. So, the team involved in clinical operations at the site should be well qualified and trained for the study. The prime objectives of CDM are to provide high-quality data to observe errors during data collection or entry, to keep a check on missing data, and to minimize the errors in data for analysis [23]. For more accurate, complete, and reliable data collection and data entry of the clinical study, several practices have been developed. This can be effortlessly achieved with the use of updated applications of the software. These applications can be used more effectively for data analysis in audits and regulatory checks with the easy detection of the discrepancies in the data [23, 24]. In clinical data management, there is a need of software databases for capturing the data in electronic form and for preparing data for electronic submissions to the regulatory authority. The databases facilitate and speed up the clinical trial process.

The development and utilization of the software is carried out as per the stages involved in the clinical trial study. The software databases for the clinical trial development are developed for the protocol designing and its execution, for IRB checks, and for the investigator relationship management. The clinical trial management stage utilizes the software databases for integrated CTMS and International Management Package for Administration of Clinical Trials (IMPACT). The patient subject and study enrollment stage in clinical trial avails the databases for their applications in investigator relationship management electronic data capture (EDC), case report forms (CRF), electronic patient diaries, etc. The databases used in the study monitoring and reporting stage have applications like clinical trial supply management, cost tracking document management, and adverse event reporting [23, 24]. The study completion and regulatory filing stage use databases for data analysis and reporting, regulatory submission assembly, communication and review, custom solutions, etc.

For accurate statistical analysis, good quality data are required. The data should satisfy the parameters defined in the protocol and should comply with it. The exclusion of the data can be done which does not comply with the specifications provided in the protocol. Regulatory agencies use to check such type of data also.

The updation and new development in the software and databases used in clinical trials provide more reliable and timely data for conducting an effective and successful clinical trial study. The new developments are with features like new design tools, commercial databases, highly secure, confidential, time saver, and easy to use. The good clinical data management practices deal with the data acquisition, privacy, electronic data capturing, case report form (CRF) printing, preservation of CRF, data storage, validations, and many more [25]. There is a separate quality assurance team to monitor the study and adherence to SOPs while conducting the clinical trial.

10.2.5 Electronic Case Report Forms (e-CRF)

An electronic case report form (e-CRF) is an electronic-based system. The CRF is developed by the sponsors to collect the data during the clinical trial. The sponsors utilize CRF for the data collection of the volunteers participated in a clinical trial. CRF includes all the details of the volunteers including the adverse events (AE). The size of the CRF depends on the condition of the patient and data collected during the clinical trial over a period of months. CRF contains all the electronically captured data which includes detailed information about the referral protocol, SOPs, and other documents utilized for a particular clinical trial [25]. Monitoring collected data and managing the results of the trial are the responsibilities of the sponsor.

CRF records all the data collected during the participation of patients in a clinical trial. Generally, these data contain information regarding patient including the study number assigned by the investigator with their medical records in detail. The supervision of the process is performed by the Institutional Review Board. The accuracy of CRF data is a major goal of the sponsor. The regular audits are carried out by the sponsor to monitor and to maintain the correct data in CRF and to minimize the errors (Fig. 10.4).

After the approval of protocols and CRF for a clinical trial, explanations and notation of the CRF are prepared. The approval of blank CRF and preparation of other required documents need to be carried out. Further, to check the database set up, it is filled up with the test or dummy data. Then the database quality is examined

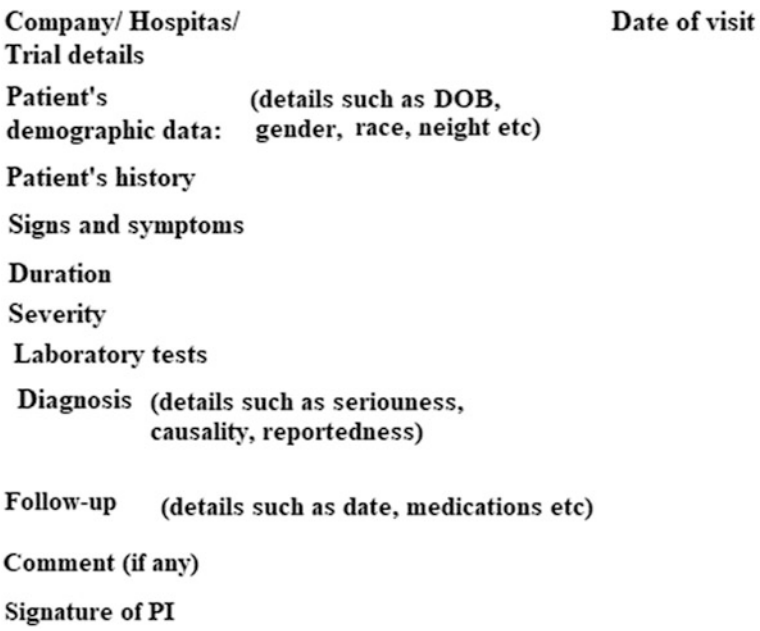


Fig. 10.4 Generalized case report form (CRF) template

as per the requirements. After quality control checks of the database, the dataset is evaluated for validation. After completion of validation, the report is checked for the discrepancies and their management. Lastly, the statistical analysis is carried out and the final report will be generated [26].

10.2.6 Pharmacovigilance/Drug Safety Systems

Pharmacovigilance is post-marketing surveillance data. The drug safety teams maintain the data in safety databases. All the adverse events (AE) or any allergies for the marketed drug are reported, collected, and maintained in safety databases [26]. The medical reviewers use databases like Medical Dictionary for Regulatory Activities (MedDRA), WHO Drug Dictionary Enhanced (WHO DDE), and WHO Adverse Reactions Terminology (WHOART) maintained by Uppsala Monitoring Centre (UMC). These databases are used for reporting, coding, and interpreting all the adverse events reported. These databases help to analyze the narration by the patient and segregate it into low-level term (LLT), preferred term (PT), high-level term (HT), and system organ class (SOC).

The adverse drug reactions (ADR) due to high dose, missed dose, any toxicity or allergies related to the specific marketed drug are reported in the safety database. Sometimes ADR is not related to the drug. Thus, the causality assessment is important in pharmacovigilance to determine the relationship of ADR with the drug. There is a separate tab for causality assessment in a safety database with options yes or no. The ADR can be serious and may result in death. The different seriousness criteria are mentioned in the database like death, permanent disability, congenital anomaly, hospitalization, and prolonged hospitalization for the serious adverse events (SAE) reported by the patient. All the adverse events are analyzed by healthcare professionals for seriousness and causality and reported. All the serious and non-serious events reported by the patients are recorded in the database after assessment by health care professionals like doctors, dentists, nurses, or pharmacists. Sometimes adverse events are reported voluntarily by patients. This type of reporting is called as spontaneous reports. All the reports are reported to the sponsors. The follow-up records are also reported and analyzed when it is required and recommended.

All the reports are regularly audited by the sponsors and regulatory authorities. A safety database team records the data, manages it, stores it, analyzes, and reports the data to the sponsor [27]. There are a number of safety databases developed for the data management in PV to report AE, ADR, and SAE. The updation and modifications in the database help the sponsor in reporting the adverse events in a more effective way and it saves time. Some examples of safety database systems are Argus, Oracle AERS, etc.

10.2.7 Electronic Data Capture (EDC)

An electronic data capture (EDC) is a computerized system specially developed for the data collection in electronic format in a clinical trial. This database system replaced the paper-based traditional methods of data collection [28]. This method is simpler, efficient as well as cost-effective. EDC database system also eliminates data discrepancies and minimizes queries and errors. These database systems are widely used by contract research organizations (CRO) and pharmaceutical companies for clinical research data collection. The EDC systems can be used in all the clinical trial phases but particularly provides benefits in phase IV, that is, post-marketing surveillance phase.

One can log into the database by using their own specific and confidential access details and can fill the data electronically on the CRF page. This CRF is further processed by the data management team. No field should be blank or incomplete. All the fields should be completely and correctly filled to avoid any error and further discrepancies in data management. For EDC technology utilization, there is a need for a particular infrastructure with computers, databases, and internet facilities at each site [29, 30]. All the data can be centralized by connecting all the study sites via network connectivity for easy communication and management. EDC systems are used for the survey for behavioral studies or health promotion clinical research. Ecological momentary assessment (EMA) is a new approach for data collection. More recently, the study investigators are collecting the data with the help of wearable devices that are synced with the smart phone applications [31]. There are many software databases that are used in the clinical trials to manage the data securely and keep in an easy accessible format.

10.3 E-Technologies Used in Clinical Trials for Data Management

All types of life science research organizations are receiving the advantages of the use of software in clinical trial management. As these software databases are utilized in every operation involved in the clinical study from data entry, statistical analysis and validation to regulatory submission, and archiving of the clinical trial. These databases are utilized in all the phases of a clinical trial with varying numbers of participants. These databases have the potential to operate with a huge amount of data. So, they play an important role in multicenter trials.

10.3.1 Softwares Used in Clinical Data Management

Some examples of the software databases used for clinical data management are ORACLE CLINICAL, MACRO, RAVE, CLINTRIAL, eClinicalSuite, EZentry, etc. Some best clinical trials management softwares that can be listed are IBM Clinical development, Ege CTMS, MasterControl CQMS, EXPeRTeClinical,

Table 10.4 Key features and functions of some of the databases

Sr. no.	Software	Application	References
1	Medrio	Fast, simple, and affordable system for electronic data capture (EDC)	[34, 35]
2	Castoredc	Affordable and easy to use Used in electronic data capture (EDC), eCRF for data collection	[34]
3	Flexdatabases	Storage and processing of the clinical data from investigators, subjects, and sites	[34]
4	Clinion	Streamlines clinical trials by validating, intuitive EDC, and manages multiple studies	[34, 35]
5	cambridgecognition.com	Provides cognitive assessment software for clinical trials, academic research, and healthcare provision	[34]
6	CRFweb	Offers eCRF, ePRO etc. with MedDRA coding Simple and fast to set up and use	[35]
7	clirinx.com	Provides a web-based CTMS, EDC, ePRO, and Patient Diary and informatics services for clinical trials and research	[34]
8	mednetstudy.com	Provides eClinical and EDC solutions for clinical trial management	[34]
9	Ripple	Manages patient recruitment in a clinical trial	[35]
10	EDGE	Provided faster access to real time study data	[35]

BioClinica CTMS, Ofni Clinical, Cliniplus, CANTAB, Data LabsXC, OpenClinica, Clinical Research IO, Trial master®, LifeSphere CTMS, Allegro CTMS, OnCore Enterprise Research, ActiTrack, Castor EC, Clinical Conductor CTMS, BSI CTMS, STARLIMS, TargetHealth, myClin, Clinical Studio, etc. [32, 33].

Some databases can be enlisted as Clinical software developed by SAS solution, Business Intelligence software by Cognos, Symmetric software by symmetric life sciences, OpenClinica software by Akaza Research, DMSYS software by the organization Sigamsoft International, Clinical software by Progeny Software, LLC., software EZ—entry by EpiData software system, Oracle Clinical software by Oracle, CLIN-E2E software by TCS, etc. The major drawbacks in the use of the software are the cost as they are very expensive and only need to train the technical person to operate it [32, 33]. The key features and functions of some of the databases are summarized in Table 10.4.

10.3.2 Digital Wearable Medical Health Devices

The wearable medical or health devices are manufactured with the goal of easy to use and comfort in wearing by the patients. These devices help in easy monitoring of patients to give more accurate and frequent data collection of the body responsible for the diagnosis, treatment, and prevention of the diseases. The devices can be tied up to the clothing, shoes, watches, eyeglasses, etc. [36]. These devices can be synced

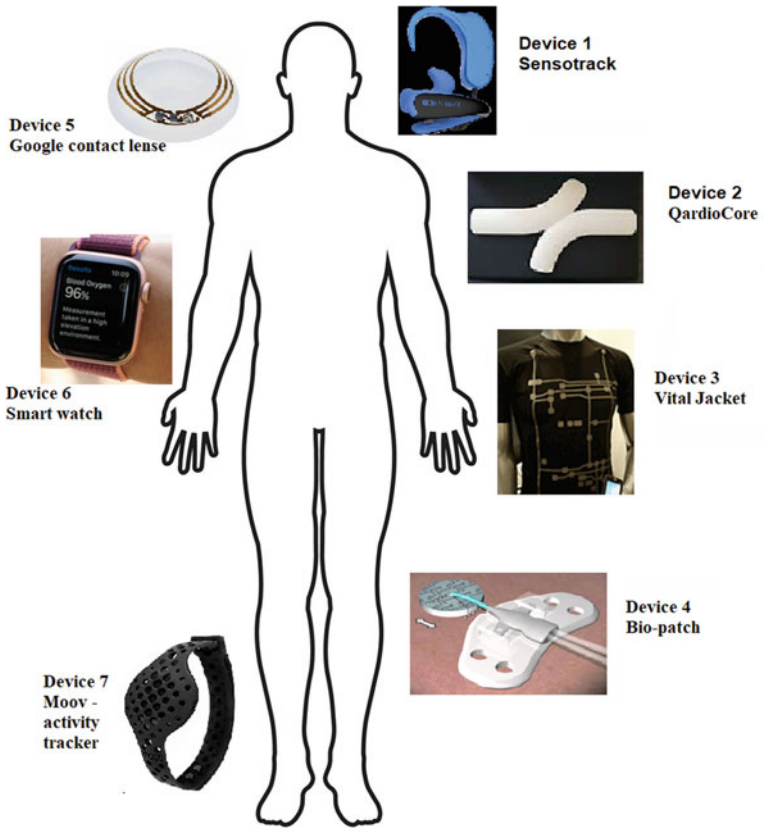


Fig. 10.5 Examples of some wearable medical health devices

with smart phones or laptops to collect and save the data. These data help in the easy interpretation of the outcome of the study after statistical treatments. Wearable devices are used for more effective patient health management and disease management [36, 37]. These technologies can make a direct influence on decisions during the clinical trial.

Figure 10.5 represents advanced wearable devices. The device 1 is the SensoTRACK ear sensor which tracks fitness activity from the ear. This single device can measure the heart rate, respiration rate, oxygen saturation, and blood pressure. The device 2 represents QardioCore which is an electrocardiogram (ECG) monitoring device. The device 3 is the Vital Jacket® t-shirt which can help in monitoring and detection of cardiac health of a patient through ECG monitoring. The device 4 shows the BioPatch which detects and prevents skin infections. The device 5 represents the Google Contact Lens which assists in the measurement of glucose level in eye tears. The device 6 is the Smartwatch which can act as an activity tracker, thermometer, heart rate, and blood pressure monitoring device. The

device 7 shows the Moov which is an activity tracker. This device can count steps walked in a day, speed of activities, calories burned, measure activity levels, tracks your geo location, determine the body posture, etc. This also measures sleep patterns, heart rate, and blood pressure [37, 38].

There are some challenges with these devices and smart phone applications. The authenticity of the results and data is one of them. When the e-technologies and electronic devices are being used in clinical trials, there is a need to standardize, optimize, and regulate them for their reliability, consistency, responsiveness, and specificity, the relevance of the results, and the data they produce. The security maintenance, that is, privacy and confidentiality of the participant, in a clinical trial is also a challenge [37, 38]. These issues need to be addressed for improvement in user acceptance, modified functions, and an increase in practical applicability.

10.3.3 Health Apps

Nowadays, other than computers and laptops, use of mobile, smart phone devices, tablets, and other advanced wearable devices is very ordinary. Wearable devices are worn on the skin or they are kept close to the skin. When these devices are in contact with skin, they can determine and examine the body signals. The devices communicate and interpret the signals into data which give feedback regarding the vital signs of the person wearing the device. Recently, smart watches and activity trackers are more popular [39].With the ease of using health applications, they are utilized substantially. So, the development in the advancement of these technologies is rapid. The anticipation regarding their use and efficiency is more.

Smart phones and wearable devices can also be utilized to assess the location of patients through the global positioning system (GPS). The patients can also be monitored for the study compliance or any other side effects. The communication of patients or volunteers involved in the clinical trial to investigators is becoming well-liked through the internet. Data collection at any point of time is possible through mobile technologies [40]. In mobile technology advancements, Apple launched its ResearchKit software in 2015, with the purpose of medical and health research.

There are many health applications that are available on a smart phone. For example, apps like BetterDoctor, Medscape, radiology, Vitals, iTriage, MedPage, MedCalc, CareFinder, Medvana, Drug Trials, Pedicine, Zocdoc, FreeCare, etc., are also available to help the patients for regular health checkups in an easy way [40, 41].

There are separate guidelines for use of medical health apps for protected health information (PHI) which are subjected to the federal Health Insurance Portability Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health Act (HITECH). There are provisions of different punishments including fine and penalty for the data security issues. Also, the FDA has formulated separate guidelines for using electronic source data in clinical trials [41]. The data generated by the use of wearable devices are challenging as they are huge in amount.

But it also opens some future opportunities for artificial intelligence (AI) techniques applications.

10.4 Big Data in Clinical Trial

The term big data in clinical trial management refers to the data collected via electronic technologies and databases. In clinical research, data collection ranges from protocol development, patient recruitment to statistical analysis of data for outcome, and conclusion. At every stage, collection of a huge number of data is carried out for more efficient and effective clinical trials with better quality results. It is difficult for investigators and sponsors to deal with the generated data as these data are in large amounts and complex. To manage and process the data with the use of traditional processing tools for data management or database management tools becomes difficult.

Some new analytical approaches are utilized to optimize and manage the big data to provide a more effective clinical trial [42]. The mining of big data reduces the risk involved in the clinical trial. These approaches utilize previous study data to optimize the study design and analysis of the data. These approaches predict the modeling with EHR and other databases. The big data accelerate the clinical trial efficiency with improved trial design, recruitment of patients, selection of the study site, and monitoring. These approaches empower the investigators, sponsors, and regulatory agencies in the process of drug discovery and drug marketing.

With the advanced technologies, there is huge data collection that is carried out and the data can be in a structured or unstructured format. The whole data are organized in a specific comparative format to simplify the analysis. The technologies streamline the process of storage and analysis of real-world data. This handles the size, frequency, and the specific format of the data. This also resolves the query or any review comments in data. This eases the handling of big data and its integration into the analysis [42]. This provides a better and rapid analysis of the big data. The use of structured query language (SQL) for data extraction is a very important step in the preparation of data for analysis [43]. The structured use of this technique saves a lot of time in processing. It also saves space in the hard disk.

10.5 Artificial Intelligence (AI) in Clinical Trial

Artificial intelligence-based clinical data management systems can improve the overall clinical trial process. It offers different methods for easy data collection, data management, and data storage for further use. AI assists in process automation to reduce errors, insights into the data analysis, and also help in critical decision making for efficient the clinical trial process [43, 44].

AI is a powerful technology for business improvement. One should understand the nature of the task which needs to be completed with the use of AI and the strength and challenges involved [45]. It is important to set the desired standards into

software in order to save time and resources. Advances in AI and its applications in clinical study trials are still in progress and promise the better healthcare management. So, promoting innovation in AI-based technologies and its utilization is important [46]. The AI-based tool should be adopted globally in clinical trial studies. Thus, these techniques accelerate and provide efficient trials as it intensifies the improved quality data collection and analysis.

10.6 Data Collection and Data Management

The data collection at every step in the clinical study is the prime important process in each and every clinical trial. In all the phases of clinical trials, data are collected and final outcome of the study is concluded based on the analysis of the data. A typical sequence of data collection while conducting the clinical trial can be summarized as the data entry at the study site, resolution of the query if any. The data from the site are reviewed and managed by the medical review team, clinical data management team, and medical coding of the events is carried out and then reconciliation is done [47]. These data are further reviewed by the statistical analysis team and medical reviewers. The resolution of the quires generated is done. The database selected by sponsors and investigators are used for further statistical reports.

In clinical trial management system, data collection is a challenge as data are huge. So, the data collection is carried out using e-technologies and computer. All the advanced softwares which collect, manage, and store the most accurate data are used in clinical management. The data storage can be done in computer hard drive, cloud storage, etc. [48]. The data collection and storage became more easy and accurate with the use of computers and electronic forms of documents (Table 10.5).

10.7 Advantages of E-Technologies in Clinical Trials

The electronic technologies and its advances are rapidly changing the way of conducting clinical trials. The advanced communications, knowledge transfer, and trouble-resolving techniques have overcome the restrictions in the traditional aspect of dealing with data and conducting a clinical trial. These advancements broaden the scope, opportunities, and applications of clinical research. The role of web applications and smart phone-based applications are widely becoming popular [47, 48]. The advantages of the utilization of computer technologies including softwares in clinical study trials can be summarized as follows:

- The computer applications and its advancement in clinical studies improve efficiency in the overall process.
- The e-technology promotes more research and development in the clinical study sector.

Table 10.5 Summary of technologies in handling clinical trial data

Sr. no.	Technology	Applications	Step in clinical data management	References
1	Internet of things (IoT), e-mails Wearable devices	Data collection Contracts: Informed consent, survey	Collection and storage of the data	[33, 48, 49]
2	Artificial intelligence (AI)	Automation in data processing Machine learning	Processing of structured data, Analysis of the data based on keywords	[33, 48, 49]
3	Differential privacy	Subset of AI, identifies clinical trial site Protects patient privacy by de-identification	Storage and easy access of the data	[48, 49]
4	Encryption	Subset of AI, Analysis of sensitive and protected data	Storage and analysis of the data	[48, 49]
5	Machine learning	Subset of AI, analysis of structural data with the use of statistical methods	Analysis of the data	[48, 49]
6	Block chain	Subset of AI, Provide accurate documentation	Collection and storage of the data	[48, 49]
7	Deep learning	Subset of AI, predictive model for structured data	Analysis and processing of the data	[33, 48, 49]
8	Cloud computing	Manages the large scale data	Collection and storage of the data	[33, 48, 49]

- The patient recruitment including contacting, screening, and enrollment into the trial becomes easy with the use of computerized registries, EHR, etc.
- Tools used to have effective and timely communication between sponsors, investigators, and participants of the clinical trial.
- With easy, effective, and communicating often with the patients, the retention of the participants in the trial is improved.
- With the use of electronic documentation systems, data collection, storage, and management process become smooth and error-free.
- The smart phone apps and advanced wearable technology permit easy and automatic collection of data in huge numbers.
- The e-technology and tools reduce the cost involved in the process as it reduces the number of personnel or other staff involved in the trial due to automation.

10.8 Challenges and Limitations of E-Technologies in Clinical Trials

There are some challenges for using computers in clinical trials. These challenges can be conquered with proper designing and some precautions. Thus, there is a need for further more research to deal with the limitations [47, 48]. The challenges and disadvantages of the utilization of e-technologies including softwares in clinical study trials can be summarized as:

- Firstly, there is a need for suitable infrastructure for e-technology utilization in clinical trials. As the patients are screened, recruited, and monitored by using computers, the facility set up should be specific and proper.
- All the advanced automated databases are expensive.
- For utilizing updated databases, there is a need for well qualified and trained personnel with the knowledge of the healthcare research, that is, clinical studies as well as computers, that is, database to be used for the data collection in a clinical trial.
- The anonymous intervention in electronic data collection is also a challenge as it is difficult to identify the staff who worked on it. To tackle this problem, user authentication and protected sign-in into the database are needed for effective data collection.
- The analysis of big data is a challenge. The accuracy of data is doubtful in big data. It is difficult to identify how data are collected, which information of data are to be captured, and how to analyze it.
- The electronic formats should have more effective, meaningful questions for accurate data collection.
- Incorporation of computerization in clinical studies creates the challenge of confidentiality issues. The privacy policy in a clinical trial may hamper because of e-technology and tools. Thus, there are possibilities of violation of regulatory guidelines provided by IRB, regulatory agencies, and other organizations involved with the clinical study trials. The privacy and confidentiality of the participants is a major concern.

10.9 Future Directions

The advancements in the e-technology and tools used in the clinical trials come up with opportunities for extra research. Further research in e-technology utilization in clinical trials will improve the design of the database, collection and handling of data, and analysis of the data collected during the trial. The latest or updated technology needs to be tested and validated for the expected performance and should prove their benefits over the traditional or currently available tools in clinical trials. Big data are always a challenge for healthcare management personnel. Future research opportunities will provide validity of the tools and more effective clinical

trials with valid results and outcomes. Also, there is a need for research on the availability of infrastructure to analyze big data.

The notable increase in the employment of health monitoring devices is predicted by Market Research Future. Also, the smart phone-based, internet-based, some smart wearable healthcare management devices will be more popular in the future. Future research in the advancement and usability of applications of these devices will be helpful for clinical trials. This will give more improvements in efficiency and cost-effective clinical trials. Thus, the need to lower the cost of healthcare and improve the health management of the participants involved in the clinical trial will be targeted.

These innovations are able to improve overall public health. There should be comprehensive efforts to construct some advanced innovations with more safety, efficiency, accuracy, and effectiveness in the clinical development process.

References

1. Jagdale SC, Hable AA, Chabukswar AR (2020) Chapter 124: Protocol development in clinical trials for healthcare management. Section 15: Medical Ethics and Law. IGI Global Publisher, pp 1797–1814
2. Thorat SB, Banarjee SK, Gaikwad DD, Jadhav SL, Thorat RM (2010) Article 019—clinical trial: a review. Int J Pharm Sci Rev Res 1(2):101–106
3. Bandameedi R (2016) Provenance of computers in pharmacy. Clin Pharmacol Biopharm 5(1):1–7. https://doi.org/10.4172/2167-065X.1000153
4. Arditi C, Rege-Walther M, Durieux P, Burnand B (2017) Computer-generated reminders delivered on paper to healthcare professionals, effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 7:CD001175. https://doi.org/10.1002/14651858.cd001175.pub4
5. Doolan DF, James BC (2003) The use of computers for clinical care: a case series of advanced U.S. sites. J Am Med Inform Assoc 10(1):94–107. https://doi.org/10.1197/jamia.M1127
6. Koop A, Mosges R (2002) The use of handheld computers in clinical trials. Control Clin Trials 23:469–480. https://doi.org/10.1016/S0197-2456(02)00224-6
7. Smietana K, Siatkowski M, Moller M (2016) Trends in clinical success rates. Nat Rev Drug Discov 15:379–380. https://doi.org/10.1038/nrd.2016.85
8. Wong CH, Siah KW, Andrew WL (2000) Estimation of clinical trial success rates and related parameters. Biostatistics 1–14
9. Hay M (2014) Clinical development success rates for investigational drugs. Nat Biotechnol 32: 40–51. https://doi.org/10.1038/nbt.2786
10. Institute of Medicine (US) Committee on Improving the Patient Record, Dick RS, Steen EB, Detmer DE (eds) (1997) The computer-based patient record: revised edition: an essential technology for health care. National Academies Press (US); 3, Computer-Based Patient Record Technologies, Washington, DC. https://www.ncbi.nlm.nih.gov/books/NBK233053
11. Trial expert (2020) Software and data management tools in clinical trials. Credevo Articles
12. Murugesan D, Banerjee R, Kumar GR (2016) Clinical data management importance in clinical research. Asian J Pharma Clin Res 9(2):59–62. https://doi.org/10.22159/ajpcr.2016.v9s2.13940
13. List of clinical trial registries From Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/List_of_clinical_trial_registries
14. Listing of Clinical Trial Registries (2015). https://www.hhs.gov
15. Doyle-Lindrud S (2015) The evolution of the electronic health record. Clin J Oncol Nurs 19(2):153–154

16. Sahneya R, Sharma M (2018) Electronic health records: a general overview. Curr Med Res Pract 8:67–70. https://doi.org/10.1016/j.cmrp.2018.03.004
17. Seymour T, Frantsvog D, Graeber T (2012) Electronic health records (HER). Am J Health Sci Third Quarter 3(3):201–210. https://doi.org/10.19030/ajhs.v3i3.7139
18. Odekunle FF (2016) Current roles and applications of electronic health record in the healthcare system. Int J Med Res Health Sci 5(12):48–51
19. Shaik M, Kumar PS, Ali L (2020) Importance of informed consent form and use of electronic informed consent form. Int J Pharma Sci Rev Res 61(2):Article No. 05, 23–27
20. Lunt H, Connor A, Skinner H, Brogden G (2019) Electronic informed consent: the need to redesign the consent process for the digital age. Intern Med J 49:923–929. https://doi.org/10.1111/imj.14339
21. Gupta UC (2013) Informed consent in clinical research: revisiting few concepts and areas. Perspect Clin Res 4(1):26–32. https://doi.org/10.4103/2229-3485.106373
22. Gangarapu P, Reddy JP, Sundaram AM, Sabareesh M, Prabhakaran V (2015) Regulatory considerations in clinical data management. Int J Curr Pharm Clin Res 5(3):143–152
23. Krishnankutty B, Bellary S, Kumar NBR, Moodahadu LS (2012) Data management in clinical research: an overview. Indian J Pharmacol 44(2):168–172. https://doi.org/10.4103/0253-7613.93842
24. Lu Z, Su J (2010) Clinical data management: current status, challenges, and future directions from industry perspectives. Open Access J Clin Trials 2:93–105. https://doi.org/10.2147/OAJCT.S8172
25. Bellary S, Krishnankutty B, Latha MS (2014) Basics of case report form designing in clinical research. Perspect Clin Res 5(4):159–166. https://doi.org/10.4103/2229-3485.140555
26. Zhengwu L (2010) Technical challenges in designing post-marketing eCRFs to address clinical safety and pharmacovigilance needs. Contemp Clin Trials 31(1):108–118. https://doi.org/10.1016/j.cct.2009.11.004
27. Szarfman A, Tonning JM, Doraiswamy PM (2004) Pharmacovigilance in the 21st century: new systematic tools for an old problem. Pharmacotherapy 24:1099–1104. https://doi.org/10.1592/phco.24.13.1099.38090
28. Frandsen M, Walters J, Ferguson SG (2014) Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. Nicotine Tob Res 16:247–251. https://doi.org/10.1093/ntr/ntt157
29. Khatri C, Chapman SJ, Glasbey J, Kelly M, Nepogodiev D, Bhangu A, Fitzgerald JE (2015) Social media and internet driven study recruitment: evaluating a new model for promoting collaborator engagement and participation. PLoS One 16:1–11. https://doi.org/10.1371/journal.pone.0118899
30. Ryan GS (2013) Online social networks for patient involvement and recruitment in clinical research. Nurse Res 21:35–39. https://doi.org/10.7748/nr2013.09.21.1.35.e302
31. Lu Z (2010) Electronic data capturing technology for clinical trials. IEEE Eng Med Biol Mag 29(2):95–102. https://doi.org/10.1109/MEMB.2009.935726
32. Kaur S, Gazali, Singh I (2017) Artificial intelligence based clinical data management systems: a review. Inform Med Unlocked 9:219–229. https://doi.org/10.1016/j.imu.2017.09.003
33. Rosa C, Campbell ANC, Gloria MM, Brunner M, Winstanley EL (2015) Using e-technologies in clinical trials. Contemp Clin Trials 45(00):41–54. https://doi.org/10.1016/j.cct.2015.07.007
34. Smith S (2016) Best clinical trial management software, small business software reviews. Services Insight and Resources. https://cloudsmallbusinessservice.com
35. Best Clinical Trial Management Software. https://www.capterra.com/
36. Izmailova ES, Wagner J, Perakslis ED (2018) Wearable devices in clinical trials: hype and hypothesis. Clin Pharmacol Ther 104(1):42–52. https://doi.org/10.1002/cpt.966
37. Yetisen AK (2018) Wearables in medicine. Adv Mater 30:1–26. https://doi.org/10.1002/adma.201706910
38. Dias D, Cunha JPS (2018) Wearable health devices—vital sign monitoring, systems and technologies. Sensors 18:2414. https://doi.org/10.3390/s18082414

39. Chindalo P, Karim A, Brahmbhatt R, Saha N, Keshavjee K (2016) Health apps by design: reference architecture for mobile engagement. Int J Handheld Comput Res 7(2):34–43. https://doi.org/10.4018/IJHCR.2016040103
40. Kayyali R, Peletidi A, Ismail M, Hashim Z, Bandeira P, Bonnah J (2017) Awareness and use of mHealth apps: a study from England. Pharmacy 33:1–14. https://doi.org/10.3390/pharmacy5020033
41. Lee CV (2014) Mobile devices and apps for health care professionals: uses and benefits. Pharm Ther 39(5):356–364
42. Martin L, Hutchens M, Hawkins C (2017) Trial Watch: Clinical Trial Cycle times continue to increase despite industry efforts. Nat Rev Drug Discov 16:157. https://doi.org/10.1038/nrd.2017.21
43. Walczak S (2018) The role of artificial intelligence in clinical decision support systems and a classification framework. Int J Comput Clin Pract 3:31–47. https://doi.org/10.4018/IJCCP.2018070103
44. Nayak VS, Khan MS, Shukla BK, Chaturvedi PR (2016) Artificial intelligence in clinical research. Int J Clin Trials 3(4):187–193
45. Harrer S, Shah P, Antony B, Hu J (2019) Artificial intelligence for clinical trial design. Trends Pharmacol Sci 40(8):577–591. https://doi.org/10.1016/j.tips.2019.05.005
46. Pozzi G (2019) Clinical information systems and artificial intelligence: recent research trends Carlo Combi. IMIA Yearb Med Inform 28(1):83–94. https://doi.org/10.1055/s-0039-1677915
47. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J (2014) Clinical development success rates for investigational drugs. Nat Biotechnol 32:40–51. https://doi.org/10.1038/nbt.2786
48. Paul J, Seib R, Prescott T (2005) The Internet and clinical trials: background, online resources, examples and issues. J Med Internet Res 7(1):5. https://doi.org/10.2196/jmir.7.1.e5
49. Chator A, Shah KA (2019) Prepared approach to technology in clinical trials, clinical leader, Guest column. https://www.clinicalleader.com/doc/a-prepared-approach-to-technology-in-clinical-trials-0001

Artificial Intelligence and Its Applications in Drug Discovery, Formulation Development, and Healthcare

11

Debadri Banerjee, Deepti Rajput, Surojit Banerjee, and Vikas Anand Saharan

Contents

D. Banerjee · D. Rajput · S. Banerjee · V. A. Saharan (✉)
Department of Pharmaceutics, School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Dehradun, Uttarakhand, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_11

Abstract

Artificial intelligence (AI) is a vast multidisciplinary field which equip machines with cognitive powers like ability to perceive reason, learn, abstract, and act. Machine learning (ML) and deep learning (DL) are two subsets of AI. AI-powered CADD tools are widely used by scientists in decision-making at key steps in drug discovery programs. ML-based expert systems and DL artificial neural networks (ANNs) are now tremendously popular for predicting the biological activity, toxicity, physicochemical property, quality or compositions of pharmaceutical formulations, and estimating interactions of drug and its target. AI-based expert systems have been developed and reported for both conventional and novel drug delivery systems. Healthcare applications of AI involve personalized treatments, medical diagnosis, and epidemic outbreaks. In entirety, the healthcare field seeks remarkable advancements, and much collaboration among companies like pharmaceuticals, medical devices, life sciences, and AI technology is observed very recently. This chapter focuses on introducing the concepts of AI, algorithms, and their applications in advancing drug discovery, formulation development, therapy, and diagnosis of diseases. Devices based on AI algorithm and recent collaborations of AI companies and pharmaceutical companies are also presented.

Keywords

Machine learning · Deep learning · Artificial neural networks · Drug discovery · Drug development · Formulation · Pharmaceutical · Health · Industry

Chapter Objectives

Upon reading of this chapter, it is expected that the reader will be able to understand:

- Differences between AI, ML, and DL.
- How AI applications are transforming drug discovery and development?
- Applications of AI in formulation development and manufacturing of drug products.
- Applications of AI in healthcare.
- Collaborations among AI technology companies and pharmaceutical companies for speeding drug discovery and development process.
- Some recent challenges for AI.

11.1 Introduction

The ability to learn and make decisions has marked a distinction between man and the natural world. Human brain's logical thinking and decision-making abilities are superior to other life forms. Humans want to conquer the world, and this aspiration of humans created machines that can be programmed to work like the human brain to solve problems with the same or even better human intellect and logic. The aspiration of creating a machine that could think like human and take decisions like human brain came from ancient times. In the modern world, this idea has turned into reality with Artificial Intelligence (AI).

AI refers to the simulation of human intelligence in machines. AI employs various technologies that allow machines to think and act autonomously. AI empowers machines with cognitive powers like ability to perceive reason, learn, abstract, and act. Thereby, AI-powered machines mimic not only human actions but also human behavior. The AI process involves the extraction of information, selecting criteria for using information, and seeking for an impactful conclusion [1]. Although computers reached to the market in the 1950s, it took another 30 years for them to become an integral part of social and professional culture. AI and its applications were developing parallel to the growth of computers. However, the growth of AI looks phenomenal only in the early twenty-first century when data overloads on humans became so cumbersome that it could not be handled and interpreted without the help of computers and AI. Now, AI is very much visible in our day-to-day tasks. This has been made feasible with AI and advancements in associated fields like computers, Internet, and associated technologies which allow interaction of man with machines via different means. Human-machine interactions via speech recognition systems, like Amazon's Alexa, Apple's Siri, Google Assistant, Microsoft's Cortana, etc., allows voice commands to execute daily routine tasks such as listening to music, watching video, searching information, switching on or off electrical equipment, and many more. Gary Kasparov, a reigning world champion, was conquered by Deep Blue, an IBM-trained chess computer [2]. The classical game show Jeopardy was won by IBM-trained Watson System, by defeating the champions [3]. This AI concept initially applied to some game shows was later brought to the field of science and technology. The advancements in AI and associated technologies like robotics, vision systems, voice-activated systems, cloud computing, and many others have made AI-equipped machines capable to think, respond, and act like humans in different situations with or without any formal training [4].

11.2 A Brief History of Artificial Intelligence

In 1955, the term Artificial Intelligence (AI) was coined by John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon in a project proposal titled "A Proposal for the Dartmouth Summer Research Project on

Artificial Intelligence" [5, 6]. However, the seminal event at Dartmouth was organized a year later in July/August 1956, the official birth date to this new field. In the 1950s and 1960s, major works of AI were focused on heuristic search methods which made interaction of man and machine feasible. Some of the machines which used the initial concepts of AI were STUDENT (1964) and ELIZA (1966). STUDENT was a question-answering system made for solving algebra story problems [7]. ELIZA, developed by Joseph Weizenbaum at Massachusetts Institute of Technology, was a computer program for simulating human conversation which made human-machine interactions a reality for the first time [8]. This program is also considered as the birth of the first chatbot [9].

AI in drug discovery emerged as patterned recognition systems for identifying the presence of structural fragments as topological (or 2D) pharmacophores/pharmacophobes. Initial researches in this direction were nonquantitative studies for finding relationships between structure of a chemical compound and biological activity. In the 1960s, some heuristic studies were reported in biomedical literature [10, 11]. Pioneering computer-assisted identification and quantification research works of Corwin Hansch and Toshio Fujita established quantitative QSAR as a field. Hansch and Fujita discovered mathematical relationships between log *P* and biological activity [12]. Also in the 1960s, Free and Wilson developed mathematical models to anticipate complex facets of toxicity [13]. In the 1970s, Adamson and coworkers applied multiple linear regression analysis on fragment sets to investigate biological activity [14]. A new class of fragment descriptors referred to as screens or structure keys or fingerprints were developed for searching substructures in large databases of chemicals. These screens were used for similarity searching, clustering chemical databases QSAR/QSPR modeling [15]. In the 1980s, some noteworthy expert systems in medicine [16, 17] and pharmaceutical sciences [18] were developed. In the 1980s, another significant development from Klopman and coworkers came as an ML-based expert system "Computer-Automated Structure Evaluation (CASE) Program" for automatic identifying descriptors, in the structure of potential chemicals, that could be related to biological activity [18, 19]. This program was subsequently developed for finding relationships between molecule fragments and the toxicity and projected as in silico tool alternative to animal testing [20, 21]. CASE and MULTICASE programs, now available commercially, have been widely used among drug regulatory agencies and pharmaceutical companies for predicting toxicity of new molecules by identifying chemical features (biophores or alerts). Richard D. Cramer and coworkers developed 3D-QSAR methods like the comparative molecular field analysis (CoMFA) and its relative, the comparative molecular similarity indices analysis (CoMSIA) [22, 23]. In the 1990s, AI was in its bust mode due to validity concerns and the rotation of reinvention wheel because of wide scope and numerous approaches in the field [24]. However, several noteworthy algorithms and approaches like decision trees, support-vector machines, Bayesian networks, and expectation-maximization were introduced.

Early 2000s marked the reemergence of ML and DL with the rise of QSAR, molecular modeling, computing speed, big data, and high-speed Internet. QSAR studies reemerged from the mid of 2000s with multidimensional (4D-QSAR

onwards) QSAR techniques [25, 26]. The combinatorial chemistry and high-throughput screening developed in the 1990s had generated numerous public databases of molecular structure, biological and pharmacological activity, and safety data. Big data navigated to chemical space with the release of NCI Open Database and subsequent emergence of PubChem, ChEMBL, ChemSpider, DrugBank, and many other databases [27]. These databases provide free accessibility to huge chemical and biological data, but on the counter side, this has increased the workload of data retrieval and analysis which could not be possible without AI. Michael Cox and David Ellsworth were among the first to use the term big data [28]. Big data refers to the large datasets that are used to find new associations and patterns. The most well-known version of big data definition comes from IBM which states that big data can be described by the five "V"s, i.e., *volume, variety, variability, velocity, and value* [29]. Further, the omics data, generated in the early twenty-first century, is expanding exponentially with human genomic data which came in 2003.

After the 2010s, deep neural networks (DNNs) are gaining immense popularity with their better prediction ability than ML algorithms. At the famous Kaggle challenge [30] and NIH Tox21 challenge [31], DNNs as a method for QSAR established the supremacy of DNNs in predicting biological activity and AI technologies assist humans in analyzing and interpreting this huge quantum of data that a human brain could not handle and interpret.

11.3 Artificial Intelligence

The intelligence of humans or other living organisms is natural. Modern machines like computers are equipped with intelligence that is different from human intelligence and hence it is termed artificial. AI can broadly be grouped together in three different types, viz. artificial narrow intelligence (ANI), artificial general intelligence (AGI), and artificial super intelligence.

ANI refers to the intelligence that focuses on a single task, hence termed narrow, and thus it is also labeled as weak intelligence. The applications of AI which we observe in our daily lives belong to ANI like the use of Alexa/Google Assistant/Siri for listening to music, watching videos, ordering online, or searching contents on the Internet.

AGI refers to intelligence which is at par with human intelligence. In the medicine field, some examples are the application of IBM Watson/Google Deep Mind/Microsoft's Project Hanover, for structured and unstructured data analysis, in assisting doctors for clinical decisions. Such systems may have cognition and reasoning abilities of humans, but this is restricted to one task only. AGI is defined as the "the ability to achieve a variety of goals, and carry out a variety of tasks, in a variety of different contexts and environments" [32]. AGI has cognition and reasoning abilities for a wide variety of tasks. Projects like Deep Mind, Human Brain Project, and Open AI are three largest projects to achieve AGI [33]. So far no claimed AGI system has been accepted globally which can have intelligence at par to humans [34]. Artificial super intelligence (ASI) seeks to surpass human intelligence

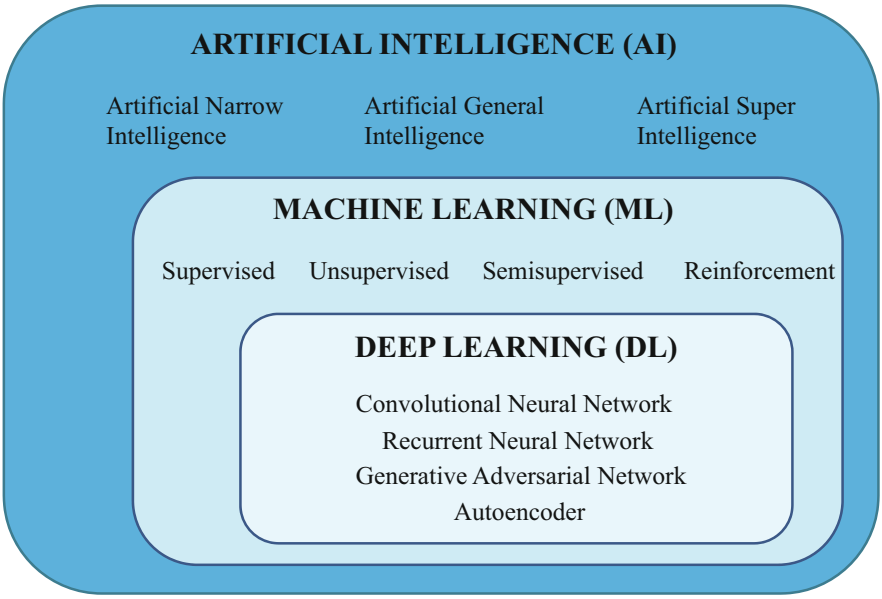


Fig. 11.1 Artificial intelligence and its subsets

and behavior. In that sense, ASI would be smarter than the best human brains in every field be it science or social skills or general wisdom [35]. ASI will evolve with emotions and experiences, very similar to humans or even better, which will evoke needs, desires, emotions, and beliefs of their own. We have observed this type of intelligence in science fiction movies in which robots enslave humanity and pose dangers to the existence of humans.

AI has its subsets called Machine Learning (ML) and deep learning (DL) (Fig. 11.1). ML is a learning ability which applies statistical methods with or without a definite programming. ML also provides systems the ability to automatically learn and improve from past experiences, just like the human brain. AI works wonderfully with different types of ML algorithms, viz. supervised, unsupervised, reinforcement, etc. [36–38]. DL on the other hand is a type of ML that uses artificial neural networks (ANNs), viz. convolutional, recurrent, generative adversarial, etc., for simulating the network of neurons of the human brain. DL has achieved great successes in various fields such as self-driven cars, speech recognition, image processing, computer vision, and drug development [39]. Discovering a new generation of information, obtaining a higher degree of precision, automated simulation and prediction, diagnosis and detection of diseases/disorders, and clinical trials designing are some of the applications of AI [40, 41]. Table 11.1 highlights some of the differences among AI and its subsets.

Table 11.1 Differences among AI, ML, and DL

Artificial intelligence (AI)	Machine learning (ML)	Deep learning (DL)
AI originated around the 1950s	ML originated around the 1960s	DL originated around the 1970s
The study of pattern recognition and mimicking human behavior by machines can be described as AI	It is a subset of AI and can be described as the study of computer algorithms which improves automatically through experience and by use of data	It is a subset of ML and AI which is based on Artificial neural networks (ANN) to mimic human brain-like behaviors
AI focuses on learning, reasoning, and self-correction	ML focuses on learning through experience without any human intervention	ML focuses on information processing pattern mechanisms to identify patterns just like the human brain
AI exhibits intelligence through decision-making	ML is an AI algorithm which allows systems to learn from data	In the case of DL, the ANN analyzes the data and provides output
Efficiency of AI is based on the efficiency provided by the ML or DL	Less efficient as compared to DL and cannot work properly for a higher amount of data	The most efficient and powerful than ML as it can easily work through a large set of data
Three broad categories of AI are ANI, AGI, and ASI	Four broad categories are supervised, unsupervised, semi-supervised, and reinforcement learning	Some deep neural networks are convolutional, recurrent, autoencoder, generative adversarial network, etc.

11.3.1 Machine Learning

The intention of developing a thinking or rational machine was to apply computing capacity of computers for finding and explaining the patterns that would otherwise be difficult for the human operators. Arthur Lee Samuel coined the term ML in 1959 and defined it as a "discipline of study that gives computers the capability to learn without being explicitly programmed" [42]. ML as a subset of AI permits devices or machines to learn from their own experiences and improve without any coding. It is a versatile area which entails statistics, algorithm complexity theories, probability theory, approximation theory, and other disciplines [43]. The main ground of ML is to introduce algorithms that absorb input data and predict output having admissible range of accuracy by applying computer analysis, learning from prior episodes, and spot patterns in data. ML, since 1959, has grown tremendously to the level that we can now observe its application in our daily life and various scientific researches like clinical trial research, speech recognition, and medical diagnosis [44–46]. ML algorithms can be grouped into four different brackets: supervised ML, unsupervised ML, semi-supervised, and reinforcement ML (Fig. 11.2).

There are some fundamental steps to structure an ML model. The steps are:

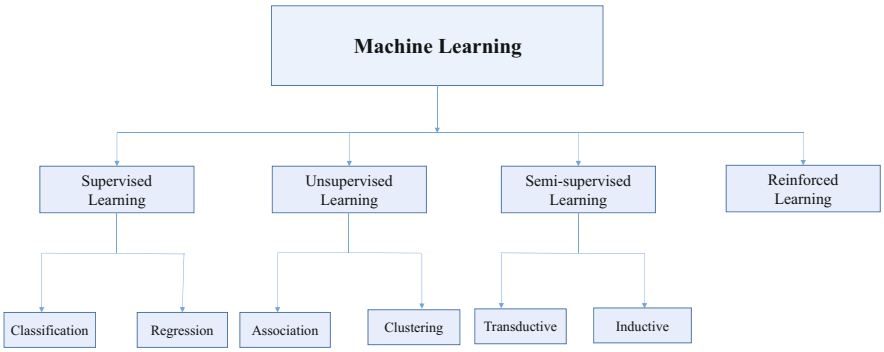


Fig. 11.2 ML, its types and subtypes

- Select and prepare a training dataset.
- Choose an algorithm to run on the training dataset.
- Train the algorithm to create a model.
- Using and improving the model.

11.3.1.1 Supervised ML

Supervised learning is the learning in the presence of a guide or overseer. It is a training in which a machine is trained using labeled data, i.e., data tagged with one or more labels like name, category, or number. Algorithms respond to a set of training data to function. There are two types of training datasets or variables, i.e., input dataset (feature or predictor) and output dataset (target) [47]. The learning is supervised with both data fed to the algorithm. When machine is provided with a new set of input data, the algorithm can map the training data to make predictions, and these predictions are optimized with the availability of more data and with more runs on predictions [48]. This process of training continues until the model achieves accurate results on the training data. When the training part is over, the supervised learning algorithm may be used to predict for new input data [49]. Supervised learning is further subdivided into two types *Classification* and *Regression*.

Classification: Classification algorithms are used to predict categorical output (e.g., whether a person has cancer). Algorithm classifies inputs into two (multiclass classification) or more than two different classes (multiclass classification). Classification algorithms may be used for choosing pertinent genes for sample classification in gene expression studies [50]. A classification algorithm will identify the smallest possible set of genes and can still achieve good predictive performance. A few classification algorithms are k-nearest neighbor (KNN), support vector machines (SVM), decision trees (DT), random forest (RF), and naïve Bayes (NB).

K-nearest neighbor (KNN): It is a simple, traditional nonparametric low learning algorithm developed by Evelyn Fix and Joseph Hodges in 1951 [51]. KNN classifies samples to its nearest neighbor (class or group) by finding the distance of the sample from its K neighbor. First and foremost, its work is to identify the K points in training data that are very close to the test value and then it calculates the distance between all

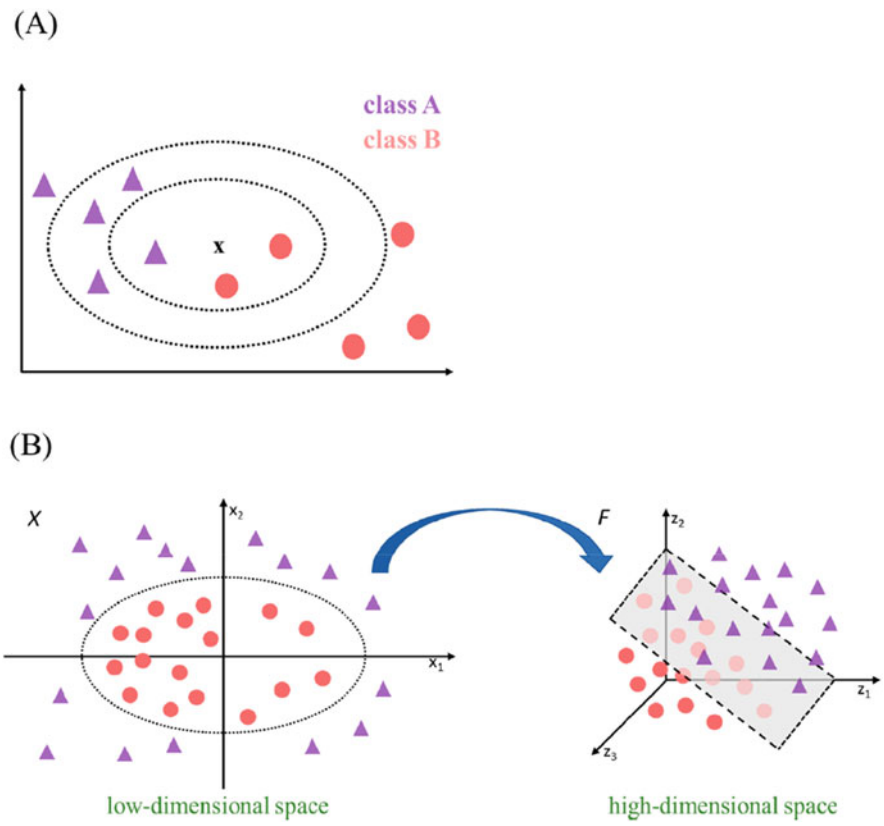


Fig. 11.3 Schematic representations of (**a**) k-nearest neighbors and (**b**) support vector machine algorithms (Reproduced with permission from Xin Yang et al. 2019) [52]

the categories (Fig. 11.3a). The test value is owned by the category whose distance is the least. It uses the entire dataset as the training set, rather than splitting the dataset into a training set and test set [53]. Mostly KNN is used as a classification algorithm, but it can also be used to solve regression problems.

Support vector machines (SVMs): SVMs were developed in 1990 at AT&T Bell Laboratories by Vladimir Vapni and colleagues [54, 55]. SVMs offer the most robust prediction methods, being based on statistical learning frameworks. SVMs are applied in various domains like multimedia information retrieval and bioinformatics [56]. It can be used to solve both classification and regression problems. SVM algorithms create the decision boundary or best line or hyperplane for segregating n-dimensional space into classes so that new data points can be correctly categorized (Fig. 11.3b).

Decision Trees: Decision trees are one of the most easy and popular nonparametric ML algorithms for solving classification problems or decision-making. Decision tree flowchart looks like natural trees to have a starting point from root which further

spreads into different branches [57]. The instances are sorted based on their feature values. Decision trees comprise nodes and branches. Nodes represent tests or attributes or questions and branches represent outcomes (Fig. 11.4a). The decision node helps in making decision and has numerous branches. On the other hand, leaf nodes are the output of those decisions, no further branches are there.

Random Forest (RF): It is based on the concept of **group learning or combined learning** which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model* [58]. RF basically contains n decision trees having a different set of hyper-parameters and trained on different subsets of data (Fig. 11.4b).

Naïve Bayes: It is a classification method based on the Bayes' theorem. Bayes theorem gives the probability of occurrence of an event given that the probability of the other event that has already occurred [59]. Bayes' theorem in mathematical terms is stated as Eq. 11.1:

$$P(A/B) = P(B/A)P(A) \div P(B) \tag{11.1}$$

where A and B are events; $P(A/B)$ is the probability of event A given evidence B has already occurred; $P(B/A)$ is the probability of event B given evidence A has already occurred; $P(A)$ and $P(B)$ are independent probabilities of event A and B.

Naïve Bayes assumes independence among features and equal contribution of each feature to the outcome. There are three types of naïve Bayes classifiers: Gaussian, Bernoulli's, and multinomial. Gaussian assumes features are normally distributed. Multinomial classifier uses discrete counts and feature vectors that represent the frequencies. Bernoulli's classifier requires feature vectors to be independent binary (0 and 1). Naïve Bayes is a core technique in pattern recognition [60] and widely used for applications such as text classification, keyphrase extraction, and medical diagnosis [61].

Regression: Regression model produces numerical values; its goal is to predict a continuous number as regression is a different task than classification. It is a predictive statistical process where models can find the interrelation between dependent and independent variables [62]. Some regression algorithms are linear regression, logistic regression, multivariate, and lasso regression.

Linear regression: It is one of the most used regression algorithms. Linear regression presumes a linear relationship between input and output [62]. A variable from the dataset is chosen to predict the output variables (future values). Linear regression algorithm is used if the labels are continuous.

Logistical regression: As it predicts the probability, it is an S-shaped curve that can take any real value number, and its output value lies between 0 and 1. Logistical regression may fail to perform multiple decision boundaries [63].

Multivariate regression: The multivariate regression is one of the widely used supervised machine learning algorithms to predict the response variable for a set of explanatory variables. With more than one predictor variable in a multivariate regression model, the model is referred to as multivariate multiple regression model [64].

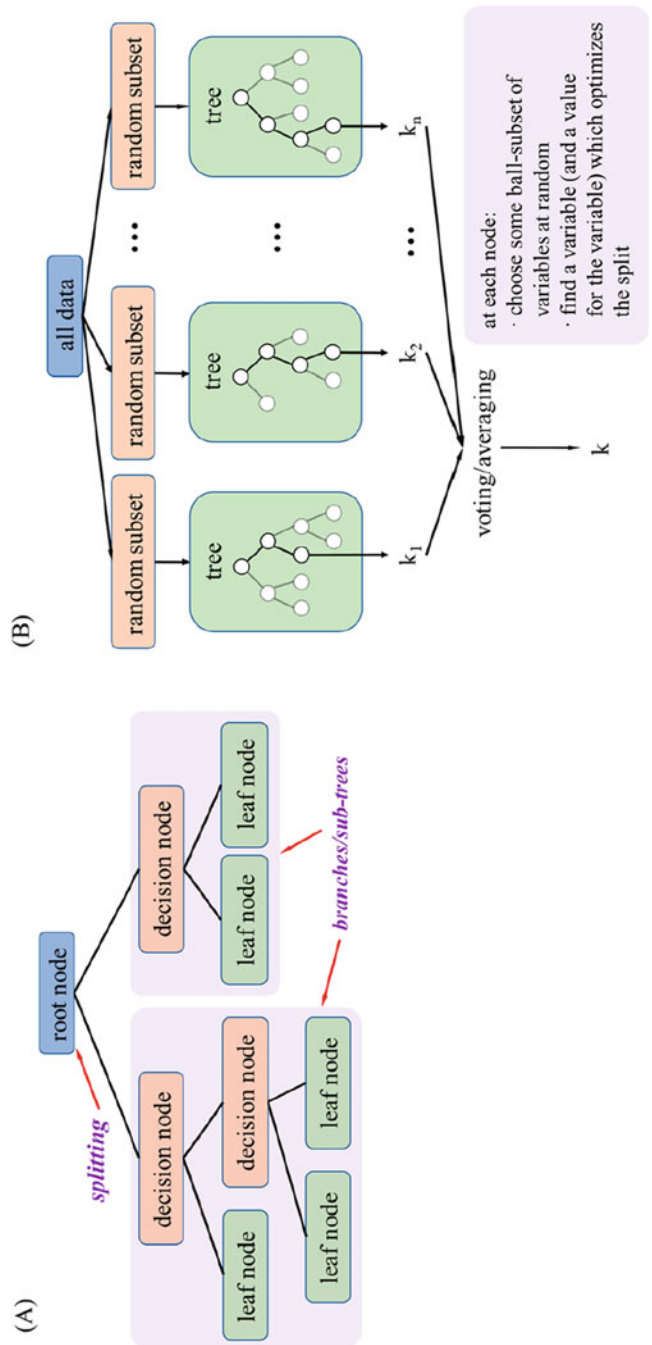


Fig. 11.4 Schematic diagrams of the (**a**) decision tree and (**b**) random forest algorithm. (Reproduced with permission from Xin Yang et al. 2019) [52]

LASSO regression: LASSO (least absolute shrinkage and selection operator) imposes a constraint on the model parameters in a manner that can shrink regression coefficients for some variables toward a zero. The goal of lasso regression is to obtain the subset of predictors that minimize prediction error for a quantitative response variable [65].

11.3.1.2 Unsupervised ML

Sometimes when one has to find hidden patterns from the dataset and there is no availability of labeled data, unsupervised ML techniques are employed. Unsupervised learning models are not under any supervision, and models are not at all trained using information and hence referred to as nonclassified or unlabeled [66]. As the human brain learns new things, unsupervised ML behaves similarly. Unsupervised learning algorithms can function freely to learn new things about the data and give us some new interesting findings on that just like our brain does in new challenges. Classification and regression are the problems where we cannot apply unsupervised ML because, unlike supervised learning, we have only given the input data but no output data. The goal of unsupervised learning is to find the fundamental format of the datasets and sort out the unsorted information according to patterns, differences, and similarities that represent the datasets in compressed format.

Unsupervised ML can be categorized into two types: *clustering and association*. *Clustering* means grouping of data on the basis of some pattern or similarity [62]. A*ssociation type of unsupervised ML* is used for finding association between the variables in larger databases and allows associating the data objects inside large databases.

K-means clustering: It is the simplest algorithm that uses an unsupervised learning method to solve known clustering issues [67]. It divides the whole datasets into K clusters (Fig. 11.5). It requires two inputs: number of clusters (k) and training set ($m = x_1, x_2, x_3 \ldots . x_m$)

Principle component analysis (PCA): PCA helps in diminishing the dimensionality of the dataset while keeping the variations to the maximum extent and correlating various variables to each other. PCA is used to make data easy to explore

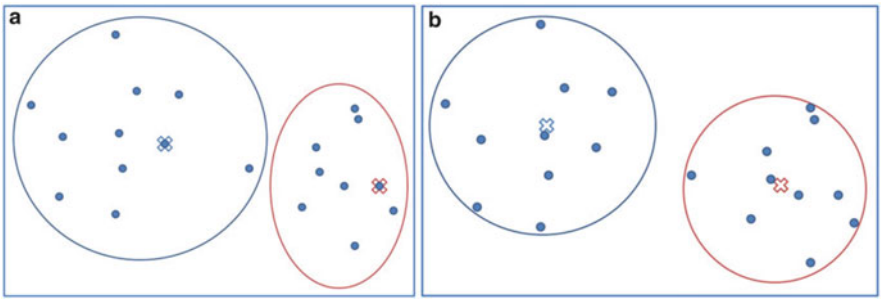


Fig. 11.5 *K*-means clustering example on a set of points, with $K = 2$. The clusters are initialized by randomly selecting two points as centers. Reproduced from Jin et al. 2017 (Copyright with Springer Nature) [67]

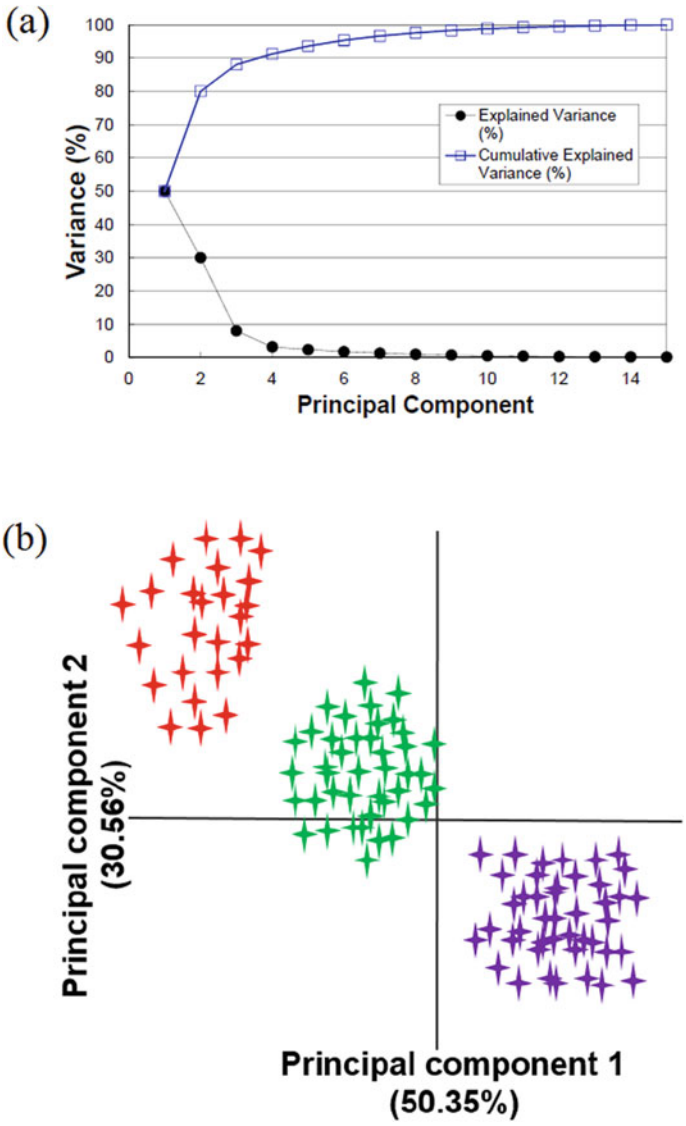


Fig. 11.6 PCA of fruit juices (orange, lemon, and grape) on the basis of chemical composition and antioxidant activity. (**a**) Represents the number of PCs, cumulative explained variance, and the explained variance. (**b**) Represents the projection of samples on the factor-plane (red: orange juice; green: lemon juice; violet: grape juice). PC1 retained about 50% of variation, PC2 explained another 30% of variability, and PC3 and PC4 explain only 11% of variance and unable to differentiate juices. (Reproduced with permission from Granato et al. 2018) [68]

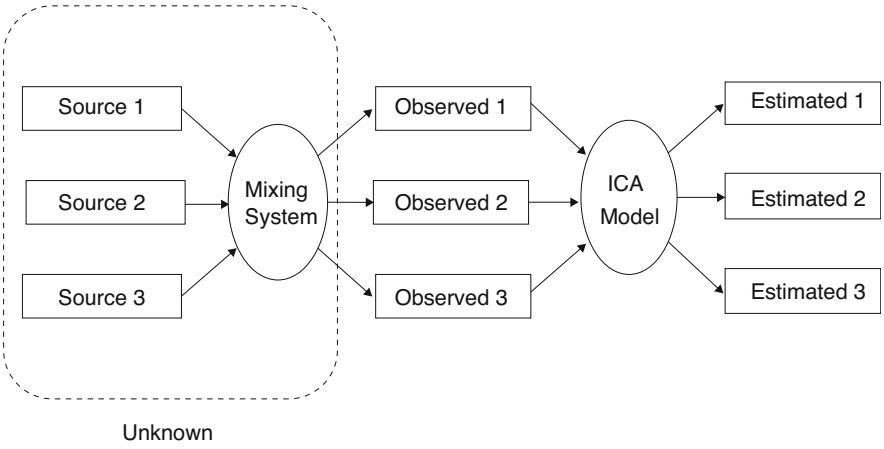


Fig. 11.7 Illustration of independent component analysis to solve blind source separation problem (Reproduced with permission from Vieira et al. 2020) [71]

and visualize by reducing the number of variables (Fig. 11.6). This is done by capturing the variance in the data into a new coordinate system (new set of variables) with axes called "principal components" [69]. First principal component has maximum variations inherited from the original components.

Independent component analysis (ICA): ICA extracts hidden factors (components) underlying in the sets of inputs (variables, measurements, and signals) by converting inputs to a new set that is maximally independent [70]. In other words, ICA is used to convert mixed variables to independent components assuming that these variables are non-Gaussian and statistically independent. Another key highlight of the ICA is that the number of inputs and outputs are equal. ICA algorithm generates an unmixing matrix that estimates values from the sources; these estimates are usually called independent components (ICs) (Fig. 11.7).

Apriori algorithm and FP growth algorithm: This algorithm is designed to work on a transactional database for mining frequent datasets to generate association rules [72]. Algorithm uses a Hash Tree or breadth-first search method to find the item set [73]. Its applications include market basket analysis, medical diagnosis, and to find drug reactions for patients. FP growth algorithm is an improved version of apriori algorithm for representing the database in a structure similar to tree, known as FP (frequent pattern) tree [74]. FP growth method is relatively faster than the apriori algorithm.

Self-organizing map (SOP): It is a type of ANN that is trained using unsupervised learning to obtain a low-dimensional map, often two-dimensional, from the input vectors. In a self-organizing process, the processing elements of the network are made competitive by fixing a winning criterion and the weights to vectors are updated [75]. Figure 11.8 shows a two-dimensional SOM. The best-matching unit (BMU)/node/neuron is the vector that has greatest similarity with the input. The data in the map are associated with the winning node and correlates to neighboring nodes

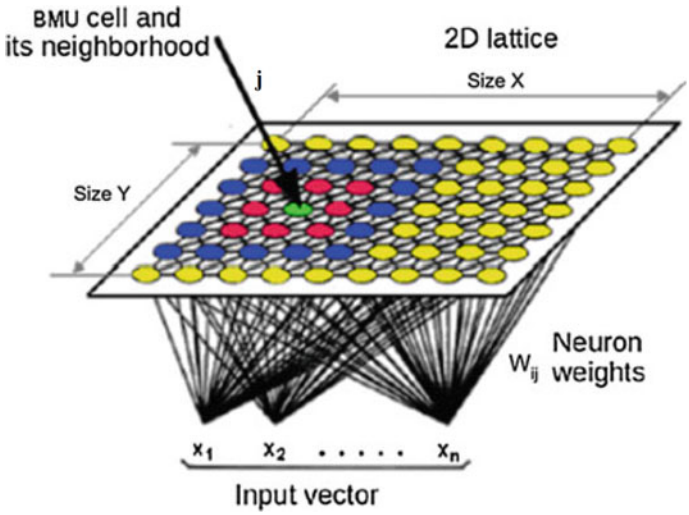


Fig. 11.8 An SOM with input vectors, a winner cell, and its neighborhood. Copyright with Springer [76]. Best matching unit is the winner cell/node/neuron (green) correlated with neighboring cells (red, blue, and yellow)

[76]. With time, both the learning rate and the radius of the neighborhood decreases. SOM is a dimensionality reduction method applied on multidimensional data for clustering. SOM was first introduced by a Finnish professor Teuvo Kohonen in the 1980s, and therefore, it is also called as Kohonen map [77]. The self-organizing map used to identify features existing in the problem is referred to as self-organizing feature map (SOFM). The SOM relates to the classical vector quantization (VQ), as it distributes input data items using a finite set of models [78].

Hierarchical clustering: Hierarchical clustering builds a hierarchy of clusters where each cluster is distinct from the other and objects of each cluster have a high degree of similarity to each other [79]. In practice, *agglomerative hierarchical clustering* involves merging of objects into clusters and then merging similar clusters sequentially. The other way is to use *divisive* approach wherein objects are kept initially in one big cluster and then successively splitting clusters to obtain smaller clusters. Appropriate metric (e.g., Euclidean distance) and a linkage criterion guide agglomeration/division of clusters. Most frequently used approach is agglomerative. Hierarchical clustering is represented as a dendrogram (Fig. 11.9)

11.3.1.3 Semi-supervised ML

One major shortcoming of supervised learning is the need of an extensive amount of human-labeled datasets for training. Labeled datasets are sometimes not available and getting it labeled by professionals may incur huge costs, for example, drug discovery, computer-aided diagnosis, and part-of-speech tagging. Sometimes, there is no lack of labeled data, but unlabeling of some data may provide additional information relevant for prediction. In such instances, semi-supervised ML is used

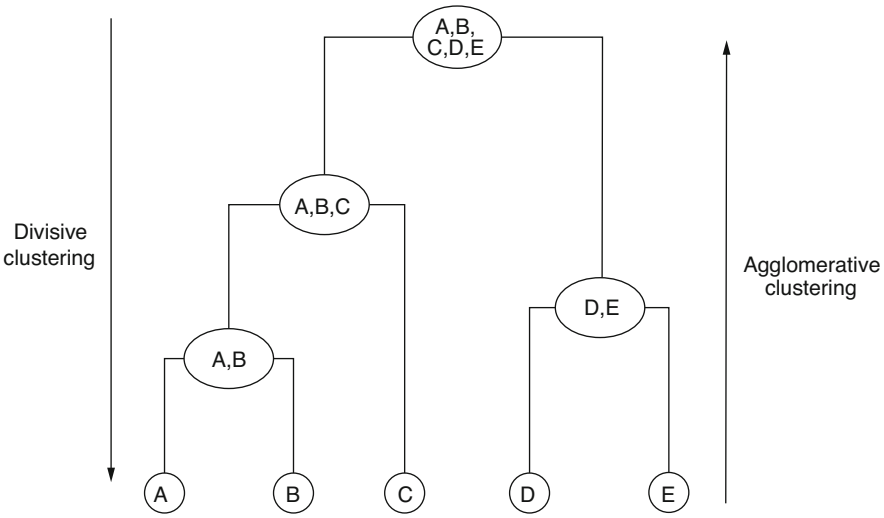


Fig. 11.9 Hierarchical clustering. Reproduced from Zhao et al. 2013 Copyright with Springer [80]

for model building. Semi-supervised ML uses a small amount of labeled data with a large amount of unlabeled data [81]. On the basis of the primary objective of the approach, semi-supervised classification methods are divided into two groups: transductive and inductive learning [82]. Transductive learning involves both unlabeled and labeled data to build a procedure competent of providing label predictions for data points in the entire input space. In more simple terms, transductive learning builds a model that fits the training and testing data points it has already observed [83]. Inductive learning builds a generic model where any new data point would be predicted, based on an observed set of training data points. The approaches applied for unlabeled data usage categorize them into wrapper methods, unsupervised preprocessing, and intrinsically semi-supervised methods. Readers are directed to a review [82] for more details.

11.3.1.4 Reinforcement ML

In reinforcement learning, there is an intelligent agent, generally a computer, to interact with the environment. The agent learns to behave in an environment by performing actions. Unlike supervised learning, the agent also learns without any labeled data and waits for the result since no labeled data is there, so agent has to learn by its experiences [44]. As it is a feedback-based ML technique, the agent (computer) gets positive feedback (rewards) for good or accurate outcome and negative feedback for bad or inaccurate results. When an agent in reinforcement learning gets maximum positive rewards, the performance of the agent starts to improve and it works effortlessly. Reinforcement learning is one of the most widely useful tools in robotics and game playing where decision-making is sequential. One of the examples of reinforcement learning is a robot moving its arm.

11.3.2 Artificial Neural Network and Deep Learning

The human brain has a complex network of neurons that are connected via synapses. Billions of neurons in human brain are interconnected with synapses (Fig. 11.10a). Neurons transmit information to and from the brain and throughout the human body.

ANN is an artificial computational model that is analogous to the biological neuronal network of the human brain. ANN can follow any paradigm be it supervised or unsupervised learning. ANN consists of artificial neurons processing units connected by coefficients (weights) [84]. ANN structure has a set of nodes organized into several layers, viz. one input layer, one or more hidden layers, and

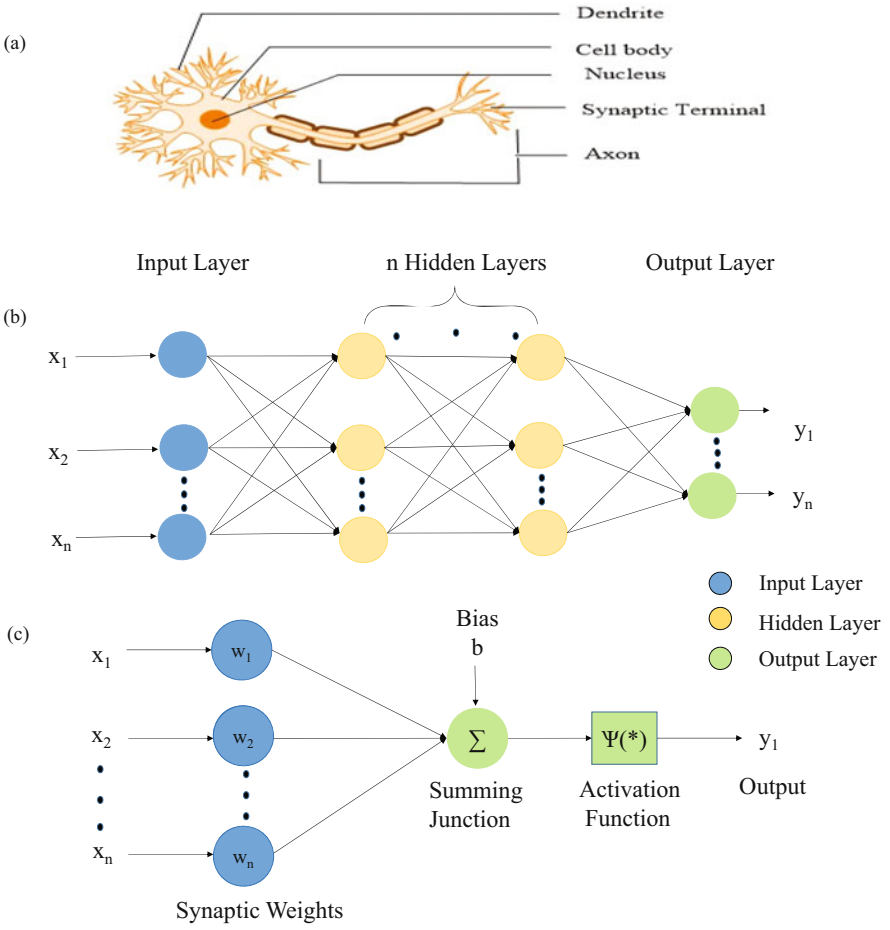


Fig. 11.10 Human and artificial neurons. (**a**) Human neuron. (**b**) Architecture of artificial neural network. (**c**) Mathematical representation of a neuron; x_1, $x_2 \ldots x_n$ and w_1, w_2, $\ldots w_n$ are their respective weights; $y_1 \ldots y_n$ are outputs; b represents an external bias applied to summing junction (Σ) to increase or lower the output signal; Ψ(*) is output signal amplitude

one output layer (Fig. 11.10b). Input layer accepts the entry of data, the hidden layer (s) processes the input, and the results or outcomes are generated by the output layer. Number of nodes in the input and output layers are determined using the number of independent and dependent variables, respectively. Number of hidden layers and nodes in each layer depends on the complexity of the problem. Many ANN models consist of only one hidden layer; meanwhile, more than one hidden layer can be implemented for modeling the complex problems [85]. *Each* neuron is characterized by its weight and *bias (w and b), but hidden* and output *layer* neurons also possess activation *functions* (Fig. 11.10c). Weights and biases represent connections of nodes of one layer with the nodes of the other layer. A bias node in the input layer or hidden layer increases the flexibility of the model. Weight and bias are applied when inputs are transmitted to the neuron. Activation function defines the output of a node. ANN can solve problems related to tabular data, image data, and text data [86]. ANN can also be known as universal function approximators as it is capable of learning any nonlinear function [87]. Most commonly used neural networks are Feed Forward Neural Network (FFNN), Multiple Layered Perceptron (MLP), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), AutoEncoder (AE), and Generative Adversarial Networks (GAN).

DL, a subset of ML, uses neural networks that can grasp and learn from a large and different amount of data [88]. DL uses ANN with two or more hidden layers, deep neural networks (DNN), to solve complex problems. DL is an automatic learning procedure which has been widely embraced in many fields of science, business, security, management, education, and other fields for their applications in computer vision, speech recognition, pattern recognition, natural language processing, and recommendation systems. ML techniques need domain expertise to draw feature extractor (Fig. 11.11). However, DL itself serves as a feature extractor whose work is to convert a low-level feature to a higher-level feature. Also, DL has the ability to find out specific minute variations, which allow these methods to be more accurate that the other ML methods. Predictability performance of DL is much better than the other ML methods (e.g., SVMs and RFs) [89]. But DL requires large training datasets when compared to ML. This can be because there are improvements in algorithms of AI, computers, and the availability of larger datasets.

11.3.2.1 Feed Forward Neural Networks and Multiple Layered Perceptron

Feed Forward Neural Network (FFNN) is the most common neural network (Fig. 11.12a). Information flow in this neural network is only one way, that is, in forward direction. Based on the number of hidden layers, these may be single layered or multiple layered. Single-layered FFNN do not have any hidden layer. Multiple Layered Perceptron (MLP) consists of one or more hidden layers allowing multiple stages of computations and information processing (Fig. 11.12b). Each deeper layer of MLP is a nonlinear function of each earlier layer. MLPs are extensively used for finding complex nonlinear relationships in data. FFNN has been used for computer vision, object recognition, and speech recognition systems.

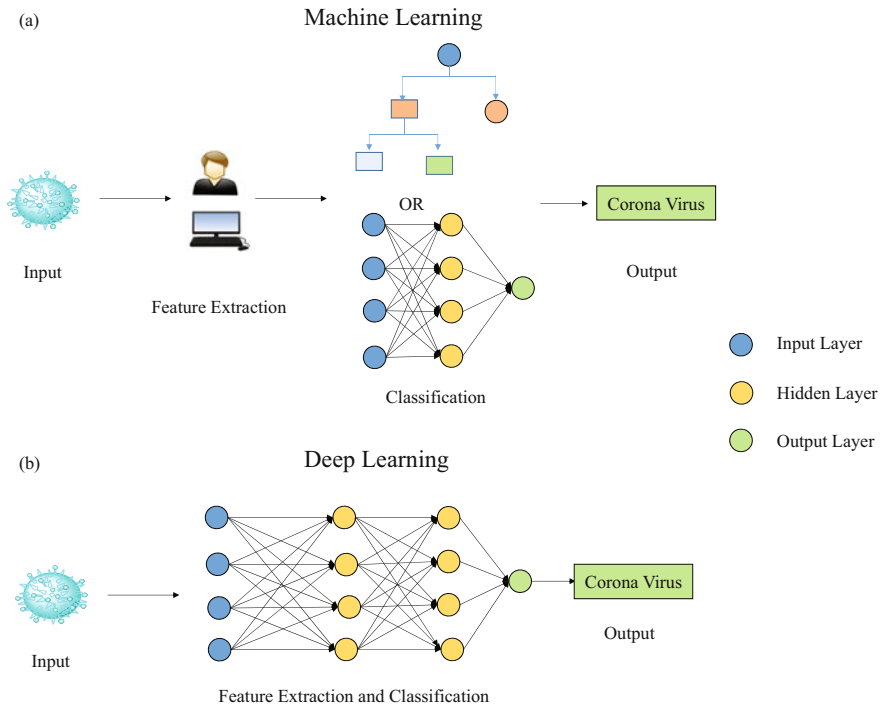


Fig. 11.11 Comparison between machine learning (ML) and deep learning (DL)

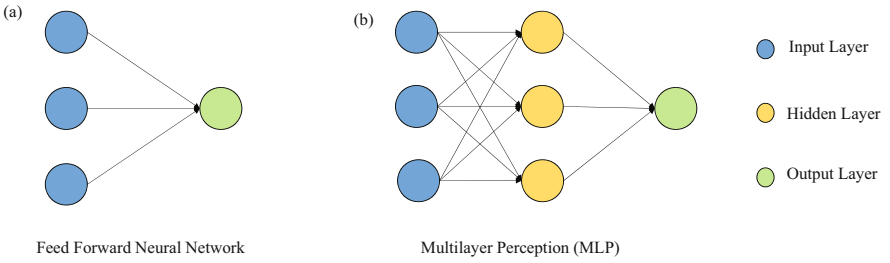


Fig. 11.12 Feed Forward Neural Network (FFNN) and Multiple Layer Perceptron (MLP)

11.3.2.2 Recurrent Neural Network

RNN signal flow is characterized with cycles (loops) between nodes within a hidden layer [90]. Cycles/loops are connections to lead back to neurons in the same layer representing recurrence of operations (Fig. 11.13). This allows RNN to perform a broader range of tasks like text prediction and language generation. Conventional RNNs have the limitations of short-term memory and are difficult to train. Long Short-Term Memory (LSTM) networks, a new type of RNN, enables to perform tasks involving longer-term memory. RNNs have gained foothold in sequential data

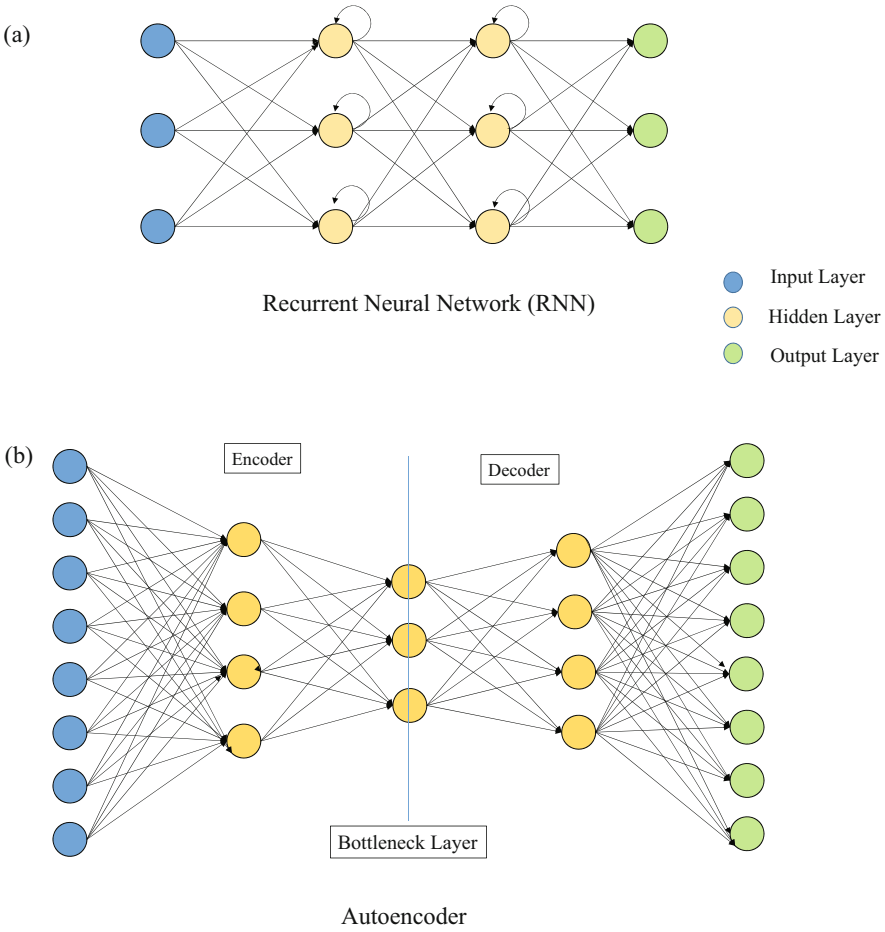


Fig. 11.13 Illustration of recurrent neural network and autoencoder

analysis for natural languages processing problems such as speech recognition, natural language generation, text recognition and text prediction, genome sequences, clinical time series, etc.

11.3.2.3 Convolutional Neural Network

CNN, also known as ConvNet, is a variant of the feed forward MLP in which the inputs are grouped spatially into hidden nodes [90]. Deep CNN works by modeling small pieces of information and amalgamating them in deeper networks (multiple hidden layers) [91]. CNN can detect features without any human supervision. CNN contains one or more convolutional layers that can be either pooled or entirely connected (Fig. 11.14). Convolution layers use filters to extract features from the input image and generate feature maps and introduce nonlinearity (ReLU; rectified

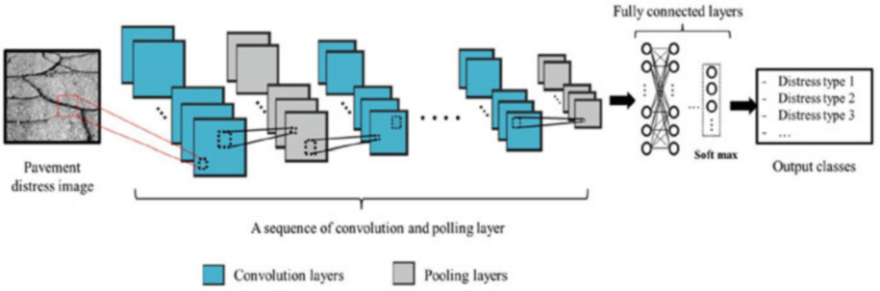


Fig. 11.14 CNN architecture for detection and classification of pavement surface distress (images). Reproduced with permission from Ranjbar et al. 2020 [92]

linear unit) to the feature maps output. Feature map records a region of an image. Pooling layer reduces the spatial size of feature map output, for example, systematically halve the width and height of feature map. After several convolutional and max pooling layers, pooled feature maps are flattened to a single column that facilitate the vector to be processed further by a fully connected (hidden layers) ANN for final classification. At last, the soft max function generalizes the logistic function for prediction [93]. CNN helps in identifying images of objects clearly, tells the location of the object as well as its connection with other objects in an image, and therefore applied in self-driving cars and machine vision [94]. Graph convolution networks are special types of neural networks that can be put in structured data in the form of graphs. Because of CNN's image recognition feature, it works well with pathology images because there is a large number of viable pixels that can be used for training from a single biopsy or incision [95] (Fig. 11.14).

11.3.2.4 Auto-Encoder

Auto-Encoder (AE) neural networks are designed to learn complex nonlinear relationships among data [96]. Simplest AE is built on feed forward and nonrecurrent networks that work like MLP (Fig. 11.12d). Output layer and input layer have the same dimensionality and the same number of nodes. Instead of being trained to predict, output value an AE is trained to produce an output that matches with its own input. Autoencoders are used for data compression, feature extraction, image generation, denoising, and colorization.

11.3.2.5 Generative Adversarial Networks

Generative adversarial networks (GANs) can produce or create new data based on real data. The GAN technique is a confrontational game between two DNNs: one is the generator and the other is the discriminator [97]. The generator generates new (fake) data and the discriminator distinguishes the real data from the new (fake) data created by the generator (Fig. 11.15). Steps are repeated several times to get better performance out of both the generator and the discriminator. GAN frameworks have wide applications in image generation, face detection, 3D object creation, computer

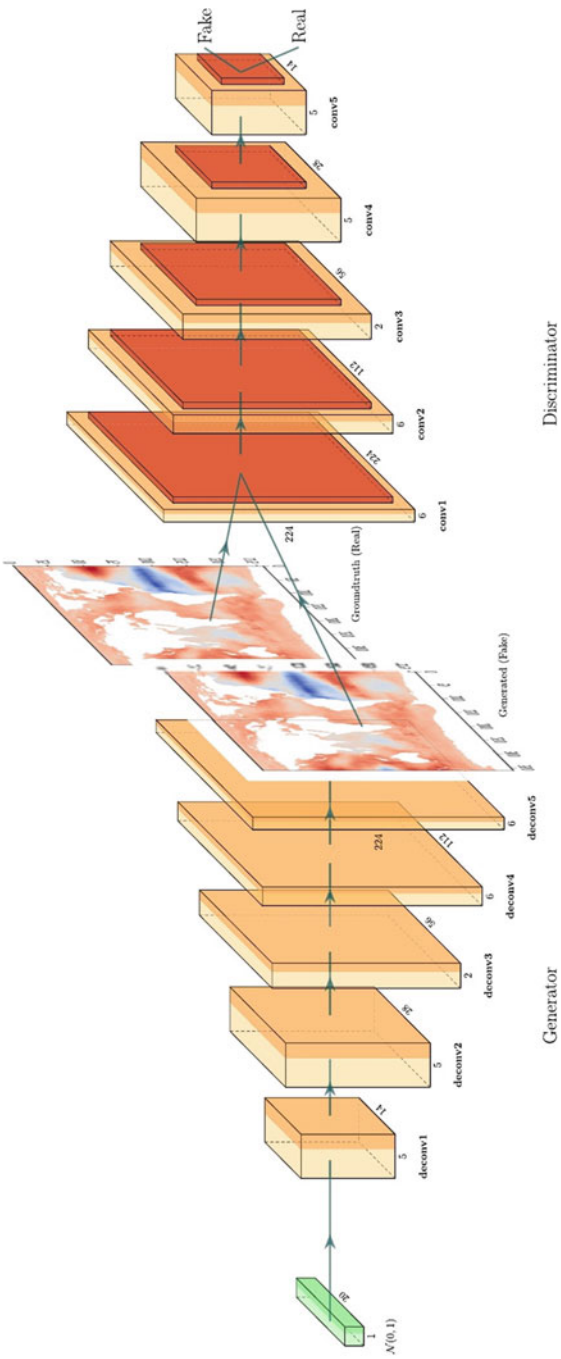


Fig. 11.15 An example architecture of a generative adversarial network. Reproduced with permission from Reimers et al. 2020 [98]

vision, text transferring, traffic control, medical image generation/classifying, and drug discovery [97].

11.4 AI in Drug Discovery and Development

The drug discovery process has several key steps (Fig. 11.16) wherein AI plays significant role [99]. Initially, high-throughput in vitro assays allow preliminary selection of compounds. Subsequently, counter-screens are used to remove compounds with undesirable properties or false positives. ADMET screening is performed on compounds for predicting absorption, distribution, metabolism, excretion, and toxicity. Structures based or ligand-based drug design approaches are used to design molecules with adequate characteristics. Designed molecular are subsequently synthesized. Synthesized compounds are tested in vitro for confirming quality. Qualified synthetic compounds are processed for preclinical studies in experimental animals, and at last, clinical trials are performed on humans.

The drug discovery progress of a new drug costs an average of $ 868 million and takes over a period of more or less 10 years [100–102]. Chemical, pharmaceutical, and life science industries are using DL and ML to analyze, interpret, and predict

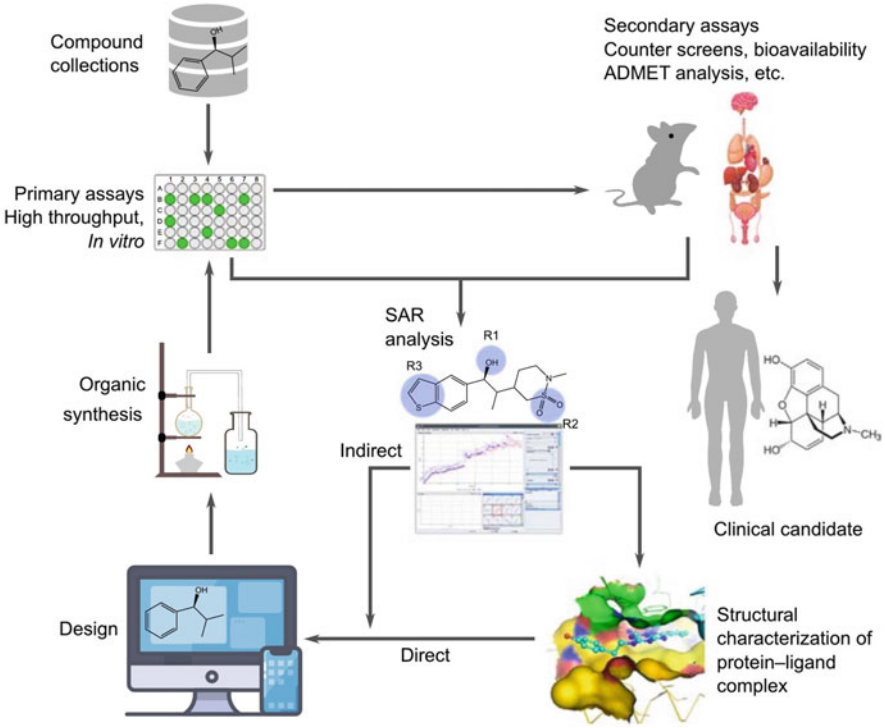


Fig. 11.16 Drug discovery process. Reproduced with permission from Chan et al., 2019 [99]

huge chemical and omics data for speeding up their drug discovery programs. Computers are powered with AI-based algorithms to receive and analyze the data for investigating chemical structure, physicochemical properties, biological activity, and their interrelationships in different in vitro and in vivo environments. More specifically, AI tools (Table 11.2) helps in predicting quantitative structure-property relationship (QSPR), quantitative structure-activity relationship (QSAR), data analysis, in vivo data, pharmacokinetics, and biological activity [119–121]. ANN modeling may also be applied for designing the molecular structures of organic compounds and predicting their physicochemical characteristics [122]. The prediction of pharmacokinetic profile of potential molecules using AI models reduces attrition rate in preclinical/clinical trials [123]. Allied AI disciplines like natural language processing, computer vision, and robotics have helped in amplifying the available information and increasing the reproducibility of experiments. DNNs may be adapted for learning joint data representation from several omics data types. This will help in combining information from different experiments and will provide precise and better-informed predictions. With advancement of AI, larger datasets can easily be accessed and more semi- and unsupervised ML applications are now able to produce new biological hypotheses [124]. These biological hypotheses may include potential novel pharmacogenomics markers and drug targets based on information obtained from large omics dataset.

11.4.1 Drug Screening

Scientists look forward to AI for improving the success rates of molecules.AI is used to predict target protein/receptor structure and drug protein/receptor interactions. Some proteins/receptors are overexpressed while others are underexpressed in disease states. Identifying correct biochemical events in the disease state is essential for correcting disease. To design a new drug molecule, the structure of protein/receptor is primary requirement. AI helps in predicting the 3D structure of the protein/receptor molecules and eases the structure-based drug discovery [125]. An AI tool, like AlphaFold, analyzes the distance between the adjacent amino group and corresponding angle of the peptide bond to predict the 3D structure of the target protein. A study was carried out by using MATLAB-assisted nonlinear three-layered neural network toolbox which predicted the 2D structure of the protein. The accuracy of predicting the 2D protein structure was about 62.72% [126].

Second most crucial step of designing a new drug molecule is predicting the drug–protein interactions. The interaction between the drug and the protein, that is, a receptor, is necessary for the efficacy of the newly discovered drug. Various AI tools, like AtomNet, predicts drug–ligand interaction [127]. The virtual screening of drug molecule is briefed below [128].

Selection of a target: The small molecular compounds target four large molecules such as proteins, polysaccharides, lipids, and nucleic acids. Often ion channels, enzymes, and receptors are selected as targets for correcting human biochemistry in the disease state.

Table 11.2 Some AI tools used in drug discovery

AI tool	AI approach	Application	Ref
Hit Dexter	Randomized trees classifiers (ML approach)	Prediction of small molecule for positive response in biochemical assays. Available at https://nerdd.univie.ac.at/hitdexter/	[103, 104]
HitPick	B-score method (hit identification) and combination of Laplacian-modified naïve Bayesian target models and 1-nearest-neighbor (1NN) similarity searching (target prediction)	Hit identification and target prediction of chemical screening using ChEMBL bioactivity data. Available at http://mips.helmholtz-muenchen.de/hitpick/cgi-bin/index.cgi?content=hitIdentification.html	[105, 106]
Chemputer	Modular robotic system equipped with AI	Automation of chemical synthesis and analysis	[107]
PotentialNet	Multistaged spatial gated graph CNN	Molecular property prediction like ADMET, solubility, protein-ligand binding affinity, etc.	[108]
REINVENT	Open-source Python application; PyTorch as a deep learning engine, and RDKit version as a chemistry engine	De novo design of small molecules; publicly available at https://github.com/MolecularAI/Reinvent	[109]
AlphaFold	CNN	Predicts 3D structure of protein with high accuracy	[110]
DeepTox	DNNs comprising SVM, RF, and elastic nets	Prediction of toxicity of drugs and environmental chemicals	[111]
The Polypharmacology Browser (PPB2)	Combination of nearest neighbor searches and naïve Bayes	Prediction and used as a target prediction tool. It also computes ligand similarities. Available at https://ppb2.gdb.tools/	[112]
ORGANIC	GAN	Molecular generation tool used to create molecules with desired properties; inverse design chemistry	[113, 114]
DeepDDI	DNN	Prediction of drug–drug interactions (and drug–food constituent interactions)	[115]
CASE Ultra	QSAR and ML	Prediction of preclinical toxicity of potential molecules	[116]
ADMETlab	QSAR regression or classification models using		[117]

(continued)

Table 11.2 (continued)

AI tool	AI approach	Application	Ref
	six different machine learning algorithms	Prediction of ADMET; available at http://admet.scbdd.com/	
Mol-CycleGAN	GAN with reinforcement learning	Molecule design for a desired physicochemical or structural property	[118]

Compound database preparation: A complete database of biological and chemical properties of the compound molecules along with their structures is readily available from databases like PubChem, ZINC, ChemBank, DrugBank, ChemSpider, etc. Dedicated software equipped with AI algorithms facilitate collection of information from various databases.

Selection of docking software: Docking software, such as AutoDock and MolDock, are available for studying the interaction of potential molecular compounds with its target [128]. Selection of appropriate software as per objectives of the protocol plays a key role governing outcomes. For large-scale docking (1 million compounds), experts prefer Linux-based high-throughput virtual docking.

Scoring system: Docking predicts the preferred orientation of the ligand molecule as it binds with the target macromolecule. Based on the orientation and binding energy of the ligand and the receptor, a docking score is calculated.

Biological experiment verification and clinical study: Once the highest docking score of the compound is obtained, it is then verified by in vitro and in vivo experimental methods. Once it passes all the preclinical studies, clinical trials are performed on the candidate compound for determining its safety and efficacy on humans.

AI is applied to predict physicochemical properties like partition coefficient, solubility, permeability, degree of ionization, pH, etc., as these properties govern pharmacokinetics of the drug [126]. Either structure of the drug or SMILES (Simplified Molecular Input Line Entry System) strings may be used to evaluate electron density and potential energy with the help of AI.

11.4.2 Predicting Toxicity

AI can predict the on- and off-target effects of a drug and in vivo safety profile even before they are synthesized. Off-target effects are shown when the drug does not bind at its intended target. Platforms such as DeepTox, used in prediction of toxicity rate, and PrOCTOR, which predicts the toxicity probability in clinical trials, uses DNN [111]. DeepTox and PrOCTOR receive raw training data which in turn supplies prediction for new data. PrOCTOR is trained using RF model, and it takes molecular features, drug-like properties, target-based features to generate PrOCTOR score. This score predicts whether the drug will pass or fail the clinical

trial. The prediction rate can be improved if a refined and bigger dataset on toxicity and therapeutic profile of different sets of compounds is available.

11.4.3 Drug Repurposing

Drug repurposing is more easy and attractive with the help of AI. Applying an existing therapeutic to a new disease allows direct Phase II clinical trials. Various in silico methods can be used in prediction of pharmacological data and drug repurposing using transcriptomic data of biological systems [129]. These are generally based on high-level representation of data using DNN, which is an adaptive multilayer system performing different data transformations. In a study, it was found that DNN could classify complex drug actions by locating the mechanism of the pathway [130]. DNN is also capable of classifying drugs based on their functional class, efficacy, and toxicity. Moreover, GAN helps in more precise drug designing and repurposing by converting text descriptions into photo-realistic images.

11.4.4 Polypharmacology and Drug–Drug Interactions

Polypharmacology is the use of drugs to treat multiple diseases. Design to achieve polypharmacology involves prediction of bioactivity using molecular fingerprints for multiple disease pathways or targets. Different database like Ligand Expo, KEGG, ChEMBL, and Multiple Target Ligand Database (MLTD) provide crystal structures, molecular pathways, drug targets, binding affinity, and biological activities [131]. AI can mine data and analyze these databases for designing pharmacological agents for multiple targets. The freely accessible polypharmacology browser (PPB) is a multi-fingerprint browser for predicting multiple targets of molecules [132]. Recently a computational platform, DeepDDI, was developed for better understanding of drug–drug interactions and recommending alternative drugs for the intended clinical use [1].

11.4.5 Clinical Trials

In clinical trials, ML has several potential applications as clinical trials are long lasting and costly [133]. Some of the uses of AI in clinical trial are as follows:

- It helps to identify candidates for clinical trials.
- It helps in finding the best suited sample size for increasing efficiency of clinical trial.
- It uses electronic medical records for cost-effective data handling and analysis.
- It can also be used for remote monitoring and secured access to real-time data.

Maintaining smart electronic health records is easier and more useful with AI [134]. Document classification using SVM and optical characterization helps in collection and digitalization of medical health information. MATLAB's ML handwriting recognition and Google's Cloud Vision API are two of the prominent examples of such innovations. The MIT clinical ML group is working on next-generation intelligent electronic health records for developing AI for diagnostic, clinical decision, and personalized treatment suggestions [135].

ML keeps an eye on biological and other signals for any sign of harm or death of subjects. ML applications increase clinical trial efficiency by finding best sample sizes, addressing and adapting to differences in sites for patient recruitment and using electronic medical records to reduce data errors [136].

11.4.6 Pharmacokinetics

Most of the traditional processes to identify ADMET properties are costly, limited, and very much time taking. AI has proven itself for the modification of models for absorption, protein binding, metabolism, elimination, and toxicity. Use of AI tools like ADMETLab, ADMET predictor, preADMET, etc. may guide synthesis of drugs having desired pharmacokinetic profile. AI tools for prediction of pharmacokinetics allows the user to input the physicochemical/structural characteristics of the compound, and subsequently the added detection model is operated in the background for prediction of pharmacokinetics. AI can thus reduce efforts, time, and money spent on compounds which might fail due to pharmacokinetic issues.

In one study, six AI techniques, ANN, SVM, KNN, Probabilistic Neural Network (PNN), Linear Discriminant Analysis (LDA), and PLS, were employed for prediction of pharmacokinetics [137]. A sundry set of informative data of 5370 drugs was used as input. A total of 29 detection models were originated, among which 24 simulation models were applied for human intestinal absorption, plasma protein binding, and renal clearance and 5 models were applied for detection of metabolism. Performances of all the 29 detection models were evaluated for sensitivity and training (cross-validation) to set accuracy, test (validation) set accuracy, precision, etc. The evaluation of performances considered that the detection models originated with SVM with radial function was better than other algorithms.

AI into the realm of protocol design in oncology was attempted with adapted Monte-Carlo tree search algorithms [138]. Initially, a PK/PD model was developed with population data for pharmacokinetics. PK/PD model was later employed to determine optimal chemotherapy regimen. Algorithm was employed for optimizing drug administration of temozolomide taking into consideration of its PK/PD model.

SVMs and ANNs are widely applied tools to investigate nonlinear relationships in population PK modeling [139]. The relative precision of SVMs, ANN, and other methods for PK modeling was compared in a study [139, 140]. ANN models gave

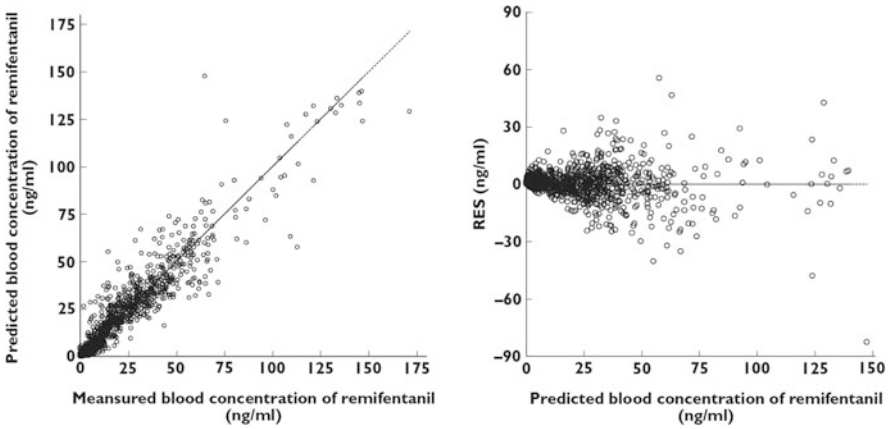


Fig. 11.17 Goodness-of-fit plots. Left: Comparison between experimental and ANN model predicted values of measured blood concentration of remifentanil. Right: Residuals (RES) versus ANN model predicted blood concentration of remifentanil. Reproduced from Kang et al., 2007 [139]

précised detections of blood concentrations of remifentanil when compared to a nonlinear mixed effects model. Precise detections of blood concentrations of remifentanil when paired with an electroencephalographic factor could be utilized to produce a modified pharmacodynamics (PD) model for estimating the consequence of remifentanil on the central nervous system (CNS). Figure 11.17 shows the application of ANN in predicting blood concentrations for remifentanil.

11.5 AI in Formulation Development and Pharmaceutical Manufacturing

Drugs are required to be formulated in suitable dosage forms that are administered to patients through suitable routes [141, 142]. AI has transformed the traditional trial and error method of formulation development to a rational development process with QbD. DoE and optimization of formulations, key determinants of QbD, have been proven very effective with the help of AI-based models. AI tools, ANN and expert systems, investigate complex variables and predict optimum values for investigated variables for best formulation or process. Applications of AI in manufacturing of drug products may be categorized into preformulation, formulation designing and optimization, and production technology.

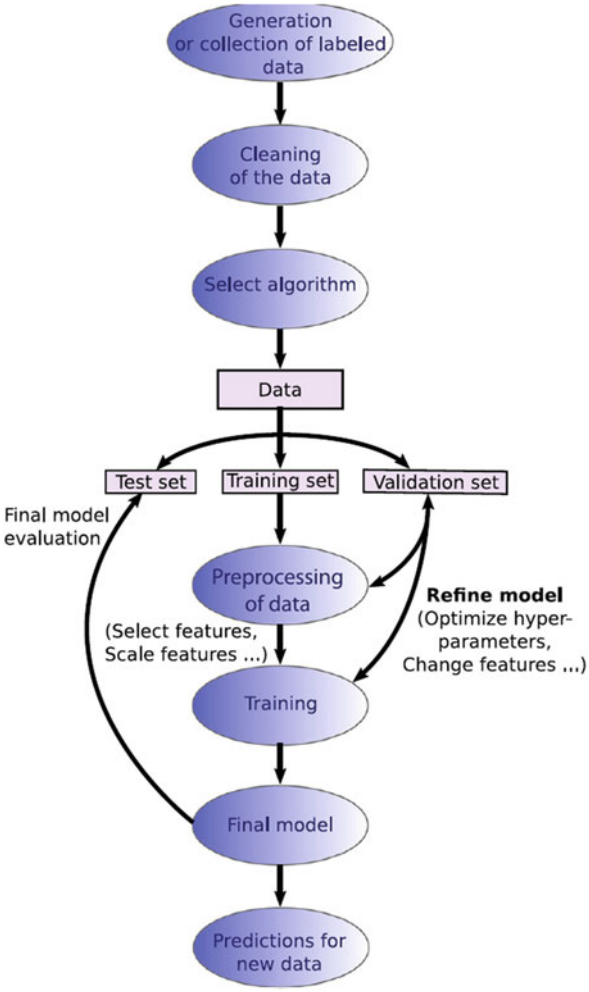


Fig. 11.18 Workflow of ANN modeling and prediction (supervised learning). Reproduced from Schmidt et al. 2019. Copyright with Springer. (https://doi.org/10.1038/s41524-019-0221-0)

11.5.1 ANN Modeling

ANNs are extensively used for modeling complex variables to find their nonlinear relationships among themselves and interactions among variables. Some pros and cons of using ANNs are listed in Box 11.1. The workflow of ANN modeling is presented in Fig. 11.18.

Box 11.1 Advantages and Disadvantages of ANN

Advantages	Disadvantages
• Easy model construction • Access to have multiple training algorithms • Less formal statistical knowledge required • Capable of investigating complex relationships among variables • Capable to investigate all possible interactions among variables • Capable of capturing nonlinearities in variables • Handle large amount of datasets • Robust system: can handle noise and ambiguous data, prone to error or complete	• Blackbox nature; nontransparent and predictions are difficult to explain • Computationally intensive • Require many trials and errors for building a good model • Sharing is difficult • Over-fitting issues: bias, perform poor on test data

11.5.1.1 Collection and Labeling of Data

The first step is the selection or collection of data from literature sources and/or experiments (prospective or retrospective) and labeling with one or more meaningful labels so that an ML model can learn from it. Labels define certain properties or characteristics of the data. Labeled data is checked for its consistency and accuracy.

11.5.1.2 Cleaning of the Data

Cleaning of missing, possible erroneous, or inconsistent data points (outliers), using criteria based on the knowledge and objective of the model, is essential for accurate predictions. Sometimes this cleaning may be required for a specific type of algorithm, for example, some algorithms are more robust than others for handling data with outliers, some algorithms (such as the RF family) do not support null values in their input, while algorithms like KNN or naïve Bayes can handle null values [143].

11.5.1.3 Selection of Algorithm

Selecting algorithm is a most crucial step which involves order selection and input selection. The simplest method of order selection is the use of incremental order selection which starts with a small number of neurons and increases to more complex algorithm until any stopping criteria are met [144]. The algorithm returns to the neural network where optimal order is attained.

Selection of inputs require in-depth knowledge of the problem domain. Standardizing, normalizing, and PCA are some general preprocessing methods for input selection [145]. Input selection algorithms automatically extract characteristics of the dataset for best generalization. Growing inputs, pruning inputs, and genetic algorithms are widely used input selection algorithms.

ANN architecture consists of input layer, hidden layers, output layer, interconnections of neurons, weight functions, activation functions, and transfer functions. Multilayer perceptron (MLP), RNN, PNN, generalized regression neural networks, radial basis function networks, and time-delay neural networks are the few of the renowned architectures [145].

11.5.1.4 Data Grouping

This step involves the grouping of data into three sets, viz. training set, test set, and validation set [145]. Generally, the largest set is used as training set and smallest set is used as test set. Training data is used to modify or adjust the weights in the ANN to produce the desired outcome. The test set is used to evaluate the developed model for accuracy, while the validation set is required to determine the end point of training of the model to achieve the minimum error.

11.5.1.5 Training

The training process involves calibration of neural network using inputs and outputs. Overfitting and underfitting are two concerns which affect performance of the neural network. Underfitting may be addressed with an increase in the number of epochs, but use of more epochs may result in overfitting. Epoch is one complete pass/one cycle of an entire dataset through the neural network. Training errors and model testing procedure are compared to find out the optimal number of epochs [145].

11.5.1.6 Postprocessing and Finalization of Model

The developed model is subjected to several tests and validation to analyze, describe, and to improve the performance of the developed model. Comparison of results of test, training, and validation sets guides postprocessing activities. For statistical comparison of these results, RMSE (root mean square error), percentage volume error, and correlation are generally used [145]. If unsatisfactory results are obtained, modifications in the development model may require changes in the number of hidden layers/neurons, weights and biases, or transfer functions [145].

11.5.1.7 Prediction with Developed Model

The final model works by predicting inputs to get desired outputs for new datasets. This prediction may be compared with actual experimental results.

11.5.2 Preformulation Studies

Physicochemical properties of drugs, excipients, polymers, and their combinations play a key role in pharmaceutical formulation development. ANNs have been widely explored in assisting preformulation studies of pharmaceutical formulations. Table 11.3 shows some preformulation studies where ANNs have been applied. Solubility and permeability of drugs are two most important physicochemical properties which affect formulation development and bioavailability. ANNs have been used for predicting physicochemical properties of potential molecules from

Table 11.3 Applications of ANN in preformulation studies

AI tool	Formulation/preformulation characteristics	Input	Output	Outcome of study	Ref.
Kohonen's SOMs (ViscoverySOMine software)	Tablets prepared by direct compression; 11 different disintegrants were tested for disintegration of tablets	Wetting time, water absorption ratio, particle size, morphological observation, swelling property, and relaxation time	Disintegration time of tablets	Classified the disintegration actions of test disintegrants into four distinct clusters	[146]
ANN (CAD/Chem)	Polymer blends (HPMC, PVP, HPC, carrageenan, sodium alginate) in water for glass transition temperature	Matrix polymer composition	Glass transition temperatures, viscosity, water uptake	ANNs accurately predicted outputs with a low % error (0–8%) of prediction	[147]
BT algorithm and multiple regression	Tablets prepared by direct compression method for tensile strength	Particle size distribution, Hausner ratio, moisture content, elastic recovery, and molecular weight of 81 drugs	Tensile strength	BT model had high performance	[148]
ANN (Matlab 6.1) and regression model	Solubility of drugs in water–cosolvent mixtures	Experimental solubility of solutes in water–cosolvent systems (35 datasets)	Solubility	ANN was superior to the regression model	[149]
ANN	Hydrotropic solubilization of indomethacin in water	Experimental data, together with various known and computed physicochemical properties	Solubility in water	In silico screening tool for drug/hydrotrope systems using ANN	[150]
ANN and MLR models	Skin permeability of new chemical entities	Skin permeability, Abraham descriptors of R_2 (excess molar refraction), the dipolarity/polarizability, the overall or effective hydrogen-bond acidity and	Skin permeability	Better prediction of skin permeability with ANN	[151]

(continued)

Table 11.3 (continued)

AI tool	Formulation/preformulation characteristics	Input	Output	Outcome of study	Ref.
		basicity, and the McGowan (215 datasets)			
DNN model and LASSO models	Permeability of small drug-like molecules across lipid membranes	Molecular descriptors and fingerprints	Membrane permeability	DNN model using molecular fingerprints can help develop a more accurate mapping	[152]
MLP and SVM	Predicting CNS permeability of drug molecules	MW, surface area, volume, log P, HLB, CNS activity (+/−), H 3d, H donor, H acceptor	CNS permeability	SVM algorithm was superior	[153]
ANNs, (BrainMaker Professional) and KNN	Pharmaceutical fingerprinting of samples of L-tryptophan (LT) (API) on the basis of HPLC trace impurity pattern	899 data entries extracted from each HPLC chromatogram; 253 chromatograms	Classifier developed	ANN with 46 inputs was superior to all other classifiers with 93% accuracy	[154]
SVM; four different types of ANN, namely, backpropagation (BPNN), genetic BPNN Mind Evolutionary Algorithm-Based BPNN (MEA-BPNN), and Extreme Learning Machine (ELM)	Powder compactability of powder blends for tensile strength and brittleness	Raw material attribute inputs; Conc. of material for shell and core (% wt/v), type of material for core, shell	Tensile strength and brittleness index	ANN algorithms were more capable of handling convoluted and nonlinear patterns of dataset	[155]

molecular fragments, topological indices, and descriptors calculated by semi-empirical quantum chemical methods [122]. For approved drugs, also the solubility and permeability estimation and their improvements play an important role in formulation development of new drug products. Physicochemical properties of polymers and polymer combinations have been modeled using ANNs.

In one study, ANNs were trained for modeling solubility of drug in water-cosolvent mixtures using 35 experimental datasets [149], and the developed model demonstrated superiority to the regression model. In another study, enhancement of solubility of indomethacin, a poorly water-soluble drug, was modeled using ANNs [150]. Solubility of indomethacin was estimated in the presence of hydrotropes by HPLC and UV detection methods. Experimental data, known physicochemical properties of hydrotropes, and computed properties were used to train ANN for predicting the solubilization of indomethacin. The trained ANN had good accuracy in predicting solubilization of indomethacin allowing computational screening of hydrotropes.

Good permeability of drugs ensures absorption of drugs. Poor permeability has always remained a great concern to a formulation scientist. The search of good permeation enhancers is now being assisted by AI. In one such study, SVM algorithm was found better than ANN in predicting CNS permeability of drug molecules when trained on identical datasets consisting of 179 CNS active molecules and 145 CNS inactive molecules [153]. DNN models using molecular fingerprints have been recently studied for modeling and predicting lipophilicity of drug-like molecules [152]. ANN model was developed using some molecular properties and penetration coefficients from carbon nanotube membranes to predict skin permeability [156]. In a different study, ANN was used to model and predict skin permeability of new chemical entities [151]. ANN model gave improved prediction of skin permeability when compared to multiple linear regression (MLR) models.

Crystal forms of active pharmaceutical ingredients (API) affect physicochemical properties like solubility, melting point, shelf life, bioavailability, density, and other physicochemical properties. Characterization of solid state has always remained a prime objective of preformulation studies and formulation development. ANNs have been employed for the identification, evaluation, and quantification of the crystal forms of unknown API samples based on the proper preprocessing of the X-ray powder diffraction patterns, attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy diffuse reflectance FTIR (DRIFT) spectral patterns, and other analytical techniques [157–160]. ANN and KNN are also performed as classifiers for pharmaceutical fingerprinting of bulk pharmaceuticals to discriminate among source of drug (manufacturer), detect changes in the manufacturing process, and to confirm instances of fraud or counterfeiting [154].

Polymer properties affect the loading and release of drug from different carriers like nanoparticles, microparticles, beads, matrix tablets, etc. Interaction of polymer with solvents also governs selection of appropriate manufacturing methods and/or processing conditions like mixing rate, temperature, etc. ANN modeling was used to predict the glass transition temperatures, water uptake, and viscosities of polymer

blends (HPMC, PVP, HPC, carrageenan, sodium alginate) in water [147].With a small set of experimental data (15 experiments) and validation set (9 experiments), the developed ANN model was capable of predicting at a very low error (0–8%).

SOM clustering was applied to classify the disintegration actions of test disintegrants [146]. Boosted tree algorithm revealed that d10 (diameter of powder particles at the 10th percentile), moisture content, and elastic recovery were the most crucial factors for tensile strength of directly compressed tablets [148].

11.5.3 Expert Systems for Preformulation and Formulation Designing

Expert systems, also known as are rule-based expert systems and knowledge-based expert systems, are the simplest form of AI and uses knowledge-based rules to solve a particular problem [161–163]. Expert systems acquire knowledge from human experts and covert this knowledge into hardcoded rules. Hardcoded rules are applied onto input data to derive output.

Expert system (Fig. 11.19) is typically subdivided into five major components. They are knowledge base acquisition facility, knowledge base, inference engine, explanation facility, and user interface [164]. The knowledge base acquisition facility helps the expert system to gather knowledge and data from various sources. The knowledge acquisition acts as an interface between experts and the expert system to gather knowledge about the specific field. From the knowledge base acquisition facility, the data are transferred to the knowledge base. Knowledge base can be described as databases which contain the rules and facts. To improve

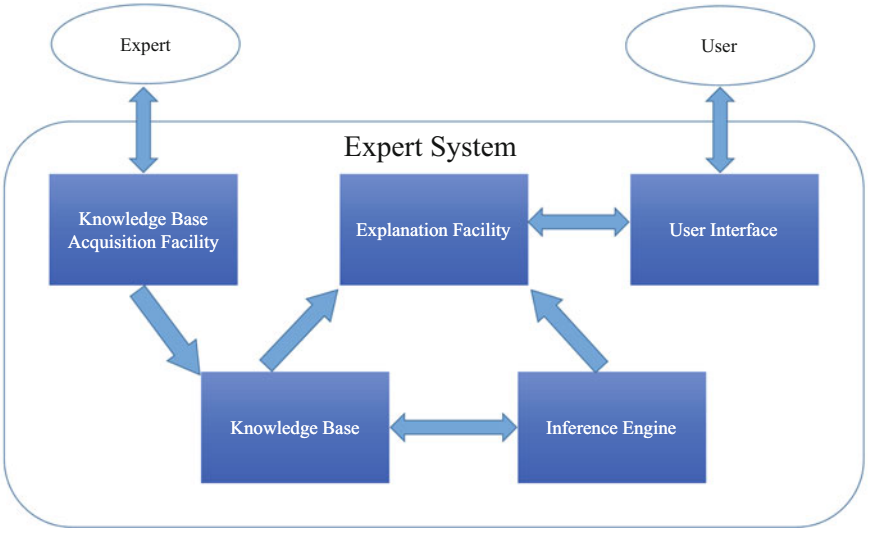


Fig. 11.19 Architecture and workflow of a knowledge-based expert system

the working of an expert system, knowledge should be given to the knowledge base about the knowledge it already contains. This is known as meta-knowledge or knowledge about the knowledge. From the knowledge base, the rules and facts are transferred to the inference engine. The inference engine works on the reasoning and providing answers based on prior knowledge [165]. It extracts knowledge from the knowledge base and provides solution to problem. It is divided into two types based on their method of working. One is backward chaining and the other is forward chaining. In backward chaining, the solution is found first and then the data is extracted to support the solution [166], whereas in forward chaining, the data is searched first and on the basis of the data the solution is given [167]. Backward chaining is target driven, whereas forward chaining is data driven. The fourth component is the explanation facility which explains the derived solution. It justifies the recommended solution and rejection of alternatives by the system. The last component of an expert system is the user interface [168]. This allows a nonexpert user to interact with the system. It serves as a medium of communication between the user and the system by presenting solutions in natural language. This generally consists of graphs and menus and eases the understanding to users and developers.

Cadila Expert System (tablets) [169], Zeneca system (tablets), Expert tab systems (fluidized bed granulation for tableting), CAPEX (capsule) [170], and Sanofi System (capsule) are some rule-based systems for pharmaceutical dosage forms [162, 163]. These expert systems have been designed for recommending quantitative and qualitative composition of desired characteristics. The knowledge base is created from the knowledge of experts. The acquired knowledge is used to structure production rules. The physicochemical, mechanical properties of selected drug and excipients and other data are entered into the system to obtain recommendations from the expert system. Such expert systems help formulation scientists in saving time and cost by predicting compositions of dosage forms. Knowledge base generated from these systems remained easily accessible, available consistently, and enriched itself with the updated information from experts. Some expert systems developed recently for optimizing pharmaceutical formulations are summarized in Table 11.4.

Some other rule-based expert systems, like CAPEX and ESFppop, utilize ANN in their prediction module. CAPEX (CAPsugel Expert) integrates backpropagation neural network and a rule-based expert system for formulation development of Biopharmaceutical Classification System (BCS) class II drugs [170, 171]. This expert system was developed through continuous research and expert panel suggestions (industrial, academic, and regulatory experts from the United States, Europe, and Japan). Its architecture involves integration of four different modules, viz. formulation module, prediction module, parameter adjustment module, and control module (Fig. 11.20). Formulation module encodes knowledge and experience of formulation experts in the form of a decision tree. Formulation module further consists of three different decision modules, viz. content uniformity, direct fill, and ordered mixing. Prediction module of this expert system is a backpropagation learning system. After completion of training, the ANN predicts the output as a feed forward method. Parameter adjustment module allows human to

Table 11.4 Expert systems for screening of excipients and formulation designing

Expert system	Formulation/ dosage form	Application	Organization	Ref.
CAPEX	Hard gelatin capsules	Formulation designing and prediction of dissolution rate	University of Maryland Baltimore County	[171, 172]
ESFppop	Push-pull osmotic pump (ppop) of poorly water-soluble drugs	Predict composition of push-pull osmotic tablets and drug release	Shenyang Pharmaceutical University, China	[173]
SeDeM expert system	Matrix tablets of theophylline	SeDeM expert system conceived and applied to assess compressibility of powder mixtures	University of Barcelona, Spain	[174]
SeDeM expert system	ODTs	Screening of excipients on the basis of Index of Good compression (IGC) for preparing ODTs	University of Barcelona, Barcelona, Spain	[175]
SeDeM-ODT expert system	ODTs	New model that provides the Index of Good Compressibility and Bucodispersibility (IGCB index) to assess fast in vitro disintegration of ODTs prepared by direct compression	Novartis Pharmaceutical-Spain and University of Barcelona, Barcelona, Spain	[176]
SeDeM expert system	Drug excipient powder blend	Assess suitability of theophylline and lactose blends for direct compression	University of Seville, Spain	[177]
SeDeM expert system	Matrix tablets of theophylline for sustained release	Assess suitability of biodegradable polyurethanes and theophylline blends for direct compression	University of Seville, Spain	[178]
SeDeM expert system	Medicated chewing gum tablets of lyophilized lysozyme	Assess powder blends for its suitability for direct compression	Goethe-University Frankfurt, Germany	[179]
SeDeM and SeDeM-ODT	Preformulation studies of pediatric ibuprofen ODT tablets	Compressibility and bucodispersibilty of ibuprofen-Ludiflash blends; optimized ODT tablets	University of Medicine and Pharmacy Tirgu Mureş, Romania	[180]

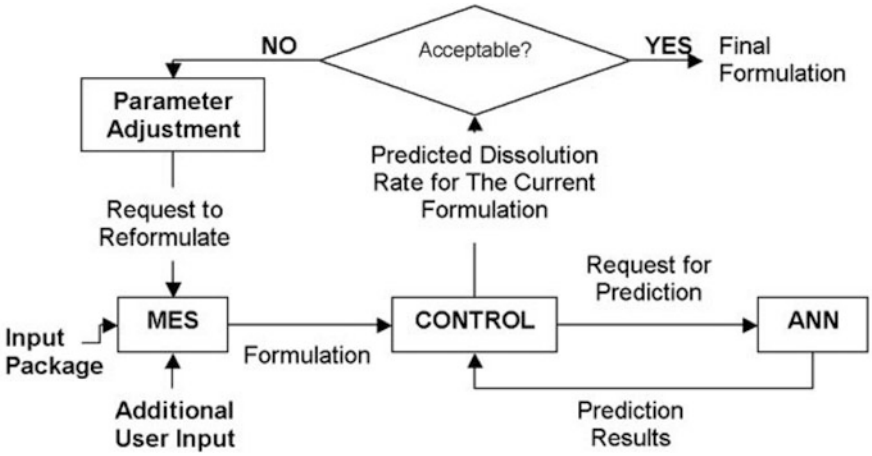


Fig. 11.20 CAPEX Expert System. (a) Integration of different modules (Figure Reproduced from [170]). Copyright with Springer

adjust formulation parameters as per the need. Control module integrates all modules and other parts of the system. CAPEX use involves a cycle of (re)formulation-prediction until the human formulator is satisfied with the predicted dissolution rate. This expert system can run on a Windows environment which has further improved its popularity and acceptance.

ESFppop uses SQL Server for database management system (DBMS), backpropagation neural network for prediction module, man-machine interface by VB.NET, and formulation design module based on prediction model for drug release. The workflow for designing push-pull osmotic pumps of poorly water-soluble drugs is shown in Fig. 11.21.

The SeDeM expert system was first conceptualized by Carreras and colleagues at the University of Barcelona [181]. This expert system classifies 12 physical properties of powders into 5 groups, viz. dimension, compressibility, flowability, lubricity/stability, and lubricity/storage. It can assess the suitability of excipients for direct compression, determine the amount of excipients, and classify directly compressible excipients. Thereby, SeDeM expert system serves as a good decision support prefomulation tool for tablets. Recently SeDeM expert systems have been modified and applied to other dosage forms like ODTs [182], sustained release matrix tablets [178], and chewing gums [179]. SeDeM-ODT uses three additional parameters and groups them into one additional group named disgregability (representing disintegratability of ODTs) [176]. SeDeM expert system has established itself as a tool in the QbD approach for screening of excipients and a quality control tool for assessing the suitability of powder blends for direct compression. A virtual knowledge base for SeDeM expert system containing data for 12 parameters for various excipients is available at iCTM knowledge base [182, 183].

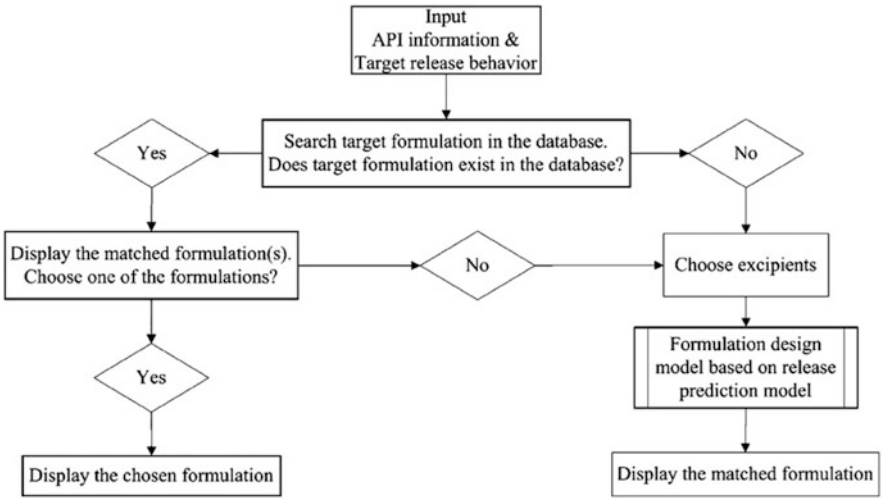


Fig. 11.21 Formulation design system workflow for ESFppop [173]

11.5.4 ANN for Formulation Development and Optimization

Application of ANNs in formulation development can be grouped into two main classes: screening of excipients for their desirable properties and the optimization of compositions of pharmaceutical formulations. Recent studies from Table 11.5 confirms that ANN are widely used for formulation development and optimization of the traditional as well as advanced drug delivery systems, such as tablets [189], minitablets [201], solid dispersions [184], pellets [191], ODTs [187], printlets [192, 193], bilosomes [194], liposomes [195], hydrogels [196], nanoparticles [198], monoclonal antibodies [199], and many more. A number of controlled-release drug delivery systems, such as oral [89, 190, 200], transdermal [188], injectable [199], have also been designed and optimized with the support of ANN models. But the complexities of in vivo absorption pose considerable challenges to such applications of ANN.

Probably for the first time, Han and colleagues developed a DNN-based model (Fig. 11.22) for predicting formulations of pharmaceutical dosage forms [187]. Datasets for 145 formulations of ODTs prepared by direct compression were obtained from the Web of Science database. Inputs for the developed model were molecular parameters of drug, amount of drugs, amount of excipients, type of encoded excipients, manufacturing parameters, and disintegration time. Dissimilarity algorithm was used to select data for test and validation sets. Higher accuracy (80%) was achieved for test data with DNN when compared to ANN (75% accuracy) for predicting disintegration time. Predictive approach using DNN and ANN was developed to assess quality of formulations. In a similar study, DNN model outperformed six other machine learning models, MLR, PLSR, SVM, ANNs, RF, k-NN, for prediction of *in vitro* characteristics of oral sustained release matrix tablets and fast disintegrating films [89].

Table 11.5 Applications of ANN in formulation development and optimization

AI tool	Formulation/dosage form	Inputs	Output	Outcome	Ref.
ANN-based expert system (PharmCAD expert system)	Ketoprofen solid dispersions (SD) and physical mixtures	7 inputs; type and preparation technology as well as qualitative and quantitative composition of SD and physical mixtures (PM)	Drug dissolution	ANNs functioned well as decision support system and data-mining tool	[184]
ANN-based expert system (PharmCAD expert system)	Solid dispersion	16 inputs; MW of drug and two carriers, amount of drug and two carriers, Connectivity Index (CI) of drug and two carriers, formulation type, preparation technology, cooing conditions, rpm paddle/basket, pH of dissolution media, time of dissolution	Amount of drug dissolved	PharmCAD expert system for deriving knowledge from empirical data	[185]
Fuzzy logic-based expert system	Freeze-dried peptide and protein-based formulations	Off coloring, image entropy, cake collapse, and light saturation were inputs	Quality of cake (quality attributes collapse, glassiness, uniformity, and color)	High-throughput formulation screening by image analysis	[186]
ANN and DNN model	ODTs	Molecular parameters of drug, amount of drugs, amount of excipients, type of encoded excipients, manufacturing parameters, and disintegration time	Disintegration time	DNN performs well in all three datasets with over 80% accuracy	[187]
DNN, MLR, PLSR, SVM, ANNs, RF, k-NN		Molecular parameters of drug, amount of drugs,			[89]

(continued)

Table 11.5 (continued)

AI tool	Formulation/dosage form	Inputs	Output	Outcome	Ref.
	Oral sustained release matrix tablets and fast disintegrating films	amount of excipients, type of encoded excipients, manufacturing parameters, and disintegration time	In vitro characteristics; disintegration and drug release	DNN model as superior to other machine learning models	
ANN model	Topical matrix patches of diclofenac sodium	Time, chitosan amount, and carrageenan amount	Drug release and the ex vitro skin permeation kinetics	• ANN-predicted outputs with reasonable accuracy	[188]
• INForm V.4 ANN for neural networks, FormRules V.3.32 for neurofuzzy logic, and INForm V.4 GEP	Ramipril tablets prepared by the direct compression	HPMC, lubricant type, lubricant concentration	Tablet weight, friability, disintegration time, dissolution	Support decision-making processes; optimized formulation was within the design space	[189]
MLP (ANN Neural Power® version 3.1)	Mesalamine matrix tablets by wet compression	9 inputs; amount of excipients	Friability, thickness, hardness, weight variation, content uniformity, dissolution	ANN-aided optimized formulation	[190]
MLP and Box Behnken Design	Multiple-unit prednisone pellet system	MCC concentration, SSG concentration, spheronization time and extrusion speed	Aspect ratio, drug release at different time points, yield	ANN was recommended as complement RSM for optimization (Box Behnken design)	[191]
Supervised MLP (STATISTICA 7.0 Neural Networks and MATLAB R2014b) and D-Optimal mixture design	Cross-linked polymeric ibuprofen printlets by digital light processing (DLP)	Amount of PEGDA, PEG 400, and water	In vitro drug release at different time points	Both ANNs performed better than D-optimal mixture design	[192]
A generalized regression neural network (GRNN) (TIBCO Statistica® Software), and self-	DLP-based 3D-printed tablets of atomoxetine	Data from 23 experiments; input variables were: tablet thickness and drug loading	Release rate after 15, 30, 60, 120, 240, and 360 min	ANN predictive models for atomoxetine release rate developed	[193]

organizing map (SOM) (Peltarion® software)					
ANN	Transdermal terbutaline sulfate-loaded bilosomes in gel thin film hydration method	Soybean phosphatidylcholine, cholesterol, sodium deoxycholate, chitosan	Particle size, entrapment efficiency, and drug release after 24 h	AN model for optimizing terbutaline sulfate-loaded bilosomes	[194]
ANN and RSM	Flexible nano-liposomes containing pilocarpine HCl	Amount of pilocarpine, sodium deoxycholate, and water content	Entrapment efficiency	ANN prediction was better	[195]
ANN and RSM	Lamotrigine containing thermosetting hydrogel for intranasal administration	Amount of Poloxamers 407 and 188, Carbomer 974P	Viscosity, pH, sol-gel transition, and drug release	Multilayered perceptron was more accurate in predictions	[196]
ANN (Matlab R2018a) and PLS	An extended-release formulation containing drotaverine	Theoretical drug and HPMC content	Predict the dissolution profile	Prediction of the ANN models was more accurate than PLS	[197]
ANN and MLR	Diclofenac sodium gelatin nanoparticles, nanoprecipitation, and two-step desolvation	Poloxamer content, gelatin content, and glutaraldehyde concentration	Particle size, zeta potential, and entrapment efficiency	Pruned ANN prediction for optimized formulation was better than MLR	[198]
ANN and PLS	Monoclonal antibodies	pH and salt concentration	Melting temperature, aggregation onset temperature, interaction parameter	RSM and subsequent ANN application for optimization	[199]
ANN models and PLS regression	An extended-release tablet formulation	Composition of the tablets and the tableting compression force	In vitro dissolution	ANN served better	[200]
• ANNs, particle swarm optimization (PSO) ANNs, and genetic programming	Mini-tablets	Small, medium, large, lactose, starch 1500, 200Pharma, R972V	Powder blend's bulk and tapped density, Carr's compressibility index,	PSO-ANNs fitting was adequate fit to all responses simultaneously	[201]

(continued)

Table 11.5 (continued)

AI tool	Formulation/dosage form	Inputs	Output	Outcome	Ref.
			Kawatika's fitting parameters (a, 1/b) minitablets' avg. weight, weight variation, and relative density		

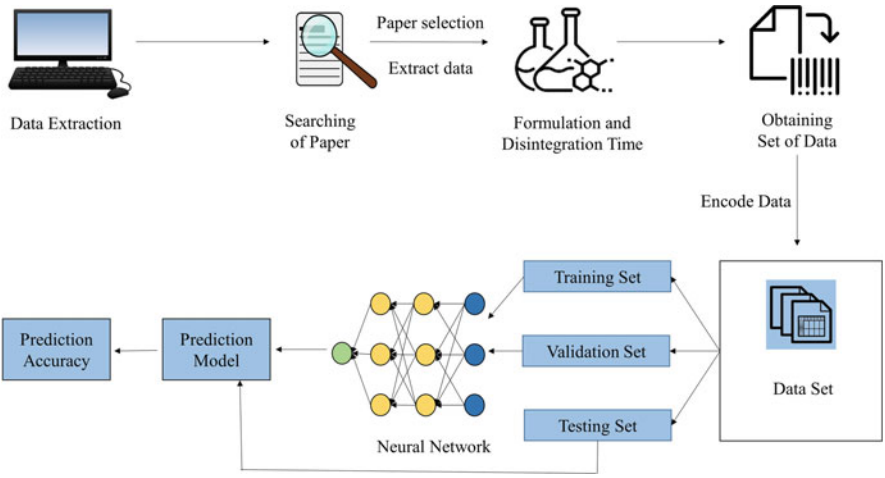


Fig. 11.22 A representative DNN-based model for predicting pharmaceutical formulations

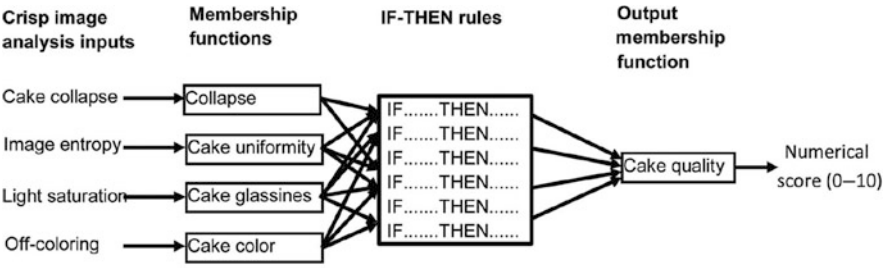


Fig. 11.23 Overview of the developed fuzzy logic system for evaluating cake quality [186]

An expert system based on fuzzy logic inference system was developed (Fig. 11.23) for assessment of cake quality of lyophilized formulations [186]. MatlabR fuzzy logic toolbox of MathWorksR R2012b used image analysis results as inputs and Mamdani's fuzzy inference method. Off coloring, image entropy, cake collapse, and light saturation were inputs to the expert system. Membership functions were used to fuzzify crisp image analysis inputs. If-then rules, as per the inputs, were constructed to obtain fuzzy outputs. The output was defuzzified after receiving results from all rules. A numerical score on a scale of 10 reflected the quality of cake. The developed expert system was found satisfactory allowing high-throughput assessment of the cake quality.

Formulation composition, preparation technology, and processing parameters were modeled in a single ANN model and used as a decision support system (PharmCAD system) in developing pharmaceutical formulations [184]. PharmCAD system applied backpropagation ANN with one to four fully interconnected hidden layers and fuzzy network to predict in vitro drug dissolution from solid dispersion formulations. Formulation type (solid dispersion/physical mixture), preparation

technology (evaporation/melting; cooling on ice/ambient temperature), qualitative parameters (carrier class) and quantitative composition of SD and physical mixtures (PM) (drug amount), and time of dissolution test were taken as input variables and output was drug dissolution. A reasonable performance of the developed model was achieved with overall generalization error close to 8% of maximum output value. In one previous study with the same PharmCAD system but with 16 inputs, specialized models were developed for a particular class of carriers, for example, macrogols, or for soluble drugs [185].

11.5.5 Production Technology

AI has established its profound use in manufacturing of dosage forms and drug delivery systems. Process monitoring, process modeling, process control, and planning and scheduling are assisted with the help of AI for finding patterns, optimizing, self-organizing, and mapping of the manufacturing processes (Table 11.6). Manufacturing processes for wet granulation [203], tableting [205, 206, 208], extrusion process (liposomes) [207], and extrusion spheronization [202] have been modeled with ANN and various other ML models. AI methods learn from its knowledge base, that is, training data, and apply the training to ascertain relationships among input variables to predict outputs within the design spaces, as required by QbD framework. In one study, DNN model and PAT were synergistically used to monitor a continuous tablet manufacturing line [203]. Predictions from DNN were helpful in obtaining sharp views of process understanding where PAT data was noisy. Additionally, DNN was recommended as a backup monitoring tool if in case one of PAT sensor fails.

ANN modeling was utilized for investigating critical raw material attributes and critical process parameters of a pelletization process (extrusion/spheronization) which can influence the drug dissolution and subsequent developing a capsule manufacturing process [202]. Retrospectively aggregated datasets from 29 industrially manufactured capsules filled with pellets were used to train ANN models. ANN investigations through sensitivity analysis revealed the greatest influence of particle size of drug and processing variables: water temperature, final torque, and spheronization time on drug dissolution after 30 min. The developed ANN model was successful in predicting drug dissolution with good accuracy.

ANN model was investigated for its implementation into the PAT concept as an integral part of the QbD approach for manufacturing tablets by slugging, direct compression, and dry granulation on a roller compactor [205]. The study found that modeling was comparatively expensive to trial and error methods, but the model has the advantage of its use in predicting optimal settings for new batches of tablets, and thereby it reduces the time for optimizing the machine for manufacturing a new batch and also reduces capping in manufactured tablets.

ANN modeling and PLS-based models were used to investigate the variability in tableting and disintegration time of high drug load (>60%) tablets prepared by a high shear wet granulation method [206].Historical 3-year data from 95 industrial-

Table 11.6 Role of ANN and other machine learning approaches in optimizing various manufacturing processes

AI tool	Production technology	Input	Output	Outcome of study	Ref.
MLP (Peltarion Synapse software); PLS and PCA	Extrusion spheronization for manufacturing pellets	75 input parameters; CMA (κ-carrageenan); and CPP (Extrusion spheronization)	Drug release after 30 min (Q30)	Developed QbD-based control strategy for capsules filled with pellets	[202]
DNN	A GMP continuous wet granulation line for table manufacturing	Particle size distribution (D10, D50, D90), contents of tablets, content in feed frame, content after dryer, content after blender	Loss on drying	Synergy between PAT and process data science creates a superior monitoring framework	[203]
Decision tree, support vector machine, bagging on decision trees, AdaBoost on decision trees	Tablet production	Numerous raw material attributes and process parameters	Batch quality class (high, medium, rejected)	Developed meta-classifier for assisting the human operator to control the manufacturing process	[204]
ANN and fuzzy models	Tablet production	Compression force, tableting speed Particle size and others CPP, CMA	Capping coefficient	Significantly improve process optimization	[205]
ANN	Industrial-scale tablet (wet granulation) production (historical dataset acquired for 95 batches)	Raw material properties, granule properties, process evolution data, water quantity, tableting speed, disintegration time	Tableting speed, disintegration time, and water amount	Reliable prediction ability was obtained	[206]
	Manufacturing ultra-	Amount of cholesterol,	Size before extrusion, zeta	ANN showed much better	[207]

(continued)

Table 11.6 (continued)

AI tool	Production technology	Input	Output	Outcome of study	Ref.
MPN and MLR model	deformable nanoliposome formulations of timolol	amount of edge activator (mg) phase in which drug was added, addition of stearylamine, type of edge activator	potential, polydispersity index, percentage of drug entrapped; all parameters before extrusion and after extrusion	predictive ability than the MLR model	
ANN	Tablets for BCS Class IV drugs	Drug PSD, namely D10, D50, and D90; tablet hardness, impeller speed, mesh size for sieving the dried granules, granulation time, and granulation liquid amount	In vitro dissolution; Q (10 min), Q (30 min), and Q (60 min)	ANN model was successfully applied to study the impact of CPPs and CMAs on quality of tablets	[208]

scale batches was used to build and train models. Prediction of output for tableting speed and disintegration time revealed better performance of ANN compared to PLS for tableting speed where ANN model was superior due to better ability of ANN to handle nonlinearity in the data. However, PLS prediction was better in case of disintegration time.

Three different ANN programs for neural networks, viz. FormRules V.3.32 for neurofuzzy logic, INForm V.4 gene expression programming, and INForm V.4 ANN for neural networks, were utilized for investigating the impact of critical quality attributes on the overall quality of ramipril tablets prepared by direct compression [189]. Knowledge space and design space limits for various inputs and outputs were estimated using ANN models. ANN models were considered good to investigate input variables for their multidimensional interactions and to keep these variables into a design space. Optimized formulation was prepared and the quality attributes were found within the design space and the knowledge space.

Feed forward backpropagation network ANN model and MLR using first-order polynomial equation were compared to study the effects of various qualitative and quantitative formulation parameters in preparing timolol-loaded liposomes and predicting best formulation for each particular response variable [207]. Developed models for responses such as polydispersity index, vesicle size, percentage of timolol entrapment, and zeta potential were trained satisfactorily by ANN for all training algorithms and architectures. ANN models were found more accurate in

predicting formulation variables. However, in regard to predictive ability, no significant differences were observed in the validation process.

A meta-classifier was developed as QbD cum PAT tool for assisting human operator to control the tablet manufacturing process [204]. Meta-classifier assimilated rule construction, with the help of human experts and ML-based methods in a single system, improved the classification accuracy. An application module was used to visualize and guide the manufacturing process. A human-machine interface allowed for simulating changes in the manufacturing process. Meta-classifier combined all four ML algorithms (decision trees, SVM, bagging on decision trees, AdaBoost on decision trees) and was trained similarly as individual ML algorithms were trained. Meta-classifier, tablet classifier, four ML-based algorithms, and human majority voting were tested for the assessment of a small database for predicting quality of tablets. A single decision tree had the least accuracy for classification. Only AdaBoost was able to score better than the human-modified rules (tablet classifier) as it was built on a combination of many decision trees. Developed meta-classifier scored best with 99.7% of accuracy and found suitable for its use as an interactive aid to a human operator.

11.6 Healthcare Applications

Healthcare application of AI may be divided into three different levels: personalized treatment, diagnosis by clinicians, and epidemic predictions. FDA approves AI/ML-based softwares as medical devices (SaMD). FDA's "AI/ML-Based Software as a Medical Device Action Plan" has suggested five-part action plan, that is, tailored regulatory framework for AI/ML-based SaMD, Good Machine Learning Practices (GMLP), patient-centered approach, regulatory science methods to assess algorithm bias and robustness, and real-world performance for facilitating approval of SaMDs [209]. Some FDA-approved AI-based devices are presented in Table 11.7. An online database for FDA-approved AI algorithm may be accessed for more details [210, 211].

11.6.1 Personalized Treatment

Personalized treatments are now monitored with the use of AI by physicians for more accurate pathological assessments and by providing treatment decision support at the right time in cases of chronic diseases like cancer, diabetes, etc. AI allows physicians to limit the choices of diagnosis and helps to assess risk to patients on the basis of symptoms or genetic information.

A meta-analysis of published research articles between January 2000 and may 2018 was conducted to calculate the diagnostic accuracy of different ML algorithms for breast cancer risk calculations [212]. Five ML algorithms SVM, ANN, decision tree, naïve Bayes, and KNN were able to predict breast cancer. Meta-analysis

Table 11.7 FDA-approved AI/ML-based devices

Medicinal field	Devices (company, date of approval)
Cardiology	Imbio RV/LV Software (Imbio, LLC, 2021) Caption Interpretation Automated Ejection Fraction Software (Caption Health, 2020) Eko Analysis Software (Eko Devices, Inc., 2020) AI-ECG Platform (Shenzhen Carewell Electronics Ltd., 2019) EchoGo Core (Ultromics Ltd., 2019) Arterys Cardio DL (ARTERYS, Inc., 2018) EchoMD (Bay Labs, Inc., 2018)
Endocrinology	Guardian Connect System (Medtronic MiniMed, Inc., 2018) DreaMed Advisor Pro (DreaMed Diabetes Ltd., 2018)
Radiology	MEDO-Thyroid (Medo.AI, 2021) LVivo Software Application (DiA Imaging Analysis Ltd., 2020) FastStroke, CT Perfusion 4D (GE Medical Systems SCS, 2020) TransparaTM (Screenpoint Medical B.V, 2019) Deep Learning Image Reconstruction (GE Medical Systems, LL, 2019) HealthPNX (Zebra Medical Vision Ltd., 2019) Advanced Intelligent Clear-IQ Engine (AiCE) (Canon Medical Systems Corporation, 2019) SubtleMR (Subtle Medical, Inc., 2019) AI-Rad Companion (Pulmonary) (Siemens Medical Solutions USA, Inc., 2019) EchoMD Automated Ejection Fraction Software (Bay Labs Inc., 2018) SubtlePET (Subtle Medical, Inc., 2018)
Neurology	Viz ICH (Viz. AI, inc., 2020) Accipiolx (MaxQ AI Ltd., 2018) Icobrain (Icometrix,2018) ContaCT (Viz.AI, 2018) EnsoSleep (EnsoData, Inc., 2017)
Internal medicine	FerriSmart Analysis System (Resonance Health Analysis Services Pty Ltd, 2018)
Ophthalmology	IDx-DR (Digital Diagnostics, 2018)
Emergency Medicine	Critical Care Suite (GE Medical Systems, LLC., 2019) HelathPNX (Zebra Medical Vision Ltd., 2019) BriefCase (Aidoc Medical, Ltd., July 2018) OsteoDetect (ImagenTechnologies, Inc., 2018)
Oncology	QuantX (Quantitative Insights, Inc., 2020) Transpara™ (ScreenPoint Medical B.V., 2019) cmTriage (CureMetrix, Inc., 2019) ArterysMICA (Arterys, Inc., 2018) Arterys Oncology DL (Arterys Inc., 2018) Profound™ AI Software V 2.1 (CAD, Inc, 2018)

established that the SVM algorithm was superior in calculating breast cancer risk with better accuracy than other ML algorithms.

Biosensors, devices, and mobile apps are now equipped with AI algorithms for health measurements and remote monitoring abilities. Monitoring of patients from remote locations by physicians or self-monitoring by patients may reduce overall

healthcare costs. SkinVision mobile app uses a CE-validated algorithm for the assessment of skin cancer risk from moles or lesions, skin spots, or other skin locations by simply uploading images from android or IOS smartphones (https://www.skinvision.com/). FDA-approved Sugar.IQ™ app is an outcome of joint research of Medtronic and IBM Watson [213]. The app collects data from Medtronic's continuous glucose monitor and tracks the effect of medications, food, and lifestyle on glucose levels in diabetic patients. IBM Watson Oncology also helps physicians in optimizing the treatment options using medical history and medication records of patients [214].

MIT Clinical Labs uses unsupervised learning algorithms for clinical predictions and studying progression of diseases like multiple myeloma, antibiotic resistance, Parkinson's disease, and other chronic illnesses [215].

Project Hanover of Microsoft is using ML algorithms for precision treatment of cancer at the Knight Cancer Institute (Oregon Health and Science University) and personalized drug therapy for Acute Myeloid Leukemia (AML) [216].

11.6.2 Medical Diagnosis

The diagnosis of disease involves grading the degree and progression of the disease which needs clear data from the patient. To improve the accuracy of diagnosis and treatment of the disease, AI plays a crucial role. The clinical potential of ANN was first explored by William G. Baxt, Department of Medicine, University of California [217]. He developed an ANN which accurately diagnoses acute myocardial infarction. From ultrasound, MRI scanning, to electrocardiogram and electroencephalogram, ANNs are used to diagnose diseases. Some other relevant uses of ANN are in the diagnosis of appendicitis, abdominal pain, and retained stone in bile duct. A classification algorithm called Post Assure Index was developed to classify changes in prostate whether benign or malignant [218]. This model had an accuracy rate of 90% with sensitivity and specificity of 81% and 92%, respectively. Fuzzy logic is a data processing methodology that has many applications in the field of medicine. Fuzzy logic is used in the diagnosis of leukemia cancer and is more precise than logistic regression analysis in lung cancer diagnosis using tumor markers [219]. In one study, among four classification algorithms ANN, Tree J48, Näıve Bayes, and Lazy-IBk tested for MRI images to detect brain tumor, ANN was the best and Lazy-IBk also did fairly well [220].

Medical technologies like positron emission tomography (PET), computed tomography (CT), and magnetic resonance imaging (MRI) have revolutionized the noninvasive evaluation of brain to identify the exact position of lesions in the brain or brain tumor. Manual processing of medical/brain images is a time-consuming process and may lead to errors of about 3–5% [221]. DL methods are widely used now for addressing brain-related diseases. A trained CNN with data from 129,450 images was at par with expert dermatologists in classifying skin cancer [222]. CNN has been widely used and popular for image-level diagnostics as it has achieved human-level performance in object classification tasks contained in images

[223]. DL has advanced over the past several years. AlexNet, a deep convergent neural network, has increased accuracy in classification of high-resolution images [224]. GoogLeNet, a 22-layer deep CNN developed by Google, has exceeded the human limit of image recognition accuracy [225, 226].

Researchers at Oxford University have a developed an AI-based screening test for Covid-19 [227]. This test, CURIAL AI, is optimized to give negative results with high confidence for Covid-19 within the first hour of arrival of a patient in the hospital. CURIAL algorithm requires data from blood tests, vital signs, blood gas testing, PCR for influenza and other respiratory viruses, and RT-PCR report for Covid-19. Another version of the screening test CURIAL-Rapide, which is under development, uses only CBC results and vital signs for ruling out Covid-19 within 10 min [228].

11.6.3 Epidemic Surveillance and Forecasting

AI-based systems have been applied for surveillance and predicting epidemics and issue alerts ahead of time to governments, other organizations, and public [229, 230]. BlueDot, AI platform, created a global Zika virus spread model and warned international spread of Zika virus in 2016 [231]. BlueDot uses human intelligence and AI (NLP and ML) to collect and analyze data from different sources to develop models for predicting spread of infectious diseases. BlueDot also identified Covid-19, as unusual pneumonia cases in Wuhan (China), 9 days before the WHO's alert on Covid-19 [232].

Google Flu uses search engine queries to track flu epidemic, but it underperformed due to its poor design. Recently it has resurfaced again with the use of data from Google Flu in influenza forecast systems [233].

FluSense is a contactless surveillance platform for real-time or near real-time prediction of seasonal flu and other viral respiratory outbreaks like SARS and Covid-19 [234]. FluSense uses a neural computing engine, a thermal camera, and microphone array to passively and continuously characterize cough sounds and speech in crowded places in a real-time manner.

11.7 Some Proprietary AI Technologies and Their Collaborations

Pharmaceutical companies are looking forward to AI for improved diagnostics or new drug designing, formulation development, biomarkers, identification of drug targets, improvements in therapy, and many more applications.

11.7.1 Drug Discovery and Development

The profit of pharma companies using AI reached from US$200 million in 2015 to US$700 million in 2018 and is expected to rise further up to $5 billion by 2024 [235]. With an expected 40% of growth in collaboration among companies from 2017 to 2024, it clearly shows that AI has the potential to revolutionize drug discovery and development. This paradigm has motivated pharmaceutical companies for collaborations with AI companies for seeking potential benefits of speeding up their drug discovery and development programs and boost profits. According to a recent report [236], top 10 AI-based companies on the basis of their collaboration with pharmaceutical companies are provided in Table 11.8 and top 5 are discussed below for their AI technologies.

Exscientia: Exscientia is an AI-driven pharmatech company founded in 2012 [237]. This was the first company to automate drug design and first to have an AI-designed immuno-oncology molecule entering clinical trials. It took less than 12 months for the company to bring the candidate into clinical trials. Exscientia uses its platform technology Centaur Chemist™ AI system for automated designing of drug molecules. Centaur Chemist™ learns not only from existing datasets but also from each of its design cycles. This system is very effective in identifying and assimilating complex trends to balance selectivity, potency, and pharmacokinetic criteria. This virtue results in more efficient achieving of the end goal than human learning. Exscientia have till date collaborated with many leading pharmaceutical companies like Sanofi, Roche, Evotec, Bayer, and GSK.

Atomwise: Atomwise is a California-based company founded in 2012 and formerly known as Chematria Atomwise. Atomwise has its own proprietary CNN-based DL technology (AtomNet) for hit discovery, toxicity prediction, and lead optimization. AtomNet represents a protein-ligand pair as a set of three-dimensional images with channels of carbon, oxygen, nitrogen, and other atom types. Atomwise has collaborated with many leading pharmaceutical companies like Lilly, Bayer, GC Pharma, and Hansoh Pharma.

Cyclia: Cyclia is a Toronto based AI company founded in 2013 and works for data-driven drug discovery [239]. Matchmaker, the deep learning platform launched in 2018, helps in proteome wide evaluation for determining complex drug polypharmacology in real time, that is, on and off target interactions. Matchmaker has been trained not only on millions of drug–target interactions but also on thousands of protein 3D structures. Matchmaker can screen across human proteome in less than one second. Cyclia's Pareto-optimal modeling (POEM), another supervised algorithm, helps in building reliable ADMET models on molecules. AstraZeneca and Bayer have partnered with Cyclia for better and efficient drug discovery.

Schrödinger: Schrödinger, a New York based company founded in 1990, has several platform technologies for computational drug discovery [240]. DeepChem/AutoQSAR is a DL platform of Schrodinger. DeepChem integration with autoQSAR allows application of DL algorithms to predict QSAR models on large datasets. FEP+ (free energy perturbation calculation software)-predicted affinities or

Table 11.8 Some recent collaboration of AI companies with pharmaceutical companies

AI company	AI platform	Major collaboration with pharmaceutical company	Date of collaboration	Use	Ref.
Exscientia	Centaur Chemist	Bayer	January 2020	Identify and optimize novel lead molecule for oncological and cardiovascular disease	[237]
		Sanofi	May 2017	To find bispecific small molecules for diabetes and comorbidities	
		GSK	April 2019	Finding novel drug molecules for targeting pathways of chronic obstructive pulmonary disorders	
Atomwise	AtomNet	Lilly	June 2019	To develop drugs on novel protein targets	[238]
Cyclia	MatchMaker	Bayer	November 2018	Pharmacokinetic property prediction and multitargeted drug design	[239]
		Merck	December 2018	Elucidate mechanism of action, safety profiles of investigational small molecules	
Schrodinger		Bayer	January 2020	Codevelop de novo design technology to accelerate drug discovery	[240]
		AstraZeneca	September 2019	To develop advanced computing technology for drug discovery.	
Insilico Medicine	Pharma.AI	GSK	August, 2017	To identify novel biological pathways	[241]
Iktos	Makya™	Pfizer	March 2021	De novo designing software for	[242]

(continued)

Table 11.8 (continued)

AI company	AI platform	Major collaboration with pharmaceutical company	Date of collaboration	Use	Ref.
				multiparametric optimization	
Biovista	COSS™	Astellas	December 2015	Identifying new indications for a number of undisclosed compounds	[243]
Numerate	Algorithm-driven drug discovery platform	Takeda	June 2017	Drug discovery for oncology, gastroenterology, and CNS disorders	[244]
Berg	bAlcis®	AstraZeneca	August 2017	Evaluation of novel targets for neurodegenerative disorders	[245]
		Sanofi	October 2017	Assess potential biomarkers for seasonal flu vaccine performance	
Benevolent	Benevolent Platform®	AstraZeneca	April 2019	NN-based platform for treatment of chronic kidney disease and idiopathic pulmonary fibrosis	[246]
		Janssen	November 2016	License the rights to develop and manufacture clinical stage drug candidates	

Glide (a virtual screening program) docking scores can be used to sample potential molecules with the help of ML methods of Deep-Chem/AutoQSAR platform. In April 2021, NVIDIA, in partnership with Schrödinger, has brought Schrödinger platform and NVIDIA Clara™ Discovery (AI frameworks of NVIDIA) together on NVIDIA DGX SuperPOD, cloud-native, multi-tenant AI supercomputer [247]. Pharmaceutical companies can now run Schrödinger platform on NVIDIA DGX SuperPOD, which is available to install in a colocation facility or on premises. Schrodinger has also announced collaboration with AstraZenaca to extend computational modeling solutions to biologics [248].

Insilico medicine: Insilico medicine, a Hong Kong based AI company founded in 2014, applies DL, big data, and genomics for in silico drug discovery for cancer and

antiaging [241]. Insilico Medicine applies GAN and reinforcement learning for development of novel molecular structures with known/unknown targets. In partnership with LifeExtension.com, the company has introduced a range of nutraceutical products using the advanced bioinformatics techniques and DL. Young.AI is a consumer-facing application for longevity science. Boehringer Ingelheium has come forward to collaborate with Insilico Medicine for identifying new drug targets. Insilico Medicine has announced collaboration with Astellas Pharma for their expertise in drug discovery.

11.7.2 Formulation Development and Manufacturing

AI nowadays plays a prominent role in not only drug designing and development but also in drug delivery and formulation development [249, 250]. The development of AI systems, like expert systems, MLs, and DNNs, for formulation designing and predicting quality or vice versa has transformed the hit and trial approaches to rational formulation development programs.

Metis Pharmaceuticals has developed an AI-based formulation platform technology for optimizing pharmaceutical formulation in high dimension design space and across multiple dosage forms [251]. This proprietary AI system uses a de novo generator of formulation conditions for bulk formulations including nano carriers followed by prediction of properties through an ML prediction system. Predicted formulations are checked on high-throughput experimental platforms and subsequently through in vitro tests. The outcomes of the in vitro test are fed to the proprietary database for learning and/or selection of other/suitable excipients. AI platform, MedAI, also provides formulation development opportunities [252]. Both of these AI formulation development systems use datasets obtained from published sources (Web of Science, etc.), experimental results, in silico molecular descriptors, physicochemical properties, evaluation parameters, neural networks, regression models, and evaluation of models for their iterative learning cycles and subsequent prediction.

Quartic.ai and Bright Path Laboratories have signed an agreement for collaborative development of an AI system for manufacturing of small molecule drugs and APIs [253]. Bright Path Lab's spinning tube-in-tube technology for scaling up of API manufacturing from bench scale to commercial scale has received support from Quartic AI expertise for DoE for process development, validation, and transfer of technology.

FabRX has recently developed M3DISEEN, a novel AI platform based on ML methods, for optimization of the process parameters of fused deposition modeling (FDM) 3D printing method [254, 255]. With this platform, accuracies of 76% for the printability and 67% for filament characteristics have been achieved. This platform is publicly available at http://m3diseen.com and requires input of material name, proportion of material (%), and selection of hot melt extruder and 3D printer.

11.7.3 Healthcare

Google's DeepMind, an AI system, has performed better than human experts in predicting breast cancer prediction from mammograms at very early stages of diseases [256]. Researchers at University College London Hospital in collaboration with Google's DeepMind have developed a 3D U-Net architecture that is at par with experts in differentiating between healthy and cancerous tissue, a segmentation process in head and neck cancer patients [257]. This development speeds up the segmentation process ensuring no healthy structures are damaged and increases the accuracy of radiotherapy planning. Royal Free London NHS Foundation Trust and Google's DeepMind AI collaboration has led to the creation of a digitally enabled care pathway to detect and treat acute kidney disease in hospitalized patients with advantages of reducing chances of cardiac arrest and healthcare costs [258].

IBM Watson is a supercomputer which uses DeepQA architecture for analyzing natural language questions to find relevant answers and their justifications from knowledge gained from combination of databases and existing natural language texts. IBM Watson, has the ability to review the genetic and oncology data of patients to diagnose cancer [259]. This AI system has diagnosed leukemia in a woman, who was unsuccessfully treated for acute myeloid leukemia, by testing oncological data from over 20 million cases [260]. Apart from Watson for Oncology (Memorial Sloan Kettering Cancer Center), some other successes of IBM Watson collaborations are clinical trial matching (Mayo Clinic), Sugar.IQ app (Medtronic), and Watson for Genomics (University of North Carolina, Quest Diagnostics, and others) [261]. Recently, IBM Watson and EBSCO Information Services have collaborated for developing a combined clinical decision support suite "DynaMed +Micromedex with Watson at www.dynamedex.com" for bringing evidence-based drug and disease information at a single source which can help informed clinical decisions [262]. IBM Micromedex with Watson uses natural language processing (NLP), a more conversational approach to bypass keyword searches, to searching peer-reviewed drug content from EBSCO's DynaMed.

P1vital® Predicting Response to Depression Treatment (PReDicT) Test is a potential medical device which can diagnose and improve the management of depression in clinical practice by using predictive analysis [263]. PReDicT clinical investigation are ongoing in an FDA-approved clinical trial [264].

11.8 Some Challenges

ML and DL have shown a significant success in the field of pharmaceutics, drug delivery, and drug discovery. However, ML and DL studies are lacking in implementation of state-of-the-art algorithms, so researchers are still limited to regression modeling [41]. The ML tools are easy to use and are widely applied in health researches leading to significant number of products in the market. On the other hand, DL tools are gaining their grounds with the development of tools like PyTorch and TensorFlow and their acceptability by drug regulatory/healthcare agencies

[43]. However, DL models require high-level programming skill and great knowledge about algorithms. Sometimes, DL-based algorithms lack sufficient data for reliable model development. Depending on a limited dataset may not adequately train algorithms leading to inaccurate predictions. The only solution to this problem is the use of validated simulated tools which can explore wide datasets. More practical implementation of AI tool demands an easy-to-use interface. To create an easy-to-use interface, researchers need more deep knowledge on programming which is a present challenge.

Hypes are more than validations of clinical outcomes and true realizations of AI-based devices. Most of the AI-based tools or devices for clinical decision support are still in research. Regular reporting, monitoring, and audits are required for their quality safety, efficacy, and transparency [265]. Some other challenges for implementing AI in healthcare are model transparency, integration of data from different modalities, model security, model bias, and federated learning [266]. Apart from these challenges, ethical issues like informed consent, data privacy, liability and algorithmic fairness, and transparency could not be ignored [267]. All these challenges have a great influence on the market reach of such AI systems. More improved algorithms and databases are required in the field of health and pharmaceutical sciences [268]. We can find only a limited number of AI-based tools that are approved by health regulators.

11.9 Conclusion

Recent technological advancements in domains like algorithms, expert systems, human-machine interface, computing speed, big data, cloud technologies, and robotics are driving the growth of AI. AI systems tend to become an embedded element of CADD systems for faster drug discovery and development. Expert systems, both ML based and ANN based, are in practice in pharmaceutical industries for continuous manufacturing of pharmaceutical products. Predicting either compositions of formulations or quality of formulations is reducing the need for actual wet-lab experiments for screening of excipients and optimization of formulations or processes. In healthcare, AI is assisting physicians for personalized treatment of diseases like cancer, diabetes, and cardiovascular diseases. AI-based systems are capable of analyzing genomes for predicting future diseases. Availability of large chemical databases, clinical research databases, and omics databases are increasing the potential role of data hungry DL approaches in drug discovery and development. All these developments in AI are revolutionizing pharmaceutical and pharmacogenomics researches that will have a wide impact on life sciences and health science.

References

1. Mak KK, Pichika MR (2019) Artificial intelligence in drug development: present status and future prospects. Drug Discov Today 24:773–780. https://doi.org/10.1016/j.drudis.2018.11.014
2. Campbell M, Hoane AJ, Hsu FH (2002) Deep blue. Artif Intell 134:57–83. https://doi.org/10.1016/S0004-3702(01)00129-1
3. On 'Jeopardy!' Watson Win Is All but Trivial—The New York Times. https://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html. Accessed 24 May 2021
4. Kettering MS (2015) Artificial intelligence. Big Data, and Cancer 1–2. https://doi.org/10.1001/jamaoncol.2015.1203.Conflict
5. McCarthy J, Minsky ML, Rochester N, Shannon CE (2006) A proposal for the Dartmouth summer research project on artificial intelligence. AI Mag 27:12–14. https://doi.org/10.1609/aimag.v27i4.1904
6. Artificial Intelligence (AI) Coined at Dartmouth | Celebrate Our 250th. https://250.dartmouth.edu/highlights/artificial-intelligence-ai-coined-dartmouth. Accessed 6 May 2021
7. Bobrow DG (1964) Natural language input for a computer problem solving system
8. Joseph W (1996) ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 9:36–45. https://doi.org/10.1145/357980.357991
9. Story of ELIZA, the first chatbot developed in 1966. https://analyticsindiamag.com/story-eliza-first-chatbot-developed-1966/. Accessed 6 May 2021
10. Krivohlavy J (1961) [Heuristic value of factor analysis]. Act Nerv Super (Praha) 3:318–325
11. Leucotia T (1961) Heuristic glimpses into the evolution of radiation therapy; particularly in reference to gynecologic cancers. Am J Roentgenol Radium Ther Nucl Med 85:3–20
12. Hansch C, Fujita T (1964) p-σ-π Analysis. A method for the correlation of biological activity and chemical structure. J Am Chem Soc 86:1616–1626. https://doi.org/10.1021/ja01062a035
13. Free SM, Wilson JW (1964) A mathematical contribution to structure-activity studies. J Med Chem 7:395–399. https://doi.org/10.1021/jm00334a001
14. Adamson GW, Bawden D (1975) A method of structure-activity correlation using wiswesser line notation. J Chem Inf Comput Sci 15:215–220. https://doi.org/10.1021/ci60004a006
15. Baskin I (2008) Chapter 1. Fragment descriptors in SAR/QSAR/QSPR studies, molecular similarity analysis and in virtual screening. In: Varnek A (ed) Chemoinformatics approaches to virtual screen. Royal Society of Chemistry, Cambridge, pp 1–43. https://doi.org/10.1039/9781847558879-00001
16. Sandell HS, Bourne JR (1985) Expert systems in medicine: a biomedical engineering perspective. Crit Rev Biomed Eng 12:95–129
17. Duda R, Shortliffe E (1983) Expert systems research. Science 220:261–268. https://doi.org/10.1126/science.6340198
18. Klopman G, Contreras R (1985) Use of artificial intelligence in structure-activity correlations of anticonvulsant drugs. Mol Pharmacol 27:86–93
19. Klopman G, Macina OT, Levinson ME, Rosenkranz HS (1987) Computer automated structure evaluation of quinolone antibacterial agents. Antimicrob Agents Chemother 31:1831–1840. https://doi.org/10.1128/AAC.31.11.1831
20. Rosenkranz HS, Klopman G (1988) Case, the computer-automated structure evaluation system, as an alternative to extensive animal testing. Toxicol Ind Health 4:533–540. https://doi.org/10.1177/074823378800400411
21. Klopman G (1985) Predicting toxicity through a computer automated structure evaluation program. Environ Health Perspect 61:269–274. https://doi.org/10.1289/ehp.8561269
22. Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967. https://doi.org/10.1021/ja00226a005

23. Marshall GR, Cramer RD (1988) Three-dimensional structure-activity relationships. Trends Pharmacol Sci 9:285–289. https://doi.org/10.1016/0165-6147(88)90012-0
24. Yang X, Wang Y, Byrne R, Schneider G, Yang S (2019) Concepts of artificial intelligence for computer-assisted drug discovery. Chem Rev 119:10520–10594. https://doi.org/10.1021/acs.chemrev.8b00728
25. Vedani A, Zumstein M, Lill MA, Ernst B (2007) Simulating α/β selectivity at the human thyroid hormone receptor: consensus scoring using multidimensional QSAR. ChemMedChem 2:78–87. https://doi.org/10.1002/cmdc.200600212
26. Lill MA (2007) Multi-dimensional QSAR in drug discovery. Drug Discov Today 12:1013–1017. https://doi.org/10.1016/j.drudis.2007.08.004
27. Brown N, Ertl P, Lewis R, Luksch T, Reker D, Schneider N (2020) Artificial intelligence in chemistry and drug design. J Comput Aided Mol Des 34:709–715. https://doi.org/10.1007/s10822-020-00317-x
28. Cox M, Ellsworth D (1997) Managing big data for scientific visualization. ACM Siggraph 97:1–17
29. The 5 V's of big data—Watson health perspectives. https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/. Accessed 6 May 2021
30. Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V (2015) Deep neural nets as a method for quantitative structure–activity relationships. J Chem Inf Model 55:263–274. https://doi.org/10.1021/ci500747n
31. Mayr A, Klambauer G, Unterthiner T, Hochreiter S (2016) DeepTox: toxicity prediction using deep learning. Front Environ Sci 3. https://doi.org/10.3389/fenvs.2015.00080
32. Goertzel B (2014) Artificial general intelligence: concept, state of the art, and future prospects. J Artif Gen Intell 5:1–48. https://doi.org/10.2478/jagi-2014-0001
33. Baum SD (2017) A survey of artificial general intelligence projects for ethics, risk, and policy. SSRN Electron J. https://doi.org/10.2139/ssrn.3070741
34. Fjelland R (2020) Why general artificial intelligence will not be realized. Humanit Soc Sci Commun 7:10. https://doi.org/10.1057/s41599-020-0494-4
35. Bostrom N (2020) Ethical issues in advanced artificial intelligence. In: Mach. Ethics Robot Ethics, Routledge, pp 69–75. https://doi.org/10.4324/9781003074991-7
36. Bishop CM (2013) Model-based machine learning. Philos Trans R Soc A Math Phys Eng Sci 371:20120222. https://doi.org/10.1098/rsta.2012.0222
37. Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB et al (2017) Deep learning in medical imaging: general overview. Korean J Radiol 18:570–584. https://doi.org/10.3348/kjr.2017.18.4.570
38. Vopham T, Hart JE, Laden F, Chiang YY (2018) Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. Environ Heal A Glob Access Sci Source 17:1–6. https://doi.org/10.1186/s12940-018-0386-x
39. Howard J (2013) The business impact of deep learning 1135. https://doi.org/10.1145/2487575.2491127
40. Russell S, Dietterich T, Horvitz E, Selman B, Rossi F, Hassabis D et al (2015) Letter to the Editor: research priorities for robust and beneficial artificial intelligence: an open letter. AI Mag 36:3. https://doi.org/10.1609/aimag.v36i4.2621
41. Hassanzadeh P, Atyabi F, Dinarvand R (2019) The significance of artificial intelligence in drug delivery system design. https://doi.org/10.1016/j.addr.2019.05.001
42. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H (2018) eDoctor: machine learning and the future of medicine. J Intern Med 284:603–619. https://doi.org/10.1111/joim.12822
43. Jiao Z, Hu P, Xu H, Wang Q (2020) Machine learning and deep learning in chemical health and safety: a systematic review of techniques and applications. https://doi.org/10.1021/acs.chas.0c00075
44. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K et al (2019) A guide to deep learning in healthcare. Nat Med 25:24–29. https://doi.org/10.1038/s41591-018-0316-z

45. Bakator M, Radosav D (2018) Deep learning and medical diagnosis: a review of literature. Multimodal Technol Interact 2:47. https://doi.org/10.3390/mti2030047
46. Chaki J, Dey N, Shi F, Sherratt RS (2019) Pattern mining approaches used in sensor-based biometric recognition: a review. IEEE Sens J 19:3569–3580. https://doi.org/10.1109/JSEN.2019.2894972
47. Schrider DR, Kern AD (2017) Supervised machine learning for population genetics: a new paradigm. Trends Genet xx:1–12. https://doi.org/10.1016/j.tig.2017.12.005
48. Fabris F, de Magalhães JP, Freitas AA (2017) A review of supervised machine learning applied to ageing research. Biogerontology 18:171–188. https://doi.org/10.1007/s10522-017-9683-y
49. Deo RC (2015) Machine learning in medicine. Circulation 132:1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593
50. Díaz-Uriarte R, Alvarez de Andrés S (2006) Gene selection and classification of microarray data using random forest. BMC Bioinformatics 7:1–13. https://doi.org/10.1186/1471-2105-7-3
51. Cunningham P, Delany SJ (2020) k-nearest neighbour classifiers, 2nd edn (with Python examples). https://doi.org/2004.04523
52. Yang X, Wang Y, Byrne R, Schneider G, Yang S (2019) Concepts of artificial intelligence for computer-assisted drug discovery. https://doi.org/10.1021/acs.chemrev.8b00728
53. Nidheesh N, Abdul Nazeer KA, Ameer PM (2017) An enhanced deterministic K-Means clustering algorithm for cancer subtype prediction from gene expression data. Comput Biol Med 91:213–221. https://doi.org/10.1016/j.compbiomed.2017.10.014
54. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297. https://doi.org/10.1007/BF00994018
55. Farhat NH (1992) Photonit neural networks and learning mathines the role of electron-trapping materials. IEEE Expert Syst Their Appl 7:63–72. https://doi.org/10.1109/64.163674
56. Wang MFZ, Fernandez-Gonzalez R (2017) (Machine-)Learning to analyze in vivo microscopy: support vector machines. Biochim Biophys Acta Proteins Proteomics 1865:1719–1727. https://doi.org/10.1016/j.bbapap.2017.09.013
57. Girard CI, Warren CE, Romanchuk NJ, Del Bel MJ, Carsen S, Chan ADC et al (2020) Decision tree learning algorithm for classifying knee injury status using return-to-activity criteria. In: Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS, 2020-July, pp 5494–5497. https://doi.org/10.1109/EMBC44109.2020.9176010
58. Wei L, Ventura S, Lowery M, Ryan MA, Mathieson S, Boylan GB et al (2020) Random forest-based algorithm for sleep spindle detection in infant EEG. In: Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS, 2020-July, pp 58–61. https://doi.org/10.1109/EMBC44109.2020.9176339
59. Berrar D (2018) Bayes' theorem and naive bayes classifier. Encycl Bioinforma Comput Biol ABC Bioinforma 1–3:403–412. https://doi.org/10.1016/B978-0-12-809633-8.20473-1
60. Lewis DD (1998) Naive (Bayes) at forty: the independence assumption in information retrieval. pp 4–15. https://doi.org/10.1007/BFb0026666
61. Xu S (2018) Bayesian Naïve Bayes classifiers to text classification. J Inf Sci 44:48–59. https://doi.org/10.1177/0165551516677946
62. DeGregory KW, Kuiper P, DeSilvio T, Pleuss JD, Miller R, Roginski JW et al (2018) A review of machine learning in obesity. Obes Rev 19:668–685. https://doi.org/10.1111/obr.12667
63. Field A (2012) Logistic regression. Discov Stat Using SPSS 404:731–735
64. Akilli A, Gorgulu O (2020) Comparative assessments of multivariate nonlinear fuzzy regression techniques for egg production curve. Trop Anim Health Prod 52:2119–2127. https://doi.org/10.1007/s11250-020-02226-5
65. Alhamzawi R, Ali HTM (2018) The Bayesian adaptive lasso regression. Math Biosci 303:75–82. https://doi.org/10.1016/j.mbs.2018.06.004

66. Roohi A, Faust K, Djuric U, Diamandis P (2020) Unsupervised machine learning in pathology. Surg Pathol 13:349–358. https://doi.org/10.1016/j.path.2020.01.002
67. Jin X, Han J (2017) K-medoids clustering. Encycl Mach Learn Data Min:697–700. https://doi.org/10.1007/978-1-4899-7687-1_432
68. Granato D, Santos JS, Escher GB, Ferreira BL, Maggio RM (2017) AC SC. Trends Food Sci Technol. https://doi.org/10.1016/j.tifs.2017.12.006
69. Livesay DR (ed) (2014) Protein dynamics, vol 1084. Humana Press, Totowa. https://doi.org/10.1007/978-1-62703-658-0
70. Lee J-M, Yoo C, Lee I-B (2004) Statistical process monitoring with independent component analysis. J Process Control 14:467–485. https://doi.org/10.1016/j.jprocont.2003.09.004
71. Vieira S, Hugo W, Pinaya L, Garcia-dias R, Mechelli A, Andre S (2020) Multimodal integration. 1:283–305. https://doi.org/10.1016/B978-0-12-815739-8.00016-X
72. Jhang KM, Chang MC, Lo TY, Lin CW, Wang WF, Wu HH (2019) Using the apriori algorithm to classify the care needs of patients with different types of dementia. Patient Prefer Adherence 13:1899–1912. https://doi.org/10.2147/PPA.S223816
73. Chang R, Liu Z (2011) An improved apriori algorithm. In: Proc. 2011 Int. Conf. Electron. Optoelectron., IEEE, pp V1-476–V1-478. https://doi.org/10.1109/ICEOE.2011.6013148
74. Han J, Pei J, Yin Y, Mao R (2004) Mining frequent patterns without candidate generation: a frequent-pattern tree approach. Data Min Knowl Discov 8:53–87. https://doi.org/10.1023/B:DAMI.0000005258.31418.83
75. Kohonen self-organizing feature map—Javatpoint. https://www.javatpoint.com/artificial-neural-network-kohonen-self-rganizing-feature-map. Accessed 24 May 2021
76. Galutira EF, Fajardo AC, Medina RP (2019) A novel Kohonen self-organizing maps using exponential decay average rate of change for color clustering. pp 23–33. https://doi.org/10.1007/978-981-13-6031-2_28
77. Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69. https://doi.org/10.1007/BF00337288
78. Kohonen T (2013) Essentials of the self-organizing map. Neural Networks 37:52–65. https://doi.org/10.1016/j.neunet.2012.09.018
79. Sarstedt M, Mooi E (2014) Cluster analysis. pp 273–324. https://doi.org/10.1007/978-3-642-53965-7_9
80. Halkidi M (2009) Hierarchial clustering. Encycl. Database Syst. Springer US, Boston, pp 1291–1294. https://doi.org/10.1007/978-0-387-39940-9_604
81. Camargo G, Bugatti PH, Saito PTM (2020) Active semi-supervised learning for biological data classification. PLoS One 15:1–20. https://doi.org/10.1371/journal.pone.0237428
82. van Engelen JE, Hoos HH (2020) A survey on semi-supervised learning. Mach Learn 109:373–440. https://doi.org/10.1007/s10994-019-05855-6
83. Inductive vs. Transductive Learning | by Vijini Mallawaarachchi | Towards Data Science. https://towardsdatascience.com/inductive-vs-transductive-learning-e608e786f7d. Accessed 24 May 2021
84. Damiati SA (2020) Digital pharmaceutical sciences. AAPS PharmSciTech 21. https://doi.org/10.1208/s12249-020-01747-4
85. Achanta AS, Kowalski JG, Rhodes CT (1995) Artificial neural networks: implications for pharmaceutical sciences. Drug Dev Ind Pharm 21:119–155. https://doi.org/10.3109/03639049509048099
86. Beresford R (2000) Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. 22:717–727
87. Das MK, Chakraborty T (2016) ANN in pharmaceutical product and process development. Elsevier Inc. https://doi.org/10.1016/B978-0-12-801559-9.00014-4
88. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T (2018) The rise of deep learning in drug discovery. Drug Discov Today 23:1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039

89. Yang Y, Ye Z, Su Y, Zhao Q, Li X, Ouyang D (2019) Deep learning for in vitro prediction of pharmaceutical formulations. Acta Pharm Sin B 9:177–185. https://doi.org/10.1016/j.apsb.2018.09.010
90. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP et al (2018) Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface 15: 20170387. https://doi.org/10.1098/rsif.2017.0387
91. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G et al (2019) Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 18:463–477. https://doi.org/10.1038/s41573-019-0024-5
92. Ranjbar S, Nejad FM, Zakeri H, Gandomi AH (2020) Computational intelligence for modeling of asphalt pavement surface distress. INC. https://doi.org/10.1016/b978-0-12-818961-0.00003-x
93. Ranjbar S, Nejad FM, Zakeri H, Gandomi AH (2020) Computational intelligence for modeling of asphalt pavement surface distress. In: New Mater. Civ. Eng., Elsevier, pp 79–116. https://doi.org/10.1016/B978-0-12-818961-0.00003-X
94. Currie G, Hawk KE, Rohren E, Vial A, Klein R (2019) Machine learning and deep learning in medical imaging: intelligent imaging. J Med Imaging Radiat Sci 50:477–487. https://doi.org/10.1016/j.jmir.2019.09.005
95. Janowczyk A, Madabhushi A (2016) J Pathol Inform Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. https://doi.org/10.4103/2153-3539.186902.
96. Yu N, Yu Z, Pan Y (2017) A deep learning method for lincRNA detection using auto-encoder algorithm. BMC Bioinformatics 18:511. https://doi.org/10.1186/s12859-017-1922-3
97. Aggarwal A, Mittal M, Battineni G (2021) Generative adversarial network: an overview of theory and applications. Int J Inf Manag Data Insights 1:100004. https://doi.org/10.1016/j.jjimei.2020.100004
98. Reimers C, Requena-Mesa C (2020) Deep learning—an opportunity and a challenge for geo- and astrophysics. Elsevier Inc. https://doi.org/10.1016/b978-0-12-819154-5.00024-2
99. Chan HCS, Shan H, Dahoun T, Vogel H, Yuan S (2019) Advancing drug discovery via artificial intelligence. Trends Pharmacol Sci 40:592–604. https://doi.org/10.1016/j.tips.2019.06.004
100. Adams CP, Van Brantner V (2006) Market watch: estimating the cost of new drug development: is it really $802 million? Health Aff 25:420–428. https://doi.org/10.1377/hlthaff.25.2.420
101. Phrma.org (2015) Biopharmaceutical research & development: the process behind new medicines. Pharm Res Manuf Am 24
102. Reiss R (2020) Transforming drug discovery through Artificial Intelligence 6–10
103. Stork C, Chen Y, Šícho M, Kirchmair J (2019) Hit Dexter 2.0: machine-learning models for the prediction of frequent hitters. J Chem Inf Model 59:1030–1043. https://doi.org/10.1021/acs.jcim.8b00677
104. Stork C, Wagner J, Friedrich N-O, de Bruyn KC, Šícho M, Kirchmair J (2018) Hit Dexter: a machine-learning model for the prediction of frequent hitters. ChemMedChem 13:564–571. https://doi.org/10.1002/cmdc.201700673
105. Liu X, Vogt I, Haque T, Campillos M (2013) HitPick: a web server for hit identification and target prediction of chemical screenings. Bioinformatics 29:1910–1912. https://doi.org/10.1093/bioinformatics/btt303
106. Awale M, Reymond J-L (2017) The polypharmacology browser: a web-based multi-fingerprint target prediction tool using ChEMBL bioactivity data. J Cheminform 9:11. https://doi.org/10.1186/s13321-017-0199-x
107. Steiner S, Wolf J, Glatzel S, Andreou A, Granda JM, Keenan G et al (2019) Organic synthesis in a modular robotic system driven by a chemical programming language. Science 363: eaav2211. https://doi.org/10.1126/science.aav2211

108. Feinberg EN, Sur D, Wu Z, Husic BE, Mai H, Li Y et al (2018) PotentialNet for molecular property prediction. ACS Cent Sci 4:1520–1530. https://doi.org/10.1021/acscentsci.8b00507
109. Blaschke T, Arús-Pous J, Chen H, Margreitter C, Tyrchan C, Engkvist O et al (2020) REINVENT 2.0: an AI tool for de novo drug design. J Chem Inf Model 60:5918–5922. https://doi.org/10.1021/acs.jcim.0c00915
110. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T et al (2020) Improved protein structure prediction using potentials from deep learning. Nature 577:706–710. https://doi.org/10.1038/s41586-019-1923-7
111. Mayr A, Klambauer G, Unterthiner T, Hochreiter S (2016) DeepTox: toxicity prediction using deep learning. 3. https://doi.org/10.3389/fenvs.2015.00080
112. Awale M, Reymond JL (2019) Polypharmacology Browser PPB2: target prediction combining nearest neighbors with machine learning. J Chem Inf Model 59:10–17. https://doi.org/10.1021/acs.jcim.8b00524
113. GitHub—aspuru-guzik-group/ORGANIC: code repo for optimizing distributions of molecules. https://github.com/aspuru-guzik-group/ORGANIC. Accessed 5 May 2021
114. Sanchez-Lengeling B, Outeiral C, Guimaraes GL, Aspuru-Guzik A (2017) Optimizing distributions over molecular space. An objective-reinforced generative adversarial network for inverse-design chemistry (ORGANIC). ChemRxiv 1–18. https://doi.org/10.26434/chemrxiv.5309668.v3
115. Ryu JY, Kim HU, Lee SY (2018) Deep learning improves prediction of drug–drug and drug–food interactions. Proc Natl Acad Sci U S A 115:E4304–E4311. https://doi.org/10.1073/pnas.1803294115
116. Saiakhov R, Chakravarti S, Klopman G (2013) Effectiveness of CASE ultra expert system in evaluating adverse effects of drugs. Mol Inform 32:87–97. https://doi.org/10.1002/minf.201200081
117. Dong J, Wang N-N, Yao Z-J, Zhang L, Cheng Y, Ouyang D et al (2018) ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. J Cheminform 10:29. https://doi.org/10.1186/s13321-018-0283-x
118. Maziarka Ł, Pocha A, Kaczmarczyk J, Rataj K, Danel T, Warchoł M (2020) Mol-CycleGAN: a generative model for molecular optimization. J Cheminform 12:2. https://doi.org/10.1186/s13321-019-0404-1
119. Gobburu JVS, Chen EP (1996) Artificial neural networks as a novel approach to integrated pharmacokinetic-pharmacodynamic analysis. J Pharm Sci 85:505–510. https://doi.org/10.1021/js950433d
120. Değim T, Hadgraft J, Ilbasmiş S, Özkan Y (2003) Prediction of skin penetration using artificial neural network (ANN) modeling. J Pharm Sci 92:656–664. https://doi.org/10.1002/jps.10312
121. Gobburu JVS, Shelver WH (1995) Quantitative Structure–Pharmacokinetic Relationships (QSPR) of beta blockers derived using neural networks. J Pharm Sci 84:862–865. https://doi.org/10.1002/jps.2600840715
122. Taskinen J, Yliruusi J (2003) Prediction of physicochemical properties based on neural network modelling. Adv Drug Deliv Rev 55:1163–1183. https://doi.org/10.1016/S0169-409X(03)00117-0
123. Sakiyama Y (2009) The use of machine learning and nonlinear statistical tools for ADME prediction. Expert Opin Drug Metab Toxicol 5:149–169. https://doi.org/10.1517/17425250902753261
124. Kalinin AA, Higgins GA, Reamaroon N, Soroushmehr S, Allyn-Feuer A, Dinov ID et al (2018) Deep learning in pharmacogenomics: from gene regulation to patient stratification. Pharmacogenomics 19:629–650. https://doi.org/10.2217/pgs-2018-0008
125. Wan F, Zeng J (2016) Deep learning with feature embedding for compound-protein interaction prediction. BioRxiv 086033. https://doi.org/10.1101/086033
126. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK (2021) Artificial intelligence in drug discovery and development. Drug Discov Today 26:80–93. https://doi.org/10.1016/j.drudis.2020.10.010

127. Tian K, Shao M, Wang Y, Guan J, Zhou S (2016) Boosting compound-protein interaction prediction by deep learning. Methods. https://doi.org/10.1016/j.ymeth.2016.06.024
128. Liu B, He H, Luo H, Zhang T, Jiang J (2019) Artificial intelligence and big data facilitated targeted drug discovery. https://doi.org/10.1136/svn-2019-000290
129. Galbusera F, Niemeyer F, Seyfried M, Bassani T, Casaroli G, Kienle A et al (2018) Exploring the potential of generative adversarial networks for synthesizing radiological images of the spine to be used in in silico trials generative adversarial networks for. Front Bioeng Biotechnol 6:1–9. https://doi.org/10.3389/fbioe.2018.00053
130. Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. https://doi.org/10.1021/acs.molpharmaceut.6b00248
131. Chen C, Wu M, Cen S, Wu J, Zhou J (2017) MTLD, a database of multiple target ligands, the updated version. Molecules 22:1–8. https://doi.org/10.3390/molecules22091375
132. Awale M, Reymond JL (2017) The polypharmacology browser: a web-based multi-fingerprint target prediction tool using ChEMBL bioactivity data. J Cheminform 9:1–10. https://doi.org/10.1186/s13321-017-0199-x
133. Nayak VS, Khan MS, Shukla BK, Chaturvedi PR (2016) Artificial intelligence in clinical research. Int J Clin Trials 3:187. https://doi.org/10.18203/2349-3259.ijct20163955
134. Lin WC, Chen JS, Chiang MF, Hribar MR (2020) Applications of artificial intelligence to electronic health record data in ophthalmology. Transl Vis Sci Technol 9:13. https://doi.org/10.1167/tvst.9.2.13
135. Kostic EJ, Pavlović DA, Živković MD (2019) Applications of artificial intelligence in medicine and pharmacy—ethical aspects. Acta Medica Median:128–137. https://doi.org/10.5633/amm.2019.0319
136. Harrer S, Shah P, Antony B, Hu J (2019) Artificial intelligence for clinical trial design. Trends Pharmacol Sci 40:577–591. https://doi.org/10.1016/j.tips.2019.05.005
137. Kumar R, Sharma A, Siddiqui MH, Tiwari RK (2017) Prediction of human intestinal absorption of compounds using artificial intelligence techniques. Curr Drug Discov Technol 14:244–254. https://doi.org/10.2174/1570163814666170404160911
138. Houy N, Le Grand F (2018) Optimal dynamic regimens with artificial intelligence: the case of temozolomide. PLoS One 13:1–15. https://doi.org/10.1371/journal.pone.0199076
139. Kang SH, Poynton MR, Kim KM, Lee H, Kim DH, Lee SH et al (2007) Population pharmacokinetic and pharmacodynamic models of remifentanil in healthy volunteers using artificial neural network analysis. Br J Clin Pharmacol 64:3–13. https://doi.org/10.1111/j.1365-2125.2007.02845.x
140. Poynton MR, Choi BM, Kim YM, Park IS, Noh GJ, Hong SO et al (2009) Machine learning methods applied to pharmacokinetic modelling of remifentanil in healthy volunteers: a multi-method comparison. J Int Med Res 37:1680–1691. https://doi.org/10.1177/147323000903700603
141. Hussain AS, Shivanand P, Johnson RD (1994) Application of neural computing in pharmaceutical product development: computer aided. Drug Dev Ind Pharm 20:1739–1752
142. Moussa T, Arabia S, Elkatatny S, Arabia S, Mahmoud M, Arabia S et al (2018) Development of new permeability formulation from well log data using artificial intelligence approaches. J Energy Resour Technol. https://doi.org/10.1115/1.4039270
143. Chibani S, Coudert F-X (2020) Machine learning approaches for the prediction of materials properties. APL Mater 8:080701. https://doi.org/10.1063/5.0018384
144. Neural networks tutorial: model selection | Neural Designer. https://www.neuraldesigner.com/learning/tutorials/model-selection. Accessed 21 May 2021
145. Salah Alaloul W, Hannan QA (2020) Data processing using artificial neural networks. In: Dyn. Data Assim.—Beating Uncertainties. IntechOpen. https://doi.org/10.5772/intechopen.91935
146. Onuki Y, Kosugi A, Hamaguchi M, Marumo Y, Kumada S, Hirai D et al (2018) A comparative study of disintegration actions of various disintegrants using Kohonen's self-organizing maps. J Drug Deliv Sci Technol 43:141–148. https://doi.org/10.1016/j.jddst.2017.10.002

147. Ebube NK, Owusu-Ababio G, Adeyeye CM (2000) Preformulation studies and characterization of the physicochemical properties of amorphous polymers using artificial neural networks. Int J Pharm 196:27–35. https://doi.org/10.1016/S0378-5173(99)00405-6
148. Hayashi Y, Oishi T, Shirotori K, Marumo Y, Kosugi A, Kumada S et al (2018) Modeling of quantitative relationships between physicochemical properties of active pharmaceutical ingredients and tensile strength of tablets using a boosted tree. Drug Dev Ind Pharm 44: 1090–1098. https://doi.org/10.1080/03639045.2018.1434195
149. Jouyban A, Majidi M-R, Jalilzadeh H, Asadpour-Zeynali K (2004) Modeling drug solubility in water–cosolvent mixtures using an artificial neural network. Farmaco 59:505–512. https://doi.org/10.1016/j.farmac.2004.02.005
150. Damiati SA, Martini LG, Smith NW, Lawrence MJ, Barlow DJ (2017) Application of machine learning in prediction of hydrotrope-enhanced solubilisation of indomethacin. Int J Pharm 530: 99–106. https://doi.org/10.1016/j.ijpharm.2017.07.048
151. Chen L, Lian G, Han L (2007) Prediction of human skin permeability using artificial neural network (ANN) modeling. Acta Pharmacol Sin 28:591–600. https://doi.org/10.1111/j.1745-7254.2007.00528.x
152. Chen G, Shen Z, Li Y (2020) A machine-learning-assisted study of the permeability of small drug-like molecules across lipid membranes. Phys Chem Chem Phys 22:19687–19696. https://doi.org/10.1039/D0CP03243C
153. Doniger S, Hofmann T, Yeh J (2002) Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms. J Comput Biol 9: 849–864. https://doi.org/10.1089/10665270260518317
154. Welsh WJ, Lin W, Tersigni SH, Collantes E, Duta R, Carey MS et al (1996) Pharmaceutical fingerprinting: evaluation of neural networks and chemometric techniques for distinguishing among same-product manufacturers. Anal Chem 68:3473–3482. https://doi.org/10.1021/ac951164e
155. Lou H, Chung JI, Kiang Y-H, Xiao L-Y, Hageman MJ (2019) The application of machine learning algorithms in understanding the effect of core/shell technique on improving powder compactability. Int J Pharm 555:368–379. https://doi.org/10.1016/j.ijpharm.2018.11.039
156. Ilbasmis-Tamer S, Tugcu-Demiroz F, Degim IT (2017) Carbon nanotube membranes to predict skin permeability of compounds. Pharm Dev Technol 22:606–616. https://doi.org/10.1080/10837450.2016.1221430
157. Agatonovic-Kustrin S, Glass BD, Mangan M, Smithson J (2008) Analysing the crystal purity of mebendazole raw material and its stability in a suspension formulation. Int J Pharm 361: 245–250. https://doi.org/10.1016/j.ijpharm.2008.04.039
158. Agatonovic-Kustrin S, Rades T, Wu V, Saville D, Tucker I (2001) Determination of polymorphic forms of ranitidine–HCl by DRIFTS and XRPD. J Pharm Biomed Anal 25:741–750. https://doi.org/10.1016/S0731-7085(01)00375-2
159. Agatonovic-Kustrin S, Wu V, Rades T, Saville D, Tucker I (2000) Ranitidine hydrochloride X-ray assay using a neural network. J Pharm Biomed Anal 22:985–992. https://doi.org/10.1016/S0731-7085(00)00256-9
160. Agatonovic-Kustrin S, Tucker I, Schmierer D (1999) Solid state assay of ranitidine HCl as a buk drug as active ingredient in tablets using DRIFT spectroscopy with artificial neural networks. Pharm Res 16:1477–1482. https://doi.org/10.1023/a:1018975730945
161. Grosan C, Abraham A (2011) Rule-based expert systems. pp 149–185. https://doi.org/10.1007/978-3-642-21004-4_7
162. Rowe RC, Colbourn EA (2006) Computers in pharmaceutical formulation. In: Comput. Appl. Pharm. Res. Dev. Wiley, Hoboken, pp 677–701. https://doi.org/10.1002/0470037237.ch28
163. Deb PK, Al-Attraqchi O, Al-Qattan MN, Raghu Prasad M, Tekade RK (2018) Applications of computers in pharmaceutical product formulation. In: Dos. Form Des. Parameters. Elsevier, pp 665–703. https://doi.org/10.1016/B978-0-12-814421-3.00019-1

164. Tolun MR, Sahin S, Oztoprak K (2016) Expert systems. In: Kirk-Othmer Encycl. Chem. Technol. Wiley, Hoboken, pp 1–12. https://doi.org/10.1002/0471238961.0524160518011305.a01.pub2
165. Kusiak A, Chen M (1988) Expert systems for planning and scheduling manufacturing systems. Eur J Oper Res 34:113–130. https://doi.org/10.1016/0377-2217(88)90346-3
166. Al-Ajlan A (2015) The comparison between forward and backward chaining. Int J Mach Learn Comput 5:106–113. https://doi.org/10.7763/IJMLC.2015.V5.492
167. Herawan Hayadi B, Bastian A, Rukun K, Jalinus N, Lizar Y, Guci A (2018) Expert system in the application of learning models with forward chaining method. Int J Eng Technol 7:845. https://doi.org/10.14419/ijet.v7i2.29.14269
168. Liebowitz J (1995) Expert systems: a short introduction. Eng Fract Mech 50:601–607. https://doi.org/10.1016/0013-7944(94)E0047-K
169. Ramani KV, Patel MR, Patel SK (1992) An expert system for drug preformulation in a pharmaceutical company. Interfaces (Providence) 22:101–108. https://doi.org/10.1287/inte.22.2.101
170. Wilson WI, Peng Y, Augsburger LL (2005) Generalization of a prototype intelligent hybrid system for hard gelatin capsule formulation development. AAPS PharmSciTech 6:E449–E457. https://doi.org/10.1208/pt060356
171. Kalra G, Peng Y, Guo M, Augsburger LL (2002) A hybrid intelligent system for formulation of BCS Class II drugs in hard gelatin capsules. In: ICONIP 2002—Proc 9th Int Conf Neural Inf Process Comput Intell E-Age, vol 4, pp 1987–1991. https://doi.org/10.1109/ICONIP.2002.1199021
172. Wilson WI, Peng Y, Augsburger LL (2005) Generalization of a prototype intelligent hybrid system for hard gelatin capsule formulation development. AAPS PharmSciTech 6:449–457. https://doi.org/10.1208/pt060356
173. Zhang Z, Dong H, Peng B, Liu H, Li C, Liang M et al (2011) Design of an expert system for the development and formulation of push–pull osmotic pump tablets containing poorly water-soluble drugs. Int J Pharm 410:41–47. https://doi.org/10.1016/j.ijpharm.2011.03.013
174. Suñé Negre JM, Roig Carreras M, Fuster García R, Hernández Pérez C, Ruhi Roura R, García Montoya E et al (2005) New galenic preformulation methodology for the characterization of substances in relation to their viability for compression: SeDeM Diagram | Science. technology pharm; 15 (3): 125-136, Jul-Sep. 2005. illus, tab | IBECS. Sci Technol Pharm 15:125–136
175. Aguilar-Díaz JE, García-Montoya E, Pérez-Lozano P, Suñe-Negre JM, Miñarro M, Ticó JR (2009) The use of the SeDeM Diagram expert system to determine the suitability of diluents-disintegrants for direct compression and their use in formulation of ODT. Eur J Pharm Biopharm 73:414–423. https://doi.org/10.1016/j.ejpb.2009.07.001
176. Aguilar-Díaz JE, García-Montoya E, Suñe-Negre JM, Pérez-Lozano P, Miñarro M, Ticó JR (2012) Predicting orally disintegrating tablets formulations of ibuprophen tablets: an application of the new SeDeM-ODT expert system. Eur J Pharm Biopharm 80:638–648. https://doi.org/10.1016/j.ejpb.2011.12.012
177. Galdón E, Casas M, Gayango M, Caraballo I (2016) First study of the evolution of the SeDeM Expert System parameters based on percolation theory: monitoring of their critical behavior. Eur J Pharm Biopharm. https://doi.org/10.1016/j.ejpb.2016.10.004
178. Campiñez MD, Benito E, Romero-azogil L, Aguilar-de-leyva Á, García-martín MDG, Galbis A et al (2017) Development and characterization of new functionalized polyurethanes for sustained and site-specific drug release in the gastrointestinal tract. Eur J Pharm Sci. https://doi.org/10.1016/j.ejps.2017.01.017
179. Zieschang L, Klein M, Jung N, Krämer J, Windbergs M (2019) Formulation development of medicated chewing gum tablets by direct compression using the SeDeM-Diagram-Expert-System. Eur J Pharm Biopharm 144:68–78. https://doi.org/10.1016/j.ejpb.2019.09.003
180. Sipos E, Oltean AR, Szabó Z-I, Rédai E-M, Nagy GD (2017) Application of SeDeM expert systems in preformulation studies of pediatric ibuprofen ODT tablets. Acta Pharm 67:237–246. https://doi.org/10.1515/acph-2017-0017

181. Hernández Pérez C, Carmona MM. New galenic preformulation methodology for the characterization of substances in relation to their viability for compression: SeDeM diagram—Dialnet. https://dialnet.unirioja.es/servlet/articulo?codigo=1310683. Accessed 21 May 2021
182. Dai S, Xu B, Shi G, Liu J, Zhang Z, Shi X et al (2019) SeDeM expert system for directly compressed tablet formulation: a review and new perspectives. Powder Technol 342:517–527. https://doi.org/10.1016/j.powtec.2018.10.027
183. Dai S, Xu B, Zhang Z, Yu J, Wang F, Shi X et al (2019) A compression behavior classification system of pharmaceutical powders for accelerating direct compression tablet formulation design. Int J Pharm 572:118742. https://doi.org/10.1016/j.ijpharm.2019.118742
184. Mendyk A, Jachowicz R (2005) Neural network as a decision support system in the development of pharmaceutical formulation—focus on solid dispersions. Expert Syst Appl 28:285–294. https://doi.org/10.1016/j.eswa.2004.10.007
185. Mendyk A, Jachowicz R, Błasiak A (2003) Neural expert system for pharmaceutical formulation development—focus on solid dispersions. Neural Networks Soft Comput 480–485. https://doi.org/10.1007/978-3-7908-1902-1_73
186. Trnka H, Wu JX, Van De Weert M, Grohganz H, Rantanen J (2013) Fuzzy logic-based expert system for evaluating cake quality of Freeze-dried formulations. J Pharm Sci 102:4364–4374. https://doi.org/10.1002/jps.23745
187. Han R, Yang Y, Li X, Ouyang D (2018) Predicting oral disintegrating tablet formulations by neural network techniques. Asian J Pharm Sci 13:336–342. https://doi.org/10.1016/j.ajps.2018.01.003
188. Lefnaoui S, Rebouh S, Bouhedda M, Yahoum MM (2020) Artificial neural network for modeling formulation and drug permeation of topical patches containing diclofenac sodium. Drug Deliv Transl Res 10:168–184. https://doi.org/10.1007/s13346-019-00671-w
189. Aksu B, Paradkar A, de Matas M, Özer Ö, Güneri T, York P (2012) Quality by design approach: application of artificial intelligence techniques of tablets manufactured by direct compression. AAPS PharmSciTech 13:1138–1146. https://doi.org/10.1208/s12249-012-9836-x
190. Khan AM, Hanif M, Bukhari NI, Shamim R, Rasool F, Rasul S et al (2020) Artificial Neural Network (ANN) approach to predict an optimized ph-dependent mesalamine matrix tablet. Drug Des Devel Ther 14:2435–2448. https://doi.org/10.2147/DDDT.S244016
191. Manda A, Walker R, Khamanga S (2019) An artificial neural network approach to predict the effects of formulation and process variables on prednisone release from a multipartite system. Pharmaceutics 11:109. https://doi.org/10.3390/pharmaceutics11030109
192. Madzarevic M, Vulovic S, Djuris F et al (2019) Optimization and prediction of ibuprofen release from 3D DLP printlets using artificial neural networks. Pharmaceutics 11:544. https://doi.org/10.3390/pharmaceutics11100544
193. Stanojević G, Medarević D, Adamov I, Pešić N, Kovačević J, Ibrić S (2020) Tailoring atomoxetine release rate from DLP 3D-printed tablets using artificial neural networks: influence of tablet thickness and drug loading. Molecules 26:111. https://doi.org/10.3390/molecules26010111
194. El Menshawe SF, Aboud HM, Elkomy MH, Kharshoum RM, Abdeltwab AM (2020) A novel nanogel loaded with chitosan decorated bilosomes for transdermal delivery of terbutaline sulfate: artificial neural network optimization, in vitro characterization and in vivo evaluation. Drug Deliv Transl Res 10:471–485. https://doi.org/10.1007/s13346-019-00688-1
195. Zhao F, Lu J, Jin X, Wang Z, Sun Y, Gao D et al (2018) Comparison of response surface methodology and artificial neural network to optimize novel ophthalmic flexible nanoliposomes: characterization, evaluation, in vivo pharmacokinetics and molecular dynamics simulation. Colloids Surfaces B Biointerfaces 172:288–297. https://doi.org/10.1016/j.colsurfb.2018.08.046
196. Melamane S, Walker RB, Khamanga SMM (2020) Formulation optimization of smart thermosetting lamotrigine loaded hydrogels using response surface methodology, box benhken

design and artificial neural networks. Drug Dev Ind Pharm 46:1402–1415. https://doi.org/10.1080/03639045.2020.1791163
197. Galata F, Könyves M, Szabó C et al (2019) Fast, spectroscopy-based prediction of in vitro dissolution profile of extended release tablets using artificial neural networks. Pharmaceutics 11:400. https://doi.org/10.3390/pharmaceutics11080400
198. Koletti AE, Tsarouchi E, Kapourani A, Kontogiannopoulos KN, Assimopoulou AN, Barmpalexis P (2020) Gelatin nanoparticles for NSAID systemic administration: quality by design and artificial neural networks implementation. Int J Pharm 578:119118. https://doi.org/10.1016/j.ijpharm.2020.119118
199. Gentiluomo L, Roessner D, Augustijn D, Svilenov H, Kulakova A, Mahapatra S et al (2019) Application of interpretable artificial neural networks to early monoclonal antibodies development. Eur J Pharm Biopharm 141:81–89. https://doi.org/10.1016/j.ejpb.2019.05.017
200. Nagy B, Petra D, Galata DL, Démuth B, Borbás E, Marosi G et al (2019) Application of artificial neural networks for Process Analytical Technology-based dissolution testing. Int J Pharm 567:118464. https://doi.org/10.1016/j.ijpharm.2019.118464
201. Barmpalexis P, Karagianni A, Karasavvaides G, Kachrimanis K (2018) Comparison of multilinear regression, particle swarm optimization artificial neural networks and genetic programming in the development of mini-tablets. Int J Pharm 551:166–176. https://doi.org/10.1016/j.ijpharm.2018.09.026
202. Vidovič S, Horvat M, Bizjak A, Planinšek O, Petek B, Burjak M et al (2019) Elucidating molecular properties of kappa-carrageenan as critical material attributes contributing to drug dissolution from pellets with a multivariate approach. Int J Pharm 566:662–673. https://doi.org/10.1016/j.ijpharm.2019.06.016
203. Roggo Y, Jelsch M, Heger P, Ensslin S, Krumme M (2020) Deep learning for continuous manufacturing of pharmaceutical solid dosage form. Eur J Pharm Biopharm 153:95–105. https://doi.org/10.1016/j.ejpb.2020.06.002
204. Gams M, Horvat M, Ožek M, Luštrek M, Gradišek A (2014) Integrating artificial and human intelligence into tablet production process. AAPS PharmSciTech 15:1447–1453. https://doi.org/10.1208/s12249-014-0174-z
205. Belič A, Škrjanc I, Božič DZ, Karba R, Vrečer F (2009) Minimisation of the capping tendency by tableting process optimisation with the application of artificial neural networks and fuzzy models. Eur J Pharm Biopharm 73:172–178. https://doi.org/10.1016/j.ejpb.2009.05.005
206. Mathe R, Casian T, Tomuță I (2020) Multivariate feed forward process control and optimization of an industrial, granulation based tablet manufacturing line using historical data. Int J Pharm 591:119988. https://doi.org/10.1016/j.ijpharm.2020.119988
207. León Blanco JM, González-R PL, Arroyo García CM, Cózar-Bernal MJ, Calle Suárez M, Canca Ortiz D et al (2018) Artificial neural networks as alternative tool for minimizing error predictions in manufacturing ultradeformable nanoliposome formulations. Drug Dev Ind Pharm 44:135–143. https://doi.org/10.1080/03639045.2017.1386201
208. Simões MF, Silva G, Pinto AC, Fonseca M, Silva NE, Pinto RMA et al (2020) Artificial neural networks applied to quality-by-design: from formulation development to clinical outcome. Eur J Pharm Biopharm 152:282–295. https://doi.org/10.1016/j.ejpb.2020.05.012
209. IMDRF SaMD Working Group (2013) Software as a Medical Device (SaMD). Int Med Device Regul Forum 1–9
210. Benjamens S, Dhunnoo P, Meskó B (2020) The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. Npj Digit Med 3:118. https://doi.org/10.1038/s41746-020-00324-0
211. The Medical Futurist. https://medicalfuturist.com/fda-approved-ai-based-algorithms/. Accessed 21 May 2021
212. Nindrea RD, Aryandono T, Lazuardi L (2018) Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: a meta-analysis. Asian Pac J Cancer Prev 19:1747–1752. https://doi.org/10.22034/APJCP.2018.19.7.1747

213. How IBM Watson overpromised and underdelivered on AI health care—IEEE spectrum. https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care. Accessed 21 May 2021
214. Zhou N, Zhang C, Lv H, Hao C, Li T, Zhu J et al (2019) Concordance study between IBM Watson for oncology and clinical practice for patients with cancer in china. Oncologist 24: 812–819. https://doi.org/10.1634/theoncologist.2018-0255
215. Research | MIT Clinical ML. http://clinicalml.org/research/. Accessed 21 May 2021
216. Faggella D (2019) 7 applications of machine learning in pharma and medicine | Emerj. Emerj
217. Baxt WG (1990) Use of an artificial neural network for data analysis in clinical decision-making: the diagnosis of acute coronary occlusion. Neural Comput 2:480–489. https://doi.org/10.1162/neco.1990.2.4.480
218. Babaian RJ, Fritsche HA, Zhang Z, Zhang KH, Madyastha KR, Barnhill SD (1998) Evaluation of ProstAsure index in the detection of prostate cancer: a preliminary report. Urology 51:132–136. https://doi.org/10.1016/S0090-4295(97)00574-8
219. Schneider J, Bitterlich N, Velcovsky HG, Morr H, Katz N, Eigenbrodt E (2002) Fuzzy logic-based tumor-marker profiles improved sensitivity in the diagnosis of lung cancer. Int J Clin Oncol 7:145–151. https://doi.org/10.1007/s101470200021
220. Al-Ayyoub M, Husari G, Darwish O, Alabed-Alaziz A (2012) Machine learning approach for brain tumor detection. ACM Int Conf Proc Ser 4–7. https://doi.org/10.1145/2222444.2222467
221. Segato A, Marzullo A, Calimeri F, De Momi E (2020) Artificial intelligence for brain diseases: a systematic review. APL Bioeng 4:041503. https://doi.org/10.1063/5.0011697
222. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:115–118. https://doi.org/10.1038/nature21056
223. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S et al (2015) ImageNet large scale visual recognition challenge. Int J Comput Vis 115:211–252. https://doi.org/10.1007/s11263-015-0816-y
224. AlexNet: the Architecture that Challenged CNNs | by Jerry Wei | Towards Data Science. https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951. Accessed 5 May 2021
225. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Proceedings of the IEEE conference on computer vision and pattern recognition. In: Going deeper with convolutions, pp 1–9. Online available at https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Szegedy_Going_Deeper_With_2015_CVPR_paper.pdf
226. Deep Learning: GoogLeNet Explained | by Richmond Alake | Towards Data Science. https://towardsdatascience.com/deep-learning-googlenet-explained-de8861c82765. Accessed 5 May 2021
227. Soltan AAS, Kouchaki S, Zhu T, Kiyasseh D, Taylor T, Hussain ZB et al (2021) Rapid triage for COVID-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test. Lancet Digit Heal 3:e78–e87. https://doi.org/10.1016/S2589-7500(20)30274-0
228. AI test rules out a COVID-19 diagnosis within one hour in Emergency Departments—Institute of Biomedical Engineering. http://www.ibme.ox.ac.uk/news-events/news/ai-test-rules-out-a-covid-19-diagnosis-within-one-hour-in-emergency-departments. Accessed 21 May 2021
229. Naseem M, Akhund R, Arshad H, Ibrahim MT (2020) Exploring the potential of artificial intelligence and machine learning to combat COVID-19 and existing opportunities for LMIC: a scoping review. J Prim Care Community Heal 11. https://doi.org/10.1177/2150132720963634
230. Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J (2020) How big data and artificial intelligence can help better manage the COVID-19 pandemic. Int J Environ Res Public Health 17:3176. https://doi.org/10.3390/ijerph17093176

231. Bogoch II, Brady OJ, Kraemer MUG, German M, Creatore MI, Kulkarni MA et al (2016) Anticipating the international spread of Zika virus from Brazil. Lancet 387:335–336. https://doi.org/10.1016/S0140-6736(16)00080-5
232. An AI epidemiologist sent the first warnings of the coronavirus | WIRED. https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/. Accessed 21 May 2021
233. Kandula S, Shaman J (2019) Reappraising the utility of Google Flu Trends. PLOS Comput Biol 15:e1007258. https://doi.org/10.1371/journal.pcbi.1007258
234. Al Hossain F, Lover AA, Corey GA, Reich NG, Rahman T (2020) FluSense. Proc ACM Interactive, Mobile, Wearable Ubiquitous Technol 4:1–28. https://doi.org/10.1145/3381014
235. Big data and artificial intelligence in pharmaceutical manufacturing | Pharma Focus Asia. https://www.pharmafocusasia.com/information-technology/bigdata-artificial-intelligence-pharmaceutical-manufacturing. Accessed 21 May 2021
236. AI for drug discovery, biomarker development and advanced R & D landscape overview (2020). Deep Pharma Intelligence. Online. available at https://analytics.dkv.global/deep-pharma/AI-for-Drug-Discovery-2020.pdf
237. Exscientia | AI Drug Discovery | Pharmatech. https://www.exscientia.ai/. Accessed 4 May 2021
238. Atomwise—better medicines faster. https://www.atomwise.com/. Accessed 4 May 2021
239. Drug discovery with an AI-augmented platform—cyclica. https://www.cyclicarx.com/. Accessed 4 May 2021
240. Schrödinger | Schrödinger is the scientific leader in developing state-of-the-art chemical simulation software for use in pharmaceutical, biotechnology, and materials research. https://www.schrodinger.com/. Accessed 4 May 2021
241. About | Insilico Medicine. https://insilico.com/about. Accessed 4 May 2021
242. Iktos—Artificial Intelligence for new drug design—Iktos. https://iktos.ai/. Accessed 4 May 2021
243. Biovista | Drug Positioning and Prioritization—Home. https://www.biovista.com/. Accessed 4 May 2021
244. Numerate to use AI platform in developing drugs for Takeda. https://www.genengnews.com/topics/translational-medicine/numerate-to-use-ai-platform-in-developing-drugs-for-takeda/. Accessed 4 May 2021
245. BERG | Back to Biology for a Healthier Tomorrow. https://www.berghealth.com/. Accessed 4 May 2021
246. BenevolentAI | Because it matters. https://www.benevolent.com/. Accessed 4 May 2021
247. NVIDIA Partners with Schrödinger to further accelerate drug discovery worldwide | NVIDIA Newsroom. https://nvidianews.nvidia.com/news/nvidia-partners-with-schrodinger-to-further-accelerate-drug-discovery-worldwide. Accessed 21 May 2021
248. Schrödinger announces expanded collaboration with AstraZeneca to extend computational modeling solutions to biologics | Schrödinger, Inc. https://ir.schrodinger.com/news-releases/news-release-details/schrodinger-announces-expanded-collaboration-astrazeneca-extend. Accessed 21 May 2021
249. Roberts RJ, Rowe R (1998) Intelligent software for product formulation. CRC Press
250. Aguilar JE, García RF, Landin M (2013) Formulation tools for pharmaceutical development. Woodhead Publishing Series in Biomedicine
251. AI drug formulation | Metis Pharmaceuticals. https://www.metispharmaceuticals.com/. Accessed 24 Apr 2021
252. Drug Formulation Development—MedAI. https://aimed.protheragen.com/drug-formulation-development.html. Accessed 21 May 2021
253. Collaboration to use AI to produce critical pharmaceutical ingredients. https://www.outsourcing-pharma.com/Article/2020/05/11/Collaboration-to-use-AI-to-produce-critical-pharmaceutical-ingredients. Accessed 24 Apr 2021

254. Elbadawi M, Muñiz Castro B, Gavins FKH, Ong JJ, Gaisford S, Pérez G et al (2020) M3DISEEN: a novel machine learning approach for predicting the 3D printability of medicines. Int J Pharm 590:119837. https://doi.org/10.1016/j.ijpharm.2020.119837
255. M3DISEEN—an AI pharmaceutical software for 3D printing of medicines—FabRx. https://www.fabrx.co.uk/2020/09/28/m3diseen-an-ai-pharmaceutical-software-for-3d-printing-of-medicines/. Accessed 24 Apr 2021
256. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H et al (2020) International evaluation of an AI system for breast cancer screening. Nature 577:89–94. https://doi.org/10.1038/s41586-019-1799-6
257. Nikolov S, Blackwell S, Zverovitch A, Mendes R, Livne M, De Fauw J, Patel Y, Meyer C, Askham H, Romera-Paredes B, Kelly C (2018) Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. arXiv preprint arXiv:1809.04430. https://arxiv.org/pdf/1809.04430. Accessed 12 Sep 2018
258. Powles J, Hodson H (2017) Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl) 7:351–367. https://doi.org/10.1007/s12553-017-0179-1
259. Luxton DD (2019) Number 2: E131-137 CASE AND COMMENTARY should Watson be consulted for a second opinion? AMA J Ethics 21:131–137
260. Zauderer MG, Gucalp A, Epstein AS, Seidman AD, Caroline A, Granovsky S et al (2014) Piloting IBM Watson Oncology within Memorial Sloan Kettering's regional network. J Clin Oncol 32:e17653–e17653. https://doi.org/10.1200/jco.2014.32.15_suppl.e17653
261. How IBM Watson overpromised and underdelivered on AI health care—IEEE spectrum. https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care#LinkToAIHealthTable. Accessed 21 May 2021
262. IBM Watson Health and EBSCO Information Services Collaborate to Launch Integrated Clinical Decision Support Solution—Mar 11, 2020. https://newsroom.ibm.com/2020-03-11-IBM-Watson-Health-and-EBSCO-Information-Services-Collaborate-to-Launch-Integrated-Clinical-Decision-Support-Solution. Accessed 21 May 2021
263. Kingslake J, Dias R, Dawson GR, Simon J, Goodwin GM, Harmer CJ et al (2017) The effects of using the PReDicT Test to guide the antidepressant treatment of depressed patients: study protocol for a randomised controlled trial. Trials 18:558. https://doi.org/10.1186/s13063-017-2247-2
264. Predicting Response to Depression Treatment (PReDicT)—Full Text View—ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT02790970. Accessed 21 May 2021
265. Esmaeilzadeh P (2020) Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med Inform Decis Mak 20:170. https://doi.org/10.1186/s12911-020-01191-1
266. Wang F, Preininger A (2019) AI in health: state of the art, challenges, and future directions. Yearb Med Inform 28:016–026. https://doi.org/10.1055/s-0039-1677908
267. Gerke S, Minssen T, Cohen G (2020) Ethical and legal challenges of artificial intelligence-driven healthcare. In: Artif. Intell. Healthc. Elsevier, pp 295–336. https://doi.org/10.1016/B978-0-12-818438-7.00012-5
268. Ekins S (2016) The next era: deep learning in pharmaceutical research. Pharm Res 33:2594–2603. https://doi.org/10.1007/s11095-016-2029-7

Robotic Automation of Pharmaceutical and Life Science Industries

12

Vikas Anand Saharan

Contents

V. A. Saharan (✉)
Department of Pharmaceutics, School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Dehradun, Uttarakhand, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_12

Abstract

Robotic automation brings flexibility, speed, accuracy, and durability to almost all processes in industries. Robots have significantly advanced from machines to collaborative robots to humanoids. Stringent safety regulations have made deployment of robots in industries safer and catalyzed the entry of collaborative robots (cobots) and humanoids in research laboratories, offices, and hospitals as assistants to humans. Robots are working at places where human cannot work or human involvement is a concern for quality issues. Internet of Things (IoT), artificial intelligence (AI), and robotic advancements are continuously decreasing human interventions in drug discovery, synthetic chemistry, and biotechnology laboratories. Robots equipped with vision guidance systems, sensors, softwares, and AI-based algorithms are re-shaping pharmaceutical and life science industries. This chapter gives a most recent update on robotic innovations for automation of manufacturing, packaging, warehousing, and laboratory processes of industries, in particular pharmaceutical and life science industries.

Keywords

Robot · Artificial intelligence · Manufacturing · Packaging · Warehousing · Palletizing · Clean room · Aseptic process · 3D printing · Personalized medicine · Radiopharmaceutical · Vaccine · Biotechnology

Chapter Objectives

Upon reading of this chapter, it is expected that reader will be able to understand

- Different types of industrial robots
- Salient features of pharmaceutical clean robots
- Role of robots in manufacturing dosage forms and personalized drug delivery systems
- Robotic automation of fill-finish lines for manufacturing parenteral products
- How robots installed within isolators and restricted access barrier systems (RABS) are redefining aseptic processing?
- Collaborative robots assisted research laboratories.
- End of the line and in-line packaging/pelletizing solutions by robots
- Automatic mobile robots for automation of activities within warehouse

12.1 Introduction

Cambridge dictionary defines robot as "a machine controlled by a computer that is used to perform jobs automatically" [1]. This is a very broad definition which includes machine, software, and the system's ability to perform tasks on its own. International Organization for Standardization (ISO) defines industrial robots as "an automatically controlled, reprogrammable, multipurpose manipulator programmable in three or more axes, which may be either fixed in place or mobile for use in industrial automation applications" [2]. Reprogramming of robots allows change in motions or auxiliary functions of robots without the need for any physical alteration and this functionality confers them great flexibility and adaptability to perform various tasks. Apart from its overall structure, other necessary components of a robot include power source, sensor, actuator, manipulator, and controller. Most commonly used power sources are batteries. Sensors are equipment to measure parameters like sound, temperature, light, acceleration, proximity, force, etc. Actuator provides means for movement of either the robot itself or its parts. Manipulators have functions similar to the human arm and, therefore, possess several joints and links. Manipulators are fixed to a supporting base at one place while at another place attached to an end-effector which is a free end to perform tasks of gripping and holding just like the human palm and fingers. Controllers, analogous to the human brain, are software coupled with hardware which gets several inputs from sensors and adjusts outputs of the robot through the manipulator and end-effector. Initial robots were stationary but modern robots have locomotion abilities like hopping, snaking, or walking.

Robots offer accuracy, speed, flexibility, and durability for highly repetitive and high volume tasks in industries. Use of robots is recommended for areas which are hazardous to humans, for example, radiopharmaceutical manufacturing/laboratory, cytotoxics, and places where human movements shall be restricted to excel clean environment, for example, aseptic areas, vaccines, and parenteral manufacturing. Camera and sensors coupled with AI algorithms are extensively used in navigating robots and enable sensing bar-code/color/weight of products to confer track and trace capability for specific handling/packaging requirements on single/multiple products during manufacturing/packaging. Digitized operational model based on most recent developments in collaborative robotics, AI, big data, distributed cloud-based architectures, and interconnectivity leads to the smart pharmaceutical industry referred to as Industry 4.0/Pharma 4.0 [3].

Robots have gained access in almost all industries, including pharmaceutical and life science industries. Manufacturing, packaging, warehousing, and laboratories are areas which have observed significant influx of automation through robotic advancements. Lots of robotic advancements, in particular relevant to automation, in these areas are highlighted in this chapter to explore the ever-rising role of robots in manufacturing of drug products.

12.2 Types of Robots

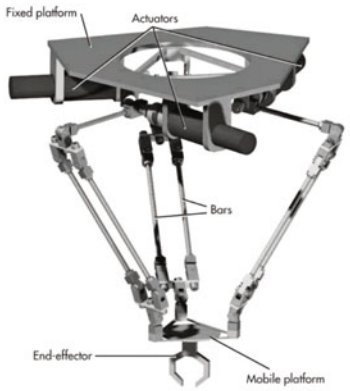
Industrial robots belong to eight different types, viz. parallel, SCARA, articulated, Cartesian, collaborative, autonomous mobile robots, exoskeletons, and humanoids. Parallel robots (like delta robot), SCARA, Cartesian, and articulated robots are widely deployed for performing manufacturing and packaging operations in industries. Industrial robots may be installed standalone or integrated into manufacturing/packaging lines [3]. Standalone industrial robots, 4–7 axis articulated arm robots, are end of line installations for pick and place operations like palletizer. Integrated robots or machine centric robots work synchronously with other machines and the whole assembly works efficiently as per Industry 4.0 concepts [4]. A single controller can control multiple robots installed in a system to achieve integration and improve operational efficiency. A tripod delta robot (e.g., ABB's FlexPicker) can be installed for picking filled and labeled vials from one conveyor belt and placing it in the primary package on another conveyor belt followed by its packing by other machines.

12.2.1 Parallel (Delta) Robots

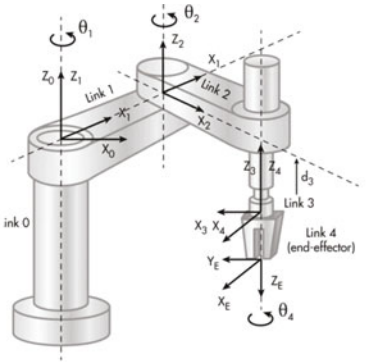
This type of robot comprises a platform mounted on legs. These systems provide more precision (e.g., for insertion) than arm robot. Delta robot is an exact type of parallel robot (Fig. 12.1a). Delta robot consists of three arms connected to a base platform via universal joints [5]. The key design feature is the use of parallelograms in the arms, which restrict the movement of the end platform [6]. Lightweight passive arms are capable of moving with great speed. Actuators are fixed at the base platform. Most common applications of delta robots are packaging, material handling, and high precision assembling. Delta robots have popular usage in picking and packaging in factories because they can be quite fast, some executing up to 300 picks per min. Merck, for example, is using a Fanuc delta robot on a bottling line to place dispenser caps onto bottled allergy medications [7]. Ten variants of the bottle can be run on the system, and the only robot line-change requirement is to select the appropriate program on the robot controller.

12.2.2 SCARA Robots

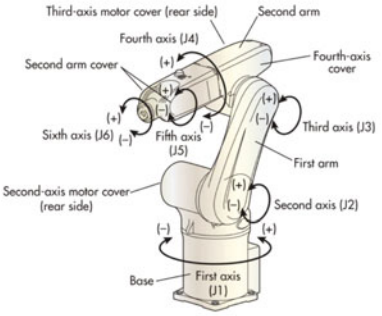
SCARA (selective compliance assembly/articulated robot arm) can move freely, but these movements are restricted to only one single geometrical plane [5]. SCARA has a cylindrically shaped work envelope. SCARA robots can move in three axes (x, y, and z) and additional fourth axis movement is conferred with the movement of an end effector (Fig. 12.1b). The movements resemble the movements of the human hand to extend the arm and to retract it by folding up. Most SCARAs have been designed for movements in four axes, hence, the term "four-axis robot" and "SCARA" are often used interchangeably [8]. However, SCARA with three, five, and six axis movements have also been designed. SCARA robots may also be integrated in a ceiling mounted long stroke Cartesian-robot for saving floor space



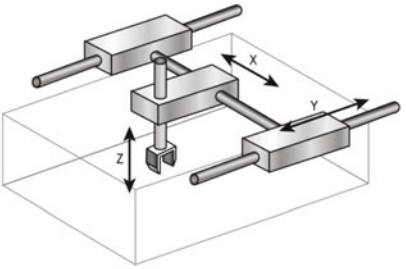
(a)



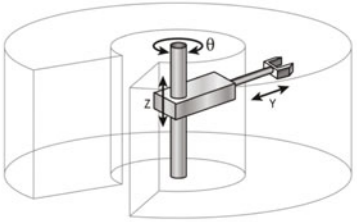
(b)



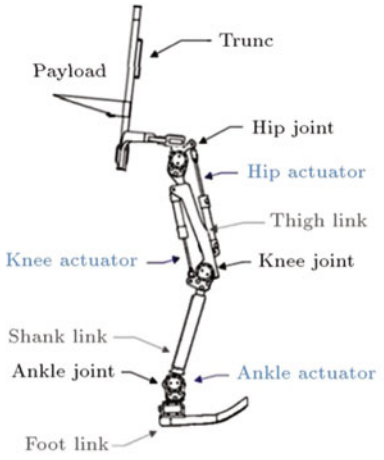
(c)



(d)



(e)



(f)

Fig. 12.1 Schematic diagrams of robots. (**a**) Parallel robot (Delta), (**b**) SCARA robot, (**c**) articulated robot, (**d**) Cartesian robot, (**e**) cylindrical robot (Courtesy: Machine Design) (**a–e**) Reproduced from [13] (**f**) Exoskeleton Reproduced with permission from [14]

and eliminating need of conveyor in some applications. Most common applications of SCARAs include precise, fast, and repetitive point to-point movements for palletizing, machine loading, and assembling [6].

12.2.3 Articulated Robots

Articulated robots have three to seven degrees of freedom due to rotary joints which allows them to bend back and forth (Fig. 12.1c). In comparison to SCARAs, articulated robots have both horizontal and vertical joints which allow greater freedom of movement [8]. Action of articulated robots mimics a human arm and hand and allows them to perform almost any task in a spherical work envelope [5]. Six-axis articulated robots, the most common articulated industrial robots, have two vertically rotating wrist joints, two vertically rotating forearm joints, and two horizontally rotating joints located at its base. With all these six joints, the movements of articulated robots resemble the dexterity of human forearm and wrist. Articulated robots, capable of working in seven axes, are efficient to perform on objects placed (position) in all three axes X, Y, and Z at various angles (orientation) [9]. Motion of a six-axis robot on a linear path (seventh axis) also referred to as linear robot transfer units or shuttle system or robot positioning systems confers facility for performing multiple tasks with the avoidance of rotation of parts and vibrations [10]. Articulated robots have been employed for wide range of applications in assembling, arc or spot welding, painting, material handling, and palletizing [6].

12.2.4 Cartesian/Linear/Gantry Robots

Cartesian robots (Fig. 12.1d) move in at least three orthogonal axes (x, y, and z at 90° angles of each other as per Cartesian coordinate) [8]. These three axes do not rotate but perform coordinated motion through a common motion controller and hence these systems are also referred to as linear robots. Their linear mobility simplifies robot control [6]. An attached wrist may allow fourth degree of freedom for rotational movements. A Cartesian robot has a cube-shaped work envelope. Cartesian robots can be mounted overhead and thus saves floor space. Due to their restricted range of motions these robots are low cost and often incorporated in machines/automated systems dedicated to a single purpose like assay testing. With increasing productivity and decreasing drug development time, these are mainly applied in radioactive, fluorescent, and luminescent analysis. Apart from these applications, these robots can also perform sample selection and scrutiny of the resulting data.

Gantry robots are configured with two base axes (x) and a spanning second axis (y). Such a configuration allows longer stroke lengths for carrying larger payloads than Cartesian robots and prevents cantilevering of the second axis (y). Two y or two z axes may be included for increasing stiffness and load capacity of the gantry.

12.2.5 Cylindrical Robots

Cylindrical robots (Fig. 12.1e) have vertical and horizontal linear axes and rotary axes at its base and at the wrist [5]. One rotary joint confers rotational motion from its base while one prismatic joint allows movement in linear motion. These robots have a cylindrical shape work envelope. Cylindrical robots have a rigid structure which requires clearance at the rear of the arm. These robots are best suited to general pick and place applications, assembling, and machine tending.

12.2.6 Collaborative Robots

Collaborative robots are equipped with software and machine learning algorithms in order to make them more interactive with humans and support humans in performing assigned tasks [6]. These robots are programmed for great flexibility and do not require safety fences around them. Collaborative robots are mostly single robot installations with simple and discrete input/output interfaces that lower installation and programming costs. Collaborative robots are currently deployed for incidental work such as materials handling and assembling operations. In logistics and warehousing, picking, shifting, and supportive works like kitting and pre-retail services can be assisted with collaborative robots.

12.2.7 Autonomous Mobile Robots (AMRs) and Autonomous Guided Vehicles (AGVs)

AMRs and AGVs are not installed or fixed at one place rather these are mobile machines that use either fixed path or a navigation system for their mobility [6]. Camera or laser-based navigation system is onboard in AMRs while AGVs depend on external guided tracks based on magnetic tape, wire, or rails mechanisms. AMRs and AGVs have been used in logistics, delivery, movements of boxes/pellets/tools/materials in warehouses, distribution centers, manufacturing intralogistics, agriculture, hospitals, and retail.

12.2.8 Exoskeletons/Human-Robot-Hybrids

Exoskeleton (Fig. 12.1f) is a robot connected to the human body to increase human strength and capacity to perform heavy-duty or ergonomically challenging tasks like lifting and carrying heavier objects [6]. Exoskeletons/Human-Robot-Hybrids technology is relatively new. They can be employed in industrial warehouses to support worker movements in lifting/carrying processes. Powered exoskeletons for upper and lower extremities have potential applications in rehabilitation, rescue and disaster relief, human performance augmentation in military defense, ergonomic support for reducing loads on spine, hips, and shoulders when lifting heavy weights at work.

12.2.9 Humanoid Robots

Humanoid robots resemble humans in their appearance. Humanoid robots are gaining wide acceptance for their use in manufacturing, delivery, research, acting, space exploration, search/rescue operations, education, personal assistance, caregiving, health care, and public relations [11]. Most common industrial applications involve delivery and supply, receptionist, research, and collaborative works. Humanoid chewing robot, better called as humanoid chewing simulator, has been recently developed for evaluating drug release from chewing gums [12].

12.3 Robots in Pharmaceutical Industries

Robots and other machines aim to achieve automation of boring and repetitive tasks in almost any type of hostile/non-hostile/hygienic/unhygienic environment and thereby reduce human intervention to its minimum. Automaton protects human operators from the occupational hazards of the industrial processes and ensures quality of products by protecting them from humans who might be a source of contamination. Operator safety is of utmost concern while manufacturing cytotoxic, radiopharmaceutical, allergic, and highly potent pharmacological products [15]. Manufacturing of parenterals, ophthalmics, vaccines, biologics, and biotechnological products require clean room conditions where human interventions shall be minimized. Robotic automation is an answer to achieve both of these objectives. Advantages and disadvantages of robotic automation are provided in Box 12.1.

Box 12.1 Advantages and disadvantages of robots

Advantages	Disadvantages
Flexibility	Safety issues if it gets rogue
Accuracy	Expensive (initial installation cost)
Precision	Return on investment (not always)
Reliability	Too complex to implement
Reproducibility	Ease of use
Ruggedness	Acceptance
Efficiency	Require expertise for installation, maintenance, and repair
Quality	
Speed and time saving	
Continuous work tirelessly	
Pandemic free manufacturing	
Easy redeployment	
Work in environment where humans cannot work	
Safety (if complies to safety standards)	

Despite the availability of economical industrial robots, automation is highly cost-intensive. Return on investment is expected through flexibility, energy savings, increased quality, and high-speed production which require a careful evaluation [16]. Robots have been extensively used in automation of pharmaceutical manufacturing, parenteral fill/finish lines, aseptic processing, packaging, and warehousing. With the rise of FDA compliant cleanroom robots, their applications in automating parenteral manufacturing, aseptic areas, vaccines, cytotoxics, and research laboratories are rising tremendously. Warehouses and logistics also look forward to robotic advancements for handling, carrying, and transportation of heavy objects. Box 12.2 provides recent applications of robots in automating various processes in pharmaceutical industries. Table 12.1 enlists prominent robotic companies and their products.

Box 12.2 Applications of robots in pharmaceutical manufacturing, packaging, and warehousing and laboratories

Manufacturing	Packaging and warehouse	Laboratory
Automated manufacturing of tablets, capsules, pellets, implants, medical devices, and other dosage forms Additive manufacturing/3D printing of dosage forms and implants Vaccine production Producing personalized medicines Parenteral fill-finish systems Preparing radiopharmaceuticals Aseptic syringe or vial dispensing, filling, and capping system Pick and place operations on an automated fill/finish lines	Packing of dosage forms into primary package Packing primary package into cartons (vertical/horizontal cartoning systems) Packing cartons into case Placing cases onto a pallet Multiple product differentiation on a single packaging line Single product packing differentiation on a single packaging line Automated pill bottle assembly, labeling, and packaging systems Automated syringe assembly, inspection, and preparation for packaging Picking/placing/transport of cases/cartons in warehouse	Preparing/collecting samples (auto-sampling) with extreme precision and accuracy Transportation of microtiter plates between instruments Implementing test protocols with high speed and accuracy Pick, place, and loading operation for analysis Cleaning operations Laboratory or pilot scale manufacturing/packing operations

Table 12.1 Robotic companies and their products

Robotic company	Head quarter	Some robotic products	Reference (URL)
ABB	Zurich, Switzerland	Articulated robots (IRB 1100, IRB 1300), SCARA (IRB 910INV, 910SC), Dual Arm YUMI IRB 14000, Single Arm YUMI IRB 14050, Collaborative SWIFTI CRB 1100, GoFa CRB 15000, IRB 360 FlexPicker Delta	https://new.abb.com/
The Yaskawa Electric Corporation	Kitakyushu, Japan	Articulated robot, double arm robots, collaborative robot (cobot), biomedical single/double arm robot (automated cell culture system, anticancer drug preparation support system, automated bacteria test system)	yaskawa-global.com/product/robotics
KUKA (Midea Group)	Bavaria, Germany	KR SCARA, KR AGILUS, LBR iiwa, KR Cybertech Nano, KR Iontec, KR 40 PA, KR Quantec, Linearroboter	https://www.kuka.com/
The Fanuc Corporation	Yamanashi Prefacture, Japan	Fanuc LR Mate 200id clean room robots, Funac M-1iA delta robot	https://www.fanuc.co.jp/
Kawasaki Heavy Industries	Tokyo, Japan	Hollow wrist clean robots (medical and pharmaceuticals; MS005N, NC004V and MC004N), dual arm collaborative SCARA (duAro), 6-axis vertical articulated robots, 7-axis articulated robots	https://robotics.kawasaki.com/
Epson Robots (Seiko Epson Corporation)	Tokyo, Japan	SCARA, 6-axis, all-in-one series robots	https://epson.com/industrial-robots-factory-automation
Denso Corporation	Aichi, Japan	4-, 5-, and 6-axis robots, integrated robots (ceiling mounted, flexible SCARA), cobotta collaborative robot, medical/pharmaceutical robot (VS-S2 Series)	https://www.densorobotics.com/
Mitsubishi Electric Corporation	Tokyo, Japan	Vertical (articulated), horizontal (SCARA), collaborative (Assista), environment resistant (for foods and medicinal products), software, robot hand tools, 3D vision sensor	https://www.mitsubishielectric.com
Honda Robotics	Tokyo, Japan	ASIMO, UNI-CUB, walking assist device, task-performing robot arm, high-access survey robot	https://global.honda/innovation/robotics.html
Marchesini Group S.p.A	Bologna, Italy	Packaging lines for vaccines; aseptic liquid filling/sealing/capping/assembling/handling/stoppering lines for injectable drugs; stick pack line for antiacid and analgesic	https://www.marchesini.com/

(continued)

Table 12.1 (continued)

Robotic company	Head quarter	Some robotic products	Reference (URL)
		(Robocombi 380; robotic pick-and-place unit), automatic inspection machines for pre-filled syringes, Integra 320 (robotic blister line)	
Universal Robots A/S (Denmark)	Odense, Denmark	UR3e, UR5e, UR10e, UR16e Collaborative Robot Arm/Industrial Robot	https://universal-robots.com
Shibuya Corporation	Ishikawa, Japan	Sterility testing isolator, steriliztion/decontamination systems, filling systems (vials, ampoules, syringe, nested syringe, opthalmic solution), aseptic powder filling systems, various other filling/packaging lines	https://www.shibuya.co.jp/en/
Stäubli	Pfaffikon, Switzerland	Industrial robots (SCARA, 6-axis), HelMo mobile robots, and AGV, cobots, Stericlean robot (TX2-40, TX2-60)	https://www.staubli.com/en-us/robotics/product-range/
Comau	Turin, Italy	Integrated robots, software, e.DO modular open source robot	https://www.comau.com/en
Omron Robotics	California, US	Omron autonomous mobile robots (LD series), industrial robots (parallel, SCARA, articulated), collaborative robots, robot controllers, software	https://robotics.omron.com/

12.4 Automation of Pharmaceutical Manufacturing

Three-dimensional (3D) printing is an additive technology of preparing objects in a repeated layer on a layer fashion to obtain three-dimensional objects. Pharmaceutical dosage forms of any shape and size can be prepared using 3D printers. 3D printers have 3 degrees of freedom as their print-head, supported by gantry, moves in x and y directions and base moves in z direction. Not only external features of a dosage form but also internal architecture like porosity can be designed precisely. 3D printing allows preparing tablets, films, wafers, and medical devices like stents. 3D printing also offers benefits of customizing dosage forms as per the need of an individual, for example, color, shape, and polypills, thereby bringing host of personalization and customization options to dosage forms. Furthermore, robots have also been deployed in groups for automated manufacturing of personalized dosage forms containing multiple drugs as per the requirement of the prescription.

12.4.1 3D Printing of Dosage Forms

FabRx, a UK-based company, has recently introduced (2020) its compact and portable 3D printer Medimaker™ (Fig. 12.2) for personalized medicines like polypills, chewable tablets, and orally disintegrating tablets [17]. This machine uses computer-aided design (CAD) software for customized layer-by-layer production of 3D-printed tablets (Printlets™). FabRx's Printlet™ technology is capable of producing flexible dosage forms with customized dose, drug release, shape, and appearance [18]. Medimaker is equipped with multiple nozzle printing systems and in-process quality control systems. Its specialized software gives access to qualified personnel through a fingerprint access control. A fitted camera monitors tablet printing process and in line quality control process assists detection of any fault during manufacturing of tablets. Three printing nozzles cater the need for different manufacturing requirements. Medimaker™ prints tablets at the rate of 28 tablets in about 8 min. FabRx is marketing this 3D printer for production of small batches of tablets for research and development, preclinical studies, clinical studies, and personalized medicine.

Aprecia Pharmaceuticals uses ZipDose technology and Z Corp 3D printing process, licensed from MIT, for fusing powder containing drug with binder solution for manufacturing FDA-approved Spritam (levetiracetam) orally disintegrating tablets since 2015 [19]. Triastek, a Chinese pharmaceutical company, has recently (Feb. 2021) received IND 505(b)(2) approval for T19, a 3D-printed drug product for the treatment of rheumatoid arthritis [20]. Triastek relies on its fully automated melt

Fig. 12.2 Medimaker™ (Source: FabRx). Reproduced from [17]

extrusion deposition technology (3D printing technology) integrated with real-time process analytical technology for producing modified release dosage forms for a precise control on drug release similar to a chronotherapeutic drug delivery system.

12.4.2 Personalized Medicines

Multiply Labs has recently introduced FDA compliant robotic factory for industrial scale production of personalized medicines [21]. Personalized prescription is placed on a try which is grabbed by a robotic arm which places it on a filling station. Filling robot fills capsules with the prescribed quantity of the formulation. When all capsules are filled, robotic arm moves the tray to a weighing station where quality robot checks all capsules for precise dosage strength. The process repeats itself till the capsule contains all the drugs mentioned in the personalized prescription. Multiply Labs applies parallel manufacturing wherein filling robots and quality robots are placed in parallel and operate simultaneously. The robotic arm moves trays from one filling robot to another and also from filling robots to quality robots. A single manufacturing unit of size 128 sq. ft. is capable of producing 10,000 personalized capsules per shift of 8 h (Fig. 12.3).

The Arcelis platform technology (Argos Therapeutics now acquired by SCM LifeScience) used two, five-axis robotic arms in the production of personalized immunotherapies. Robots were deployed for production of mRNA from a patient's own tumor [7]. The produced mRNA was subsequently used as antigen to patient's own dendritic cells in a cellular processing equipment. Antigen-loaded dendritic cells were formulated into the patient's plasma for administration via intradermal injection [22].

12.5 Automation of Parenteral Manufacturing and Aseptic Processes

Manufacturing of vaccines, biologics, biosimilars, cytotoxics, radiopharmaceuticals, and parenteral drug products requires aseptic conditions. Drug regulatory agencies are looking forward to minimizing or eliminating human interventions in aseptic areas and parenteral/biotech manufacturing as humans are the biggest source of contamination. Therefore, pharmaceutical/life science industries are aggressively seeking alternatives of human operators from the parenteral manufacturing processes. Replacement of humans with robots in aseptic areas and parenteral fill/finish lines is, therefore, highly in demand. Self-injected therapies through autoinjectors are further increasing the adoption of robotic manufacturing. Isolator technologies heavily rely on use of robots. Robotic solutions provide pharmaceutical manufacturers with a flexible, faster, and cost-effective method of filling different container formats via the same filling line/platform. Robotic automatic solutions quickly respond to the growing demand for customized packaging, quick format, small batches and product changeovers together with less human (user) intervention

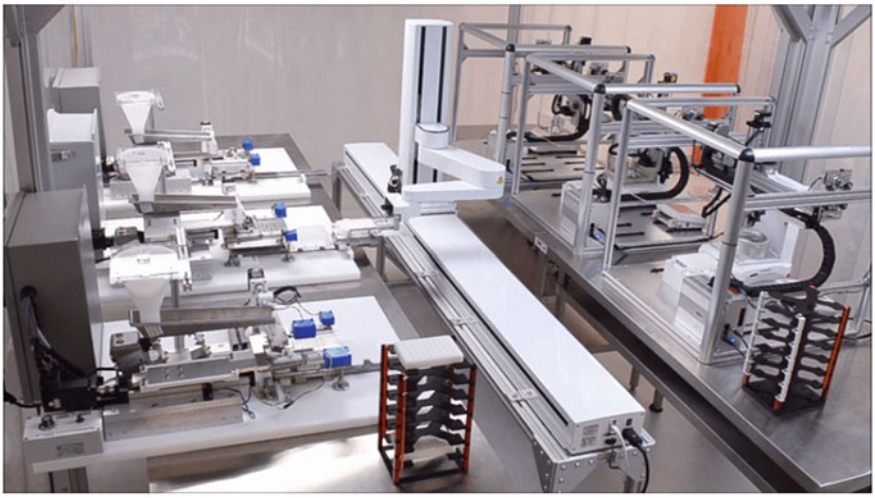


Fig. 12.3 Reproduced with permission from Multiply Labs. Source: Multiply Labs. Reproduced with permission from [21]

during the filling process. Robotic automation through isolators, RABS, and fill-finish systems excel at protecting working personnel from toxic substances and simultaneously assuring aseptic conditions with minimum human intervention. With precise and accurate handling, robotic automation also reduces glass-to-glass and glass-to-metal contact which prevents breakage of containers.

12.5.1 Pharmaceutical Clean Room Robots

Pharmaceutical robots are constructed from FDA-certified materials, for example, stainless steel, and designed in a manner to prevent deposition of dirt/dust, ingress of liquids and leakage of greases. FDA compliant coating and sealing makes such robots resistant to the action of alkaline and acidic cleaning liquids. Such robots could be sterilized safely with either exposure of UV light or hydrogen peroxide or wiped with hydrogen peroxide water. Pharmaceutical/food grade (e.g., NSF H1 certified) grease also aids cleanliness. In some robots cleanliness is secured with the use of double sealings at joints to prevent leakage of grease and designing wrist with a combined metallic and a resinous gear that does not require supply of grease [23]. Smooth surface, rounded exterior, sealed joints, and no external screws or especially designed bolts ensure efficient daily cleaning and prevent deposit of dirt and dust. As per ingress protection rating system IP67 rating and onwards confer adequate protection from dust and liquids to meet clean room requirements of pharmaceuticals. To facilitate easy cleaning and streamlining motion of robotic arm, bottom located control-cable connector and internal embedding of wires are preferred. Computational system validation including qualification, data integrity,

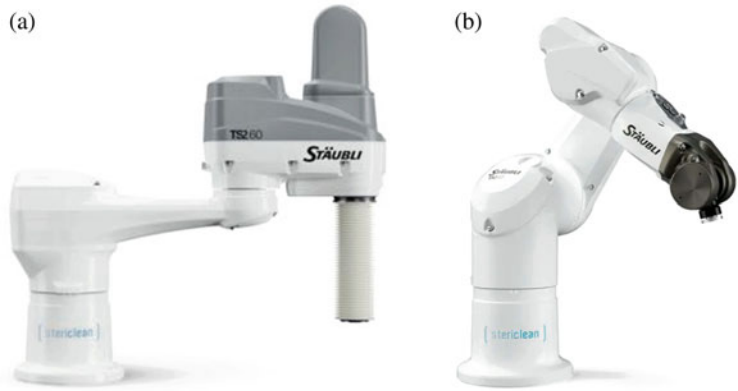


Fig. 12.4 Aseptic robots from Stäubli. (**a**) Stericlean SCARA Source: Stäubli. Reproduced with permission from [25]. (**b**) Stericlean six-axis robot. Source: Stäubli. Reproduced with permission from [26]

audit trail, and implementation of pharmacopoeial methods are some other additional requirements to be met by robots for pharmaceutical industries [24]. Two aseptic robots from Stäubli are shown in Fig. 12.4.

An epoxy white IP65 version with covered *Z*-axis spindle of SCARA robot (FANUC SCARA SR-12iA) in compliance to cleanliness class ISO5 is presented in Fig. 12.5a. The figure shows FANUC compact six-axis robot (Fig. 12.5b) with the approximate size and reach (911 mm) of a human arm and standard IP67 (FANUC LR Mate 200iD/7LC) a reliable mini robot for process automation in many industries.

12.5.2 Aseptic Isolators/Restricted Access Barrier Systems

An aseptic isolator suits for small-scale manufacturing of sterile products in pharmaceutical and biotech industries. Prefilters, HEPA filters, and catalytic converters are part of the air handling system [29, 30]. Connected isolators may have different pressure zones. Catalytic converters release filtered air in the room or filtered air may be re-circulated into the isolator. Positive/negative pressures may also be maintained as per the requirement in these aseptic isolators. Stainless steel and glass made isolators and restricted access barrier systems (RABS) are extensively used for small-scale manufacturing of parenteral products. Stainless steel and glass made isolators and RABS are easy to decontaminate with hydrogen peroxide and other oxidizing chemicals (H_2O_2). Aseptic isolators fitted with robots minimize or eliminate human interventions and thus may be constructed with or without gloves. Isolators fitted with interchangeable L-flange make them flexible for filling different types of containers such as vials, ampoules, cartridges in the same isolator. Inflatable gasket is used to seal L-flange to the isolator and a pressure inside the system makes

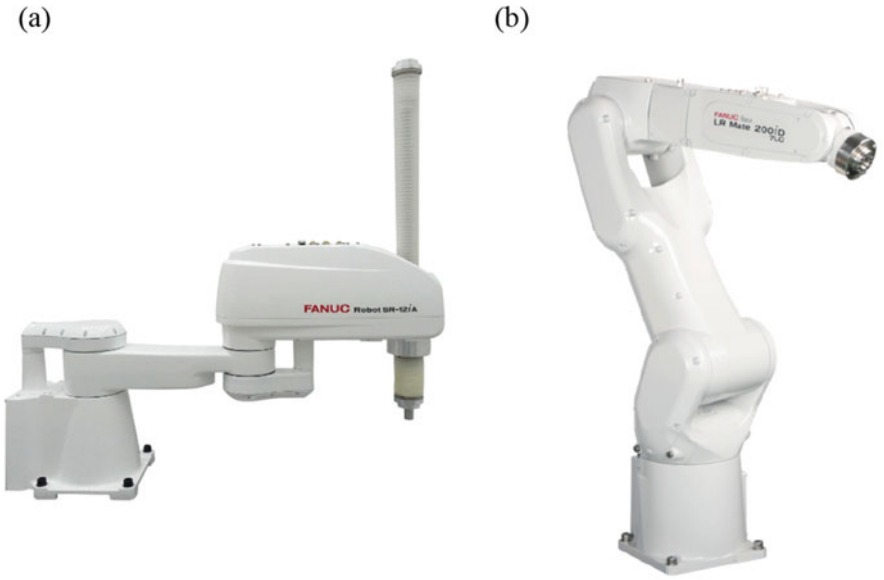


Fig. 12.5 Cleanroom robots from FANUC. (**a**) FANUC SCARA SR-12iA, Environmental option (IP 65 rating for body, wrist. and J3 arm) Source: FANUC Reproduce with permission from [27]. (**b**) FANUC LRMate200iD/7LC Source: FANUC Reproduce with permission from [28]

a perfect seal during the manufacturing process. L-flange has all devices for filling and transfer of liquids/solids, containers, and closures. Production may be started after decontamination of all parts with hydrogen peroxide. A control system for robots is equipped outside the isolator for recording real-time data from integrated robots and other sensors.

VarioSys® production systems are combinations of an isolator made by SKAN and machine modules made by Bausch + Ströbel and Harro Höfliger [31]. VarioSys® is suitable for a wide range of pharmaceutical manufacturing which ranges from small batches to clinical samples and production of batches in one installation in compliance with FDA requirements. The Bausch + Ströbel's DDM 9105 machine module is a fully automatic module for tub opening and denesting. This module is fitted with two Staubli stericlean robots mounted on an L-flange (Fig. 12.6a) of the Pharmaceutical Safety Isolator (hence named PSI-L) [29]. The robot opens tubs using a heating frame. A second robot removes the ready to use vials from the nest (denesting) and transfers them to the next module (e.g., SFM5105) for nested filling and closing (Fig. 12.6b) [32]. Quick decontamination unit (Sara airlock) with Bausch+Ströbel's DDM 9105 machine module and SFM5105 module offers complete automation of in-nest opening, denesting, filling, stoppering, crimp capping, and tray loading of ready to use vials (Fig. 12.6c). This aseptic line may also be adapted for ready to use syringes and an additional freeze drying module to expand its utility. Further, Bausch+Ströbel's L-shaped trolley-shaped interchangeable working table (L-flange) has provided a great flexibility for quick change of equipment on

Fig. 12.6 (**a**) Basic L Flange on webpage. (**b**) Staubli stericlean robots are mounted on the L-flange of the PSI-L isolator from brochure. (**c**) VarioSys® module DDM 9105 for tub opening and denesting (Bausch + Ströbel) fitted in Isolator PSI-L (SKAN) [32]. Reproduced with permission from [33]. Source: SKAN and Bausch + Ströbel Maschinenfabrik Ilshofen GmbH + Co

the same line for different manufacturing processes. A robotic L-flange is shown in Fig. 12.5a, wherein two Staubli stericlean robots mounted on the L-flange of the PSI-L isolator are shown in Fig. 12.6b.

Crystal® L1 Robot Line (Fig. 12.7) has been designed for aseptic filling of liquids and freeze-dried parenterals, including biohazards and cytotoxics, in ready to fill AT-Closed Vials®, open vials, and syringes when these containers are used as nests [36]. Staübli stericlean TX-40 robot (Fig. 12.7b, c) is mounted on a L-flange of the PSI-L isolator of SKAN. This robotic line efficiently fills 0.1–50 mL liquid at a capacity up to 600–700 units/h.

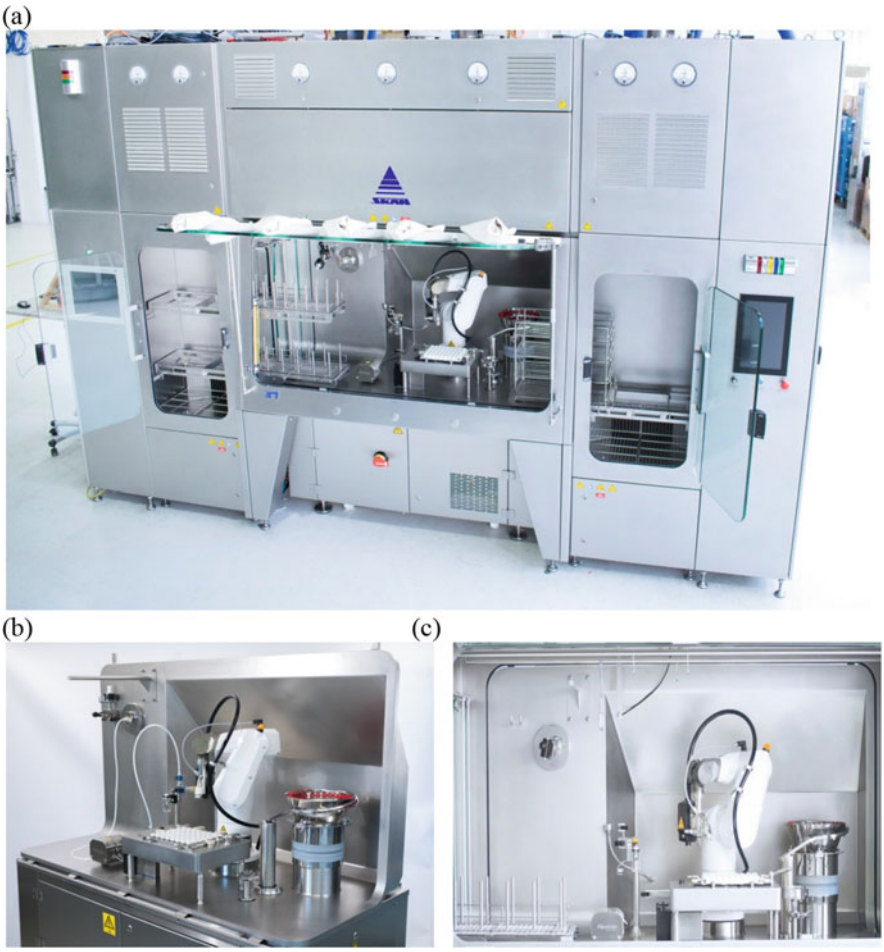


Fig. 12.7 (**a**) Cystal L1Robot Line on L-flange in PSI-L isolator by SKAN. (**b**) Cystal L1 Robot Line. (**c**) Close view of Staübli stericlean TX-40 robot. Reproduced with permission from Add [34, 35]. Source: Aseptic Technologies

Vanrx's robotic aseptic filling workcells utilize pre-sterilized, ready-to-fill nested vials/syringes/cartridges, for robotic manufacturing and incorporating automation in-process controls and environmental monitoring. These workcells are devoid of glove parts and rely on ready-to-use containers and closers avoiding the need of aluminum crimp caps which are a common source of particle contamination in parenterals. Gloveless isolator technology eliminates the need for conveyors, vibratory bowls, star wheels inside workcells thereby minimizing particle contaminations. Further, this technology allows containers and closures to be handled only by the robots. These aseptic work cells fill nested vials, cartridges, cylinders, and close with press-fit vial closures, stoppers, and plungers. Microcell vial filler is developed for

filling vials for personalized medicine and clinical trials. SA25 aseptic filling workcell is a flexible filling system which can fill vials, syringes, and cartridges with added functionality of lyophilization capacity in addition to filling. Roche/Genentech, Singota Solutions, WuXi Biologics, ADMA Biologics, FUJIFILM Diosynth Biotechnologies, Emergent BioSolutions, Amgen, Wildlife Pharmaceuticals are using SA25 aseptic filling workcell for manufacturing of their products. Adaptive Phage Therapeutics has adopted Microcell vial filler for PhageBank products which are collection of viruses to kill the most dangerous antibiotic resistant superbugs.

Bausch + Ströbel KCP series uses clean room robots for filling and closing parenteral products [37]. Laminar air flow remains unaffected with the robotic arms due to their location below the containers. Fully robotic manufacturing ensures minimum human intervention allowing handling of highly potent substances without any hazard to the operator. Vials of 7–52 mm in diameter and height of 30–94.5 mm and syringes of 0.5–20 mL volume can be filled, closed, sealed, and inspected on this line [38]. Vaporized hydrogen peroxide can be used for decontamination of robots and working space on this machine.

12.5.3 Robotic Fill/Finish Systems

ESS Technologies' TaskMate (Model SF20) is equipped with two FANUC LR 200iD clean glass robot and a restricted access barrier system for filling and capping syringes/vials in a sterile environment [38, 39]. Empty syringes/vials are manually fed through a sterile rapid transfer port into the vibratory hopper which gently drops syringes/vials onto a flex feeder with a powered, backlit, vibratory conveyor [40]. First clean class robot uses vision and a custom designed vacuum style end-of-arm tool to pick syringe/vial from a flexible feeding system and transfer it to the filling station. The second robot picks a cap from the tray and places it into the torque station. The syringe/vial is transferred to the torque station from the fill station by a servo driven pick and place unit. Filled and capped syringes descend a divided gravity chute to a discharge bin while faulty syringes are discharged to the opposite side of the chute for inspection. Filling volume ranges from 50 μL to 60 mL at speeds up to 20 vials per min or 15 syringes per min. Changeover generally can be accomplished in 10–15 min.

AqVida and Steriline co-designed RVFCM50, a robotic isolator filling and capping machine, is capable to handle vials of 2–100 mL at a speed up to 50 vials/min [39]. Robotic arms move below the height of the filling needles to avoid disruption of airflow. This filling/capping machine meets ISO 14644-1 guidelines as it has only one sample port for this compact unit which is less than 2 m^3. Glove ports are located on either side of the machine to allow for complete access to the line. This machine is custom designed to meet the requirements of sterile parenteral filling of oncology products of AqVida. With this machine AqVida won the 2018 CPhI Excellence in Pharma Award for Manufacturing Technology and Equipment [41].

AST's (Automated Systems of Tacoma) GENiSYS R is a flexible modular system designed to automate aseptic filling and closing of ready-to-use, nested, presterilized vials, syringes, and cartridges in strict accordance with cGMP requirements [42]. This robotic fill finish system can be integrated with either RABS or isolator to ensure aseptic conditions as required in ISO 7 and ISO 5 environments. AST uses TX2-40 and TX2-60 models of Stäubli stericlean robots for handling critical steps in each stage of the aseptic processes, viz. bag and tub opening, filling, stoppering, sealing/closing, and reject handling [43]. The GENiSYS R is optimized for small batch filling for clinical and commercial applications in compounding, drug manufacturing, personalized medicine, cell and gene therapy development, cytotoxic drug processing, and other areas. Vials of size 2–50 mL, syringes of size 0.5–50 mL, and cartridge sizes of 1–20 mL can be filled at 20 units per min.

AST's ASEPTiCell® Robotic Production Fill-Finish System has been designed for producing clinical and commercial drug products in vials, syringes, and cartridges at a capacity of up to 100 containers per min [42]. It is equipped with two vision guided Stäubli stericlean robots multiple RABS and aseptic isolators modules for bag and tub opening, automatic filling and closing, lyophilization, and vial capping [44].

12.5.4 Radiopharmaceuticals

Theodorico 2 from Comecer Group is a robotic dispenser for filling, dose calibration, and dispensing of radiopharmaceuticals in open or closed vials [45]. The dispensing chamber is equipped with a robot for the handling of the vials in all stages of dispensing. Lead 75 or 100 mm shielding coupled with robotic automation ensures significant reduction in radiation exposure to operators. An autoclave for final sterilization of filled radiopharmaceutical is also available as an accessory. Integrated sensors and calibrators ensure accurate filling and dosing of radiopharmaceuticals. This robotic dispenser is a good example of application of advanced technologies like robotics, isolators, automatic inspection systems, and automated filling systems brought together for aseptic processing of radiopharmaceuticals [15].

12.5.5 Mobile Autonomous Robots for Cleanroom Monitoring

PM group (www.pmgroup-global.com) in partnership with Novartis, University College Dublin, and Lonza aims to develop autonomous robots for monitoring the environment in cleanrooms [42]. This project is funded by Enterprise Ireland. These autonomous robots may either find their own path or predefined motion paths to locate and expose petri dishes, pick and place sample plates, take surface samples with contact plates, sample viable air, and sample particulate air. Some tasks may require use of vision systems in robots while others may be performed without a

vision system. Lonza's MODA software is employed for workflow, scheduling, and data analysis of the environmental monitoring samples [42].

PM group is also working with Technical University Dublin for developing self-sanitizing air locks for robot's independent transfer of materials in cleanrooms. The project seeks possibilities of using UV light, e-beam, or VHP as a replacement to isopropyl alcohol wipes. In a third project PM group is automating the requirements of quality control laboratories with mobile and static robots [46].

12.6 Laboratory Automation

Robots organize the most routine and repetitive part of the laboratory like centrifuging, aliquoting, and automating routine chemistry, immunoassay, hematology, and urinalysis [47]. The systems are most often enabled by bar codes that tell robotic elements within various instruments what to do. Robotic arm not only can process more tests without fatiguing but also handle tinier amounts of liquid, far less than a person could. With the advancements of robotics and AI, robotic interventions in the laboratory are increasing. AI will allow the robots to start building on this data and running their own experiments, thus further pushing science forward.

Robots entered pharmaceutical laboratories in the early 1980s. The Zymate cylindrical robot competed with Perkin-Elmer's Masterlab which was equipped with a Mitsubishi robot [16]. Waters tablet processing system and Zymark tablet processing workstation II were developed for content uniformity analysis of tablets. Perkin Elmer Masterlab robot arm (with five degree of freedom), Zymate system, Josco Smartarm (SCARA), and Perkin-Elmer Masterlab robot system were some initial robots which offered arm like movements for automation of laboratory activities [48]. CTC Analytics PAL system is a Cartesian robot for automation of simple laboratory activities like mixing, dilution of samples, and filling solution into loop of an autoinjector for analyzing samples [16]. The PreciseFlex 400/300 are laboratory autosampler SCARA robots developed for life science/laboratory automation integrators.

RoboDis II (Erweka) is a robotic dissolution system which offers automation of the entire drug dissolution process from setup to the cleaning process [49]. Automated dissolution testing involves setup of apparatus, media preparation, sampling, filtration, ultraviolet/visible (UV-Vis) spectrophotometer/HPLC analysis of samples, and automated cleaning process. Disso.NET software manages all these steps and controls robot arm for transferring dissolution samples to the analytical instruments. A real-time collection of samples and collection of data allows the software to generate dissolution reports.

12.6.1 YuMi® Robot

YuMi® is a seven-axis collaborative robot with single or double or multiple arms (three or five) in one assembly [50]. This robot can safely work side-by-side with

humans in a laboratory, industry, or other places to get the best productivity. This collaborator robot has arms with grippers that may be custom-designed as per the need of laboratory or workplace. Unlike other robots which are guarded by fences/shields/barriers or fitted inside manufacturing/packing lines/machines, YuMi® is more independent and a step toward a humanoid type robot [51]. YuMi® holds lots of possibilities for adaptation in research laboratories, clinical laboratories, and healthcare settings. At Texas Medical Center campus, in Houston, a mobile YuMi® robot assists medical and laboratory staff with laboratory and logistics tasks in hospitals. Inside laboratories, YuMi® robots can help in test tube filling and centrifuge tending, and an IRB 1200 robot can be employed for pipetting application. In hospitals, YuMi® has the capability to dispense medicines, transport medicines to different places, and bring medical supplies/bed linen to bedrooms of patients. European Institute of Oncology in Milan has integrated YuMi® into its laboratory for collaborative immunoassay testing for viral antigens. Several repetitive operations in the immunoassay process and the washing of well-plates have been successfully assisted by YuMi®. About 77% of the serological testing for Coronavirus has been automated with the support of YuMi® robot and this cooperation helped to analyze up to 450 samples/h at Polytechnic University of Milan [52]. Copan Diagnostics is using a YuMi®-based HEPA filtered biosafety workstation to manage many microbiology processes like processing fiber swabs, blood cultures, tissue, and sterile body fluids [53]. Manually seeded plates are subsequently streaked and placed on a conveyor system for its automatic transfer to incubators.

12.6.2 The Robotic Cloud Lab

Very soon in the near future, more robots can be seen inside research laboratories than humans [51]. Humans will be supporting roles of robots rather than its reverse in traditional laboratories. Researchers will use their desktop computer to have access to the laboratory through a secure portal. Programmable robots will execute the directors of the researcher where robots may get support by a few human helping hands. With these expectations, Strateos is developing robotic remote-access laboratories for synthetic chemistry, preclinical testing, and small molecular drug discovery to cater the demand of the pharmaceutical and life science industries [54]. Fully robotic workcells like TISO, BRAVO, reagent dispensing, and PCR protocol have been developed. TISO workcell includes custom-modified freezer, refrigerator, and incubator units for enabling automated access and retrieval of stored reagents. BRAVO is a multi-channel pipetting robot capable of perdispensing liquids in 96 to 1536 well formats. Multichannel reagent dispenser workcells have been designed for high-throughput plate filling applications. PCR protocol ensures a full array of thermocycling operations applicable for both PCR and fluorescent measurement (RTqPCR) applications.

Scientists at UK's University of Liverpool have designed a robotic lab assistant (Fig. 12.8) for automating time-consuming and tedious laboratory experiments on photocatalysis of water for hydrogen production [55]. Scientists employed a six-axis

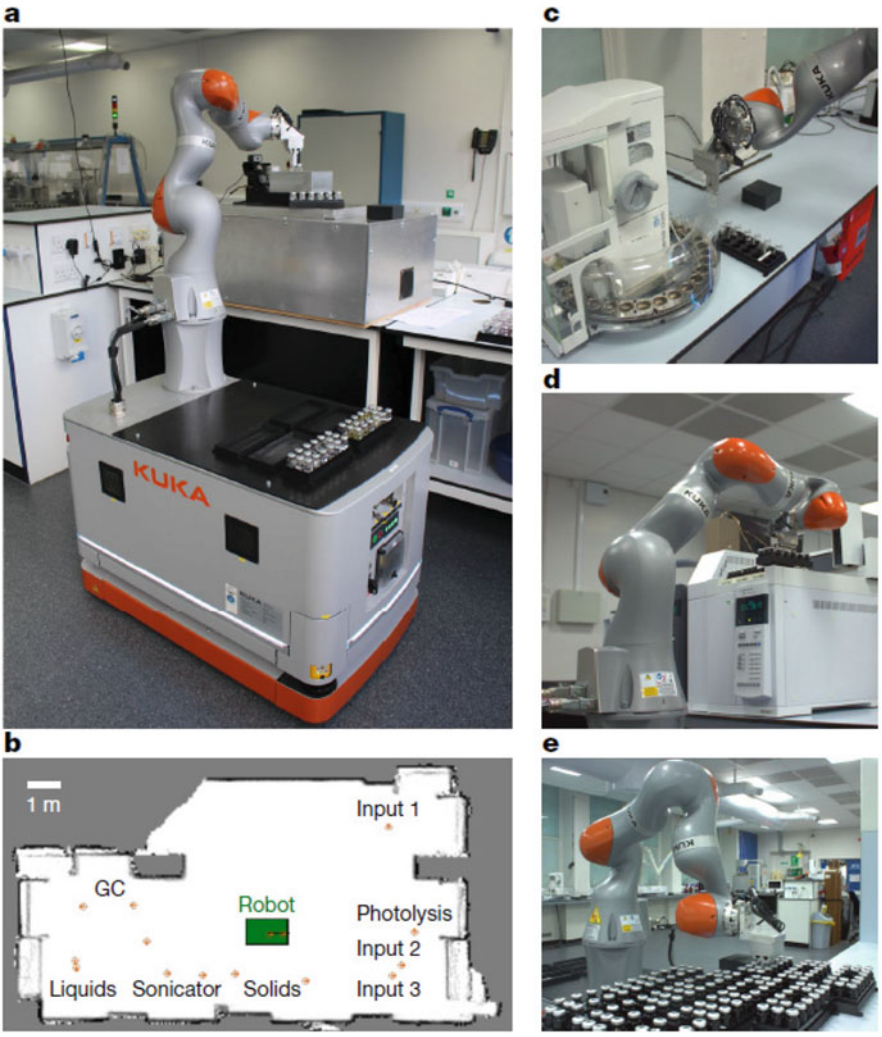


Fig. 12.8 Autonomous mobile robot (KUKA KMR iiwa) assisting in research laboratory. (**a**) Loading samples for analysis. (**b**) Laboratory map showing different stations; green rectangle indicate robot and orange crosshairs indicate recorded navigation locations; inputs 1–3 are storage areas for empty vials or completed sample racks. GC indicates gas chromatography station. (**c**) Robot loading empty sample vials into the solid-dispensing station. (**d**) Robot loading rack of samples in the analyzer (GC). (**e**) Input Station 1: robot placing analyzed samples in storing racks at station 1 [55]. Copyright with Springer Nature

dexterous KUKA mobile robot mounted on a KUKA mobile platform form base for automation of the researcher rather than the equipment. A custom designed gripper for holding 10 mL gas chromatograph sample vials, solid dispensing cartridges, and a 16-position sample rack was fitted to the robotic arm for carrying out tasks required

for the designed workflow. A Bayesian search algorithm was used to design experimental space with 10 variables and 688 experiments in 8 days to perform with the developed robotic lab assistant. Some representative robotic tasks were mixing samples in glass vials, exposing them to light, manipulating other laboratory equipment, and analyzing results using a gas chromatograph. Robotic lab assistant navigates using LIDAR (light detection and ranging) for day and night working in the laboratory. This laboratory robot may operate autonomously 7 days a week, 22 h per day as it requires only 2 h off time for recharging.

12.7 Automation of Pharmaceutical Packaging Operations

Pharmaceutical packaging protects the contents from physico-chemical deterioration and prevents microbial contamination. A good package influences the choice of consumers and assures the quality of a drug product. Labeling of drug products ensures correct identification and provides essential information, viz. composition, expiry date, brand name, etc., to the consumers. Batch number, QR code, and barcode on labels are used for tracking and tracing drug products. Drug products are packed in three different configurations, viz. primary, secondary, and tertiary. Primary packaging, also referred to as retail or consumer packaging, is the packaging which is in direct contact with the product. Primary packaging contains drug product which in addition serves to protect and/or preserve and inform the consumer. Secondary packaging, also referred to as grouped or display packaging, is required for the display of branding and logistical purposes. Primary packs are collated in groups as secondary packages to achieve convenience in storage, handling, and transportation. Secondary units are packed further in tertiary packaging to facilitate the protection, handling, and transportation of secondary packaging during loading as units and their subsequent transportation.

Robots are mostly seen at work in the packaging lines, either alongside operational lines or at the end-of-line. Packaging robots are designed to open, fill, transport, palletize, seal, code, and/or label product packaging. Robotic automation brings the benefit of speed and repeatability in packaging processes [4]. Robots provide accuracy and flexibility. Robots work more efficiently than dedicated packaging machines for some of the packaging applications like carton loading. Robots are flexible to handle a variety of primary packages during carton loading. Robotic loaders and unloaders automatically stop if the product accumulates at the discharge or there is excess of in feed.

12.7.1 Robotic Picking of Container Products

Delta robots and SCARA robots are popular for picking and placing due to their high speed and precise performance. Generally low weight and high speed are the criteria for most of the picking applications in industries. Multiple versions of robots are available for picking applications. Compact and lightweight robots are chosen for

good speed and repeatability. Robots are designed especially for high speed sorting, labeling, packing, and material handling applications. Robots with vision systems or line tracking ensure maximum accuracy in packaging.

12.7.1.1 FANUC Delta Robots for Packaging Operations

FANUC M-1iA delta robot has a unique six-axis parallel link arm design to automate difficult tasks that require extreme precision. This robot can handle very small parts and has a great versatility in industries like medical, food, and pharmaceuticals. FANUC's DR-3*i*B (Fig. 12.9a) series of robots are three-axis delta robots for high speed picking and packaging applications, especially in wash down environments and food handling. The DR-3*i*B Series (Fig. 12.9b) has the classic parallel-link design for longer reach (1600 mm), good inertia, and high payload capacity. Special design features make it easier to clean and sanitize, while meeting even the most strict food safety and drug safety regulations.

12.7.1.2 ABB's IRB 360 (FlexPicker)

ABB's IRB 360 robots, three-axis delta robots also known as the FlexPicker, are designed for quick picking and optimized for packing applications. For food and pharmaceutical industries, stainless washdown (IP69K), and stainless cleanroom (IP54) versions have all stainless steel parts, viz. delta plate, telescopic shaft, arm system parts, and spring [58]. The ball joints are lubricant-free and side entered cables allows easy cleaning with minimum risk of water ingress into the robot. This variant also features integrated vision software, large grippers for efficient handling,

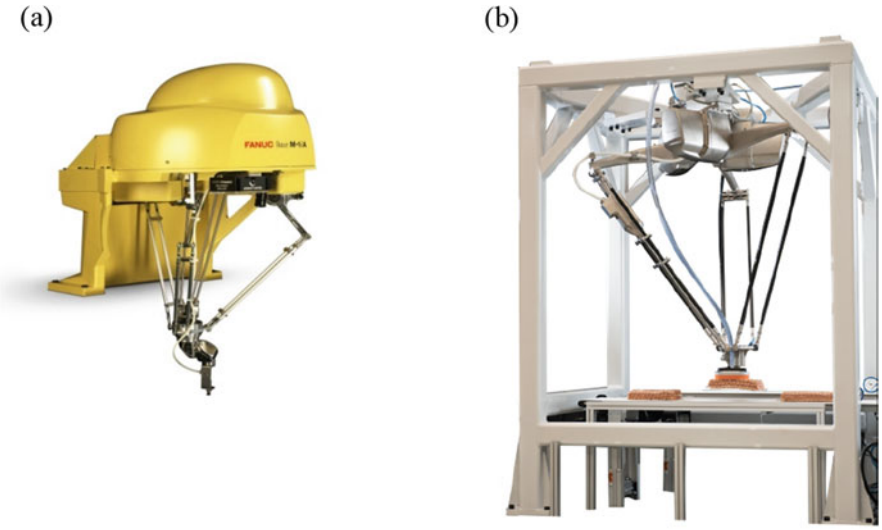


Fig. 12.9 FANUC Delta robots. (**a**) FANUC M-1*i*A-Delta Robot. Reproduced with permission from [56] (**b**) FANUC DR-3iB Food Grade Delta Robot. Reproduced with permission from [57] Source: FANUC

and integrated control of indexing belts [59]. The IRB 360 robots have payload capacity up to 8 kg and a speed up to 200 picks per min.

12.7.1.3 IRB 390 FlexPacker

This four- or five-axis delta robot is the latest offering of ABB's. IRB 390 has a payload capacity up to 15 kg, 35% faster than the IRB 360-8/1130 and 45% more reachable volume. This version has been designed for customers in pharmaceutical, foods, logistics, and consumer-packaged goods industries. IRB 390 suits the requirements for secondary packaging like shelf ready packaging and retail ready packaging. FDA compliance materials for constructing these robots and use of NSF H1 food grade lubricants make them suitable for food and pharmaceutical industries. FlexPacker features PickMaster Twin software for vision-guided random flow picking and packing applications [60]. The software supports up to ten cameras for guiding accurate position and inspection through a powerful color vision system. This robot supports customized packaging, vertical packing, and high-speed, high-variation sorting and on-demand picking in logistics and e-commerce fulfillment centers. IRB 390 FlexPacker and IRB 360 FlexPicker in combination can provide a complete solution to pre-sorting and group packing requirements of industries [61].

12.7.2 Robotic Palletizing

Palletizing refers to loading an object such as a carton on a pallet in a defined pattern. A robotic palletizer provides flexibility, speed, accuracy, reliability in loading and unloading of objects with minimum time spent on teaching and training. End of arm tooling (EOAT) advancements have improved robot palletizing tasks remarkably. Robotic solutions for palletizing are revolutionizing the material handling with layer forming and inline palletizing, mixed case palletizing, and layer depalletizing and palletizing each have their own specific benefits [62].

Robots are extensively employed for picking and placing the container packs in cases and cases are packed into pallets. Robotic end of arm tools (EOATs) are designed in a manner to pick a wide range of primary, secondary, and tertiary packaging. Vacuum, clamp, spatula, full layer sweep are some EOAT's. Vacuum EOATs are most common and efficient tool for picking cases and cartons. Clamp tools are second most popular EOATs which can offer great flexibility and speed in packaging. Spatula provides a great support to the product and hence a good choice for packages which are not rigid. Full layer sweepers are the largest EOATs which carry heaviest loads but albeit at lower speeds.

12.7.2.1 SERPA Pelletizers Equipped with FANUC Robots

FANUC four-axis (SCARA) electro servo driven robots with vision guidance/inspection systems have been equipped in automatic, single cell, and double cell pelletizers by Sherpa Packaging [63]. These palletizers are characterized with compact footprint, integrated controller, and web-based software tools for monitoring, diagnosis, and remote connectivity. Pelletizing robots may use end of arm tools

like grippers and clamps for handling cases, heavy duty claw for palletizing bags, and vacuum grippers that can handle multiple products at once. Fanuc robots have been integrated into Serpa Packaging Systems as vertical cartoning system for pharmaceutical vials [64].

12.7.2.2 AstraZeneca Case Study

AstraZeneca has installed several FlexCell systems (Flexpack Robotics) equipped with ABB's IRB 140 and ABB IRB 4400 robots in their plants at Umeå, Gärtuna, and Snäckviken in Sweden [65]. IRB 140 and IRB 4400 are six-axis industrial robots with payload capacities of 6 kg and 60 kg, respectively. In FlexCell systems, IRB 140 robot uses a custom designed gripper having suction pads for holding shrink wrapped packages and places these packages in a cardboard box, 16 at a time until it is filled with 160 in a cardboard box. A different machine in the cell is used to tape the boxes shut. The second robot of FlexCell system IRB 4400 robot erects the cardboard boxes, lifts with suction pads and brought it close to an inkjet printer for printing product information, barcode, and batch number. A barcode reader checks the barcode. Subsequently, ABB IRB 4400 robot palletizes the checked boxes in pallet areas.

12.7.2.3 URIO: A Collaborative Robotic Pelletizer

Smart Robotics' URIO Palletizer is a cobot integrated palletizing system. The URIO is also TÜV and CE certified for safety and complies with ISO norms. Being a cobot, it is extremely user friendly, cooperative, and employees on the floor can work easily with the system. It requires an area (footprint) of only 3 m^2 for efficient handling of a wide variety of boxes on pallets of different sizes compatible with fork, vacuum, customized 3D printed grippers [66]. The vacuum gripper is suitable for picking one or multiple boxes up to 8 kg from the top. Fork gripper is used for handling open boxes and boxes with a loose lid up to 6 kg. Tailored gripper designs are also available. This smart palletizer can stack up boxes up to a height of 230 cm; however, this depends on the stacking pattern and the product. Bayer, a pharmaceuticals company, has installed one such cobot integrated smart palletizer at its Grenzach plant in Germany [67].

12.8 Warehouse Automation

Autonomous mobile robots (AMRs) with their dynamic and flexible technologies enable pharmaceutical companies to automate their internal material transportation in a simple and easy way. Most modern AMRs are equipped with cameras, sensors, and software for safe and obstacle free navigation of AMRs to their destination [7]. AMRs can identify its surroundings and take the most efficient route to its destination, safely avoiding obstacles and people, which allow the robots to work safely next to humans. AMRs can be connected to a network via Wi-Fi which allows easy access to the user friendly interface from any mobile, tablet, or computer device. Areas like warehousing, production, and packaging are dynamic and agile

due to continuous transport of raw materials section to production area, production area to packaging section, finished goods to warehouse, etc. [68].

12.8.1 MiR200

MiR200 robot (Mobile Industrial Robots, Denmark) transports materials up to 200 kg at a maximum speed of 1.1 m/s between the warehouse and the production under clean room conditions [69]. Customized top modules like racks, bins, lifts, conveyors, or even a collaborative robot arm may be mounted on this robot as per the demand of the application. Wide range of customized top modules offers great flexibility for deployment and redeployment of the robot for different tasks [70]. The MiR200 robot can safely move around obstacles and people, in and out of elevators and through doorways. CAD files of the working area can be used in a simple web-based interface to program this robot. Argon Medical Devices implemented a MiR200 AMR which can be controlled via a mobile phone to transport instruments from the cleanroom to the warehouse [70]. With this implementation, valuable time of employees is saved as the employees from the cleanroom environment are no longer needed to un-gown each time for transporting something.

12.8.2 Mini™

Mini™ (Bluebotics, a Switzerland-based company) is a compact, low particle emission AMR platform which is driven by Autonomous Navigation Technology (ANT®) for seamless deployment and functioning of multiple units to function in the same area [71]. Mini™ is ISO5 cleanroom certified and equipped with automatic charging and autonomous handling of obstacles. It can carry a payload of 150 kg and move at a maximum speed of 1.5 m/s. The mini™ UVC uses UV C light for disinfecting large space on time, every time. A small fleet of Stöcklin EAGLE-ANT AGVs transport materials around the pharmaceutical facility, including quality control samples [72].

12.8.3 Omron LD/HD Mobile Robots

The Omron LD/HD Mobile Robots are a self-navigating Autonomous Mobile Robots (AMRs) for manufacturing, cleanrooms, warehousing, and laboratories [73]. Omron AMR speed ranges from 0.9 to 1.8 mm/s and payload capacity are 60–130 Kg (LD Series) and 1500 Kg from HD series. An intelligent fleet management system is named Fleet Operations Workspace (FLOW) which keeps a complete control and improves productivity, throughput, and traceability. Flow monitors mobile locations of the robot, guides movements of the fleet, and takes tract of job requests to ensure peak efficiency.

12.8.4 Yujin YRL3 Series

Yujin YRL3 series AMR are equipped with light detection and ranging (LiDAR) for detection of objects, navigation, and simultaneous localization and mapping of surrounding area up to 20 m [74]. The YRL3 employs precise laser sensors for 270° horizontal and 90° vertical scanning to detect objects and surrounding environment. LiDAR's fundamental principle is based on direct ToF (Time of Flight) and designed to collect and measure Cartesian coordinates (x, y, z) of objects enabling identification of ranges, angles, and intensities. Yujin YRL series has IP67 rating and products are classified on the basis of measurement distance (5, 10, 20 m).

12.8.5 Squid

BionicHive, an Israel-based company, is developing Squid as a warehouse logistics robot [73, 75]. Squid works synchronously with other fleet members and possesses three dimensional movement capabilities inside warehouse. This robot can pick boxes weighing 35 pound or less from any location from floor to ceiling (20 m height) and thereby automate package/case handling from receiving to shipping. Squid comes with RTTM (real-time traffic management) core that lays a linear programming-based model fed to a solver. The solver is supported with a heuristic algorithm and optimizes pick and place jobs through a self-learning controller.

12.9 Conclusion

Pharma 4.0/Industry 4.0 sees robots as workforce 4.0. Most of the Cartesian and SCARA robots have been aggressively replaced with more sophisticated six and seven-axis articulated robots. For high speed automation of repetitive tasks in manufacturing, industries rely on standalone robots fitted in their manufacturing line while end-of-line installed robots pack cases and palletize them for logistic/warehousing. Inside warehouse AMRs and AGVs are reducing the role of humans in picking, placing, transporting, and storage of cases/boxes. Cleanroom robots installed within isolators are minimizing human interventions in manufacturing parenterals, cytotoxics, and radiopharmaceuticals from small scale to industrial scale. The use of collaborative robots inside research laboratories is trending day by day, and the day is not far when we shall observe more robots than humans in laboratories and humans will be simply assisting robots in performing high end tasks. Now innovators are planning for research laboratories on cloud. Robotic systems are becoming central to aseptic processing and R&D productivity increases. Collaborative double arm movable robots with vision and AI are assisting humans in laboratories and pharmacies and supplying medicines, foods, and daily necessities to humans in hospitals. Collaborative robots equipped with artificial intelligence have progressed enough to be developed as humanoids. AI humanoids like Sophia and Shibuya Mirai are getting citizenship in countries like Saudi Arabia and Japan,

respectively. Advancements in software, vision and sensor technologies, and AI are making robots smarter day by day.

12.10 Credible Online Resources for Further Reading

URL	Description	What to refer?
https://ifr.org/	International Federation of Robotics	News, standards, case studies, position papers, international symposium
https://www.robotics.org/	Robotic Industries Association (USA)	Industry insights, news, articles for education
https://www.jara.jp/	Japan Robot Association	Strategies, members, news, reports, white papers, etc.
https://www.eu-nited.net/eunited+aisbl/robotics/eunited-robotics-european-robotics-association-.html	euRobotics AISBL Brussels-based international non-profit association for all stakeholders in European robotics	Strategies, members, reports, white papers, etc.
http://www.past.kros.org/	Korea Robotics Society	News, members, reports, white papers, journals, etc.
https://schunk.com/	Schunk Company (Germany): automation equipment supplier	Clamping, gripping and other tools for robots
https://www.unigripper.com/	Tepro Machine Pac and System AB company (Sweden): gripper supplier	Grippers for automatic handling
https://www.festo.com/	Festo, a Germany-based company: automation equipment suppliers	Gripper, handling systems, vacuum handling; automation of tablet, parenteral and cream production
https://www.rnaautomation.com/; https://www.rna.de/	Rhein-Nadel Automation (Germany): automation equipment suppliers	Customized handling, assembling and inspection systems
https://www.cognex.com/	Cognex Company (United States): automation equipment suppliers	Automating primary and secondary packaging of vials, vaccine; tablet and pill inspection and barcode system

References

1. ROBOT | meaning in the Cambridge English Dictionary. https://dictionary.cambridge.org/dictionary/english/robot. Accessed 27 Mar 2021
2. ISO 8373:2012(en), Robots and robotic devices—Vocabulary. https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en. Accessed 27 Mar 2021

3. Binggeli L, Heesakkers H, Wölbeling C, Zimmer T. Pharma 4.0™: hype or reality? | pharmaceutical engineering. In: Pharm Eng. https://ispe.org/pharmaceutical-engineering/july-august-2018/pharma-40tm-hype-or-reality. Accessed 6 Apr 2021
4. Machine-centric robotics revitalising pharma packaging. https://www.industr.com/en/machine-centric-robotics-revitalising-pharma-packaging-2474155. Accessed 27 Mar 2021
5. Wilson M. Implementation of robot systems. Butterworth-Heinemann
6. Industrial robotics: insights into the sector's future growth dynamics. https://www.mckinsey.com/~/media/McKinsey/Industries/AdvancedElectronics/OurInsights/Growthdynamicsinindustrialrobotics/Industrial-robotics-Insights-into-the-sectors-future-growth-dynamics.ashx. Accessed 27 Mar 2021
7. Markarian J. Using robotics in pharmaceutical manufacturing. In: Pharm Technol https://www.pharmtech.com/view/using-robotics-pharmaceutical-manufacturing. Accessed 28 Mar 2021
8. Selecting robots for use in pharmaceutical manufacturing. https://www.revbase.com/tt/sl.ashx?z=78fb9fde&DataID=3902998&ft=1
9. What is "Robot"? | Robotics | Products & Solution | Global Site. https://www.yaskawa-global.com/product/robotics/about. Accessed 27 Mar 2021
10. Eitel L. Designing (seventh-axis) linear motion tracks for robotic positioning. In: Linear Motion Tips. https://www.linearmotiontips.com/designing-linear-motion-tracks-robotic-positioning/. Accessed 7 Apr 2021
11. Merkusheva D. 10 humanoid robots of 2020—ASME. In: ASME Top. Resour. https://www.asme.org/topics-resources/content/10-humanoid-robots-of-2020. Accessed 7 Apr 2021
12. Alemzadeh K, Jones SB, Davies M, West N (2021) Development of a chewing robot with built-in humanoid jaws to simulate mastication to quantify Robotic agents release from chewing gums compared to human participants. IEEE Trans Biomed Eng 68:492–504
13. What's the difference between industrial robots? | Machine design. https://www.machinedesign.com/markets/robotics/article/21835000/whats-the-difference-between-industrial-robots. Accessed 7 May 2021
14. Ataei MM, Salarieh H, Alasty A. An adaptive impedance control algorithm; application in exoskeleton robot
15. Izumi Y, Akers J. The impact of automation on aseptic processing. In: Pharm Technol. https://www.pharmtech.com/view/impact-automation-aseptic-processing. Accessed 27 Mar 2021
16. North N. Robotics and laboratory automation in pharmaceuticals analysis. Encycl Anal Chem. https://doi.org/10.1002/9780470027318.a1919
17. FabRx's pharmaceutical 3D printer for personalised medicines, M3DIMAKER™, is now available!—FabRx. https://www.fabrx.co.uk/2020/04/06/fabrxs-pharmaceutical-3d-printer-for-personalised-medicines-m3dimaker-is-now-available/. Accessed 7 Apr 2021
18. INNOVATIONS IN 3D PRINTED PHARMACEUTICALS—ONdrugDelivery. https://www.ondrugdelivery.com/innovations-in-3d-printed-pharmaceuticals/. Accessed 27 Mar 2021
19. Aprecia Preps to 3D print medicine with new facility. https://3dprintingindustry.com/news/aprecia-pharmaceuticals-prepares-to-produce-3d-printed-medicines-with-new-facility-43568/. Accessed 27 Mar 2021
20. MED™ 3D printed pharmaceutical product receives IND clearance from the US FDA | Business Wire. https://www.businesswire.com/news/home/20210209005529/en/MED™-3D-Printed-Pharmaceutical-Product-Receives-IND-Clearance-From-the-US-FDA. Accessed 27 Mar 2021
21. Lights-out pharma factory: why the future of pharma production is robotic | by Zack Bright | Multiply Labs | Medium. https://medium.com/multiply-labs/lights-out-pharma-factory-why-the-future-of-pharma-production-is-robotic-ee1c5297bb8a. Accessed 27 Mar 2021
22. Argos Therapeutics | BioPharm. https://www.biopharm.saint-gobain.com/news/argos-therapeutics. Accessed 27 Mar 2021
23. Tamura T, Kurebayashi H, Tanaka Y, Sakakibara S, Nihei R, Inaba Y. High speed intelligent handling robot for food and pharmaceutical products. In: Proc. IEEE Work. Adv. Robot. Its Soc. Impacts, ARSO. pp 60–64

24. Coulon C-H (2020) Robotics of the future in the pharmaceutical industry. TechnoPharm 10: 150–157
25. 4-axis SCARA TS2-60 robot for all environments/Stäubli. https://www.staubli.com/en-in/robotics/product-range/industrial-robots/4-axis-scara-robots/ts2-60/. Accessed 7 May 2021
26. 6 axis industrial robots TX2-60 | Stäubli. https://www.staubli.com/en-in/robotics/product-range/industrial-robots/6-axis-robots/tx2-60/. Accessed 7 May 2021
27. FANUC SR-12iA—high payload and small footprint. https://www.fanuc.eu/de/en/robots/robot-filter-page/scara-series/scara-sr-12ia. Accessed 7 May 2021
28. FANUC LRMate 200iD/7LC industrial Robot. https://www.fanuc.eu/hu/en/robots/robot-filter-page/lrmate-series/lrmate-200id-7lc. Accessed 7 May 2021
29. Bassler H-J, Lehmann F (2020) Robots in aseptic isolators for pharmaceutical applications. TechnoPharm 10:254–260
30. Bassler H-J, Haefele F, Lehma F (2017) A flexible and scalable isolator platform for state of the art aseptic processing. Pharm Ind 79:725–730
31. VarioSys—flexible solutions for pharmaceutical processing. https://www.variosys.com/en. Accessed 7 Apr 2021
32. Nest processing | VarioSys. https://www.variosys.com/en/lines/nest-processing. Accessed 27 Mar 2021
33. Flexible solutions for pharmaceutical processing
34. Products | Aseptic Technologies—Safer & Easier Aseptic filling operations. https://www.aseptictech.com/products#crystal-filling-lines. Accessed 7 May 2021
35. Crystal L1 Robot Line. Aseptic Technologies. https://www.aseptictech.com/sites/default/files/aseptic_technologies_crystal_l1_datasheet_0.pdf
36. Crystal ® SL1 Robot Line Versatile Equipment Filling process
37. KCP series | Bausch + Ströbel. https://www.bausch-stroebel.com/en/machines/kcp-series. Accessed 27 Mar 2021
38. Forcinio H. Robots package parenteral products. In: Pharm Technol. https://www.pharmtech.com/view/robots-package-parenteral-products-1. Accessed 7 Apr 2021
39. Sookne K. Robotic arms make the difference in oncology vial filler and capper | healthcare packaging. In: Healthc. Packag. https://www.healthcarepackaging.com/machinery-materials/packaging-filling/article/13292734/robotic-arms-make-the-difference-in-oncology-vial-filler-and-capper. Accessed 27 Mar 2021
40. ESS Technologies: robotic aseptic syringe filler and capper From: ESS Technologies | Packaging World. https://www.packworld.com/machinery/primary-packaging/product/13372385/ess-technologies-robotic-aseptic-syringe-filler-and-capper. Accessed 27 Mar 2021
41. Robotic automation enhances fill/finish. https://www.pharmtech.com/view/robotic-automation-enhances-fillfinish. Accessed 27 Mar 2021
42. Wolton D. Aseptic manufacturing: the robots are coming. In: Cleanroom Technol. https://www.cleanroomtechnology.com/news/article_page/Aseptic_manufacturing_The_robots_are_coming/161887. Accessed 7 Apr 2021
43. Robotic precision in small batch aseptic processing | Healthcare packaging. https://www.healthcarepackaging.com/machinery-materials/aseptics-isolators-rabs/article/21199482/robotic-precision-in-small-batch-aseptic-processing#previous-slide. Accessed 27 Mar 2021
44. Russel J. Robotics and isolator-barrier systems: the perfect combination for advanced aseptic processing. In: Innov. Pharm. Technol. http://www.iptonline.com/articles/public/AST.pdf. Accessed 7 Apr 2021
45. THEODORICO 2 robotic dispenser for radiopharmaceuticals. https://www.comecer.com/theodorico-2-robotic-dispenser-for-radiopharmaceuticals/. Accessed 27 Mar 2021
46. Autonomous robotics in pharma manufacturing. https://www.pmgroup-global.com/news/pharmaceutical-robotics/. Accessed 7 Apr 2021
47. Robots get ready to roam in clinical labs | AACC.org. https://www.aacc.org/cln/articles/2020/october/robots-get-ready-to-roam-in-clinical-labs. Accessed 27 Mar 2021

48. Lung KR, Lochmuller CH (1986) The anatomy and function of the laboratory robot. J Liq Chromatogr 9:2995–3031
49. Fully automated dissolution system RoboDis II—ERWEKA GmbH. https://www.erweka.com/fully-automated-robodis-ii.html#. Accessed 28 Mar 2021
50. ABB demonstrates concept of mobile laboratory robot for Hospital of the Future. https://new.abb.com/news/detail/37301/abb-demonstrates-concept-of-mobile-laboratory-robot-for-hospital-of-the-future. Accessed 27 Mar 2021
51. Miller JA. Robots get ready to roam in clinical labs | AACC.org. In: AACC.org Clin. Lab News. https://www.aacc.org/cln/articles/2020/october/robots-get-ready-to-roam-in-clinical-labs. Accessed 7 Apr 2021
52. COVID-19: YuMi, a collaborative robot helps analyse a higher number of serological tests | Politecnico di Milano. https://www.mynewsdesk.com/it/politecnico-di-milano/pressreleases/covid-19-yumi-a-collaborative-robot-helps-analyse-a-higher-number-of-serological-tests-3034336. Accessed 27 Mar 2021
53. Copan to showcase robotic solution for clinical microbiology. https://roboticsandautomationnews.com/2018/06/01/copan-to-showcase-robotic-solution-for-clinical-microbiology/17507/. Accessed 27 Mar 2021
54. The Robotic Cloud Lab | Strateos. https://strateos.com/robotic-cloud-lab/. Accessed 27 Mar 2021
55. Burger B, Maffettone PM, Gusev VV et al (2020) A mobile robotic chemist. Nature 583:237–241
56. Delta Robots | FANUC M-1iA Small Delta Robot Series | FANUC America. https://www.fanucamerica.com/products/robots/series/m-1ia-delta-robots. Accessed 7 May 2021
57. Delta Robots | FANUC DR-3iB Food Grade Delta Robot Series | FANUC America. https://www.fanucamerica.com/products/robots/series/dr-3ib-series-delta-robots. Accessed 8 May 2021
58. IRB 360, M2004, Product specification. https://search.abb.com/library/Download.aspx?DocumentID=3HAC029963-001&LanguageCode=en&DocumentPartId=&Action=Launch. Accessed 27 Mar 2021
59. IRB 360 | ABB Robotics—Industrial Robots | ABB Robotics. https://new.abb.com/products/robotics/industrial-robots/irb-360. Accessed 27 Mar 2021
60. ABB to enhance robotic picking and packing portfolio to transform omnichannel order fulfilment and retail logistics. https://new.abb.com/news/detail/62221/robotic-picking-and-packing-portfolio. Accessed 27 Mar 2021
61. Mohan AM. Robotics Special Report: E-Commerce Requires new Solutions | Automation World. https://www.automationworld.com/home/article/21141597/robotics-special-report-ecommerce-requires-new-solutions. Accessed 27 Mar 2021
62. Robotic palletizing. https://www.mhi.org/solutions-community/solutions-guide/robotics. Accessed 27 Mar 2021
63. Single cell robotic palletizer for GMA, CHEP & Euro pallets. https://www.serpapackaging.com/products/robotic-palletizers/single-cell-robotic-palletizer/. Accessed 28 Mar 2021
64. Pharmacy automation systems & robotics for pharmaceutical manufacturing. https://www.fanucamerica.com/solutions/industries/pharmaceutical-manufacturing-automation. Accessed 7 Apr 2021
65. Farnsworth A. The quick cure. www.abb.com/robotics; https://packlines.com/images/EN_14-15_astrazeneca_korr2_abb_pack_108.pdf. Accessed 28 Mar 2021
66. Smart Pelletizer (2021) Product Information Brochure Smart Robotics. http://www.smart-robotics.nl/wp-content/uploads/2021/01/Smart-Palletizer-brochure-EN-2021.pdf. Accessed 28 Mar 2021
67. Bayer case—Smart Robotics—palletizing in the pharma industry. https://www.smart-robotics.nl/case/case-bayer/. Accessed 28 Mar 2021

68. Mobile robots in the pharmaceutical industry | Mobile Industrial Robots. https://www.mobile-industrial-robots.com/en/industry/life-science/mobile-robots-in-the-pharmaceutical-industry/. Accessed 28 Mar 2021
69. Mullen E. Mobile robots in the pharmaceutical industry. In: Pharm. Process. World. https://www.pharmaceuticalprocessingworld.com/mobile-robots-in-the-pharmaceutical-industry/. Accessed 7 Apr 2021
70. MiR200 more powerful autonomous mobile robot | Mobile Industrial Robots. https://www.mobile-industrial-robots.com/en/solutions/robots/mir200/. Accessed 28 Mar 2021
71. BlueBotics mini™—mobile robot platform for cleanrooms. https://bluebotics.com/engineering-services/mini/. Accessed 28 Mar 2021
72. Fighting HAIs & COVID-19 with healthcare automation solutions | BlueBotics. https://bluebotics.com/healthcare-automation-solutions/. Accessed 28 Mar 2021
73. Hart M. Box-picking robots now automating warehouses. In: Nerdist. https://nerdist.com/article/automated-box-picking-robots-warehouses/. Accessed 7 Apr 2021
74. YUJIN LiDAR ⋆ YUJIN ROBOT. https://yujinrobot.com/solutions/lidar/. Accessed 28 Mar 2021
75. BionicHIVE. https://www.bionichive.com/. Accessed 28 Mar 2021

Soft Robots for the Delivery of Drugs

13

Surojit Banerjee and Vikas Anand Saharan

Contents

Abstract

Miniaturized soft robots compliant to human tissues and organs improve efficiency of drug for both spatial and temporal targeting. Origami robots, sperm robots, nanorobots, robotic capsules, and soft multi-legged robots are some soft robots that have been investigated for their potential role in therapeutics and diagnosis. Intelligence, locomotion, and swarming behaviors of these soft robots have been used to actuate, sense, and manipulate delivery of active constituents to the target site. Use of soft and stretchable electronics, power sources, and sensors make them autonomous. Due to their excellent flexibility and adaptability, soft robots can easily navigate in the human body without causing any harm. Soft robots are rapidly progressing with the advancements in soft materials, 3D printing technologies, soft sensors, soft electronics, soft power sources,

S. Banerjee · V. A. Saharan (✉)
Department of Pharmaceutics, School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Dehradun, Uttarakhand, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_13

human-machine interactions, and new algorithms. Significant challenges to developing soft robots include miniaturization, cost intensive technology, and requirement of high level of technical expertise. This book chapter focuses on some drug delivery attempts with soft robots, in particular for targeting drugs to a particular site and sustained delivery of drugs.

Keywords

Drug delivery · Targeting · Controlled release · Locomotion · Torque · Magnetic field · Challenges

Chapter Objectives
Upon reading of this chapter, the reader will be able to understand:

- What are the differences between soft robots and hard (industrial) robots
- Different types of soft robots for drug delivery
- How soft robots are able to navigate and deliver drugs at specific sites in the human body
- Preclinical and clinical trials on soft robots
- Various challenges for the development of soft robots for their applications in drug delivery

13.1 Introduction

Soft robots are engineered mobile machines fabricated from intrinsically compliant soft and bio-friendly materials like soft polymers and elastomers [1, 2]. These robots are inspired from nature and, thereby, also referred to as bioinspired robots. Soft robots can easily operate in air, vacuum, or liquid. Due to use of soft and compliant materials in their structure, these robots are very much flexible, resilient, and can easily adapt to the surrounding environment. Deformable structure of soft robots allows muscle-like activation that mimics biological systems allowing infinite number of degrees of freedom, unlike limited degrees of freedom of hard robots. Movements and locomotion of soft robots and living organisms resemble drastically. Soft robots have been designed to move like caterpillar, worm, snake, octopus, fish, and many other living organisms [1]. Hard robots, contrary to soft robots, are made from hard metals and plastics and possess numerous rigid links and joints. Hard robots are regarded as unsafe in natural environments and in close proximity to humans. Use of soft compliant materials in soft robots considerably reduces chances of inadvertent harm while soft robots work or interact with humans or other biological systems. Comparative features of soft and hard robots are presented in Table 13.1.

The earliest research reports on the development of soft robotics came from McKibben's (1950s) braided pneumatic actuators for polio patients; McKibben's artificial muscle which was employed in different types of robot designs; Shimachi

Table 13.1 Characteristic features of soft robots and hard robots [1, 6]

Characteristic feature	Soft robots	Hard robots
Material	Soft (elastic modulus $\leq$1 GPa), deformable, flexible, and stretchable materials with reversible and variable properties (e.g., silicon rubbers and polymers like polyethylene terephthalate, polyethyleneoxide); a strategic combination of soft and hard materials may also be used	Hard (elastic modulus $\geq$1 GPa) materials with invariable properties (e.g., hard metals like stainless steel and aluminum and hard plastics like acrylonitrile butadiene styrene); hard robots may strategically employ soft components like foot pads, to absorb shock or springy joints to store/release elastic energy
Actuation	Soft fluidic actuation system to generate reaction forces; electrical, thermal, pressure difference, and magnetic force based actuation; actuators integrated and distributed throughout the robot	Hard robots are rigid-linked, with actuators for every joint; hard actuation system; actuation mechanisms are hydraulic, pneumatic, electromagnetic, electromechanical, etc.
Design features	Nature inspired or bioinspired design; appearance similar to insects, octopuses, microorganisms, etc., unprecedented adaptation, sensitivity, and agility	Appearance similar to machines; limited adaptability, sensitivity, and agility
Manufacturing	Shape Deposition Manufacturing (SDM), soft lithography, the Smart Composite Microstructure (SCM) process, 3D multimaterial 3D printing, molding, bonding, etc.	Casting, welding, assembling, and machined (similar to making machines)
Compliance with the environment	Compliance matches with the environment	Smooth contact is facilitated with feedback controlled strategies and sensors
Safety	Comparatively more safe	Unsafe, may cause harm to humans, animals and biological systems
Degree of freedom	Infinite degree of freedom	Finite degree of freedom
Adaptability	Safe, adaptive, and tolerant in unknown environments	Limited adaptability in unknown environments
Electronic, sensors, and power source	Flexible and stretchable electronic, sensor, and power source	Conventional electronic, sensor, and power source
Bioinspiration and diversity	High level of biodiversity and behavioral diversity	Low level of bioinspiration and behavioral diversity
Weight and cost	Low weight and low cost	High weight and cost
Accuracy and speed	Low accuracy and low speed	High accuracy and high speed
Fatigue	Prone to fatigue; fatigue failure	Less prone to fatigue
Applications	Exosuits, biomimicry, robotic muscles, climbing robots, edible, wearable, prosthetic, soft grippers,	Industrial automation of manufacturing, packaging, warehousing, and laboratory

(continued)

Table 13.1 (continued)

Characteristic feature	Soft robots	Hard robots
	soft actuators, rehabilitation, assistance, surgical instruments, drug delivery and diagnosis	research. Rehabilitation, entertainment, elderly care, hospital care, and many more

and Matumoto reports (1990) on soft fingers; and Suzumori and colleagues (1991) who developed flexible microactuator from silicone rubber and studied several applications [3]. Some other key developments in the area of soft robotics are electrostrictive polymer artificial muscle actuators, pneumatic bellows actuators, pneumatic rotary soft actuator, flexible pneumatic actuator, fluidic muscle, rubbertuator, flexible fluidic actuator, tentacle manipulator, elephant trunk manipulator, OctArm, Air-Octor, caterpillar robot, continuum manipulators, Clobot, and many more [3]. An octopus inspired soft robot, the Octobot (2016), was the first entirely soft, autonomous robot developed by scientists at Harvard University [4]. Octobot was fabricated with a combination of embedded 3D printing, molding, and soft lithography. Octobot was powered by a chemical reaction and controlled with a microfluidic logic that autonomously regulated fluid flow. A catalytic decomposition of an on-board monopropellant fuel supply generated gas for inflating fluidic networks downstream resulting in actuation. With time soft robots have gained abilities like jumping, peristaltic locomotion, locomotion through small crevices, arm elongation and shortening, underwater legged locomotion, pulsed-jet swimming, stretchability, climbing, bending, stiffening, adaptable grasping, morphing, self-healing, biodegradability, edibility, and resilience [5]. Significant developments have also been observed in drug delivery attempts with soft robots like DNA nanorobots, origami robots, sperm robots, and untethered magnetic millirobot. This chapter describes these soft robotic developments for site-specific delivery of drugs in a controlled manner.

13.2 Robotics in Drug Delivery

In the 1990s, robotic technologies developed enough for conceptualization of delivery of drugs in the human body [1, 2]. It was first observed in the year of 2000 that the robotic micro-systems can reach to the distant areas of the gastrointestinal tract [7–10]. From there the concept of robotic capsules originated. Since then, pharmaceutical industries and research institutions have invested a lot for soft robotic developments in drug delivery and health care. That financial impetus had provided great opportunities for developing safe robotic drug delivery platforms with potential therapeutic efficacy albeit at a cheap price due to advancements in soft computers, soft fluidic actuators, soft material 3D printing, sensors, vision systems, and nanotechnologies. The outcome came in the form of DNA nanorobots for the treatment of cancer [11, 12] and origami robots, sperm robots, and untethered

magnetic millirobot [13–17]. Soft robotics are inherently safe and offer a great control on movements within the human body [11–15]. In drug delivery approach through robots, biocompatibility and biodegradable nature of soft material is an essential prerequisite for safe and harmless therapeutic outcomes. Soft robotic approaches hold great potential for cancer therapy and as a personalized device for treatment of diseases like atherosclerosis, blood clots, etc. [11, 18, 19]. These soft robots includes origami robots, sperm robots, soft multi-legged robots, robotic patch, robotic capsule, and nanorobots [20–22].

13.2.1 Origami Robots

The term "origami" came from two words, that is, "ori" and "kami," in which the former means "folding" and the later means "paper" [23]. Origami robots are popular in the field of engineering for making objects that have complex 3D structures and suitable mechanical properties [13, 14]. Proteins function properly when they have their proper folding structure of amino acids [24]. With these concepts scientists have developed soft robots for delivery of drugs.

Origami robots consist of a flat composite sheet of cylindrical structure, that is, tiles, made of different materials like rice paper, sausage casing, etc., which are arranged in an array structure with capability of self-assembly and change of shape through folding. Tilings are contained in the sheets, which are separated by compliant joints, forming a furrow structure [25]. Among these only some definite joints allow folding while rest of the part remains inflexible. Here the term "kirigami" must be mentioned [26, 27]. It actually refers to the subclass of these origami structures, which enables the internal cuttings in the sheets [25–27]. For appropriate furrow structure it requires a suitable 3D shape. To fulfill this purpose sheets have to withstand some steps of folding sequentially and a proper plan is required for the folding of definite edge(s) of sheets at the desired angle. Apart from these, computational work and power are needed in the whole sheet along with the sensation and actuation. After preparing it, for obtaining the best result, higher level control algorithms and planning are required along with the control of lower level for the locomotion. The folding approaches need the proper design and modeling of it, which allow the degree of freedom and support the mechanism of origami. In 1996, the algorithmic strategy for developing the origami function was first introduced [28]. The algorithmic strategy provided a universal solution to the problems of folding, but it also had a demerit of material wasting. Despite this demerit, it produces many layers of folding. There is another practical algorithm, alternative to the former approach, which has universality and also guarantees the folding of each convex face flawlessly. It can be started as one convex polygon, which is unfolded. As of now fabricating origami robots requires the use of software like popupCAD which rationalizes the laminate composites for design aspects such as flexibility and structural properties [25, 29–31]. The furrow patterns for a 3D shape design are automatically engendered by origamizer. Different stages of 3D origami structure fabrication are illustrated in Fig. 13.1 [23]. Beside this software like Merlin

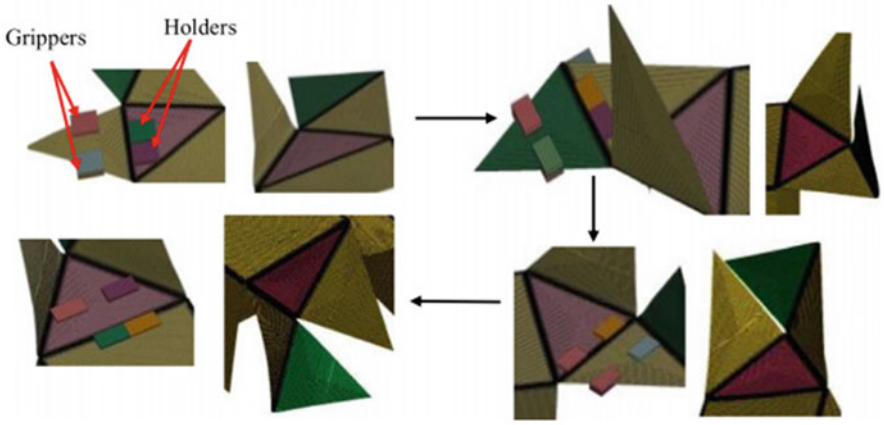


Fig. 13.1 Stages of 3D origami structure preparation. (Reproduced from Thai et al., 2018) [23]

may provide nonlinear mechanics to the origami structure and a robot compiler is used to provide accurate folding based on their functionality or their specifications along with the print design.

By using this design the MIT researchers, the University of Sheffield and the Tokyo Institute of technology presented their work at the International Conference on Robotics and Automation [32]. They have developed the design of it into a capsule structure and tested it in a pig stomach. Biodegradable and biocompatible materials, like titanium, plastic, paper or ingestible animal by-products (sausage casing), ceramic perovskite, polystyrene, polyolefin, polyvinyl chloride, polypyrrole films, etc., were used for robot modeling and an assembly having size in millimeter.

Shuhei and colleagues (2014) had investigated the mechanism of soft robot for delivering drugs into the human body. The developed soft robot was ingested in the stomach of human volunteers [32–34]. Robot successfully delivered the drug through the esophagus and reached to the site of action or the wounded area and unfolded itself from the swallowed capsule to release the drug kept, in one/more folds, at the desired site. For the bigger size, it cannot cross the pyloric sphincter, so it remained in the stomach and releases the drug for prolonged periods [35].

13.2.2 Sperm Robots

Sperm move toward the ovum and release the genetic materials after attaching with it. Using this concept scientists have recently developed a device for drug delivery, which consists of sperm and the drug contained it [36]. In vitro study has confirmed the efficacy of sperm robots for delivering doxorubicin HCl for the treatment of gynecological cancer [37].

Sperm robots consist of a motile sperm cell and a tetrapod which is a magnetic tubular microstructure (3D printed) with a drug contained in it [38]. Tetrapod was

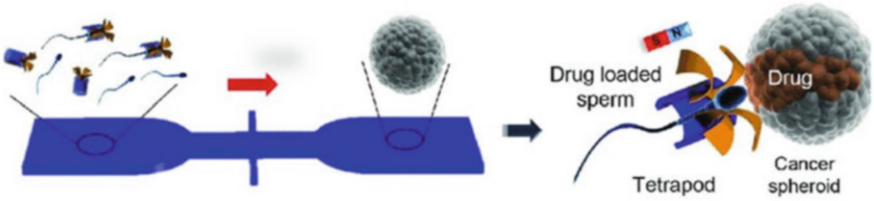


Fig. 13.2 Sperm robot with microfluidic structure delivering drug to a cancer cell. (Reproduced from Xu et al., 2018) [36]

coated with iron oxide (10 nm) with 15° tilt angle and casings were made with tubular bodies with four arms. This tetrapod was coated with a magnetic layer for controlling this device from outside the body. An external magnet was placed within a certain distance from the sperm sample (around 10 cm) [39, 40]. For this kind of design sperm robot was called the micromotor. The maximum distance between the narrowest points of four arms was kept at 4.3 μm. The dimensions were maintained according to the sperm structure. The head of the sperm robot was 1 μm thick, 4.5 μm wide, and 10 μm long [41]. Apart from these, for improving the biocompatibility of composites a supplementary layer made of titanium (2 nm) was placed between the photoresist (light-sensitive polymer like methyl methacrylate) and iron oxide. More than 50 tetrapods were connected to the sperm. Researchers had calculated the force for the movement of the device. With an applied force of 128 pN to one arm the movement was 116 nm which resulted in an adequate displacement of sperm and a force of 450 pN the movement achieved was 407 nm [36]. The in vitro release of drug through sperm robot against the cancer cell was tested in a microfluidic channel of polydimethylsiloxane [42].

In the in vitro study, the device with tetrapod moved toward the spheroids of cancer cells when magnetic force was applied in a controlled manner [36, 43]. When micromotor arrived at the spheroids, the sperm was released by a mechanical trigger and attached itself with the cells leading to release of the drug. Drug release was controlled through a polydimethylsiloxane chip. When two arms of the tetrapods knocked at the corners, the rotation of arms stopped and the sperm came out for releasing the drug within 7 s. When four arms of tetrapods knocked at the wall, the rotation of the micromotor did not stop for which the sperm came out in 12 s [36]. In various in vitro studies sperm robots effectively killed the cancer cells by 90% within 72 h [40, 43–45]. This structure of sperm robot is illustrated in Fig. 13.2 [36].

13.2.3 Soft Multi-legged Robots

Soft materials, like polyurethane, polyethylene terephthalate, polyethyleneoxide, etc., are used to fabricate multi-legged robots [15]. These robots are characterized by multiple number of legs which provide high degrees of freedom for movement and their ability to cross barriers [46, 47]. Such robot designs have allowed excellent

flexibility to various jarring environments with ultrafast locomotion speed (>40 limb length/s), ultra-strong carrying capacity (>100 own weight), and excellent obstacle-crossing ability (stand up 90° and across obstacle >10 body height).

The soft robot can be prepared by various ways, among which magnetic particles assisted modeling approach is very much noteworthy [48–50]. Soft multi-legged robot consists of solid structures made from organic polymer like polymeric organosilicon compounds (e.g., polydimethylsiloxane) and magnetic particles [15, 48]. The magnetic field activated externally interacts with the dipole moment of magnet present in soft multi-legged robots. Thereby, a magnetic torque is generated and the drug is released with the deformation of the soft materials. These robots have length less than 20 mm, 15 mm width, and 180 mm thickness [51]. Young's modulus of these soft robots should range from 10^3 Pa to 10^7 Pa [52, 53]. A low ratio of limb length to optimal use of supporting power of legs promotes spatial and temporal targeting ability of soft robotic devices. There are many multi-legged animals, which have low limb ratio and optimal power of support for good locomotion [54, 55]. Multi-legged robots which utilize a low ratio of limb length and optimal supporting power have been designed for delivering drugs. In one such study, 1:1 ratio of length of feet and supporting power is suggested for good locomotion for such robots [54].

The robot's locomotion is controlled through a magnetic field of a strong magnet [48, 56]. Feet of robot are aligned toward a specific site with the use of magnetic field which also creates the magnetic torque and force of pulling. This pulling force moves the robot toward the specific site for releasing the drug [57, 58]. The friction between legs and the ground is minimized affecting small contact time with the ground leading to negligible loss of structural flexibility. It is reported that this device is capable to crossing almost any obstacle [59–61]. A typical structure of the soft multi-legged robot is illustrated in Fig. 13.3 [55].

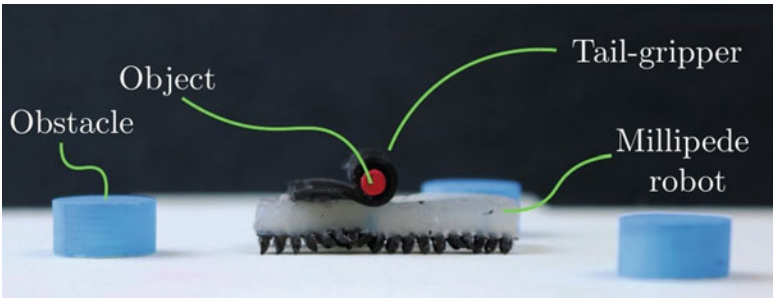


Fig. 13.3 Typical structure of a soft multi-legged robot. (Reproduced from Venkiteswaran et al. 2020) [55] (CC BY 4.0)

13.2.4 Robotic Capsules

Drug delivery through robotic capsules is a growing approach among all the approaches. Researchers have started exploring the design space of Medical Capsule Robots (MCRs) as embedded micro-systems that can operate autonomously within the human body and can diagnose, prevent, monitor, and cure diseases [62, 63]. Nowadays, wireless robotic approach for drug delivery is also gaining immense popularity among researchers. Unfortunately, this approach is not fruitful for other parts of the gastrointestinal tract, where more deadly forms of cancer usually develop like colon and the stomach [64, 65]. The main reason for this is related to the lack of advanced functionalities such as active locomotion, advanced diagnosis and tissue manipulation, biopsy sampling, or drug delivery.

The design process of MCRs is very grueling because it has to address severe cross-cutting impediments such as size (to gain non-invasive access, MCR diameter is limited to about 1 cm), power consumption (limited space for battery is available onboard), and fail safe operation (MCR operates deep inside the human body) [66, 67]. Therefore, MCR design and development requires significant skills and efforts in embedded systems, miniaturized electronics, packaging, debugging, and mechanical miniaturization of the device.

Beccani and colleagues had illustrated one design (Fig. 13.4) for robotic capsule [62]. As per their construction the robotic capsule consisted of a drug chamber, two coils, and the magnetic piston. The drug chamber (length = 6.35 mm, volume = 314.42 mm^3) was hosted in a cylindrical enclosure together with an axially magnetized cylindrical permanent magnet acting as a piston. The drug delivery capsule shell and the drug chamber were prototyped with an Objet30 3D printer

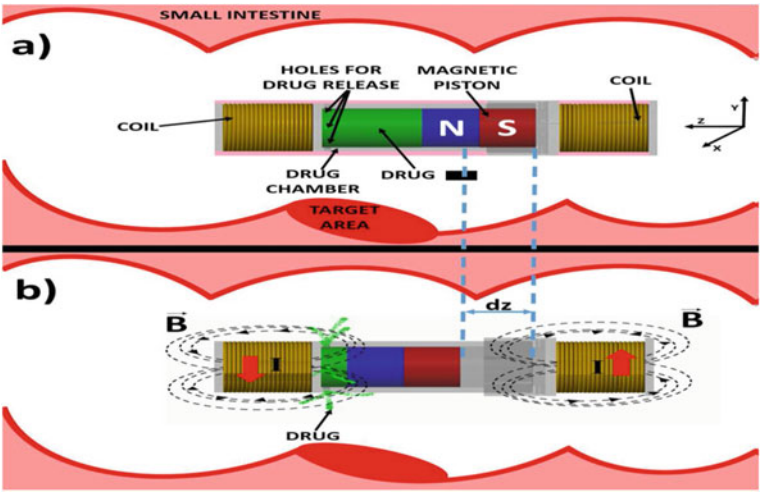


Fig. 13.4 Design and mechanism of drug release from the robotic capsule. (Reproduced from Beccani et al., 2016) [62]

with Vero white material, which was an opaque polyjet resinous material [68, 69]. The clearance between the outer diameter of the magnetic piston and inner diameter of the chamber was kept between 0.2 and 0.9 mm. This value guaranteed a low friction with the magnet resulting in no leakage of the drug and the actuating mechanism. The distal collar edge of the chamber had 12 circular holes (each with a radius of $r_h = 0.8$ mm) from where the drug was released in the surrounding environment. The number of holes and their radius had been chosen such that the drug was deployed uniformly without being affected by the capillary structure.

In the first approach of Beccani and colleagues (Fig. 13.4), the activation of coils of robotic capsule deployed the drug out of the drug chamber [62, 70, 71]. To trigger the magnetic piston, two coils were mounted at both sides of the drug chamber such that, when they were activated, the current induced two static magnetic fields aligned along the same axis, having the same magnitude, but in opposite directions. In vitro study showed that the robotic capsule was able to deploy solutions with a viscosity up to 1000 cP (higher than blood, water, and air) [72]. An in vitro trial on the robotic capsule showed excellent performance for drug delivery (Fig. 13.5) [62, 63]. Solutions of varying viscosities were filled in the capsule drug chamber

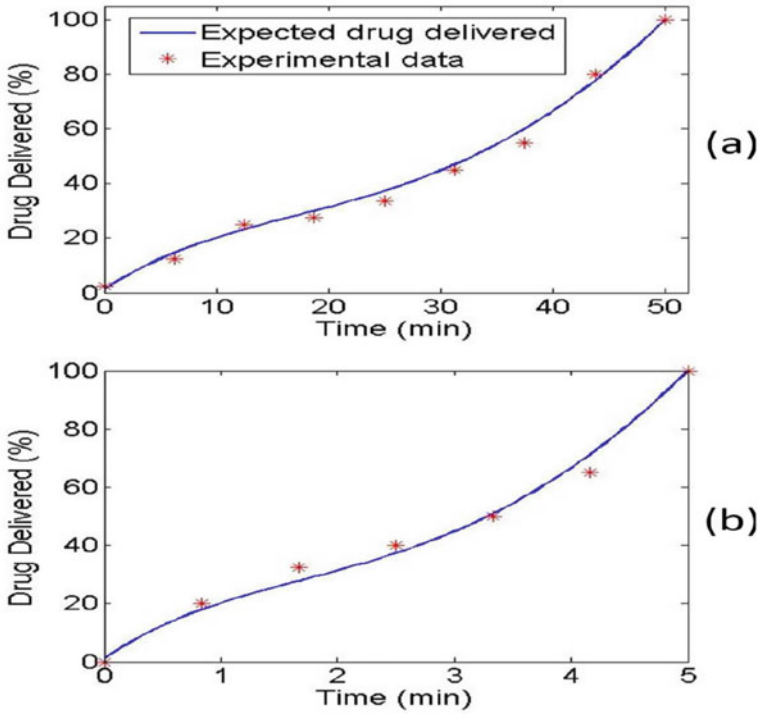


Fig. 13.5 Expected (**a**) and observed (**b**) drug delivery (%) with time from robotic capsule. (Reproduced from Beccani et al., 2016) [61]

to assess the performance of drug delivery and release from the robotic capsule. As a result, a good matching of the experimental data with the expected drug release profile was observed.

Quaglia and colleagues had illustrated one another design of a robotic capsule which utilized batteries as a power source [64]. Robotic capsule of 155 mm diameter consisted of a planar bioadhesive patch containing drug. A plate supporting the planar patch was integrated into the capsule. Moreover, for effective adhesion to take place, a large enough patch surface was available. More precisely, nearly 1 cm^2 was needed for the anchoring force to be on the order of 1 N; such a force value could be conservatively assumed to be enough for anchoring the envisioned swallowable modules [73]. By taking advantage of the actually fabricated, scaled up prototype, a 15–25 mm^2 patch supporting plate (PSP) was then designed [62, 74]. Additionally, patch exposure prior to application was minimized in order to avoid degradation of bioadhesion characteristics. Hence, the PSP was fully encapsulated within capsule case, and a release mechanism was designed for its ejection, also exploiting two ejectable shells (ES). For preserving bioadhesive properties, the capsule was required not to be watertight until the ES were opened: it sufficed that the devised sealing of the ES contained potential bodily fluids leakages; thus avoiding detrimental patch swelling leading to hampered adhesion. The related mechanism of release was found to be limited to a volume of approximately 4000 mm^3 [73, 75].

In the second approach of Quaglia and colleagues, a preloaded mechanism was adopted in order to contain power cost for patch release [63]. Three different steps of mechanism were: (1) an elastic preloading was designed for PSP holding prior to release; (2) PSP release was associated with a remotely activated triggering mechanism; and (3) patch deployment onto tissue was achieved by synchronizing PSP lift with suitable displacement of the ES [62].

Matryoshka-inspired micro-origami capsule (MIMC) was designed by assembling multiple hydrogel bilayers made from a shape shifting stimuli responsive hydrogel for enhancing drug loading and transport of drugs [76]. Matryoshka is a famous Russian stacking doll. *N*-Isopropylacrylamide, acrylamide, and polyethylene glycol diacrylate in a molar ratio of 90:10:0.5 and 85:15:2 (supporting layer) were used in formulating a thermoresponsive hydrogel. A UV light based polymerization process was used for preparing hydrogel disks. Subsequently photolithography was employed to prepare self-folding hydrogel bilayer comprising two different lower critical solution temperatures. Unfolding and refolding of bilayered hydrogel disks was achieved with a temperature change from room to body temperature. The assembled bilayers were magnetized and locomotion was achieved using eight coil electromagnetic manipulation system. Multiple hydrogel bilayers were assembled to form an MIMC. Drug leakage from hydrogel matrix at room temperature was 90%, while it significantly reduced at body temperature to 6% as the device folded up and assembled to encapsulate the drug. Diffusion of encapsulated drug up to 30% was observed from the hydrogel matrix within 1 h with its unfolding and assembling.

An oral capsule comprising luminal unfolding microneedle injector (LUMI) has been developed recently for oral delivery of biologic drugs like insulin [77]. LUMI has a tube like geometry which facilitates multiple points of contacts with the

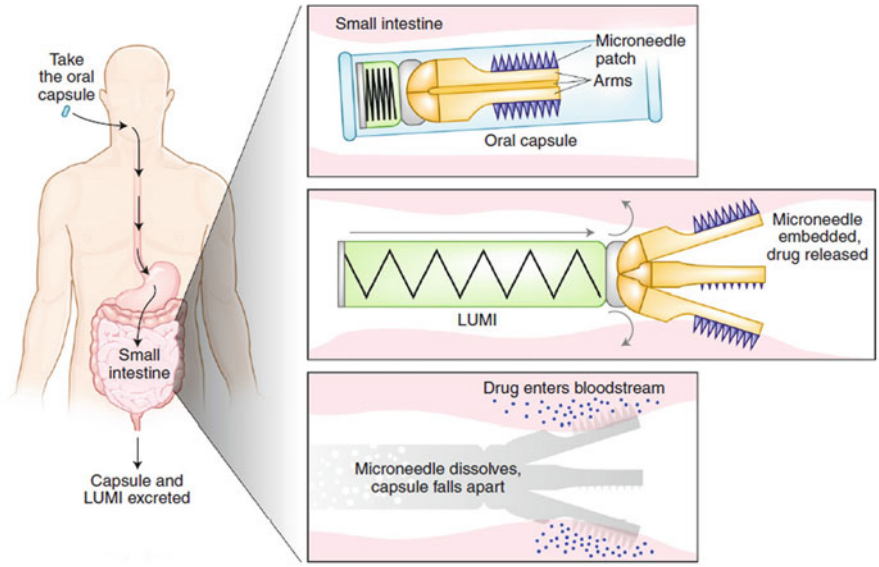


Fig. 13.6 Delivery of macromolecular drugs through intestine using microneedle patches encapsulated within oral capsule. (Reproduced from Prausnitz et al. 2019) [78]. Copyright Springer Nature

gastrointestinal tissue (Fig. 13.6). LUMI and a compressed spring are encapsulated within a custom capsule of 30 mm length and 9 mm diameter until its activation. The capsule has a pH-sensitive coating with poly(methacrylic acid-*co*-ethyl acrylate) in combination with polyethylene glycol which allows the activation of capsule within 1–5 h after reaching in an environment of pH 5.5 or higher. A compressed spring propels LUMI out of the capsule soon after its activation. The arms of LUMI are fabricated with mixtures of two biodegradable polymers polyethylene oxide and Soluplus (polyvinyl caprolactam-polyvinyl acetatepolyethylene glycol graft copolymer) which allows degradation within 24 h, in vivo and in vitro. The arms of LUMI have enough strength for about 10 min to push microneedles (1 mm in height) and inject biologic drug into the gastrointestinal tissue. Preclinical studies in swine revealed arapid pharmacokinetic profile and systemic uptake greater than 10% from LUMI loaded with insulin when compared to the systemic uptake of a subcutaneous injection, over a period of 4 h.

13.2.5 Nanorobots

Nanorobots of size 1–100 nm are diminutive contrivance which are fabricated to execute specific molecular level or cellular level or atomic level tasks in a precise manner [79]. According to nanorobotic theory, a large number of nanorobots may work together to perform microscopic and macroscopic tasks [80]. Nanorobots have

found applications in therapy and protecting the human body from various pathogenic diseases [81–83]. Some nanorobots which have been explored for their use in therapy and health care are respirocytes, microbivores, clottocytes, and pharmacytes [18].

Nanorobots have identical characteristics which are as follows:

- It should communicate with the operator by encoding messages to acoustic signals at carrier wave frequencies of 1–100 MHz [84].
- It will prevent itself, from being attacked by the immune system by having a passive, diamond exterior [85, 86].
- Nanorobots must have size in between 0.5 and 3 microns large with 1–100 nm parts [18, 87]. Larger nanorobots or nanorobots (>0.5–3 microns) will block capillary blood flow [80].
- It might produce multiple copies of it to replace worn-out units, a process called self-replication [79].

The framework of nanorobots may be designed with carbon nanotubes because of the inertness and high strength of carbon as a diamond or fullerene [84–87]. Apart from the carbon the framework can be made up with nitrogen, sulfur, hydrogen, silicon, oxygen, fluorine, etc. Ultrasonic sensors are attached around the body of nanorobots to prevent collision of nanorobots [88–90]. Folate materials are attached onto the body of nanorobots to confer specific targeting to cancer cells. Rotary motor, referred to as flagella motor, induces a magnetic field and the required torque to navigate nanorobot to the target site in the human body [91, 92].

Bachelet and colleagues used an open-source software caDNAno for designing a DNA nanorobot. The technique of making tiny shaped structures out of DNA is called DNA origami [93]. DNA can be built into a desired shape by cutting a small portion of it (staple strand) and attaching it to a long strand. Two strands of DNA bind due to the interaction between the complementary base pairs and form the desired shape. These bots look like a nano-sized open-ended barrel consisting of two halves which can be opened and closed in the form of a clamshell. The two halves are connected by molecular hinges and kept together by molecular locks or latches made of DNA double helixes. Inside the bots, there are 12 sites for attaching payload molecules [94]. On the outside, there are two positions for attaching aptamers which are short nucleotide strands with special sequences for recognizing molecules on the target cell. The drug is loaded inside the nanorobots, and secured by molecular anchors. The aptamers act as clasps, and once they recognize their target, the device opens up and releases the payload [95, 96].

Important parts of nanorobots are as follows:

- *Medicine Cavity:* A void section inside the nanorobot can carry small doses of chemicals or medicines. This site serves as a source site for slow and sustained delivery of drug to the target site or inflamed site [97].

- *Microcamera:* Nanorobots may include a miniature camera. An operator at a console will be able to steer the device while watching a live video feed, navigating it through the body manually [18, 97].
- *Energy Source:* A nanorobot with mounted electrodes could form a battery using the electrolytes present in human blood. Another option is to create chemical reactions with human blood to burn it for energy. Energy source would hold a small supply of chemicals that would become a fuel source when combined with blood [98].
- *Chemical Sensor:* Chemical nanosensors are used to avoid striking with other materials present in the human body [18].
- *Capacitor:* It is required to induce a magnetic field that pulls conductive fluids via one end of an electromagnetic pump and pushes out via the back end. Nanorobot travels in this environment just like a jet. Nonconducting materials like porcelain, mylar, mica, cellulose, Teflon, and ceramic are used as dielectric materials in capacitors [98].
- *Power Supply:* Complementary metal oxide semiconductor (CMOS) and nanocircuits are used for effective generation and supply of electromagnetic energy of 1.7 mA at 3.3 V [99].
- *Swimming Tail or Flagellum:* Nanorobots require a thrust to move around the body because in many cases it has to move towards the counter direction of flow of blood. Rotary motor induces magnetic field to generate required torque in robotic arms (also called as flagellum) for the movement of nanorobot to the target site [99].

The nanorobot design includes integrated nanoelectronics which may involve use of mobile phones. RFID (radio frequency identification device) and CMOS (complementary metal oxide semiconductor) transponder system have been used for tracking in vivo positioning, using well-established communication protocol [97, 98]. The parts of nanorobot are depicted in Fig. 13.7. Some other morphologically different nanorobots, their mechanism of action and biomedical applications of nanorobots can be found in a review paper [100].

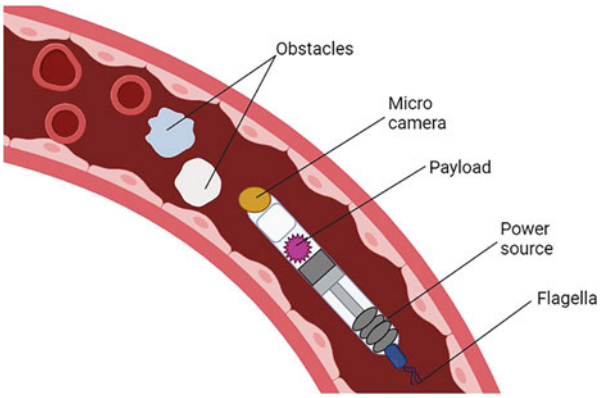


Fig. 13.7 A schematic diagram of a nanorobot in blood vessels

Normally, when the nanorobot is administered in the body, it starts to move towards the target in the liquid environment of the body. Different mechanisms of the nanorobots promotes swarming and swimming of nanorobots inside blood vessels. Due to the presence of leukocytes, erythrocytes, thrombocytes, and plasma, nanorobots may face an unstructured and cluttered environment when it is travelling in the human body [101]. Nanorobots contain sensors to clearly identify the environment, find its way, and avoid obstacles. Sensor and pre-programmed microprocessor identifies the target cells followed by binding and subsequent release of the drug contained in the nanorobot [102]. When the goal is achieved the nanorobot may be excreted via excretory routes of the body or cleared by nano-terminators, that is, active scavenger systems.

The robot should be able to differentiate between malignant tumor and healthy cells, which is most important, otherwise it may also destroy the healthy cells along with the malignant tumor cells [103, 104]. Practically, it is quite difficult for nanorobots to differentiate between a healthy cell and a malignant tumor cell or differentiating between leukemia type tumor cells and lymphoma cells. Uncertainty of the sensor, noise, uncertain information, and unknown parameters are some of the challenges yet to overcome.

13.3 Robotic Innovations Under Clinical Trial

There are many robotic innovations which have proven efficacy in preclinical trials. Among these robotic innovations, very few have reached clinical trials. One such robotic innovation which deserves a special mention here is a robotic capsule named RaniPill™.

13.3.1 RaniPill™

Rani Therapeutics, LLC ("Rani," in San Jose, California, US) has invented a robotic capsule namely RaniPill™, which is an oral delivery approach for almost any kind of drug. RaniPill™ has no electrical components, springs, or metal parts. Traditionally, parenteral route is a preferred route for administering biologic drugs like proteins, peptides, and antibodies. Rani's robotic capsule has the potential to replace subcutaneous injections for such sensitive molecules. This robotic enteric coated capsule houses a folded balloon containing reactants, a reaction valve, and microsyringe (Fig. 13.8). Drug is loaded into a hollow, dissolvable needle loaded within a microsyringe. After oral ingestion, pH-sensitive enteric coating protects the capsule as well as its contents from the harsh acidic environment of the stomach [105, 106]. Once the robotic capsule reaches into the small intestine, the reactants mix and generate carbon dioxide which in turn inflates a folded balloon to deliver a drug filled dissolvable microneedle into the intestinal wall wherein the microneedle dissolves to release the drug for absorption (Fig. 13.9). Human intestines are insensate to sharp stimuli, and thus no pain is felt by the patient.

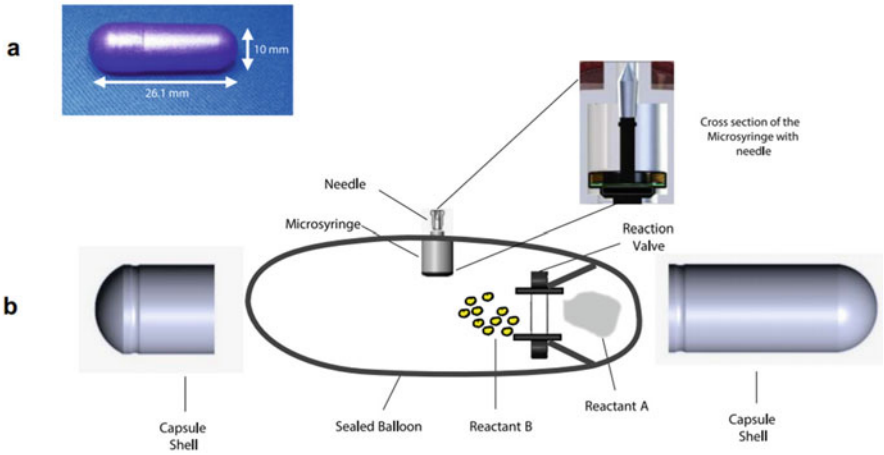


Fig. 13.8 RaniPill™ design (**a**) Assembled enteric coated RaniPill™ and (**b**) Different parts and internal design of RaniPill™. (Reproduced from Dhalla et al. 2021) [106]

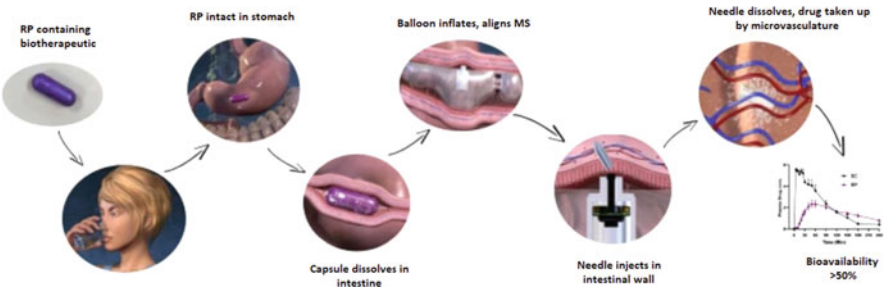


Fig. 13.9 The journey of drug encapsulated in RaniPill™ from ingestion to the systemic circulation. (Reproduced from Dhalla et al. 2021) [106]

Rani has successfully completed several preclinical studies using insulin and adalimumab as model drugs [104, 105]. These drugs were successfully delivered via the RaniPill™ into the wall of the small intestine and had PK equivalence with parenteral routes of administration. In preclinical studies it was found that the RaniPill can deliver approximately 3 mg of any kind of drug or biomolecule including proteins, antibodies, and peptides with high bioavailability rivaling subcutaneous injections.

Rani has conducted a Phase I clinical study (NCT03798912) in 62 healthy volunteers in which the robotic pills containing octreotide (a drug used to treat acromegaly) were administered to check the feasibility of this device. RaniPill™ was found to be safe, well-tolerated and delivered octreotide with absolute bioavailability of 65 ± 9% [106]. The reliability of drug delivery with RaniPill™ was 25–80% which was found to be related with the balloon size which was regarded

as a key factor for optimization to achieve best results. In a separate clinical study in 20 healthy volunteers, RaniPill™ was shown to be safe in both the fed and fasted states [105–108]. Other oral versions of molecules the company is developing include human growth hormone, Factor VIII, parathyroid hormone, TNFα inhibitors, and GLP-1 (glucagon like peptide 1) agonists.

13.4 Challenges and Future Prospects

Though robotic systems have proven itself as an efficient approach for drug delivery, integrating an ample powered source of electrical/pneumatic energy onto the body of a miniaturized functional soft robot is still a challenge. To overcome this problem, the soft grippers which are stimuli-responsive can be used [109]. Soft robots require complex fluid or air supply components like compressors for their locomotion. These components make the structure complex, for which manufacturing becomes rather difficult. The fabrication process takes much time due to the manual assembly of soft robots [1]. For this reason, it is quite difficult to give a proper geometrical shape or proper resolution. In many cases, the coils inside the robots get heated. These hamper the actuation and locomotion time period. To avoid this issue, external magnetic field is used which creates necessary torque for the actuation and locomotion. All these issues may be eliminated by fabricating soft robots using 3D printing technology. 3D printing technology provides a precise and accurate geometrical assembly to a soft robot. In the case of robotic drug delivery using magnetic field, significant challenge comes from safe use of magnetic field gradient as it has tissue damage potential. Controlling magnetic field density for safe use is very difficult. Sourcing soft materials which does not harm the human body is another important concern. 3D/4D printing allows fabrication of soft robots or their parts. Fitting a camera module, drug chamber, and a magnet in nano dimensional robot require extensive miniaturization [110]. To overcome this issue, multi-dimensional robotic devices are introduced for easy actuation, sensing, control, and communication. To perform tasks in dynamic in vivo environments and complex, cluttered environments and for materials synthesis, integration of soft-robotics innovative solutions are needed. For example, the soft grippers, when inserted from a cold state, activate within a few minutes of being exposed to the actual temperature of the body. But for few cases like targeted delivery of drug or surgery, the mechanism of triggering necessitates that the grippers close only when they reach the target site, which may occur at shorter or longer times. Hence, it is mandatory to modify grippers that can respond to other physicochemical signals such as enzymes, pH, or other biomolecular triggers, or external optical and magnetic stimuli [110–112]. Another major challenge for drug delivery via robotic devices is the reduction of the host defense mechanism of the human body [11].

For transportation of drug in the gastrointestinal tract (GI tract) soft robots are used. However, to extend their application in more remote places like the vascular system, soft robotics require further miniaturization to submicron levels. To achieve appropriate way-finding in blood vessels, Fe–Co or Ni–Co alloys mean stronger

magnetic materials or FePd nanowires should be used [113, 114]. In future, the main prominence in drug delivery will switch to medical engineering from medical science, where robotic technology, that is, robotic drug delivery system will be the insurgence. This advancement of medical sciences will bring outstanding speed, effectiveness, and relaxation from invasive administration and ease of treating emergency cases, which will reduce cost and risk [102, 114]. It is now proven by many researchers that robotics holds great potential to treat and diagnose diseases when compared to the traditional practices of delivering drugs [10, 62, 79, 81, 83]. In the near future robotics will dominate the therapy and diagnosis which at present seems like a fiction.

13.5 Conclusion

Soft robots provide a number of applications like sustained release, drug targeting, diagnosis, and treatments of diseases like cancer. Origami robots provide a unique advantage of encapsulating drug into different folds which allows delivery of drugs in multiple ways like immediate release, controlled release, or in both manners. Sperm robots are efficient means for pathway signaling and temporal targeting. Soft multi-legged robots are effective at both spatial and temporal targeting. Robotic capsules require an external magnetic field for generating torque for actuation. Robotic capsules are a unique mean of enclosing drug formulations like patches. Nanorobots have some potential applications in targeted drug delivery, pharmacokinetic monitoring of drug delivery, and treatment of cancer and arteriosclerosis. Most of these applications, at present, are investigational. Only a few applications have reached up to the clinical trials. RaniPill™, a soft robotic capsule, is in clinical trials for its potential to replace subcutaneous injections. With the collaborative endeavors of engineers, scientists, and doctors, smart and flexible in vivo communicable soft robots can perform on demand therapy in the near future.

Acknowledgments The authors would like to thank Dr. Arvinder Dhalla from Rani Therapeutics for discussion and suggestions on RaniPill™ section.

References

1. Rus D, Tolley MT (2015) Design, fabrication and control of soft robots. Nature 521:467–475. https://doi.org/10.1038/nature14543
2. Trimmer B (2013) Soft robots. Curr Biol 23:R640–R641. https://doi.org/10.1016/j.cub.2013.04.070
3. Bao G, Fang H, Chen L, Wan Y, Xu F, Yang Q et al (2018) Soft robotics: academic insights and perspectives through bibliometric analysis. Soft Robot 5:229–241. https://doi.org/10.1089/soro.2017.0135
4. Wehner M, Truby RL, Fitzgerald DJ, Mosadegh B, Whitesides GM, Lewis JA et al (2016) An integrated design and fabrication strategy for entirely soft, autonomous robots. Nature 536:451–455. https://doi.org/10.1038/nature19100

5. Laschi C, Mazzolai B, Cianchetti M (2016) Soft robotics: technologies and systems pushing the boundaries of robot abilities. Sci Robot 1:eaah3690. https://doi.org/10.1126/scirobotics.aah3690
6. Alici G (2018) Softer is harder: what differentiates soft robotics from hard robotics? MRS Adv 3:1557–1568. https://doi.org/10.1557/adv.2018.159
7. Schmidt CK, Medina-Sánchez M, Edmondson RJ, Schmidt OG (2020) Engineering microrobots for targeted cancer therapies from a medical perspective. Nat Commun 11:5618–5636. https://doi.org/10.1038/s41467-020-19322-7
8. Hockstein NG, Gourin CG, Faust RA, Terris DJ (2007) A history of robots: from science fiction to surgical robotics. J Robot Surg 1:113–118. https://doi.org/10.1007/s11701-007-0021-2
9. Tanaka K, Rangelova E, Del Chiaro M (2018) Is robotic pancreatectomy indicated for patients with pancreatic cancer? Ann Pancreat Cancer 1:16–24. https://doi.org/10.21037/apc.2018.04.01
10. Menciassi A, Quirini M, Dario P (2007) Microrobotics for future gastrointestinal endoscopy. Minim Invasive Ther Allied Technol 16:91–100. https://doi.org/10.1080/13645700701266982
11. da Silva Luz GV, Barros KVG, de Araújo FVC, da Silva GB, da Silva PAF, Condori RCI et al (2016) Nanorobotics in drug delivery systems for treatment of cancer: a review. J Mater Sci Eng A 6:167–180. https://doi.org/10.17265/2161-6213/2016.5-6.005
12. Fletcher M, Biglarbegian M (2013) Intelligent system design for bionanorobots in drug delivery. Cancer Nano 4:117–125. https://doi.org/10.1007/s12645-013-0044-5
13. Shigemune H, Maeda S, Hara Y, Hosoya N, Hashimoto S (2016) Origami robot: a self-folding paper robot with an electrothermal actuator created by printing. IEEE/ASME Trans Mechatron 21:2746–2754. https://doi.org/10.1109/TMECH.2016.2593912
14. Torelli E, Marini M, Palmano S, Piantanida L, Polano C, Scarpellini A et al (2014) A DNA origami nanorobot controlled by nucleic acid hybridization. Small 10:2918–2926. https://doi.org/10.1002/smll.201400245
15. Lu H, Zhang M, Yang Y, Huang Q, Fukuda T, Wang Z et al (2018) A bioinspired multilegged soft millirobot that functions in both dry and wet conditions. Nat Commun 9:3944–3951. https://doi.org/10.1038/s41467-018-06491-9
16. Iacovacci V, Lucarini G, Ricotti L, Dario P, Dupont PE, Menciassi A (2015) Untethered magnetic millirobot for targeted drug delivery. Biomed Microdevices 17:9962–9974. https://doi.org/10.1007/s10544-015-9962-9
17. Saha O, Dasgupta P (2018) A comprehensive survey of recent trends in cloud robotics architectures and applications. Robotics 7:47–72. https://doi.org/10.3390/robotics7030047
18. Manjunath A, Kishore V (2014) The promising future in medicine: nanorobots. Biomed Sci Eng 2:42–47. https://doi.org/10.12691/bse-2-2-3
19. Marvel JA, Bostelman R, Falco J (2018) Multi-robot assembly strategies and metrics. ACM Comput Surv 51:14–46. https://doi.org/10.1145/3150225
20. Johnson M, Chen Y, Hovet S, Xu S, Wood B, Ren H et al (2017) Fabricating biomedical origami: a state-of-the-art review. Int J Comput Assist Radiol Surg 12:2023–2032. https://doi.org/10.1007/s11548-017-1545-1
21. Khalil ISM, Magdanz V, Sanchez S, Schmidt OG, Misra S (2014) Biocompatible, accurate, and fully autonomous: a sperm-driven micro-bio-robot. J Micro-Bio Robot 9:79–86. https://doi.org/10.1007/s12213-014-0077-9
22. Buchanan R, Bandyopadhyay T, Bjelonic M, Wellhausen L, Hutter M, Kottege N (2019) Walking posture adaptation for legged robot navigation in confined spaces. IEEE Robot Autom Lett 4:2148–2155. https://doi.org/10.1109/LRA.2019.2899664
23. Thai PT, Savchenko M, Hagiwara I (2018) Finite element simulation of robotic origami folding. Simul Model Pract Theory 84:251–267. https://doi.org/10.1016/j.simpat.2018.03.004
24. Dobson CM (2003) Protein folding and misfolding. Nature 426:884–890. https://doi.org/10.1038/nature02261

25. Turner N, Goodwine B, Sen M (2016) A review of origami applications in mechanical engineering. Proc Inst Mech Eng C J Mech Eng Sci 230:2345–2362. https://doi.org/10.1177/0954406215597713
26. Xu L, Shyu TC, Kotov NA (2017) Origami and Kirigami nanocomposites. ACS Nano 11:7587–7599. https://doi.org/10.1021/acsnano.7b03287
27. Zhang K, Qiu C, Dai JS (2015) Helical Kirigami-enabled centimeter-scale worm robot with shape-memory-alloy linear actuators. J Mech Robot 7:021014. https://doi.org/10.1115/1.4029494
28. Turner N, Goodwine B, Sen M (2016) A review of origami applications in mechanical engineering. Proc Inst Mech Eng C J Mech Eng Sci. 230:1989–1996. https://doi.org/10.1177/0954406215597713
29. Tachi T (2010) Origamizing polyhedral surfaces. IEEE Trans Vis Comput Graph 16:298–311. https://doi.org/10.1109/TVCG.2009.67
30. Aukes DM, Goldberg B, Cutkosky MR, Wood RJ (2014) An analytic framework for developing inherently-manufacturable pop-up laminate devices. Smart Mater Struct 23:094013. https://doi.org/10.1088/0964-1726/23/9/094013
31. Hawkes E, An B, Benbernou NM, Tanaka H, Kim S, Demaine ED et al (2010) Programmable matter by folding. Proc Natl Acad Sci 107:12441–12445. https://doi.org/10.1073/pnas.0914069107
32. Miyashita S, Guitron S, Yoshida K, Li S, Damian DD, Rus D (2016) Ingestible, controllable, and degradable origami robot for patching stomach wounds. In: Proc.—IEEE Int. Conf. Robot. Autom., pp 909–916. https://doi.org/10.1109/ICRA.2016.7487222
33. Ornes S (2017) Inner Workings: medical microrobots have potential in surgery, therapy, imaging, and diagnostics. Proc Natl Acad Sci 114:12356–12358. https://doi.org/10.1073/pnas.1716034114
34. Zhao YX, Shaw A, Zeng X, Benson E, Nyström AM, Högberg B (2012) DNA origami delivery system for cancer therapy with tunable release properties. ACS Nano 6:8684–8691. https://doi.org/10.1021/nn3022662
35. Miyashita S, Meeker L, Tolley MT, Wood RJ, Rus D (2014) Self-folding miniature elastic electric devices. Smart Mater Struct 23:094005. https://doi.org/10.1088/0964-1726/23/9/094005
36. Xu H, Medina-Sánchez M, Magdanz V, Schwarz L, Hebenstreit F, Schmidt OG (2018) Sperm-hybrid micromotor for targeted drug delivery. ACS Nano 12:327–337. https://doi.org/10.1021/acsnano.7b06398
37. Srivastava SK, Medina-Sánchez M, Koch B, Schmidt OG (2016) Medibots: dual-action biogenic microdaggers for single-cell surgery and drug release. Adv Mater 28:832–837. https://doi.org/10.1002/adma.201504327
38. Magdanz V, Schmidt OG (2014) Spermbots: potential impact for drug delivery and assisted reproductive technologies. Expert Opin Drug Deliv 11:1125–1129. https://doi.org/10.1517/17425247.2014.924502
39. Khalil ISM, Youakim K, Sanchez A, Misra S (2014) Magnetic-based motion control of sperm-shaped microrobots using weak oscillating magnetic fields. In: IEEE Int. Conf. Intell. Robot. Syst., pp 4686–4691. https://doi.org/10.1109/IROS.2014.6943228
40. Magdanz V, Sanchez S, Schmidt OG (2013) Development of a sperm-flagella driven micro-bio-robot. Adv Mater 25:6581–6588. https://doi.org/10.1002/adma.201302544
41. Lin X, Wu Z, Wu Y, Xuan M, He Q (2016) Self-propelled micro-/nanomotors based on controlled assembled architectures. Adv Mater 28:1060–1072. https://doi.org/10.1002/adma.201502583
42. Wu Z, Lin X, Si T, He Q (2016) Recent progress on bioinspired self-propelled micro/nanomotors via controlled molecular self-assembly. Small 12:3080–3093. https://doi.org/10.1002/smll.201503969

43. Mhanna R, Qiu F, Zhang L, Ding Y, Sugihara K, Zenobi-Wong M et al (2014) Artificial bacterial flagella for remote-controlled targeted single-cell drug delivery. Small 10:1953–1957. https://doi.org/10.1002/smll.201303538
44. De Ávila BEF, Angsantikul P, Li J, Angel Lopez-Ramirez M, Ramírez-Herrera DE, Thamphiwatana S et al (2017) Micromotor-enabled active drug delivery for in vivo treatment of stomach infection. Nat Commun 8:272–280. https://doi.org/10.1038/s41467-017-00309-w
45. Magdanz V, Medina-Sánchez M, Chen Y, Guix M, Schmidt OG (2015) How to improve spermbot performance. Adv Funct Mater 25:2763–2770. https://doi.org/10.1002/adfm.201500015
46. Aoi S, Tanaka T, Fujiki S, Funato T, Senda K, Tsuchiya K (2016) Advantage of straight walk instability in turning maneuver of multilegged locomotion: a robotics approach. Sci Rep 6:30199. https://doi.org/10.1038/srep30199
47. Aoi S, Sasaki H, Tsuchiya K (2007) A multilegged modular robot that meanders: investigation of turning maneuvers using its inherent dynamic characteristics. SIAM J Appl Dyn Syst 6:348–377. https://doi.org/10.1137/060664756
48. Huang HW, Sakar MS, Petruska AJ, Pané S, Nelson BJ (2016) Soft micromachines with programmable motility and morphology. Nat Commun 7:12263. https://doi.org/10.1038/ncomms12263
49. Lin HT, Leisk GG, Trimmer B (2011) GoQBot: a caterpillar-inspired soft-bodied rolling robot. Bioinspir Biomim 6:026007. https://doi.org/10.1088/1748-3182/6/2/026007
50. Majidi C (2013) Soft robotics: a perspective—current trends and prospects for the future. Soft Robot 1:5–11. https://doi.org/10.1089/soro.2013.0001
51. Zhang H, Wang W, González-Gómez J, Zhang J (2009) Design and realization of a novel modular climbing caterpillar using low-frequency vibrating passive suckers. Adv Robot 23:889–906. https://doi.org/10.1163/156855309X442990
52. Aceituno-Cabezas B, Dai H, Cappelletto J, Grieco JC, Fernandez-Lopez G (2017) A mixed-integer convex optimization framework for robust multilegged robot locomotion planning over challenging terrain. In: IEEE Int. Conf. Intell. Robot. Syst., 24–28 Sep 2017, Vancouver, BC, Canada, IEEE. https://doi.org/10.1109/IROS.2017.8206313
53. Kim S, Laschi C, Trimmer B (2013) Soft robotics: a bioinspired evolution in robotics. Trends Biotechnol 31:287–294. https://doi.org/10.1016/j.tibtech.2013.03.002
54. Raibert MH (1986) Legged robots. Commun ACM 29:499–514. https://doi.org/10.1145/5948.5950
55. Venkiteswaran VK, Tan DK, Misra S (2020) Tandem actuation of legged locomotion and grasping manipulation in soft robots using magnetic fields. Extrem Mech Lett 41:101023. https://doi.org/10.1016/j.eml.2020.101023
56. Calisti M, Picardi G, Laschi C (2017) Fundamentals of soft robot locomotion. J R Soc Interface 14:20170101. https://doi.org/10.1098/rsif.2017.0101
57. Machado JAT (2014) An overview of legged robots. Soins Pediatr 35:30–34. https://doi.org/10.1016/j.spp.2014.09.006
58. Kano T, Ikeshita Y, Fukuhara A, Ishiguro A (2019) Body-limb coordination mechanism underlying speed-dependent gait transitions in sea roaches. Sci Rep 9:2848–2856. https://doi.org/10.1038/s41598-019-39862-3
59. Easton TA (1984) Biological model for multilegged robot control. Robot Age 6:26–29
60. Bowling AP (2006) Dynamic performance, mobility, and agility of multilegged robots. J Dyn Syst Meas Control 128:765–777. https://doi.org/10.1115/1.2229252
61. Zhang J, Diller E (2018) Untethered miniature soft robots: modeling and design of a millimeter-scale swimming magnetic sheet. Soft Robot 5:761–776. https://doi.org/10.1089/soro.2017.0126
62. Beccani M, Aiello G, De Momi E, Völgyesi P, Lédeczi Á, Valdastri P (2016) Component based design of a drug delivery capsule robot. Sensors Actuators A Phys 245:180–188. https://doi.org/10.1016/j.sna.2016.04.035

63. Mapara SS, Patravale VB (2017) Medical capsule robots: a renaissance for diagnostics, drug delivery and surgical treatment. J Control Release 261:337–351. https://doi.org/10.1016/j.jconrel.2017.07.005
64. Quaglia C, Tognarelli S, Sinibaldi E, Funaro N, Dario P, Menciassi A (2014) Wireless robotic capsule for releasing bioadhesive patches in the gastrointestinal tract. J Med Device 8:014503–014508. https://doi.org/10.1115/1.4025450
65. Moglia A, Menciassi A, Schurr MO, Dario P (2007) Wireless capsule endoscopy: from diagnostic devices to multipurpose robotic systems. Biomed Microdevices 9:235–243. https://doi.org/10.1007/s10544-006-9025-3
66. Ciuti G, Caliò R, Camboni D, Neri L, Bianchi F, Arezzo A et al (2016) Frontiers of robotic endoscopic capsules: a review. J Micro-Bio Robot 11:1–18. https://doi.org/10.1007/s12213-016-0087-x
67. Guo S, Yang Q, Bai L, Zhao Y (2018) Development of multiple capsule robots in pipe. Micromachines 9:259–274. https://doi.org/10.3390/mi9060259
68. Carpi F, Pappone C (2009) Magnetic maneuvering of endoscopic capsules by means of a robotic navigation system. IEEE Trans Biomed Eng 56:1482–1490. https://doi.org/10.1109/TBME.2009.2013336
69. Hosseini S, Mehrtash M, Khamesee MB (2011) Design, fabrication and control of a magnetic capsule-robot for the human esophagus. Microsyst Technol 17:1145–1152. https://doi.org/10.1007/s00542-011-1231-0
70. Than TD, Alici G, Zhou H, Li W (2012) A review of localization systems for robotic endoscopic capsules. IEEE Trans Biomed Eng 59:2387–2399. https://doi.org/10.1109/TBME.2012.2201715
71. Beccani M, Di Natali C, Aiello G, Benjamin C, Susilo E, Valdastri P (2015) A Magnetic drug delivery Capsule based on a coil actuation mechanism. Proc Eng 120:53–56. https://doi.org/10.1016/j.proeng.2015.08.564
72. Yim S, Jeon D (2014) Magnetic mechanical capsule robot for multiple locomotion mechanisms. Int J Control Autom Syst 12:383–389. https://doi.org/10.1007/s12555-013-0078-9
73. Woods SP, Constandinou TG (2013) Wireless capsule endoscope for targeted drug delivery: mechanics and design considerations. IEEE Trans Biomed Eng 60:945–953. https://doi.org/10.1109/TBME.2012.2228647
74. Milligan CJ, Li J, Sukumar P, Majeed Y, Dallas ML, English A et al (2009) Robotic multiwell planar patch-clamp for native and primary mammalian cells. Nat Protoc 4:244–255. https://doi.org/10.1038/nprot.2008.230
75. Yoo S-S, Rama S, Szewczyk B, Pui JWY, Lee W, Kim L (2014) Endoscopic capsule robots using reconfigurable modular assembly: a pilot study. Int J Imaging Syst Technol 24:359–365. https://doi.org/10.1002/ima.22113
76. Huang H, Tibbitt MW, Huang T, Nelson BJ (2019) Matryoshka-inspired micro-origami capsules to enhance loading, encapsulation, and transport of drugs. Soft Robot 6:150–159. https://doi.org/10.1089/soro.2018.0028
77. Abramson A, Caffarel-Salvador E, Soares V, Minahan D, Tian RY, Lu X et al (2019) A luminal unfolding microneedle injector for oral delivery of macromolecules. Nat Med 25:1512–1518. https://doi.org/10.1038/s41591-019-0598-9
78. Prausnitz MR, Gomaa Y, Li W (2019) Microneedle patch drug delivery in the gut. Nat Med 25:1471–1472. https://doi.org/10.1038/s41591-019-0606-0
79. Umai RD, Devi PB, Thiruchelvi R (2018) A review on DNA nanorobots—a new technique for cancer treatment. Asian J Pharm Clin Res 11:61–64. https://doi.org/10.22159/ajpcr.2018.v11i6.25015
80. Yang G, Sau C, Lai W, Cichon J, Li W (2014) Nanorobotic applications in medicine: current proposals and designs. Am J Robot Surg 1:4–11. https://doi.org/10.1126/science.1249098. Sleep

81. Kshirsagar N, Patil S, Kshirsagar R, Wagh A, Bade A (2014) Review on application of nanorobots in health care. World J Pharm Pharm Sci 3:472–480
82. Soto F, Chrostowski R (2018) Frontiers of medical micro/nanorobotics: in vivo applications and commercialization perspectives toward clinical uses. Front Bioeng Biotechnol 6:170–181. https://doi.org/10.3389/fbioe.2018.00170
83. Rifat T, Hossain S, Alam M (2019) A review on applications of nanorobots in combating complex diseases. Bangladesh Pharm J 22:99–108
84. Jain K (2012) Nanodevices for medicine and surgery. In: Handbook of nanomedicine, 2nd edn. Humana Press, pp 183–193. https://doi.org/10.1007/978-1-60327-319-0_6
85. Fu J, Yan H (2012) Controlled drug release by a nanorobot. Nat Biotechnol 30:407–408. https://doi.org/10.1038/nbt.2206
86. Akrout M (2018) Nanorobots in 2018: three recent advances shaping the future of research. PreScouter. https://www.prescouter.com/2018/09/nanobots-2018-recent-advances-shaping-future-research/. Accessed 12 Apr 2020
87. Peter H (2016) Diamandis. Nanorobots: where we are today and why their future has amazing potential. Singul Univ. https://singularityhub.com/2016/05/16/nanorobots-where-we-are-today-and-why-their-future-has-amazing-potential/. Accessed 12 Apr 2020
88. Popov AM, Lozovik YE, Fiorito S, Yahia L (2007) Biocompatibility and applications of carbon nanotubes in medical nanorobots. Int J Nanomedicine 2:361–372
89. Nikitin MP, Shipunova VO, Deyev SM, Nikitin PI (2014) Biocomputing based on particle disassembly. Nat Nanotechnol 9:716–722. https://doi.org/10.1038/nnano.2014.156
90. Sharma N, Mittal RK (2008) Nanorobot movement: challenges and biologically inspired solutions. Int J Smart Sens Intell Syst 1:87–109. https://doi.org/10.21307/ijssis-2017-280
91. Ali J, Cheang UK, Martindale JD, Jabbarzadeh M, Fu HC, Jun Kim M (2017) Bacteria-inspired nanorobots with flagellar polymorphic transformations and bundling. Sci Rep 7:14098–14107. https://doi.org/10.1038/s41598-017-14457-y
92. Saxena S, Pramod B, Dayananda B, Nagaraju K (2015) Design, architecture and application of nanorobotics in oncology. Indian J Cancer 52:236–241. https://doi.org/10.4103/0019-509X.175805
93. Douglas SM, Bachelet I, Church GM (2012) A logic-gated nanorobot for targeted transport of molecular payloads. Science 335:831–834. https://doi.org/10.1126/science.1214081
94. Li S, Jiang Q, Ding B, Nie G (2019) Anticancer activities of tumor-killing nanorobots. Trends Biotechnol 37:573–577. https://doi.org/10.1016/j.tibtech.2019.01.010
95. Sakai Y, Islam MS, Adamiak M, Shiu SCC, Tanner JA, Heddle JG (2018) DNA aptamers for the functionalisation of DNA origami nanostructures. Genes (Basel) 9:1–20. https://doi.org/10.3390/genes9120571
96. Katsnelson A (2012) DNA robot could kill cancer cells. Nature 1–3. https://doi.org/10.1038/nature.2012.10047
97. Kumar SS, Nasim BP, Abraham E (2018) Nanorobots a future device for diagnosis and treatment. J Pharm Pharm 5:44–49. https://doi.org/10.15436/2377-1313.18.1815
98. Jiang W, Von Roemeling CA, Chen Y, Qie Y, Liu X, Chen J et al (2017) Designing nanomedicine for immuno-oncology. Nat Biomed Eng 1:1–11. https://doi.org/10.1038/s41551-017-0029
99. Li J, de Ávila BEF, Gao W, Zhang L, Wang J (2017) Micro/nanorobots for biomedicine: delivery, surgery, sensing, and detoxification. Sci Robot 2:eaam6431–eaam6440. https://doi.org/10.1126/scirobotics.aam6431
100. Mitthra S, Karthick A, Anuradha B, Mensudar R, Ramkumar Sadhana K, Nidhya Varshini G (2016) Nanorobots—a small wonder. Biosci Biotechnol Res Asia 13:2131–2134. https://doi.org/10.13005/bbra/2374
101. Sánchez S, Pumera M (2009) Nanorobots: the ultimate wireless self-propelled sensing and actuating devices. Chem Asian J 4:1402–1410. https://doi.org/10.1002/asia.200900143

102. Patra JK, Das G, Fraceto LF, Campos EVR, Rodriguez-Torres MDP, Acosta-Torres LS et al (2018) Nano based drug delivery systems: recent developments and future prospects. J Nanobiotechnol 16:71–103. https://doi.org/10.1186/s12951-018-0392-8
103. Amir Y, Ben-Ishay E, Levner D, Ittah S, Abu-Horowitz A, Bachelet I (2014) Universal computing by DNA origami robots in a living animal. Nat Nanotechnol 9:353–357. https://doi.org/10.1038/nnano.2014.58
104. Rahul VA (2017) A brief review on nanorobots. Int J Mech Eng 4:15–21. https://doi.org/10.14445/23488360/ijme-v4i8p104
105. Rani Therapeutics (2019) First human study of "robotic" RaniPill ™ capsule to replace injections announced by rani therapeutics. PR Newswire. https://www.prnewswire.com/news-releases/first-human-study-of-robotic-ranipill-capsule-to-replace-injections-announced-by-rani-therapeutics-300800556.html. Accessed 14 May 2020
106. Dhalla AK, Al-Shamsie Z, Beraki S, Dasari A, Fung LC, Fusaro L et al (2021) A robotic pill for oral delivery of biotherapeutics: safety, tolerability, and performance in healthy subjects. Drug Deliv Transl Res. https://doi.org/10.1007/s13346-021-00938-1
107. Miller R (2019) Starts & stops: first-in-human trial of RaniPill drug-delivery device begins. Pharma Intell Inf. https://pink.pharmaintelligence.informa.com/MT124729/Starts%2D%2DStops-FirstInHuman-Trial-Of-RaniPill-DrugDelivery-Device-Begins. Accessed 14 May 2020
108. Hale C (2019) Rani Therapeutics completes first-in-human safety study of its robotic biologic pill. Fierce Biotech. https://www.fiercebiotech.com/medtech/rani-therapeutics-completes-first-human-safety-study-its-robotic-biologic-pill. Accessed 14 Apr 2021
109. Robertson S (2018) New robot could improve drug delivery in the body. News Med Life Sci. https://www.news-medical.net/news/20180926/New-robot-could-improve-drug-delivery-in-the-body.aspx. Accessed 20 Jan 2021
110. Simaan N, Yasin RM, Wang L (2018) Medical technologies and challenges of robot-assisted minimally invasive intervention and diagnostics. Annu Rev Control Robot Auton Syst 1:465–490
111. Lenaghan SC, Wang Y, Xi N, Fukuda T, Tarn T, Hamel WR et al (2013) Grand challenges in bioengineered nanorobotics for cancer therapy. IEEE Trans Biomed Eng 60:667–673. https://doi.org/10.1109/TBME.2013.2244599
112. Ghosh A, Yoon C, Ongaro F, Scheggi S, Selaru FM, Misra S et al (2017) Stimuli-responsive soft untethered grippers for drug delivery and robotic surgery. Front Mech Eng 3:1–9. https://doi.org/10.3389/fmech.2017.00007
113. Somberg JC (1989) Drug delivery system. J Clin Pharmacol 29:673–673. https://doi.org/10.1002/j.1552-4604.1989.tb03400.x
114. Tzafestas S (2018) Roboethics: fundamental concepts and future prospects. Information 9:148–152. https://doi.org/10.3390/info9060148

Online Literature Searching for Research Projects in Pharmaceutical Sciences

14

Vikas Anand Saharan, Surojit Banerjee, Swati Dobhal, Manoj Kumar Sarangi, and Anupama Singh

Contents

V. A. Saharan (✉) · S. Banerjee · M. K. Sarangi
Department of Pharmaceutics, School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Dehradun, Uttarakhand, India

S. Dobhal
Department of Pharmacology, School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Dehradun, Uttarakhand, India

A. Singh
Department of Pharmacognosy, School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Dehradun, Uttarakhand, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_14

Abstract

Computers and subsequent advancements in programming, software, databases, and internet have led to a significant change in the process of literature searching. In the pre-internet era, researchers used to visit libraries for hand searching of abstracts and journals for finding literature relevant to their research. Nowadays libraries have more e-literature like e-books, electronic abstracting and indexing databases, online journals, etc. Abstracting services of the nineteenth century have transformed themselves into platforms for providing online services for bibliographic and/or citation searching. A vast number of research articles and other primary literature have always remained a challenge to deal with. However, the advancements in software, databases, and artificial intelligence have significantly eased the process of literature searching with the availability of human curated and natural language processing databases. Literature can be indexed in these databases by either words they contain or using domain-specific or customized thesauri. This chapter focuses on identifying significantly important databases and other online literature for literature searching in the field of pharmaceutical sciences. Additionally guidance on comprehensive literature searching and electronic searching of databases are provided to novice researchers and postgraduate students.

Keywords

Database · Journals · Books · Patents · Repositories · Conference proceedings · Drug · Chemical · Traditional medicine · Information retrieval · Literature review

Chapter Objectives

Upon reading of this chapter, the reader will be able to understand:

- Difference between primary, secondary, and tertiary literature
- Some significant information sources in the field of pharmaceutical sciences
- How to perform a comprehensive literature search for any research project or for writing review/systematic review
- How to search bibliographic databases

14.1 Introduction

Technological improvements in computer, data storage, and internet have evolved literature sources to various electronic formats which are frequently linked with each other. Electronic linkages among literature sources have immeasurably increased the functionality of information resources. The textbooks are being linked to the full text of their references. Research article text links to the study's original data; index/abstract databases are linking to the full-text articles they index. Aggregation of literature sources, indexing, and their interlinking has resulted into different types of electronic databases, viz. bibliographic, abstracting, and citation.

Isolating the relevant literature in a short time from huge collection of literature sources is the crucial goal of researchers and other information seekers. Extracting the needed information from documents containing potential data is an arduous task due to information overload and time-consuming efforts [1]. Institutions/organizations have their self-hierarchies, regulations, and procedures, which are subjected to frequent expansion and change. Everyone desires to get accurate and timely information as soon as they can. Hence, to deal with a huge and expanding unstructured literature and data, automated, and semi-automated bibliographic databases equipped with search paradigms for information retrieval have been developed [2].

Online literature searching starts with conversion of query into specific keywords very carefully using natural language. Even then, search results from databases or search engines are frequently associated with nonrelevant/least relevant items. The primary issue is either the lack of adequate training and skill of the information seeker for right choice of keywords and how these keywords are connected together to build a search query. Searching electronic databases through keywords has improved considerably with subject specific or proprietary thesauri especially developed for improving the quality of search results. This chapter starts with different types and sources of literature. On the basis of experience in the field and as per latest trends, authors have compiled some significant online literature sources for more relevant search results. Comprehensive literature searching and guidance for electronic search, in two different sections of this chapter, aim to provide stepwise directions to novice researchers for effective searching of online databases.

14.2 Literature

Literature is categorized into primary, secondary, and tertiary literature. However, the differentiations among these three may be ambiguous due to their use and differences among different disciplines, viz. sciences, humanities, and social sciences. One literature may be a primary source on one occasion while the same may be secondary in other instance. A newspaper article may be both a primary and secondary literature.

14.2.1 Primary Literature

Primary literature contains the first written accounts of full descriptions of original research. Journals, patents, technical reports, thesis, dissertations, lab notebooks, preprints, and conference proceedings belong to primary literature. All of these are regarded as the best source of information on current topics. In terms of size, the primary literature is probably larger than either the secondary or tertiary literature. Newspaper and periodicals also ture in the primary literature but sometimes these may be categorized in secondary literature as per nature and use of the information.

The first communication of research data and ideas may or may not be private. If there is a financial or proprietary interest in the research, the first communication may be in the form of a patent application. Before the research's formal publication, the work may be presented as a paper or poster at a professional meeting or conference. The current evolution or transition of the journal from a paper-only publication to a paper-and-electronic publication is a significant change in the journal literature. In the electronic format, the lag time between submitting a written research report and its publications can be shortened considerably. Electronic preprint, either from the author or from the journal publisher, places the article before the reader several months ahead of paper publication release, albeit without the benefit of formal full peer review.

14.2.2 Secondary Literature

Secondary literature are publications, like review articles, abstracts, textbooks, indexing/abstracting periodicals, etc., wherein information from primary literature are collected, compiled, indexed, interpreted, and sometimes critically analyzed or commented. Existing state of the art on any topic is best described in secondary literature. Reviews, books, and treatises are secondary literatures which best describe existing state of art on a topic. Different theories, principles, ideas, and their comparisons are best to find in secondary literatures.

14.2.3 Tertiary Literature

Tertiary literature is a publication which contains information compiled or analyzed from secondary literature. Most common tertiary literature are compilations, bibliographies, encyclopedias, and handbooks. Being farthest from the primary literature, the tertiary literature characteristically is the least current and the most vulnerable to misinterpretations, biases, and inaccuracies. But just as characteristic, the tertiary is the most accessible, most comfortable to use, and perhaps the most used of all information resources. Tertiary literature is regarded as the best place to find overview on a subject or topic.

14.3 Significant Literature Sources

14.3.1 Databases

Database may be defined as the organized collection of data or structured information that is stored in a computer system for easy access, update, and retrieval. Databases are designed to offer convenience and speed for retrieving scholarly literature through the use of keywords. Databases are either freely accessible or may be accessed by pay per view or by subscription [3]. Databases contain literatures and data in different formats like metadata, html full text, full-text pdf, html abstract with links to full text, or a simple bibliography with links to abstracts and full texts, etc.

Bibliographic databases are electronic collection of references to published literatures like journals, conference proceedings, reports, books, newspapers, etc. Bibliographic databases may cater to the need of specific academic discipline like pharmaceutical sciences, for example, International Pharmaceutical Abstracts, or it may be broad in scope, for example, PubMed. Many of bibliographic databases are proprietary and available through licensing direct from indexing/abstracting agencies or through licensing/subscription through vendors. Some bibliographic databases have evolved with time as online digital libraries which also provide access through pay per view/download. Some bibliographic database also records citations which allows user to see how many times a particular publication/author/institution has been cited by others. These citation databases, for example, Web of Science, Scopus, Google Scholar, allow comparative assessment of publications, researchers, and institutions through various citation metrics like h-index, i-10 index, impact factor, etc. Some databases used for storing literatures are listed in Table 14.1.

World Health Organization (WHO), in association with other organizations, institutions, and publishers, has run different programs, collectively known as Research4Life, for improving access to high quality peer-reviewed contents to researchers in lower and middle income countries (Table 14.2). HINARI (Research for Health), ARDI (Research for Innovation), OARE (Research in the environment), AGORA (Research in Agriculture), and GOALI (Research for Global Justice) are five different programs which aim to improve research and teaching in health, life sciences, physical sciences, social sciences, agriculture, and environment sciences. Online journals and database content is available through dedicated webpages of these programs. HINARI and other such programs have the potential to contribute significantly in improving the conditions of health and life in developing countries [4]. For searching literature in pharmaceutical sciences and pharmacology, HIARI is recommended.

Physicochemical properties are notably important while formulating medicines. There are several authentic resources which may be consulted for physicochemical properties of drugs and other chemical [5]. These databases allow textual searching and search by chemical names or CAS registry number or by drawing structure of molecule. Some of these databases are listed in Table 14.3.

Table 14.1 Some databases with their brief description for online literature searching for research projects

Database	Organization maintaining	Brief description	URL
PubMed	NCBI, NIH, NLM	Bibliographic free access database with 32 million citations from biomedical literature; first released in 1996; results include MEDLINE, PubMed Central, and Bookshelf; full-text linkouts	https://pubmed.ncbi.nlm.nih.gov/ or https://pubmed.gov/
MEDLINE (MEDLARS Online)	NCBI, NIH, NLM	Bibliographic free access database with 27 million references to journal articles back to 1946; Launched by NLM in 1964 as MEDLARS Online and free public version through PubMed in 1996; searched via MESH vocabulary on PubMed; largest subset of PubMed database	https://www.ncbi.nlm.nih.gov/mesh/
EMBASE	Elsevier	Bibliographic proprietary database for Biomedical and Pharmacological literature; records from over 8500 journals (1947–Present) including MEDLINE titles; full-text indexing of drug, disease, and medical device data; Emtree subject searching (including MeSH); access via subscription from Elsevier or interfaces like OVID, Dialog, EBSCO, and others	http://www.elsevier.com/online-tools/embase
Cochrane Library	John Wiley and Sons	Contains three bibliographic databases, viz. Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials (CENTRAL), and Cochrane Clinical Answers; founded in 1993; specialist database for evidence-based medicine and health care; MeSH search and PICO Search facility; access via subscription	https://www.cochranelibrary.com/
Google Scholar	Google	Biggest academic database and search engine; searches full-text articles, released in 2004; institutional and	https://scholar.google.com/

(continued)

Table 14.1 (continued)

Database	Organization maintaining	Brief description	URL
		researcher profiles; track citations and calculate metrics h index, i-10 index	
Web of Science	Clarivate	Bibliographic and citation database; Core collection contains ten indexes including Science Citation Index Expanded (SCI-Expanded); Social Sciences Citation Index (SSCI); Arts & Humanities Citation Index (A&HCI); Emerging Sources Citation Index (ESCI); two indexes for conference proceedings, two book citation indexes, and two chemical citation indexes; proprietary multidisciplinary database; paid (subscription-based access); calculate metrics like h index, impact factor; indexing coverage from 1900 to present; contains 1.9 billion cited references from over 171 million records, 10.1 million total Open Access records; 21,100+ unique global journals covering 254 disciplines	https://clarivate.com/products/web-of-science/
Scopus	Elsevier	Abstract and citation database; launched in 2004; records from back to 1788; indexes 1.7 billion cited references dated back from 1970, 82 million+ items, 17 million author profiles, 2.34+ lakh books, 7000+ publishers; calculates citation metrics like CiteScore and h index	https://www.scopus.com/
Crossref	Crossref	Citation database; DOI assigner; 120+ million metadata records and DOIs; free access; launched in 2000	https://www.crossref.org/
Dimensions	Digital Science and Research Solutions Ltd	Citation and research analytics suite; started in 2016; contains 860 million academic citations, 106 million publications with over 1.2 billion citations openly accessible, 124 million	https://www.dimensions.ai/

(continued)

Table 14.1 (continued)

Database	Organization maintaining	Brief description	URL
		documents and 86 million articles and books and 34 million patents; citation-based indicators and Altmetric attention scores; artificial intelligence based	
International Pharmaceutical Abstracts	Clarivate	Abstracting and indexing database for comprehensive coverage of worldwide pharmacy literature; contents include 501,000 records, 800 abstracted and indexed global journals dated from 1970 to present; access by subscription or through vendors like OVID; Dialog, Proquest, etc.	https://clarivate.com
Biological Abstracts	Clarivate	Abstracting and indexing database curated for biological sciences; began service in 1926; content includes 4300 journals; includes BIOSIS indexing and MeSH terms; proprietary database; part of web of science database and also available through EBSCO	https://clarivate.com/webofsciencegroup/
Biosis Previews	Clarivate	Abstract and citation indexing; database covers preclinical and experimental researches; covers data from 1926 to current with 5363 journals and 27,905,411 records; includes BIOSIS indexing and enhanced MeSH; disease terms; access through Web of Science or other platforms like Dialog or vendors like EBSCO	https://clarivate.com/webofsciencegroup/solutions/webofscience-biosis-previews/
SciFinder (Chemical Abstracts)	CAS (American Chemical Society)	A curated database of chemical and bibliographic information; comprehensive chemical literature searching; provides access to chemical abstracts via SciFinder; launched in 1995 and web version was released in 2008. More than 54 million records from publications; dating back to the early 1800s, with continuous coverage since	https://www.cas.org/scifinder-discovery-platform

(continued)

Table 14.1 (continued)

Database	Organization maintaining	Brief description	URL
		1907; allows searching with CAS registry numbers, chemical structures, substructures, or reactions; access via direct subscription from CAS	
CAB Abstracts (SciFinder)	CABI	Bibliographic database for agriculture, environment, and all related applied life sciences; SciFinder Discovery Platform includes Scifinder^{n}, Formulus and Methods Now; access to over 10 million research records from 1973 onwards; indexing with CAB Thesaurus for easy searching and relevant results; access through CAB Direct, Web of Science, OVID, Dialog, and other vendors	https://www.cabdirect.org
Global Health	CABI	Abstracting and indexing database for publish health research and practice; containing more than 4.3 million records including over 128,000 full-text articles, access to over 400 book chapters, over 190 reviews, and over 1000 news records from 1973 to present; access via subscription through CAB Direct or through Web of Science or vendors like Ovid and EBSCO	https://www.cabdirect.org/
PubMed Central	NCBI, NIH, NLM	Free digital repository that archives open access full-text scholarly articles from biomedical and life science journals; contains 6.9 million articles; started in 2000	https://www.ncbi.nlm.nih.gov/pmc/
Mendeley	Elsevier	Crowdsourced database; launched in 2008; contains 100 million articles; reference managing online and offline software; helps to integrate either Word, OpenOffice, and LaTeX; free access and subscription plans	http://mendeley.com/

Table 14.2 WHO initiatives for free or discounted access to journal articles and other peer-reviewed contents in developing countries

Database/platform (organization)	Brief description	URL
Research4Life (WHO, STM Publishers, and other organizations)	Institutions from lower and middle income countries gets; free access or low cost online access to academic and professional peer-reviewed content; full-text access via login id and password; simple search and advanced search options	https://www.research4life.org/
HINARI (WHO in association with publishers and other organizations)	Content from 170 publishers and about 85,000 information resources are available either free or at low cost; program started in 2002; access via HINARI program page	https://www.who.int/hinari/en/

Table 14.3 Some database for seeking information on physical and chemical properties of drugs

Database	Brief description	Ref. and URL
PubChem (NCBI, NLM, NIH)	Search engine for chemicals and drugs; drawing of chemical structure; availability of smiles, pharmacokinetic, and pharmacodynamic information of the drug; open access	https://pubchem.ncbi.nlm.nih.gov/
Merck Index (Royal Society of Chemistry)	Authoritative information on chemicals; 1500+ monographs, 500+ organic name reactions and reference tables; access by subscription	https://www.rsc.org/Merck-Index/
Chemspider (Royal Society of Chemistry)	Provides chemical structure and information of drugs in free of cost; availability of QSAR of drugs; free access	http://www.chemspider.com/
Reaxys (Elsevier)	Database of retrieval of information of drugs (like chemical structure, reactions, properties of the drug, etc.) from different journals, patents, etc.; access by subscription from Elsevier	https://www.reaxys.com/

14.3.2 Online Digital Library Platforms

Most of the scientific literature is published online by various Scientific, Technical and Medical Publishers (*STM publishers*) [4]. Some publishers host journal-specific websites while others group journals together on single/multiple online platforms. Online digital libraries are the electronic platforms of publishers which provide full-text access to all types of literatures, including research articles, review articles, books, encyclopedia, etc. Researchers can visit the website/platform of the publisher and for accessing the available contents. Researchers need to enter keywords in search boxes for quick results which may be refined by year, article type, subject areas, and publication titles. Optionally, advanced search may be used for finding more relevant results which also work by limiting search results to a journal or date of publication. The details of the journal like impact factor, indexing details, guidelines for article submission, etc. are hosted at other locations but it may be linked in with the journal page on the platform [5]. Researchers can access some

Table 14.4 List of online library/platform of journal and book publishers

Online library/platform (publisher)	Contents	URL
Sciencedirect (Elsevier)	About 4 million articles; 4000+ academic journals; 30,000+ e-books	https://www.sciencedirect.com/
Springerlink, Nature, BioMedCentral and SpringerOpen (Springer Nature)	7 million+ articles; 3000+ journals; 300,000+ books, 200+ series	https://link.springer.com/; https://www.nature.com/; https://www.biomedcentral.com/; SpringerOpen.com
Wiley Online Library (Wiley)	1600+ journals; 250+ reference works; 22,000+ online books	https://onlinelibrary.wiley.com/
Taylor and Francis Online (Taylor & Francis)	Contains 4,596,000+ articles	https://www.tandfonline.com/
ACS Publications (ACS Publications)	1,300,000+ research articles; 100,000+ news stories; 35,000+ book chapters; 1000+ references and standards	https://pubs.acs.org/
SAGE Journals (SAGE)	1000+ journals	https://journals.sagepub.com/
Lippincott Research (Wolters Kluwer Health)	300+ journals	https://journals.lww.com/
Eurekaselect (Bentham Science Publishers)	100 journals in both electronic and printed formats	https://www.eurekaselect.com/

content free of cost (open access and promotional) while other content is kept behind paywalls (pay per view or through library subscription). Some publishers also provide an option of sending eTOC (electronic table of contents) to researchers after registration on their platforms. Some major online library platforms of publishers in pharmaceutical sciences are provided in Table 14.4.

14.3.3 Online Journal Databases/Platforms/Search Engines

Open access initiative has increased the number of online journals which are providing their content freely on the internet. There are various governmental and nongovernmental platforms, like J-STAGE, SCIELO, JOL, etc., which support publication of online journals and also provide full-text access to their contents (Table 14.5). Search engines on these platforms and custom search engines like JURN are making the literature searching easier for open access articles.

14.3.4 Patent Databases

Patent is the primary literature which is defined by World Intellectual Property Organization (WIPO) as "A patent is an exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing

Table 14.5 Online journal databases/platforms/search engines for searching open access articles

Open access electronic library/ platforms/search engine	Organization maintaining	Brief description	URL
Directory of Open Access Journals (DOAJ)	DOAJ, nonprofit organization	Launched in 2003; 16,231 journals and 5,945,132 article records; all articles are open access	https://doaj.org/
J-STAGE	JST (Japan Science and Technology Agency)	Since 1999 J-STAGE support Japanese societies and research organizations in publishing electronic journals; 5,217,102 articles, 3278 titles, and 25 subject areas; over 80% content is free to read	https://www.jstage.jst.go.jp/
Scientific Electronic Library Online (SCIELO)	Sao Paulo Research Foundation	Established in 1997, this platform hosts 1200+ journals	https://scielo.org/en/
African Journals Online	Sida, INASP, and other partners	Launched in 1988; it hosts 526 journals with 187,634 abstracts and 181,309 full-text articles; most of the content is open access	https://www.ajol.info/index.php/ajol
Bangladesh Journals Online	Bangladesh Academy of Sciences and Ubiquity Press	Established in 2007 by INASP; an online platform for Bangladeshi journals	https://www.banglajol.info/
JURN	Established by David Haden	Established in 2009 as a custom search engine for finding OA journals of arts and humanities. Expanded in 2014 to science, biomedical, and other fields	http://www.jurn.org/

something, or offers a new technical solution to a problem" [6]. Patent grants an exclusive right to the patent holder for commercializing it for 20 years from the date of filing and derive profits out of it. This right is exercised through licensing of the patent to manufacturers or other companies. Hence, most companies prefer not to publish their innovations in journals or conference proceedings. Excluding patent search from the literature search may result in a literature gap, sometimes. Most of the countries publish patents and patent applications in open domain through their patent offices which allow searching and free full-text access via websites. Some publicly available free patent databases are listed in Table 14.6. While commercial databases have a host of other options like patent analysis, portfolio analysis, citation in addition to full-text patent searching (Table 14.7) [7]. Readers may refer to Chap. 15 for more details on patent searching and reference for learning basic concepts on patent searching.

Table 14.6 Free patent databases at a glance

Database	Organization maintaining	Brief description	URL
Patentscope	World Intellectual Property Organization	Access to international Patent Cooperation Treaty (PCT) applications in full-text format on the day of publication; nongovernmental database	https://patentscope.wipo.int/search/en/search.jsf
esp@cenet	European Patent Office	Over 120 million patents from 1973; search available for full text and all data; patent from European and other countries (which other mention here either name or number of other countries); open access	https://worldwide.espacenet.com/
USPTO PatFT/ AppFT	US Department of Commerce	Available from 1970; patent numbers and/or classification codes required for the patents of 1970 to 1976; country-specific governmental database; open access	https://uspto.gov/
J-PlatPat	Japan Patent Office	Available from 1922; achieved its goal of reducing the amount of time it takes to issue first actions to 11 months; country-specific governmental database, open access	https://www.j-platpat.inpit.go.jp/
Patent Search and Service System	China National Intellectual Property Administration	Supports nine languages, with a collection of patent data from more than 100 countries, regions, and organizations; medicine search facility; country-specific governmental database, open access	http://pss-system.cnipa.gov.cn/sipopublicsearch/inportal/i18n.shtml
KIPRIS Patent Search	Korea Institute of Patent Information (South Korea)	Access to full-text Korean patents including PDF, TIFF formats; country-specific and governmental database	http://eng.kipris.or.kr/enghome/main.jsp
inPASS	Office of the Controller General of Patents, Designs and Trademarks (India)	Provides comprehensive search and analysis of patent literature from 1856; country-specific and governmental database	https://ipindiaservices.gov.in/publicsearch
Google Patents	Google	Search engine from Google, indexes patents and patent applications from 17+ patent offices	https://patents.google.com/

(continued)

Table 14.6 (continued)

Database	Organization maintaining	Brief description	URL
Lens.org Patent Search	Cambia (Australia); a nonprofit organization	Patents of Europe, Australia, the USA, and WIPO PCT applications can be searched	https://www.lens.org/lens/search/patent/structured?preview=true

14.3.5 Online Clinical Trial Registries

Clinical trials are the primary source of information for first-in-human interventional studies. Before the actual conduct of clinical trials, the details of these trials are provided on an open access platform for enrollment of subjects, ensuring transparency and disclosure of essential information to the public. As per international ethical concerns, registering human studies also on online clinical trial registries have been made mandatory. Sponsors of clinical trials register their clinical trials on these country-specific online registries. Most of these clinical trials registries (Table 14.8) are accessed free of cost [8]. These registries contain details of the clinical trial, which include general description, current phases of trial, inclusion and exclusion criteria, number and gender of volunteers, type of the study, information about intervention (applicable for interventional studies) or observation details (applicable for observational studies), results of the study, etc. Researchers can search clinical trial by entering specific keywords like disease, country, trial identifier number, etc. This search can be made more specific with advanced search option [9]. Through advanced search option researchers can further limit results input study type, recruitment status, inclusion or exclusion criteria, phases, name of funding agency, dates of initiation and completion, etc.

14.3.6 Drug Regulatory Agencies and Other International Organizations

There are several drug regulatory agencies and international organizations that govern drug product registration, producing, distribution, control, marketing, analysis and development, international trade, and intellectual property protection [10]. Drug regulatory agencies, for example, FDA and EMA, examine the application for new drugs or drug products for granting marketing authorizations in their territorial jurisdictions. WHO and ICH frame harmonized guidelines for quality, safety, efficacy, and other issues with medicines. WHO and WTO support the member countries (United Nations) for universal health coverage and international trade, respectively. Regulations, guidelines, approved drug products; NDA, ANDAs, INDs, etc. may be accessed from country-specific regulatory agencies. Most of the content is available free of cost from the websites. However, some content may be restricted. Harmonized guidelines are available from ICH and WHO.

Table 14.7 Some paid databases for retrieval of information from patents

Database	Organization maintaining	Brief description	URL
Derwent World Patents Index	Clarivate	Most comprehensive patent database; analyze, abstract, and manually index patents from 59 worldwide patent authorities; invention-based records; capability to search chemical patents through chemical patent index; forward and backward patent citation search; patentability search, state-of-the-art search, freedom to operate and product feature; clearance, part of Web of Science core collection; access through Web of Science	https://clarivate.com/derwent/solutions/derwent-world-patent-index-dwpi/
PatentPak (SciFinder)	CAS (American Chemical Society)	Provides access to ~18 million full-text patents searchable from 46 patent offices; integrated and access through SciFinder and STN	https://www.cas.org/products/patentpak
Patseer	Gridlogics	Coverage of 108+ patent authorities and 96M+ full-text records; powerful patent analysis tool with global searching and portfolio valuation; specialized data normalization and indexing techniques; chemical search via chemical lookup; artificial intelligence driven natural language processing engine for semantic search; portfolio and quality level metrics; access by subscription	https://patseer.com/
PatBase	Minesoft Ltd and RWS Group	Patent data from 106 patent-issuing authorities; full-text patents from 74 countries; patent searching and analysis of patent data; interactive citation explorer; litigation status; portable patent assignee search; access by subscription	https://www.patbase.com/
TotalPatent One®	LexisNexis	Search patent documents from 100 countries; full-text patents from 31 countries; semantic search, inbuilt analytics tools; patent portfolio; licensing research; citation analysis, litigation links, Scopus abstract	https://www.lexisnexisip.com/products/totalpatent-one/

(continued)

Table 14.7 (continued)

Database	Organization maintaining	Brief description	URL
		for prior art search, access via subscription	
Patsnap	Patsnap	Patentability search, state-of-the-art search, freedom to operate and assessment search; explore 114 million chemical structures linked to 125+ million patents; analytical report, licensee report, citation tracker, technology and competitor workspaces; patent landscape, patent portfolio audit, competitive intelligence, and benchmarking; access by subscription	https://www.patsnap.com/
Questel	Questel	Database of 100 million patents from 45 patent authorities; patentability search, patent invalidity search, freedom to operate; patent analysis, patent portfolio; patent data analysis; access direct purchase from Questel	https://www.questel.com/
World Traditional Medicine Patent Database	Beijing East Linden Science and Technology Co., Ltd	Database of 200,000 patents related to natural medicines or extractions, issued since 1985, from two international organizations and 20 countries; bilingual indexing (English and Chinese); formula search, formula similarity search, chemical search, natural medicine search and synonym expansion features	http://www.eastlinden.com/product_detail.aspx?nid=5&id=22
WIPS Global	WIPS Co., Ltd	AI-powdered search for patent information	https://www.wipsglobal.com

Some drug regulatory agencies and international organizations are listed in Table 14.9.

14.3.7 Institutional Repositories

Institutional repository is a library containing digital objects and metadata from an educational or other institution [11]. DSpace, Eprints, and DigitalCommons are some of the most popular repository software (Table 14.10). Since 2000, several

Table 14.8 Online clinical trial registries

Name	Organization	Brief description	URL
International Clinical Trials Registry Platform	WHO	Established in 2005; updated with contents from 17 clinical trial registries; largest database	https://www.who.int/clinical-trials-registry-platform
ClinicalTrials.gov	Nation Library of Medicine, US	Launched in 2000; studies listed in the database are from 50 US states and 220 countries	https://clinicaltrials.gov/
EU Clinical Trials Registry	European Union	Clinical trial data 2004 onwards; list of clinical trials of European Union (EU), or the European Economic Area (EEA) and some other countries	https://www.clinicaltrialsregister.eu/ctr-search/search
ISRCTN	BioMedCentral	Clinical trial registry recognized by the WHO and ICMJE. Registries of clinical trials from all countries; launched in 2000	https://www.isrctn.com/
CTRI	ICMR-National Institute of Medical Statistics	Register of clinical trials conducted in India; register of clinical trials conducted in other countries, which do not have a primary registry; launched in 2007	http://ctri.nic.in/Clinicaltrials/login.php
Chinese Clinical Trial (CHiCTR)	West China Hospital, Sichuan University, China	Established in 2005; registers clinical trials conducted in China and other countries	http://www.chictr.org.cn/
Brazilian clinical trial registry	Pan-American Health Organization	Launched in 2010; register of clinical trials conducted in Brazil	http://www.ensaiosclinicos.gov.br/
Japan Primary Registries network	Japan National Institute of Public Health	Since 2007, registers clinical trials in Japan. Comprises of three registries: UMIN-CTR, JMACCT, JAPIC	https://jrct.niph.go.jp/
Pan African Clinical Trials Registry (PACTR)	South African Medical Research Council	Since 2007 PACTR registers all clinical trial conducted in Africa	http://www.edctp.org/pan-african-clinical-trials-registry/#
Australian New Zealand Clinical Trials Registry	NHMRC and TGA	Register of clinical trials being undertaken in Australia, New Zealand, and elsewhere since 2007	http://www.anzctr.org.au/

institutions have developed their own software for managing the structure and contents of these repositories [12]. Many institutions host institutional repositories for providing online access to their dissertations, thesis, and other publications. Appropriate metadata can be added for each paper, and the researcher and an

Table 14.9 List of some international organizations and their brief description

Name of organization and headquarter	Brief description	URL
World Health Organization (WHO), Geneva, Switzerland	United Nations specialized agency for primary health care, revamping the access of essential health and drug products, and prevention of noncommunicable diseases	https://www.who.int/
International Conference on Harmonization (ICH), Geneva, Switzerland	International association, a legal entity under Swiss law, for interpretation and implementation of technical guideline for safety, efficacy, and quality of drug and drug products	https://www.ich.org/
Pan American Health Organization (PAHO), Washington, DC, USA	International health agency for the Americas which promotes health of public. Work for the prevention of communicable and noncommunicable diseases; works in emergency condition	https://www.paho.org/en
World Trade Organization (WTO), Geneva, Switzerland	Intergovernmental organization for implementing and reviewing the mechanism of trade policy, handling disputes of trade and connects with other international associations	https://www.wto.org/
Food and Drug Administration (FDA), Maryland, USA	Drug regulatory agency of the USA; regulation of food, medical devices, vaccines, blood, biologic products, cosmetics and tobacco products; post-marketing surveillance	https://www.fda.gov/
European Medicines Agency (EMA), Amsterdam, The Netherlands	Drug regulatory agency of European Union; monitors safety and efficacy of a drug product; promotes modification of drugs; stores information about the drug, drug products, blood, biologics, etc.	https://www.ema.europa.eu/en
Pharmaceuticals and Medical Devices Agency (PMDA)	Drug regulatory agency of Japan; ensuring safety, efficacy of drug products and medical devices; publishes Japanese Pharmacopoeia; post-marketing surveillance	https://www.pmda.go.jp/english/
Central Drug Standards and Control Organization (CDSCO), New Delhi, India	Drug regulatory agency of India; new drug testing and approvals; monitoring clinical trials; amendment to Drugs and Cosmetics Act and Rules	https://cdsco.gov.in/
National Medical Product Administration (NMPA), Beijing, China	Drug regulatory agency of China; regulation of registration of drugs, medical devices; giving marketing	http://english.nmpa.gov.cn/

(continued)

Table 14.9 (continued)

Name of organization and headquarter	Brief description	URL
	approval; emergency approval of drugs	
Korea Food and Drug Administration, North Chungcheong Province, South Korea	Drug regulatory agency of South Korea; monitoring safety of foods and genetically modified foods; monitoring labeling of food and drug products	https://www.mfds.go.kr/eng/index.do
United States Pharmacopeia, Rockville, Maryland, USA	Independent, scientific nonprofit organization which publishes USP-NF, USP Compounding Compendium, Food Chemicals Codex, Dietary Supplements Compendium, Pharmacopoeial Forum, etc.; prepare and sale reference standards; conducting performance verification test for USP authorized equipment	https://www.usp.org/ and https://www.uspnf.com/
European Directorate for the Quality of Medicines and Healthcare (EDQM), Strasbourg, France	EDQM is a European government organization responsible for quality standards for safe medicines; provides policies for the safe application of medicines; promotes the safety and ethics of blood transfusion	https://www.edqm.eu/

author's documents can be easily grouped. Such repositories are searchable by Google Scholar and compatible with other search engines too. Most of these repositories can be accessed freely [13]. Country-specific initiatives like American Doctoral Dissertations, British Library EThOS, Theses Canada, ShodhGanga (India), and Diva-Portal (Sweden) allow storage of e-thesis/dissertations eliminating the need to establish and maintain repository at each institution. OpenDOAR indexes open access repositories of educational and research institutions. Digital Commons Network (Elsevier/Bepress), PDQT Open (ProQuest), EBSCO Open Dissertations, and Networked Digital Library of Theses and Dissertations (NDLTD) are some resources which may be used for searching thesis/dissertations of many countries.

14.3.8 Miscellaneous Categories

Beside primary literatures, secondary and tertiary literatures are also very much important for understanding the concepts of a research topic or study. Dissertations/thesis, newspapers, and magazines are some other gray literature which are considered important for comprehensive reviews and systematic reviews. Google Books provide previews of the books where some chapters are available without any cost

Table 14.10 Institutional repositories or platforms with their salient features

Repository software/ platform	Features of the software/ repository	Universities/organization	URL and Ref.
DSpace	Limited Flexible Repository Structure, customizable metadata, hosting solution, locally installed software, community support	University of Cambridge	https://www.repository.cam.ac.uk/
		Massachusetts Institute of Technology	https://www.mit.edu/
		Harvard University	https://www.harvard.edu/
		IIT-BHU	http://idr-lib.iitbhu.ac.in:8080/xmlui/
EPrints	Flexible repository structure, hosting solution, customized metadata, customizable repository design	University of Southampton	https://www.southampton.ac.uk/
		Fred Hutchinson Cancer Research Center	https://www.fredhutch.org/en.html
		California Institute of Technology (Caltech)	https://www.caltech.edu/
Digital Commons	Flexible repository structure, simple and qualified dublincore metadata, customized metadata, customizable repository design	Georgia State University	https://www.gsu.edu/
		Ohio University	https://www.ohio.edu/
Islandora	Flexible repository structure, hosting solution, customized metadata, customizable repository design	California Institute of Technology (Caltech)	https://www.caltech.edu/
		Johns Hopkins University	https://www.jhu.edu/
Fedora	Limited flexible repository structure, hosting solution, customizable metadata, customizable repository design	Oxford University Research Archive (ORA)	http://ora.ox.ac.uk/
		University of Queensland	https://www.uq.edu.au/
Diva-Portal	Search tool for thesis and research publications	49 universities and research institutions (mostly located in Sweden)	http://www.diva-portal.org/
Shodh Ganga	Thesis/dissertation repository with browsing facility with keywords, title, researcher/guide	476+ Indian universities	https://shodhganga.inflibnet.ac.in/
ProQuest Dissertations and Theses	Largest repository of graduate dissertations and theses	5 million works; grows by 200,000 each year; across 100 countries	https://pqdtopen.proquest.com/

(continued)

Table 14.10 (continued)

Repository software/ platform	Features of the software/ repository	Universities/organization	URL and Ref.
Global™ and PDQT Open			and https://www.proquest.com/
DART-Europe E-theses Portal	Thesis of European countries can be searched free of cost	University College London, Trinity College Dublin, Oxford University, and Dartington College of Arts and others	https://www.dart-europe.org/
NDLTD Global ETD Search	Houses 6,007,017 electronic theses and dissertations from various countries; Global ethesis search; NDLTD harvests metadata from participating repositories	Networked Digital Library of Theses and Dissertations	http://search.ndltd.org/
Digital Commons Network	Thesis and other full-text scholarly articles from hundreds of universities and colleges worldwide	Elsevier (Bepress)	http://network.bepress.com/
EBSCO Open Dissertations	American doctoral dissertations and dissertations/thesis from other countries	EBSCO and BiblioLabs	https://biblioboard.com/opendissertations/
EThOS	Open access of thesis from UK	British Library	https://ethos.bl.uk/
Theses Canada	Provides open access to Canadian universities thesis	Library and Archives Canada	https://www.bac-lac.gc.ca/eng/services/theses/Pages/search.aspx
OpenDOAR	Simple and advanced search options for finding repositories by country, software, and various other means	Global Directory of Open Access Repositories	https://v2.sherpa.ac.uk/opendoar/

[14]. Several important platforms for searching books, magazines, and newspapers are provided in Table 14.11 and for software for digital publishing of magazines readers are directed elsewhere [15].

Images and videos also play an important role in illustrating the scientific concepts in research. Some image and video searching platforms (Table 14.12) are extremely useful when illustrations are required for working on presentations or writing thesis or research reports [16].

Table 14.11 Options for searching information from books, newspapers and magazines

Database/website (organization)	Brief description	URL
Google Books (Google)	Books are available in full view, preview, snippet view, and no preview format; free access	https://books.google.co.in/
Hathi Trust Digital Library (Hathi Trust)	Preserving 17+ million digitized items; free access	https://www.hathitrust.org/
ProQuest Book Search (Proquest)	Contains more than 810,000 ebooks and 1350 imprints; access by subscription	https://www.proquest.com/
Kindle eBooks (Amazon)	Require Kindle or Kindle reading application, access by purchase	https://www.amazon.com/b?ie=UTF8&node=2735182011
Google News (Google)	Include news from different fields like science, technology, sports, films, business, etc. of India as well as whole world; capable to search some pharmaceutical magazines like Pharmaceutical Technology; free access	https://news.google.com/
Microsoft News (Microsoft)	Developer allows to subscribe only some specified sources of news; free access	https://www.msn.com/en-in/news
Magzter (Magzter)	Provides newspapers and magazines of 3400 publishers in 50 languages from more than 175 countries, access by purchase or subscription	https://www.magzter.com/

Table 14.12 Some image and video search platforms

Platform/database (organization)	Brief description	URL
Google Image Search (Google)	Image search simple and advanced; free access	https://www.google.co.in/imghp?hl=en&ogbl
Bing Image Search (Microsoft)	Image search; free access	https://www.bing.com/images
TinEye (Idée, Inc.)	Reverse image search; free access	https://tineye.com/
Yandex Image Search (Yandex LLC)	Provides image, video, news and browsing	https://yandex.com/images/
YouTube (Google)	Video search, free access	https://www.youtube.com/
Vimeo (IAC)	Availability of live streaming, advertisement-less video, screen record; paid access	https://vimeo.com/

14.4 Literature Search Methods

Bibliographic database searching and supplementary search methods are two essential components of a comprehensive literature search [17].

14.4.1 Bibliographic Electronic Databases

For systematic reviews and meta-analysis of clinical trials, a comprehensive search requires the use of Cochrane Clinical Trials Register or at least two electronic databases, like PubMed or Embase, and a minimum one search method for unpublished clinical trials such as clinical trial registers, conference abstracts, theses, and contact with experts in the field [18]. Also, a comprehensive literature search should not be restricted to English language only. It is a common perception among scientists that no literature review is comprehensive as it is not possible to retrieve all literature on a topic. An optimal biomedical literature search should combine at least MEDLINE, Embase, Web of Science, and Google Scholar to achieve an overall recall of 98% or higher [19]. Table 14.13 briefly introduces significant and popular bibliographic databases for a comprehensive literature search and literature search for systematic reviews and meta-analysis of the literature. The minimum requirement which we can see here is one subject database and at least two general databases. When searching on literature for Chemistry, SciFinder may be coupled with Web of Science (Science Citation Index) and Scopus [20]. Beilstein (organic) and Gmelin (inorganic) in addition to SciFinder may be used for specific interest in a specific chemical compound. Similarly a subject specific search engine for analytical chemistry is Analytical Abstracts. In the area of clinical sciences like pharmacology and pharmacokinetics, it is recommended that Cochrane or Embase be used with other databases like Web of Science/Scopus/PubMed and Google Scholar. NICE (National Institute of Health Care and Excellence) recommends MEDLINE, Embase, Cochrane Database of Systematic Reviews (CDSR; Cochrane Reviews), Database of Abstracts of Reviews of Effects (DARE), Cochrane Central Register of Controlled Trials (CENTRAL; Clinical Trials), Health Technology Assessment (HTA) database (Technology Assessments) as core databases for literature searching on clinical evidences of pharmacological interventions [21]. Short clinical queries may be addressed through Google Scholar as it can find more relevant articles than PubMed and provide greater full-text links to articles [22]. PubMed and Google Scholar can be used synergistically if access to proprietary databases is an issue. International Pharmaceutical Abstracts are recommended for literature requirements in all fields of pharmacy and pharmaceutical sciences.

14.4.2 Searching Supplementary Literature

Apart from electronic database searching, research whether published or unpublished may also be found with one or more search methods suggested in this section.

14.4.2.1 Searching Clinical Trial Registers

Trials can be identified by searching online clinical trials registries that exist (Table 14.8). It can be a particularly useful approach to identifying unpublished or ongoing trials. Many of the registers are available on the internet and some of the larger ones, such as International Clinical Trials Registry Platform (ICTRP; https://

Table 14.13 Comprehensive and systematic reviews published in 2017–2021 reflecting the use of different bibliographic databases

Databases	Type of study	Journal and year	Ref.
SciFinder, Google Scholar and PubMed	Comprehensive review	Medicinal Chemistry 2017	[23]
PubMed, Google Scholar, Web of Science and CNKI	Comprehensive review	Phytomedicine 2018	[24]
Web of Science, SciFinder, PubMed, Elsevier, Baidu Scholar (Chinese), and CNKI (Chinese)	Comprehensive review	Pharmaceutical Biology 2018	[25]
PubMed, Embase, Cochrane, and Scopus databases	Comprehensive review, clinical	Daru 2019	[26]
International Pharmaceutical Abstracts, EMBASE, PubMed, OVID, Scopus, Google, and Google Scholar	Comprehensive review, clinical	The Journal of Pharmacy and Pharmacology 2020	[27]
SciFinder (database for the chemical literature) CAS (Chemical Abstract Service) search, web of science, Marin Lit (marine natural products research) database	Comprehensive review	Biomolecules 2021	[28]
PubMed, EMBASE, Cochrane Libraries, Cumulative Index to Nursing and Allied Health Literature, International Pharmaceutical Abstracts, Google Scholar. Gray literature was identified through use of the Canadian Agency for Drugs and Technology	Systematic review and meta-analysis, clinical	Vaccines 2016	[29]
PubMed, Scopus, and Web of Science	Systematic review and meta-analysis	Pharmaceutics 2018	[30]
Google Scholar, PubMed, HerbMed, MEDLINE, Science Direct, Scifinder Scholar, Cochrane Library, International Pharmaceutical Abstracts, EMBASE, Biological Abstracts and Commonwealth Agricultural Bureau Abstracts	Systematic review and meta-analysis, clinical	Tropical Medicine and Health 2021	[31]
MEDLINE, EMBASE, CINAHL, Scopus, Google Scholar, ProQuest Dissertations and Theses Global	Systematic review and meta-analysis, clinical	Journal of Pharmacy and Pharmaceutical Sciences 2021	[32]

www.who.int/clinical-trials-registry-platform) and US online clinical trial registry (www.ClinicalTrials.gov), include the facility to search by drug name or by condition. While some registers are disease specific, others collect together trials from a specific country or region. Pharmaceutical companies may also make information about trials they have conducted available from their websites.

14.4.2.2 Searching Patents and Patent Applications

Patent and patent applications provide some of the latest innovations in the field which shall not be ignored for a comprehensive review of literature. There are fair

chances that one cannot find even a single published article on newer innovations protected through patents. Publication in a journal may come at a later stage. Most companies patent to prosper. Table 14.6 provides a list of publically available databases out of which espacenet and Patentscope are recommended for a very quick search. If there is a need for analysis on patent data, exclusively proprietary databases from Table 14.7 may be used. Publicly available databases and proprietary databases provide simple and advanced search options for literature searching.

14.4.2.3 Searching Institutional Repositories for Thesis/Dissertations and Other Contents

One might see thesis/dissertations, preprints of journal articles, postprints of journal articles, books, book chapters, working articles, datasets, reports, conference proceedings, bibliographies stored in academic repositories of institutions. Platforms like Digital Commons Network (Elsevier/Bepress), PDQT Open (ProQuest), EBSCO Open Dissertations, and Networked Digital Library of Theses and Dissertations (NDLTD) are some good resources for finding open access dissertations/thesis (Table 14.10). For other digital contents repositories of individual institutions may be visited.

14.4.2.4 Searching Internet Sources/Web Searching

Internet searching is a useful approach for finding gray literature, like unpublished studies, reports, white papers, regulatory filings, and conference abstracts [33]. Some studies are published more informally than in a journal indexed in a bibliographic database. In such cases, searching the internet is worth considering. A structured search and well documented involve reporting of the website (URL), search terms used, and the date of searching.

14.4.3 Scanning Reference Lists/Bibliography of Key Papers

A comprehensive search should not rely solely on the use of search strings to find relevant papers but it must be complemented with scanning reference lists of key papers (both research and reviews) [34]. The use of reference lists of key papers are also an important source of finding relevant papers. Titles of the paper in the bibliography and the citation in the main body of the paper may be checked for finding relevancy of the paper. Hyperlinked bibliography from either full-text webpages or PDF of papers allows a convenient way to reach to the full-text location of listed references. Searching bibliography through concepts such as co-citation and bibliographic coupling helps in measuring affinity between papers [35]. Co-citation refers to citation of two papers simultaneously in other papers. Bibliographic coupling assumes that two papers that cite a common source (literature) are related to each other and thus it measures affinity between papers.

14.4.4 Citation Searching of Relevant Studies

Citation is referred to as the use of a paper as a reference. Citations are regularly updated in Citation databases like Web of Sciences, Scopus, Google Scholar, etc. Each citation database has its own index of citations which differs from database to database as the database counts only those citations which come from journals indexed in that database only. Citation databases can be used to find more relevant studies through the option of citation searching (like cited by option of Google Scholar and "Cited reference Search" of Web of Science) [23]. In addition to this, some publishers update webpages of papers on journal's websites with new citations as hyperlinked references lists. Citation search is search forward through time which enables its application on older papers but this search strategy cannot be applied to most recent papers [33]. Citation searching is helpful in finding possible search terms, preparing recommended reading lists, and assembling bibliography. Citation searching has its own set of disadvantages also as it can, sometimes, come up with relevant articles from different disciplines.

14.4.5 Contacting Study Authors/Experts/Manufacturers

Research teams or subject experts or manufactures may provide useful information on ongoing and unpublished studies [33]. Further, in cases of missing information for materials, like quality or grades, or protocols in a particular study or any doubt or clarity, one may seek information from authors of the study. Contacting authors of old studies may require resources and time and the authors may not respond due to change in contact details or for other reasons [33]. It is always valuable to have an advisory group for consulting at critical stages of a study. Subject experts may be identified and contacted for more informed advice. Advisory group may be constituted with methodological experts, analytical experts, preclinical studies experts, clinical experts, healthcare professionals, industry professionals, patients, and members from administrative and funding agencies. Advisory groups may also be consulted for identifying missing studies. Manufacturers may be asked for more details about their products which are sometimes used as control in preclinical and clinical studies.

14.4.6 Hand Searching of Pertinent Journals

Some studies do not appear in electronic search results as they lack appropriate search terms or not all relevant search terms were listed in titles/abstracts/keywords either due to inconsistency of indexing by indexers or authors fail to describe them in methodology [36]. Further, there are instances when journals or conference proceedings are not indexed in databases. Some databases do not index letters, editorials, or commentaries from indexed journals [33]. Supplementing electronic searching with hand searching ensures finding of such additional relevant articles

[34]. For hand searching the first step is to identify pertinent journals which contain the highest number of relevant studies. Next step involves manual page-by-page inspection of selected journals by visiting either the brick and mortar library or online library or through the website of the journal.

14.5 Electronic Searching Guidance for Bibliographic Databases

Finding relevant information from database or from the internet is not an easy task. With the availability of millions/billions of documents in a database, the task of finding relevant documents needs considerable skills and experience. However, these skills may be acquired by following steps [37–39].

14.5.1 Step 1: A Well-Focused Research Problem

A well-focused research question is the starting point from where the process of literature searching starts: Framing the research problem with adequate scientific terminology is most important as it brings the research problem into meaningful literal terms. A relevant research question may originate from a very small query of an experimental step or it may come from a long-term research project. Here is one such research query with which one can move forward to next steps of literature searching, for example, "Active Targeting of Docetaxel to Folic Acid Receptors of Cancer Cells."

14.5.2 Step 2: Identify Search Terms or Search Words or Keywords

A keyword is a word or phrase, also referred to as a Search word or search term. that is used to assist an electronic search for information using databases or search engines. Identifying keywords is one of the most important steps in the planning of a literature search. Identified words or phrases are keyed in the databases or search engines to retrieve content relevant to specific topics or questions.

Active Targeting, Folic Acid Receptor, and Cancer may be identified from the well-focused research problem mentioned as an example in Step 1.

14.5.3 Step 3: Find Synonyms for Keywords

Synonyms are words that mean the same thing or are closely related to our main search terms. Synonyms are helpful in expanding search results. Table 14.14 represents some synonyms for identified keywords.

Table 14.14 Synonyms for keywords

Active targeting	Folic acid receptor	Cancer	Docetaxel
	Folate receptor	Lymphoma	Taxotere
		Carcinoma	
		Melanoma	
		Tumor	

14.5.4 Step 4: Searching Literature with Keywords

Keywords are used with Boolean operators for finding relevant information. Boolean operators connect two or more keywords together in an electronic search. Multiple keywords with Boolean operators are used to build a relevant search string which requires intellect of a person performing electronic search.

14.5.4.1 Connect Keywords with Boolean Operators

Boolean operators such as "AND," "OR," and "NOT" are used to connect keywords and the concepts. Boolean operator "OR" gives a broader search with citations containing either of the keywords. When using "OR" in between keywords, the usual practice of placing parentheses around keywords prevents searching of keywords together. Use of "AND" narrows down the search results as the searching looks for the presence of all keywords in the literature, either together or at separate locations. By default most search engines and databases does not require use of "AND" between keywords. It is recommended that in a single search box, Boolean operators shall be capitalized otherwise they may be ignored. "NOT" is used to exclude words that are not required in search results. Hence, Boolean operators are used to find search results with one or more keywords and also exclude one more keywords.

14.5.4.2 Building Search String

Search strings are built with the use of keywords, Boolean operators, and curved brackets. Here are some of these search strings for one such research problem cited as example in Sect. 14.5.1:

for example, (folic acid receptor OR folate receptor) AND (docetaxel OR taxotere) AND (Cancer OR tumor OR lymphoma OR carcinoma or melanoma) AND (active targeting) NOT (skin cancer)

(folic acid receptor OR folate receptor) AND (docetaxel OR taxotere) AND (Cancer OR tumor OR lymphoma OR carcinoma or melanoma) AND (active targeting)

(folic acid receptor OR folate receptor) AND (docetaxel OR taxotere) AND (Cancer OR tumor OR lymphoma OR carcinoma or melanoma)

There can be *n* number of search strings which may be built with judicious use of keywords and Boolean operators. Building such search strings refines the search results to more relevant ones. Using less number of keywords in a search string

increases the number of search results while using more number of keywords reduces search results.

14.5.5 Some Search Tips

Electronic searching in databases may be modified with the use of truncation symbols, wildcards, proximity operators, and phrase searching.

14.5.5.1 Truncation Symbols

Electronic databases use different truncation symbols like the asterisk "*" or question mark "?" or dollar "$" for electronic searching of variations in words resulting in search results having different word endings. These truncation symbols increase the number of search results as they enable different words to be searched together, for example, "pred*" will find results with the words prednicarbate, prednisolone, and prednisone. Truncation requires its application very cautiously. Databases like PubMed searches only the first 600 variations in spelling [39]. Hence, search with "therap*" would not give results for "therapy" as it falls outside the first 600 variations.

14.5.5.2 Use of Wildcards

Wildcards are symbols, like #, ?, *, etc., that represent letter or letters in a word. The use of wildcards in a keyword allows searching of words with minor spelling variations, for example, the use colo?r may give search results for both color and colour. However, specific wild card symbols will vary with the database and require a careful check within the help section of the database.

14.5.5.3 Exact Match or Phrase Searching

Searching for a group of words inside a double quotation mark (phrase) in most of the databases allows an exact match search, for example, "folate receptor for cancer" will search for all four words in the same sequence in all literature available in the database.

14.5.5.4 Proximity Operators

Proximity operators are also referred to as adjacency operators. As precision maximizers they allow to specify how close keywords to be found in relation to each other. PubMed does not offer proximity operators [39]. To apply adjacency, one has to separate search terms with the ADJ operator and a number from 1 to 99. for example, The use of adjacency for "animal adj3 therapy" will retrieve search results animal therapy, therapy using animals, animal-based therapy, and animal assisted play therapy.

14.5.6 Step 5: Refining Search Results

Databases vary for features used to refine electronic searches. However, the common techniques which may be used to narrow down the search results to more relevant ones are:

- Adding more keywords for increasing the relevance.
- The use of indexing thesauri or controlled vocabulary terms to focus search. Articles in PubMed, Embase, and Cochrane Reviews are indexed by MeSH (Medical Subject Headings), Emtree terms, and PICO (Population, Intervention, Comparison, Outcome) terms, respectively, that have their own specific definitions and helps in obtaining more relevant search results.
- Limiting results to the type of study (review, systematic review, books and documents, randomized controlled clinical trial, research article, etc.).
- Limiting the search to a particular period by assigning date or date range.
- Searching for specific type of file (pdf, Word, excel, etc.) or full-text content of the database.
- Searching in particular fields, that is, title, citation, abstract, etc.
- Searching with author names or their affiliations.

14.5.7 Step 6: Adapt and Track Search Results

The key to learn searching is continuous trying till one gets what is required. Through this entire process one shall refine the search and adapt it. It is always better to store or save searches either on personal computer or in the database itself, if the database provides such a facility. PubMed and most subscription-based databases allow saving and tracking of literature searches through a personal account. Databases also provide regular alerts for new researches of a particular field for which one can register. Saved searches allow their rerun, storing of references, and setting up alerts for new researches in the field.

14.5.8 Step 7: Reference Managing

Reference managing software helps in downloading and storing search results on a personal computer or in the cloud. Endnote, Mendeley, and Zotero are some most popular reference management software. Researchers and other information seekers shall learn to use these software for creating libraries, importing/exporting citations, and formatting bibliography lists in word documents.

14.6 Conclusion

The most widely available "information" in today's world is still textual, and literature searching techniques are broadly applicable in facilitating its productive use. PubMed and Medline search with MeSH/Emtree thesauri is a gold standard in literature searching not only in the medical field but also in other biomedical and life science disciplines like pharmaceutical sciences, pharmacology, biotechnology, microbiology, biology, etc. Google Scholar has emerged as a specialized search engine for finding peer-reviewed literature. Proprietary databases like Embase and Cochrane Library are the first choice for literature searching in clinical studies. The emergence of open access journals and their online platforms has motivated traditional journal publishers to provide hybrid options for publishing articles. Access to quality literature has remained a concern for developing countries where HINARI and other initiative of Research4Lifeare providing free access or access at a low cost to peer-reviewed literature. Gray literature on internet is rising with increased demand for its organization and structured storage. Search engines technology is a very dynamic field, always looking for improvements and new ideas to satisfy the demand of searching video, image, chemical structure, chemical equations, mathematical content, and other data types.

References

1. Abdullah M, Al Zamil MGI (2018) The effectiveness of classification on information retrieval system (case study). ArXiv. http://arxiv.org/abs/1804.00566. Accessed 21 Apr 2021
2. Hider P (2007) Information access. In: Ferguson S (ed) Libraries in the twenty-first century: charting directions in information services. Elsevier Inc., pp 145–164. https://doi.org/10.1016/B978-1-876938-43-7.50009-0
3. Patience GS, Pounis G, Patience PA, Boffito DC (2018) Principles of research publication. In: Analysis in nutrition research: principles of statistical methodology and interpretation of the results. Elsevier, pp 197–230. https://doi.org/10.1016/B978-0-12-814556-2.00008-7
4. Johnson R, Watkinson A, Mabe M (2018) An overview of scientific and scholarly publishing. International Association of Scientific, Technical and Medical Publishers. https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf. Accessed 21 Apr 2021
5. Boucherie S (2018) "Predatory" vs trustworthy journals: what do they mean for the integrity of science? Elsevier. https://www.elsevier.com/connect/predatory-vs-trustworthy-journals-what-do-they-mean-for-the-integrity-of-science. Accessed 21 Apr 2021
6. World Intellectual Property Organization (2021) Patents. https://www.wipo.int/patents/en/. Accessed 20 Apr 2021
7. Clarke NS (2018) The basics of patent searching. World Pat Inf 54:S4–S10. https://doi.org/10.1016/j.wpi.2017.02.006
8. Matsuura Y, Takazawa Wlch N, Sakai T, Tsutani K (2020) Clinical trial registration, and publication in acupuncture studies: a systematic review. Integr Med Res 9:56–61. https://doi.org/10.1016/j.imr.2020.01.008
9. Atkinson LZ, Cipriani A (2018) How to carry out a literature search for a systematic review: a practical guide. BJPsych Adv 24:74–82. https://doi.org/10.1192/bja.2017.3
10. Sengar G, Tripathy P (2012) Pharmaceutical regulatory agencies and organizations around the world: scope and challenges in drug development. Pharmatutor. https://www.pharmatutor.org/

articles/pharmaceutical-regulatory-agencies-and-organizations-around-world-scope-challenges-in-drug-development. Accessed 20 Apr 2021
11. Cullen R, Chawner B (2011) Institutional repositories, open access, and scholarly communication: a study of conflicting paradigms. J Acad Librariansh 37:460–470. https://doi.org/10.1016/j.acalib.2011.07.002
12. Bankier JG, Gleason K (2014) Institutional repository software comparison. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000227115. Accessed 20 Apr 2021
13. Asadi S, Abdullah R, Yah Y, Nazir S (2019) Understanding institutional repository in higher learning institutions: a systematic literature review and directions for future research. IEEE Access 7:35242–35263. https://doi.org/10.1109/ACCESS.2019.2897729
14. Andro M, Asselin E, Maisonneuve M (2012) Digital libraries: comparison of 10 software. Libr Collect Acquis Tech Serv 36:79–83. https://doi.org/10.1016/j.lcats.2012.05.002
15. Echo (2020) Top 8 best digital publishing software for magazines & newspapers. FlipHTML5. https://fliphtml5.com/learning-center/top-8-best-digital-publishing-software-for-magazines-newspaper/. Accessed 20 Apr 2021
16. Cromey DW (2012) Digital images are data: and should be treated as such. Methods Mol Biol 931:1–27. https://doi.org/10.1007/978-1-62703-056-4_1
17. Aronson B, Glover S (2005) Journal access initiatives: where are they? Why the World Health Organization? And what is HINARI? Ser Rev 31:279–283. https://doi.org/10.1016/j.serrev.2005.09.008
18. Kim S (2018) Public chemical databases. In: Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics. Elsevier, pp 628–639. https://doi.org/10.1016/B978-0-12-809633-8.20192-1
19. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH (2017) Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst Rev 6:245. https://doi.org/10.1186/s13643-017-0644-y
20. Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J (2003) How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess (Rockv) 7:1–76. https://doi.org/10.3310/hta7010
21. Burns T, Walker MJ (2016) Searching the chemical literature: a brief guide to a comprehensive literature review (with particular reference to analytical chemistry). J Assoc Publ Analysts 44:12–17
22. Excellence NI for H and C (2012) 5 Identifying the evidence: literature searching and evidence submission. https://www.nice.org.uk/process/pmg6/chapter/identifying-the-evidence-literature-searching-and-evidence-submission. Accessed 27 Apr 2021
23. Cals JWL, Kotz D (2016) Literature review in biomedical research: useful search engines beyond PubMed. J Clin Epidemiol 71:115–117. https://doi.org/10.1016/j.jclinepi.2015.10.012
24. Singh IP, Kumar S, Gupta S (2017) Naphthyridines with antiviral activity—a review. Med Chem (Los Angeles) 13:430–438. https://doi.org/10.2174/1573406412666161228112127
25. Li X, Li X, Huang N, Liu R, Sun R (2018) A comprehensive review and perspectives on pharmacology and toxicology of saikosaponins. Phytomedicine 50:73–87. https://doi.org/10.1016/j.phymed.2018.09.174
26. Wang ZL, Wang S, Kuang Y, Hu ZM, Qiao X, Ye M (2018) A comprehensive review on phytochemistry, pharmacology, and flavonoid biosynthesis of scutellaria baicalensis. Pharm Biol 56:465–484. https://doi.org/10.1080/13880209.2018.1492620
27. Moini Jazani A, Nasimi Doost Azgomi H, Nasimi Doost Azgomi A, Nasimi Doost Azgomi R (2019) A comprehensive review of clinical studies with herbal medicine on polycystic ovary syndrome (PCOS). DARU, J Pharm Sci 27:863–877. https://doi.org/10.1007/s40199-019-00312-0
28. Khdour MR (2020) Treatment of diabetic peripheral neuropathy: a review. J Pharm Pharmacol 72:863–872. https://doi.org/10.1111/jphp.13241

29. Elissawy AM, Soleiman Dehkordi E, Mehdinezhad N, Ashour ML, Mohammadi Pour P (2021) Cytotoxic alkaloids derived from marine sponges: a comprehensive review. Biomolecules 11:258. https://doi.org/10.3390/biom11020258
30. Isenor JE, Edwards NT, Alia TA, Slayter KL, MacDougall DM, McNeil SA et al (2016) Impact of pharmacists as immunizers on vaccination rates: a systematic review and meta-analysis. Vaccine 34:5708–5723. https://doi.org/10.1016/j.vaccine.2016.08.085
31. Viseu T, Lopes CM, Fernandes E, Real Oliveira MECD, Lúcio M (2018) A systematic review and critical analysis of the role of graphene-based nanomaterials in cancer theranostics. Pharmaceutics 10:282. https://doi.org/10.3390/pharmaceutics10040282
32. Walusansa A, Asiimwe S, Kafeero HM, Stanley IJ, Ssenku JE, Nakavuma JL et al (2021) Prevalence and dynamics of clinically significant bacterial contaminants in herbal medicines sold in East Africa from 2000 to 2020: a systematic review and meta-analysis. Trop Med Health 49:10. https://doi.org/10.1186/s41182-020-00295-8
33. Centre for Reviews and Dissemination (2008) Systematic reviews: CRD's guidance for undertaking reviews in health care. www.york.ac.uk/inst/crd. Accessed 27 Apr 2021
34. Horsley T, Dingwall O, Sampson M (2011) Checking reference lists to find additional studies for systematic reviews. Cochrane Database Syst Rev 2011:MR000026. https://doi.org/10.1002/14651858.MR000026.pub2
35. Skoglund M, Runeson P (2009) Reference-based search strategies in systematic reviews. https://portal.research.lu.se/portal/files/5685284/1467090.pdf. Accessed 27 Apr 2021
36. Hopewell S, Clarke MJ, Lefebvre C, Scherer RW (2007) Handsearching versus electronic searching to identify reports of randomized trials. Cochrane Database Syst Rev 2007: MR000001. https://doi.org/10.1002/14651858.MR000001.pub2
37. University of Wollongong (2021) Literature review. https://uow.libguides.com/literaturereview/how. Accessed 27 Apr 2021
38. University of Wollongong (2021) Identifying effective keywords. https://www.uow.edu.au/student/learning-co-op/finding-and-using-information/keywords/. Accessed 27 Apr 2021
39. University of Tasmania (2021) Systematic reviews for health: 6. Phrase searching, wildcards and proximity operators. https://utas.libguides.com/SystematicReviews/PhrasesWildcardsProximity. Accessed 27 Apr 2021

Patent Searching

15

Brahmeshwar Mishra and Gunjan Vasant Bonde

Contents

B. Mishra (✉)
Department of Pharmaceutical Engineering and Technology, Indian Institute of Technology (Banaras Hindu University), Varanasi, India
e-mail: bmishra.phe@itbhu.ac.in

G. V. Bonde
Department of Pharmaceutical Engineering and Technology, Indian Institute of Technology (Banaras Hindu University), Varanasi, India

Department of Pharmaceutical Sciences, School of Health Sciences, Energy Acres, University of Petroleum and Energy Studies, Dehradun, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_15

Abstract

Patents are the exclusive rights conferred to the inventor for the innovation they brought up. The research and development resulted in a massive number of patents applications, which subsequently turns into issued patents. Every country is maintaining its patent database; some of them are freely available to the public. These are available through specific online search engines provided by a national patent office such as Espacenet by European Patent Office and InPASS by Indian Patent Office. However, data mining and retrieval of an intended patent from such a huge data demands a specialized skill set to operate search engines and knowledge about typical terms, keywords, and technical keywords pertaining to patents. This chapter provides insight into a systematic review of patent information retrieval systems in connection to the patent classification systems, types of patent searches, information sources, various search engines, and their peculiar characteristics.

Keywords

Delphion · Espacenet · InPASS · J-PlatPat · Patent classification · PatentScope · Questel · USPTO

Chapter Objectives

- Discuss and emphasize the significance of patent searching
- Provide insights on different classification systems implemented for classification of inventions
- Understand the purpose of various types of patent searches
- Recognize the precise use of various operators for efficient patent searching
- Appreciate different intellectual property databases and their salient features

15.1 Introduction

The development of technologies and creativity brings up numerous inventions with its commercial applicability. The invention demands huge efforts and cost of development by the inventors. To protect the exclusivity of the inventors and to maximize the financial return from its commercialization, patents are granted to the inventors. In return, the inventors disclose their invention for the research fraternity. Patents confer the legal rights to the inventors that exclude others to have commercial benefits over the invention. Over the years, plethoras of inventions have come up with emergence of new technologies. Patent search is a systematic strategy of search that is conducted through available repository to retrieve information of the patents and analyze the data for one's interest. A patent search and data mining through such a huge data is very cumbersome act; however, it is very crucial for updating one's knowledge and make them aware of the existing technologies. Therefore, the searching of a patent is a skilled task and demands a prior information related patent sources, their classification and the information about how to use available sources in cost and time efficient manner. The chapter provides insight on such crucial aspects of patent searching regarding sources of patents, their classification, and various ways to retrieve the data of one's interest.

15.2 What Is Patent?

Before describing patents search, let us have some basics of patent with a searching perspective. The patent is a bilateral agreement and is also a techno legal document which contains legal as well as technical information. The patent is an intellectual property right to its innovator. Intellectual property (I.P.) refers to "the property resulted from one's own intellect" or in other words these refer to "creations of the mind." World Intellectual Property Organization (WIPO) defines patent as "an exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing something, or offers a new technical solution to a problem" [1]. In other words, a patent is a document issued by a government office in the country of application. Through the legal rights of patent, an inventor exercises his/her right of exclusivity and excludes others from making use of the benefits of their invention for a limited period of time. Generally, the term of patent is 20 years albeit it may vary country-wise and needs to be maintained or keep active (in force) by the payment of renewable fees [2–5]. The patent is granted for an invention that may be a process or a product which may serve a solution to the technological problems [3]. The inventor can exploit the patented invention for either use for manufacture, import, or sell others and enjoy a monopoly in business [2, 6]. The patents are country specific and rights are exercised only in the country where patent is granted. Also, inventor can gain the financial returns for the investment one had made for a creation of such intellectual property [7]. It is a significant benefit of patent to public that the underlying invention is made available to public at the cost of patent rights to inventor. Others indulging in related activities

like research, technology, and development are also benefitted by using such available information for further development and avoidance of duplicate work and thereby promote the progress of science and useful arts.

Patents are granted only for inventions that are new, consist of inventive steps that are nonobvious, and have industrial utility [4, 8]. The patents are granted by the country where the patent is applied for. Almost every country has its own regulations related to the granting and the use of patents. After being granted, these patents are made available to public by publishing them in a national gazette or an official publication of that country and are also uploaded to online databases for ease of access. Dissemination of scientific expertise via patents provides reciprocal advantages to inventors and consumers in a manner that encourages social and economic well-being and respects privileges and responsibilities [6]. Patents are techno-legal documents where patent writers use a set of different words and phrases to explain the invention with an intention to maximize the scope of the invention. This selective use of legal and technical expression of phrases sometimes makes it difficult for a layperson to access the published application or understand the scope of the invention. In order to cluster the similar inventions the patents are classified into categories and subcategories with unique number system where each patent is provided with a unique identification number according to their hierarchy in the classification systems. Patents are classified in different types depending on the purpose of inventions such as product patent, utility patent, and process patent. In earlier days however, the manner of classification may vary country-wise but after WIPO's efforts to harmonizing the patent filing systems around the world, most of the countries now follow the same patent classification system in order to facilitate inventors around the globe.

Another aspect of patents to be understood is that whether the research done or invention under consideration can be patentable or not, usually referred as patentability of an invention. It is also very important to understand what is patentable and non-patentable. The criteria that deal with the patentability of any invention vary from country to country. Given emphasis to the Indian patent law following attributes are excluded from the scope of patents for which patents cannot be granted [4, 6].

- A mere discovery of new use or new form of patented technologies
- The discoveries of materials already in nature
- The theorems, theories, principles, and mathematical methods used in scientific fraternity
- Natural biological processes or their discoveries
- The treatment or surgical methods used for humans or animals, or diagnostic methods in the interest of public welfare. However, the products used for above purposes qualify patentability.
- Algorithms of computer program; business method
- Method of agriculture and horticulture
- Methods related to atomic energy

15.3 Why Search Patents?

A million-dollar question is why one should perform a patent search? Patent search is a systematic and comprehensive strategy of search that is conducted through available repository to retrieve an information/knowledge relevant to patents and analyze the data for one's interest. It includes not only the database of granted patents but also involves a search through patent applications and non-patent data such as journal articles, prior-art, or any relevant data pertaining to the inventions. In addition to patentability search, as discussed in above section, there are many other valid reasons for undertaking a patent search such as [2, 3, 7]:

- Inventors can assess the patentability of their invention. The analysis and comparison of retrieved data during patent search can help one to understand whether invention is patentable or not.
- Through patent search, the invention that may be similar to the one's invention is revealed. So, inventors and researchers can ensure their uniqueness of invention and can avoid the repetition of work. That is very crucial to save resources in terms of money and time. Patent search avoids any chance of infringement of granted patents.
- Patent search also helps one to know whether the invention or similar invention is already patented and in which countries the patent is in force or active. The approach will lead you to search a country of interest where you can have patent for an invention. For instance, a particular invention is patented in one country, say India, but not in other countries. Another inventor has come up with a new technology but very similar to the patented one, that is, of India. In that case, one patent office cannot grant new patent on the similar invention as the previously filed invention may act as prior art.
- Prior art search during patent search serves the inventor a ready knowledge which can guide to draft the claims for one's patent application. This strategy may reduce the chances of rejection of patent.
- It is a very valuable tool to keep an eye on the technological developments, developments brought up by your competitors in commercial market, develop the strategies for new invention of value addition to already patented technologies, and ensure your financial credit and returns from the market.
- It can act as a measure to analyze oneself where you and your inventions are standing in this world of competition and among your competitors. The approach can guide a path for your development.
- If any company wishes to adopt new technology, patent search is an option of choice to search for whom to contact for licensing/renting a patented technology. It can allow discovering a potential collaborator for your invention, for example, if one pharmaceutical company is working in the area API manufacturing and required to make co-crystallized API in which the particular company is not specialized, then they can find a potential inventor who is specialized in the particular technology through patent searching and can collaborate. This

collaboration can save money and time to develop new technology if already exist, and can get experts to serve the purpose.

- The patent searches also help to decide over the commercial value of the invention in the market and serve as a guide in determining the commercial value of your invention. The inventor can conduct patent search to have comprehensive list of patents for related inventions worldwide and their commercial value in the market through market research. The comparison of benefits and shortcomings of one's own invention with other related inventions can help to decide the commercial value of their invention.

15.4 Patent Classification System

The developments in various fields have brought up a huge number of patent applications day by day. Such data should be organized in an appropriate manner or hierarchy so that the information can be easily retrieved. Considering this need, the patents are categorized in different classes and subclasses based on various criteria such as technical applications, design, product, or process. This system is adopted by most of the patent offices and is referred to as Patent Classification System. The patent office classifies every patent received to them as applications according to the prevailing classification system. However, predominantly used classification systems are International Patent Classification (IPC), European Classification (ECLA), and United States Patents Classification (USPC) [2]. The new classification system has been developed by collaboration of United States Patent and Trademark Office (USPTO) and European Patent Office (EPO) and is referred as Cooperative Patent Classification (CPC). In following section, we have provided description of various classification systems.

15.4.1 International Patent Classification

The International Patent Classification (IPC) was established in 1971 by the Strasbourg Agreement. IPC classifies the patents in sections (A to H) with their section titles according to their areas of technology and designates them with English alphabets A to H. Further, hierarchical classification is established in class, subclass, group, and subgroup; the details are described in Table 15.1 with an example. The classification identification numbers are available on the first page of patent under the heading Int. Cl., as shown in Fig. 15.1a. The organized classification forms the basis to disseminate the information in proper manner, interrogate the state of the art in specific technology, and avoid ambiguity in results. The IPC is amended and revised each year and a new version comes into force from January 1 each year [2, 9].

Table 15.1 Description of various terms used in international patent classification system with an example

Sr. no.	Term	Description	Example
1	Section	A very broad category that indicates the contents of the section, and these are designated by capital alphabets from A to H	Section A includes patents related to human necessities (other section codes and titles are similar to CPC except section Y, as shown in Fig. 15.1)
2	Subsection	Title provides further information under the section and used to categorize into various subsections; it does not have any designation	Section A is divided into subsections like Agriculture Foodstuffs; Tobacco Personal or Domestic Articles Health; Life Savings; Amusement
3	Class	Sections and subsections are further classified into classes depending on their features and specific contents of class and symbolized by capital alphabet followed by a two-digit number	A 42 Headwear
4	Subclass	Subclass is the next hierarchical level that provides more specific and precise insights into the considered class and is designated by capital alphabet	A42 B Hats; Head coverings or A42C Manufacturing or trimming head coverings
5	Group	It precisely describes the specific field of subject matter within the considerations of subclass and is denoted by designations of subclass followed by a one- to three-digit number and the number 00 separated by the oblique stroke	A42B 1/00 Hats, Caps, Hoods
6	Subgroup	Under the scope of group, the subgroup defines the subject matter more specifically and precisely. The title may be preceded by the dots and their number defines the level of hierarchy below the group considered for search It is denoted by designations of subclass followed by a one- to three-digit number specific to its group and the number other than 00 separated by the oblique stroke	A42B 1/019 • characterized by their material A42B 1/0192 ••Paper; Cardboard A42B 1/0195 ••Antimicrobial or antibacterial

For further details of classification system, please refer the guide to International Patent Classification (https://www.wipo.int/edocs/pubdocs/en/wipo_guide_ipc_2020.pdf)

Fig. 15.1 (**a**) The first page of a patent showing different classifications (Source of screenshot: https://patentimages.storage.googleapis.com/ce/f1/e4/d74ed85d44dc05/US8313972.pdf. Accessed 19 Oct 2021); (**b**) a search result on Google patent showing different classification of the patent; a complete hierarchy of classification can be seen on mouse rollover (Source of screenshot: https://patents.google.com/patent/JP5739883B2/en. Accessed 19 Oct 2021)

15.4.2 United States Patent Classification (USPC)

The US Patent Office implements USPC to stratify the patents and other relevant documents as their national system of patent classification. There are three types of patents, viz. design patent (identified by 'D' followed by one or two-digit integer, e.g., D1), utility patent (employs one-, two-, or three-digit integer as an identifier, e.g., 438/63), and plant patents (uses 'PLT' as an identifier). Each subject matter or major technology is designated as class and subclasses. The class differentiates one major technology from another whereas the subclass describes functional and structural classification or a process of technology identified and includes the scope of a class. An alphanumeric code consisting of three digits is assigned to every class for identification followed by subclass code. In Fig. 15.1a, 438/63 identifies USPC classification, where the numeric identifiers suggest that it is a utility patent. Class 438 stands for semiconductor device manufacturing: process and subclass 63 stands for particulate semiconductor component. With the emergence of innovative technologies, the classification is revised to involve necessary amendments in classes and subclasses. However, USPC is currently in force in case of plant and design patents while other patent type follows Cooperative Patent Classification (CPC) [10, 11]. The search for USPC or CPC classification can be conducted at web address: https://www.uspto.gov/web/patents/classification/.

15.4.3 European Classification

European classification is adopted by the EPO as a variant of IPC. Unlike USPC, ECLA codes are not printed in the printed version of patents [2]. The ECLA uses the code for section, class, subclass, and group as those used in IPC, for example, B62 J11/00 where B, 62, J, 11/00 are section, class, subclass, and group codes, respectively. Further, unlike IPC, ECLA optionally adds either a number or alphabet for subgroup identification, for example, B62 J11/00 B. The ECLA has generated more subgroups as compared to IPC for incorporation of more detailed classification based on technical features of invention. However, from January 1, 2013, ECLA follows the codes and classification as CPC, as described below.

15.4.4 Cooperative Patent Classification

The Cooperative Patent Classification (CPC) is implemented from January 1, 2013. It is developed by collaboration of EPO and USPTO in light of harmonization of the existing classification systems and to avoid ambiguity [10, 12]. It is divided into major nine sections, that is, A to H and Y (refer Fig. 15.2), which are further categorized in classes depending on the specific technical area which is followed by the number identifying subclasses, groups, and subgroups (refer Table 17.1). One can find the CPC by entering keyword and also can verify by reading the definition by hitting the CPC number in the result list. CPC consists of almost 250,000 classifications. The example of CPC and identification codes is shown in Fig. 15.3.

15.5 Patent Search Type

The information available from different online or offline sources is very vast. A random search through the literature can land up with huge information. So, it is important to define a type of search, one wish to have to get relevant information. Depending upon the data required by the researchers and its comprehensiveness, the patent search can be commonly classified as state of art, novelty, freedom to operate, validity, and patent portfolio analysis. This part of chapter describes various attributes for above said types of patent search. However, pre-requisite to the patent search type is that the researcher should be provided or should have prior information about the purpose, time coverage, and the most relevant sources to search for patents pertaining to specific subject. Another important fact one should bear in mind is that the titles of some search types can be used. For example, evidence of use can be used instead of state-of-the-art searches, similarly as patentability or novelty is also substitute for each other [13].

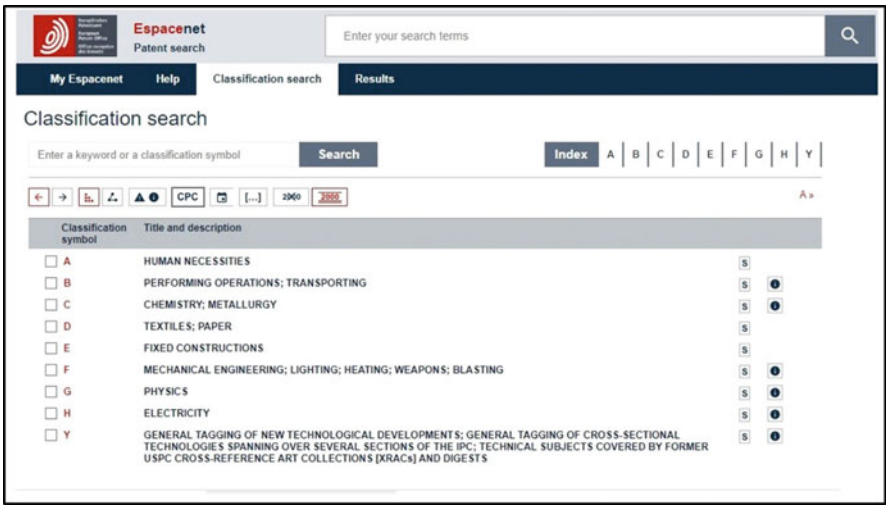


Fig. 15.2 Various sections and their descriptions as per cooperative patent classification system for patents (Source of screenshot: https://worldwide.espacenet.com/patent/cpc-browser. Accessed 19 Oct 2021)

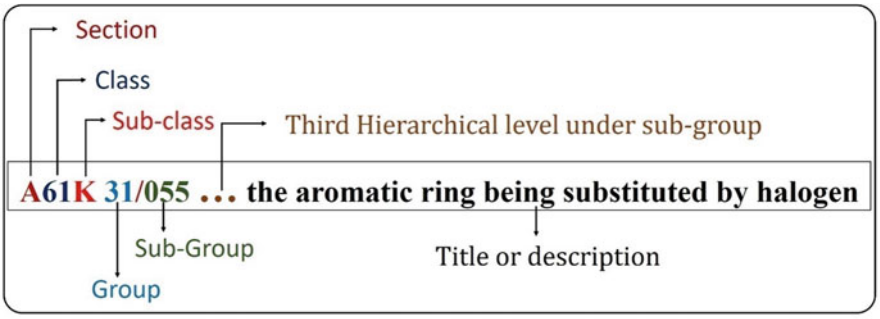


Fig. 15.3 Interpretation of level of hierarchy in CPC code for a patent

15.5.1 State-of-the-Art Patent Search (Evidence of Use Search)

This is one of the easiest ways of patent search. Generally, the goal of the state-of-the-art quest is to obtain a detailed description of a product or technology via searching all the available resources. In other terms, it is rather a compilation of all related publications released globally in a certain technological area or in the fields or patents submitted by particular applicants or innovations created by the inventors concerned. It would also be important to involve queries in all forms of literature such as research/review articles from journals, thesis, and dissertations and in "grey literature" such as instruction manuals and advertising literature [7]. Therefore, this work, indeed, is undertaken before conceptualizing and hypothesizing any research

and development project or investment is made to have an overview of the related research already undertaken in that specific area and to create novelty in one's work [14, 15]. The data collected will surely influence the appropriate selection of the topic and technology for new project and funding of a new project. The approach provides safety for infringement of the other patents and ensures exclusivity for the new inventor; two aspects are very necessary in terms of cost and time of the investing company or research and development firms.

Once the comprehensive data is retrieved, one can then classify it according to different cluster patterns with an objective to better summarize the searched prior art on the basis of inventors, assignees, patent types, and patent classification codes. On the other hand, data also helps in identification of various technologies involved therein, different competitors, and technology experts working in the similar fields. For the better understanding and analysis of such a huge data, different graphics and charts are used, referred as a patent landscape analysis [2, 16]. Numbers of tools are now commercially used for such a purpose, which can arrange the data as per various attributes like patent classification according to their type, their validity, technology, etc. However, few points to be considered for good analysis are use of the standard data, duplicity removal, and stratification of all patents according to family of patent [2].

15.5.2 Novelty (Patentability)

By definition, the research undertaken for the patent should be novel, nonobvious, and commercially applicable. Therefore, it is mandatory to ensure all three aspects through a literature search for the entire relevant prior art that may demonstrate the likelihood of getting a patent granted and does not infringe any other patent, by chance. This type of a patent search is a restricted and typical type of state-of-the-art search. The search encompasses not only the books, journals, and patent publications but also includes a search through other sources like theses and dissertations, press releases, websites, conference proceedings, technical data, etc. [2, 7].

Generally, a company intends to conduct novelty search prior to drafting the patent application with an intention to decide the scope of the proposed claims. So, the purpose for novelty search type is to verify the novelty, exclusivity, nonobviousness, and its applicability for commercial production. Hence, the result of this type is the compilation of prior art patents/publication which is narrower and very closely related to the technologies/designs/products that are to be used in the invention for which the patent is filed [17]. The search can be run by using keyword, classification codes, patent inventors or assignee, citation data, and the typical information such as drawing and chemical structure. Google patent search engine can be helpful to collect the relevant data [3, 18].

15.5.3 Freedom to Operate (Infringement, Right to Use, Clearance)

Infringement check and right to use or clearance check are the interchangeable terms that can be used for freedom-to-operate (FTO) patent search [14]. In order to prohibit anyone from copying the innovation specified in the patent, the assignee or inventors are given exclusive rights by means of a patent. Ideally, it is intended to guarantee that the one innovative product under consideration does not infringe the patents of someone else that has not yet expired [5]. Unlike state-of-the-art and novelty search, the scope of this search is limited to the already granted patents which are in force or active at the time of launch of a product by the company. Within the patent, the claims have to be evaluated for FTO since the statements are the only legally valid element of the patent application [3]. Therefore, it encompasses the data from last 20 years as generally the validity of the patents is 20 years; however, patents are country specific so it is advised that one should check the claims and validity of all the similar patents from all the countries before concluding and analyzing the data [14]. As said earlier, the term of a specific patent is fixed for finite years but term of some patents can be extended; so one should also consider this aspect before rushing to the patent application process in case the term of patent is extended by an inventor [7]. Although the claims in all the relevant data has to be properly understood and ensured not to infringe, the attention must be given to the design, process, raw materials methods, and technologies employed therein [2]. That will guarantee the uniqueness of one's claims if the process quoted in patents is also under patent as in some countries, like India where individual patents can be granted as product patent and process patent. Following attributes can be very helpful for FTO search:

- File history
- Patent family
- Patent inventors/assignee
- Validity status
- Geographical and time specificity where patents are currently in force

15.5.4 Opposition Search

As the name suggests, the essence of this type of patent search is to oppose the grant of patent or one has a belief that patent should not be granted by the authorities with a sufficient reason. The purpose of the quest for opposition is to include reasons as to why the application will not be approved. These arguments by opponent provide significant information that might have missed or may not have treated appropriately by the granting authority during the complete process of granting. The statement of objection can be filed during the process of examination and granting after the publication in official gazette. However, the time for an opponent to raise such objection may be different for different countries which may extend even after date of grant. This period is generally referred as grace period, for example, 9 months is the grace period after the date of grant for the patent granted by the EPO [7].

15.5.5 Validity

Unlike opposition search, validity search is performed at later stages of the lifespan of a patent. This type of search might be performed in light of competition and might discover the prior art which may invalidate the claims made in patent already granted and currently in force [7, 14]. The validity of particular patent is examined through this search type, so it is named as validity search. For the purpose, the search can be run through search history used and documents cited by patent examiner, and statements of examiner for granting a patent. As obvious, only those literature must be reviewed which were available prior to the date of granting the patent for opposition search. The validity of each claim should be checked; however, claims for the patent belonging to the same family may differ country-wise, therefore special attention is to be given regarding the country where the patent is in force, one is interested in.

15.5.6 Due Diligence (Patent Portfolio Analysis)

This type of a patent search is very crucial for the companies wishing to trade their patent. In other words, it is very important for the company which is willing to sell their patent and so for the buyers also. Through due diligence search, a seller company is able to decide a fair and comparative price in the market for the patented technology. On the other hand, such analysis can guide the buyers or investors assigning a fair price and negotiating over the price for the desired product or technology by considering market prices and other available alternatives. The due diligence search type deals with examining the robustness of the patent if a company's patents are robust to exclude competitors and to ensure the least probability of an infringement. Therefore, it encompasses the attributes of all three, FTO search, validity search, and patent portfolio of the company [2, 7, 14].

15.6 Strategies for Query Construction

An effective query construction is necessary for retrieving specific or desired patent documents from the millions of documents in database. The approach will limit and narrow down the search results to a desired one. Different types of operators, wildcards or truncations, nesting, and phrases can be employed for creating any permutation and combination of keywords in search query. The effective use of such strategies can effectively save the time and efforts for searching a technology from a global database. This section explains such strategies in detail with proper example [19, 20].

15.6.1 Boolean Operators

These are logic operators that can be used to combine two or more terms/keywords in a query. Examples of such operators are AND, "+," OR, NOT, ANDNOT, "−," etc. However, if no operator is used to separate the terms, default operator is considered as "AND." Some examples with their objective are given as follows:

- Steel AND pot—will retrieve any possible document containing both the keywords words.
- Steel OR pot—will retrieve any document having either or both words.
- Steel XOR pot—will retrieve only those documents containing either of the keyword but not both.
- Steel ANDNOT pot—will retrieve any document that contains steel but does not contain pot.

15.6.2 Proximity Operators

The use of proximity operators allows the search of desired keywords with a specific range of words appearing in a text. The default value of words is five; however, it can be specified using proximity operator with tilde symbol "~" followed by number specifying word limit, for example, NEAR~7 to search the keyword within the range of seven words from each other. The examples of proximity operators are NEAR, BEFORE, and ADJ for adjacent. These can efficiently be used for narrowing down the results. Near operator is not direction specific, so the words may lie irrespective of the order given in search query, for example, "steel NEAR pot" searches both the keywords within five words of each other and can retrieve results containing steel pot; pot made up of steel, etc. It will not find pot for storage. In addition, BEFORE provides order of keywords for the search whereas ADJ operator will search in order as specified in search query.

15.6.3 Wildcards

Wildcard search is conducted to search the terms that differ by single or multiple characters within single terms. Wildcards are the symbols that can be used within a term. For conducting a single character wildcard search, "?" symbol is used, for example, te?t can search for text or test. On the other hand, multiple character wildcard search uses "*" symbol either at the end or in a middle of term, for example, nano* can search for nanoparticle, nanomicelle, nanocomposite, etc. However, the query cannot be generated using wildcard operator as first character, for example, *est.

15.6.4 Parentheses and Nesting

Grouping of keywords can efficiently be done using parentheses to construct nested queries. The strategy provides efficient control over the use of the Boolean operators for a query. This strategy is generally used when a specific order and combination of keywords is desired. Complex queries can also be formulated using multiple parentheses within a single query. The example can be explained as follows:

"(Steel OR copper) and pot" retrieve results containing either steel or copper together with pot. In other words, this query searches for documents that contain steel and pot and also those containing copper and pot.

15.6.5 Phrases

Phrases are used for search query when exact match to the query is desired in the document. The phrase to be searched includes in double quotation mark, for example, "blood pressure."

15.7 Information Sources

Plethoras of patent resources are available, nowadays, for comprehensive research. Some resources like USPTO, EPO, etc. provide the information free of cost whereas there are some commercial tools available on the paid basis such as Orbit, PatBase, and Thomson Innovation [3]. Some of the professional companies working in the patent search area use such commercial tools for the retrieval of the data. Such resources provide vast sets of patent data and combination of different search attributes enabling field browsing, concise interfaces distinguishing claims/titles/abstracts from the full document, if required by the user. The use of other attributes such as clever syntax, Proximity, Boolean logic, exporting, saving, editing, and re-executing complex and detailed strategies can be employed to extract a comprehensive data and provide a fast summary in a far shorter period than freely available internet services [3, 13].

In contrast, free databases provide a more comprehensive data. Some examples are Google patents, Espacenet (by EPO), PatentScope (by WIPO), USPTO, etc. However, these sources are updating in terms of quality, pace and currency, offering inventors/applicants and scholars a decent rundown of what is in the public domain before they hire a skilled searcher or file a patent application. Some sources serve a database from a single country, for example, USPTO, whereas others can encompass the data from various countries which they cover, for example, Espacenet and PatentScope. For instance, Google patents can compile data from countries but limited to only the USA, Canada, China, Germany, EPO, and World Intellectual Property Office (WIPO). The selection of a particular resource is up to the researcher depending on the coverage of information they ask for. Being an official search engine provided by respective patent offices, Espacenet and patent scope are usually

recommended as they are free to access and have good technical capabilities for search and for export of results. In contrast to above, some sources can only provide the abstracts and bibliographic data. Patent law in-nation is specific while there exist a degree of harmonization around the board and it is helpful to be able to check for evidence across a range of countries [13, 21]. The following part of the chapter focuses on some of the free databases available on the internet and usually used for patent searching. Following attributes can be used for search, rather these can influence the selection of search tool:

- Nature of invention
  - Utility/device/design/process/product, that is, type of patent
  - Chemical structures
  - Data related to biosequence
- Search techniques desired or most appropriate
  - Natural language
  - Keywords
  - Proximity
  - Value-added indexing
  - Boolean logic
  - Similarity
  - Full-text link
  - Word truncation
  - Search term weighing
  - Chemical structure based on textual description
  - Special characters like symbols and Greek alphabet
  - Multilingual search query or language translation
- Alerting features
  - Classification codes
  - Legal status
- Analysis of results
  - Ranking the results regarding relevance to input keywords
  - Sorting of results
  - Citation mapping (citing and cited)

15.7.1 Google Patents

It is a freely available search engine dedicated for patent search and created by Google. At the cost of few clicks, Google patents can land up with huge number of patents. Google patents offer some excellent benefits like ease of searching, quick search, translation of foreign languages, ease to download full text if available, automatic grading according to closeness, and relevance of input by researcher. Once you provide a keyword in search box in Google Patents, it provides all data by just one hit. The search is run through multi-country databases and through full text if available or abstract [3]. Although it provides patents in order of their relevance to

keyword, the data is not stratified according to other necessary attributes like country, priority date, etc. For the purpose, one should look for "Advanced Search" option provided by Google Patents, available at the bottom of their webpage. Advanced Search serves you with a combination of various attributes or filters that can be seen in Fig. 15.4 in left side panel which helps in filtering and restricting the data according to the input provided. The term or keyword is highlighted in the results for easy identification. In addition, the results can be sorted either by their relevance, classification, or by the patent family. The results are in list form; however, one can choose "side-by-side" view to see the details of result you click on the same window, as shown in Fig. 15.4 [22]. Full text will open once you click on the patent of your interest. Very useful features on this page are Google prior art finder and similar option that are provided by Google; see Fig. 15.4 at right panel. Google prior art finder serves to search through non-patent literature whereas similar option provides you other patents related to the patent you had chosen [22]. The search engine has some limitations like one cannot export the list of results, result exploration is available only in online mode, and it is time consuming.

15.7.2 Espacenet

EPO extends its database freely to public by means of Espacenet search engine. It is a multi-country database providing search coverage over a wide range of patents granted in around 90 countries worldwide, for example, Albania, Australia, Canada, New Zealand, the Unites States, the United Kingdom, France, and Germany. It offers three types of patent search, viz. smart search, advanced search, and classification search. In addition, EPO also launched a mobile beta version of Espacenet for easier

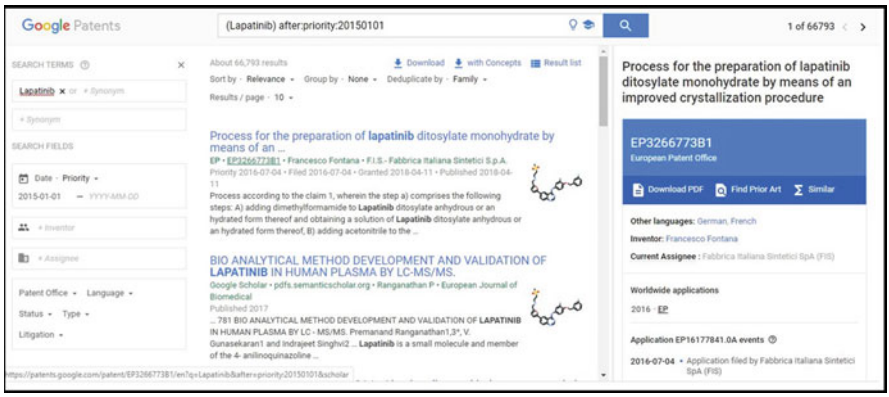


Fig. 15.4 Webpage of Google Patents. Left panel displays options for building a query of patent search. Middle panel displays the list of results you have searched for. Right panel displays brief descriptions of the selected patent from list for a quick review (Source of screenshot: https://patents.google.com/patent/EP3266773B1/en?q=(lapatinib)&after=priority:20150101. Accessed 07 May 2020)

search even on the go [23]. Smart search allows the search using keywords; 20 keywords can be used for one hit. To refine the search in light of particular interest, for example, inventor, assignee, and title, the advanced search is the best option; see Fig. 15.5. The query is searched through worldwide database, as default; however, the results can be concise by exploiting various filters available in advanced search form. Inventor or applicant names, priority dates, classes, and keywords in any order can be used for interrogation query. The query can be enhanced by using Boolean logic operators like and/or and closer proximity using "n" to signify the number of words apart. Patents in various languages can be made available; however, one has to select a required language in advanced search and should run the search again. Once you click on a particular patent, the complete data pertaining to the claims, drawings/designs, descriptions, the legal status, patent, family, and citing documents list is displayed.

Unlike Google Patents, availability of exporting the list of patents to excel format is very useful option provided by Espacenet which enables researcher to explore and analyze the patents in offline mode. Espacenet enables the result list to be navigated

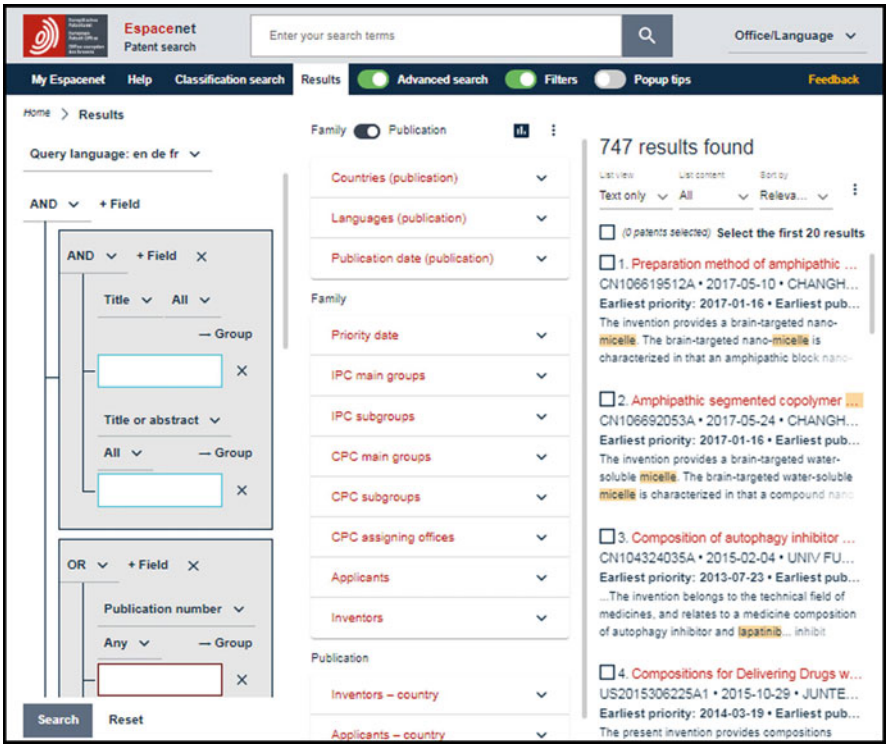


Fig. 15.5 Webpage of Espacenet advanced search in the left panel where various keywords can be input for a query by combination of logic operators. The central panel displays various filters to concise the results and the right panel displays the list of results (Source of screenshot: https://worldwide.espacenet.com/patent/search?q=lapatinib%20micelle. Accessed 07 May 2020)

and also provides document details, simultaneously. Also, history of search is saved in "My Espacenet" list for future reference [24]. Very important feature is "view chart/graph overview" available that displays the results in the form of charts and graphs depicting all the results. It is also a time saving process as it is easier to decide over whether the particular patent is duplicated in a list or a member of same patent family by reading titles and inventors in excel format [3, 21].

15.7.3 PatentScope

WIPO provides access to their database freely through a search engine called "PatentScope." The PatentScope offers links to the complete text of the International Patent Cooperation Treaty (PCT) applications as well as to the patent records of the participating national and regional patent office [25]. It is noteworthy to mention that, unlike Espacenet, PatentScope provides wide coverage of databases from various countries beyond EPO. The patents can be extracted by entering patent registration, names of inventors, keywords, and other search requirements in several languages. The PatentScope provides five types of search, viz. simple, advanced, filed combination, cross-lingual expansion, and chemical compounds; refer Fig. 15.6a and the example of results illustrated in Fig. 15.6b [26]. Field combination is good choice for a beginner where one can interrogate by using various combinations available in the view of your interest. The approach can reduce the number of spurious results and sorts results as per requirement. Keywords are explored only through claims of patents and hence very efficient choice of search for FTO [3]. Once you hit search button, the list of results are displayed having keywords highlighted and you can explore through the necessary data like descriptions, claims, etc. in the tabs provided on the front page of displayed patent. It is recommended to have a PatentScope account that allows you to save queries and export results in Excel format (up to 100 patents); in case of search of chemical compounds, one should have account albeit registration is free. Exported list in excel provide hyperlinked patent numbers for easy retrieval full text. In addition, PatentScope serves two useful features, viz. machine translation and analysis. Machine translation can provide the translation in ten languages, and in addition, one can take advantage of Google Translate option available there. Analysis feature allows to sort results as per inventors, assignee, countries, and publication dates [26]. Pitfall of PatentScope is that the complete original full text is not available as pdf for all the patents, rather they are available as a zip file of TIFF images of each page of patent document [21].

15.7.4 USPTO

It is service provided by US patent office exclusively for patents granted in their country and its partner EPO. It allows access to PatFT (Patents Full-Text and Image) and AppFT (Applications Full-Text and Image) databases. A search can be performed by using keywords and phrases; however, the use of USPC and CPC

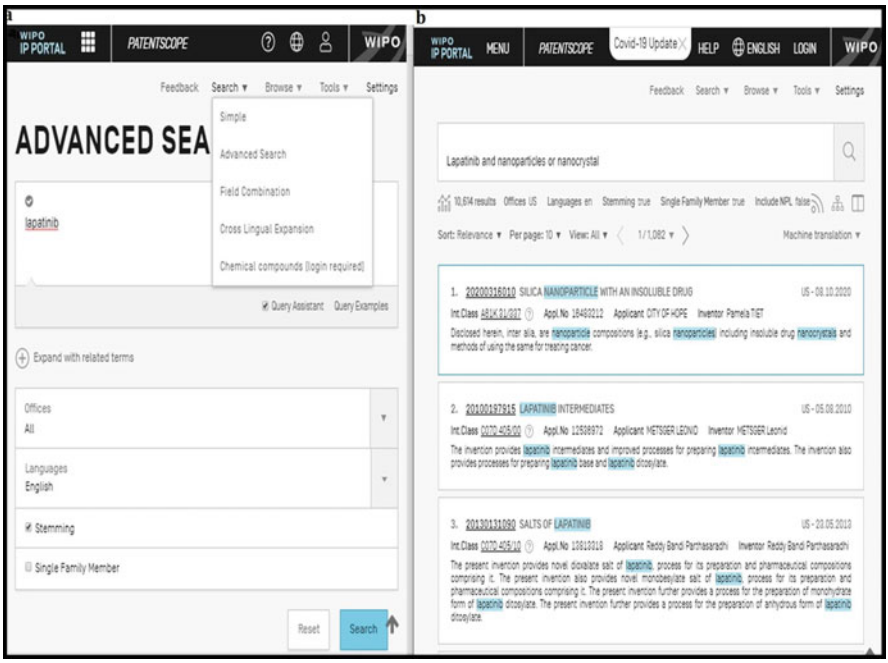


Fig. 15.6 (**a**) Webpage of PatentScope advanced search portal. In search box a query can be generated by combination keyword using logic operators. Other options allow to choose required patent office, language of patent, and the family member options to limit the scope. Search dropdown menu showing different types of search. Panel (**b**) shows total 1082 search results using lapatinib, nanoparticles, and nanocrystal as keywords separated by Boolean operators and search is limited by selecting US Patent Office and single family members. These options are available below the search box (Source of screenshot: https://patentscope.wipo.int/search/en/result.jsf?_vid=P21-KUYJ2O-06212. Accessed 19 Oct 2021)

codes is recommended to make a search more comprehensive and concise. USPTO provides quick search, advanced search, and number search; refer Fig. 15.7. In quick search, two keywords can be interrogated using logic operators whereas a query can be built up using field code and keywords, for example, ttl for title, in for inventor name, etc.; as shown in Fig. 15.8a. In addition to quick search, advanced search option allows to construct a query with multiple field codes and keywords separated by Boolean operators to refine and get required result; shown in Fig. 15.8b. Number search explores through the patent numbers you entered in query box. One can also check the status and history of the events related the patent by hitting "PatFT Status, History" button on the webpage. Full-text data is available from 1976 [27]. USPTO adopts a seven-step strategy to help researcher to refine his/her research, listed as follows [28]:

Fig. 15.7 Webpage of USPTO providing numerous options for patent search and their related information (Source of screenshot: https://patft.uspto.gov/netahtml/PTO/index.html. Accessed 06 May 2020)

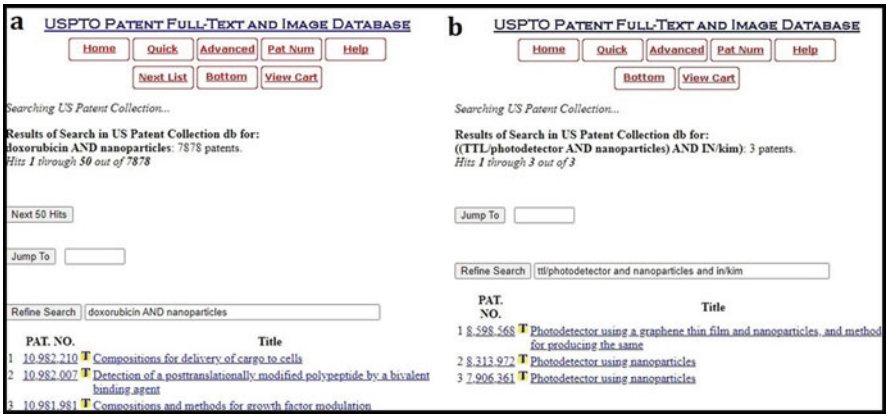


Fig. 15.8 (**a**) Result of quick search option of USPTO using doxorubicin as term 1 and nanoparticle as term 2 (Source of screenshot: https://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-adv.htm&r=0&f=S&l=50&d=PTXT&Query=doxorubicin+AND+nanoparticles. Accessed 20 Apr 2021); (**b**) result of advanced search using photodetector and nanoparticles as keywords for title and Kim for an inventor name. In this particular example, TTL and IN field codes are used for limiting the search through title and inventor name, respectfully (Source of screenshot: https://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.htm&r=0&p=1&f=S&l=50&Query=%28%28ttl%2Fphotodetector+and+nanoparticles%29+and+in%2Fkim%29%0D%0A&d=PTXT. Accessed 20 Apr 2021). The field codes for formulating a specific query are available on the advanced search page of USPTO (refer for details: http://patft.uspto.gov/netahtml/PTO/search-adv.htm)

- Brainstorm the terms: prepare a query for an invention using combination of its use, purpose.
- Use above terms/keywords using any of the search option on USPTO website and retrieve its Cooperative Patent Classification or international classification system.
- Review and verify the Classification system codes you retrieved from above step and its closeness to your query by going through the definitions of classification codes.
- Use PatFt (for published patents and images) feature to draw a data based upon the identified classification system in above step for your invention.
- Read the patents that are displayed and sort them as per your interest and narrow down the list accordingly.
- Use AppFT (for published patent applications and images) feature to draw a data based upon the identified classification system in above step for your invention.
- Also broaden the research by running the search for interested keyword through non-US patents sites to retrieve a data and analyze it.

15.7.5 Indian Patent Advanced Search System (InPASS)

InPASS is a free search engine provided by Indian Patent Office. It offers a search through the granted patents and published patent application. In addition, patents that belong to the patent family from The Patent Cooperation Treaty (PCT) patents can also be searched via InPASS, provided one should have either PCT application number or PCT publication number. InPASS provides almost 16 different fields such as publication date, title, and applicant/inventor name (illustrated in Fig. 15.9) and allows their combination by logic operators to refine the search. The patents can be explored in either published or granted patents by selecting publication type by ticking appropriate checkbox in the first search field. The list of result with application number, title, status, and e-register is displayed after hitting search. The complete document can be viewed by clicking E-register option followed by "view document" button [29].

15.7.6 Japan Platform for Patent Information (J-PlatPat)

National Center for Industrial Property Information and Training (INPIT) runs J-PlatPat search engine. The search results can be available in English. J-PlatPat serves as a database for patents, utility models, designs, and trademarks. Publications related to industrial property information issued by the Japan Patent Office (JPO) are also available. Additionally, the history information of a patent can also be assessed, which will help to follow the current progress of applications going through the examination process [30, 31].

The webpage of J-PlatPat provides the search option for all law documents, patent/utility models, designs, and trademarks. On the home page, one can enter a

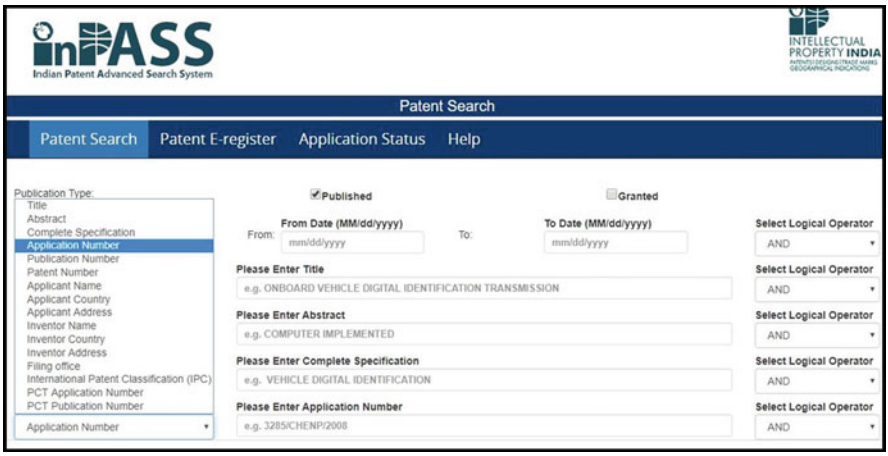


Fig. 15.9 Webpage of InPASS, Indian patent search engine. Left panel provides various options to choose through, middle panel is for the input of keywords, and right panel allows selecting the logic operators to build a query by combination of keyword entered (Source of screenshot: https://ipindiaservices.gov.in/publicsearch. Accessed 06 May 2020)

keyword or number and select the category of one's interest for searching a database through a simple search option. The results of a simple search display all the hits that are categorized in a different list of patent/utility models, designs, and trademarks, and one can select the list as desired. The comprehensive result list displays information of a patent that includes document number, application number, filing and publication date, title, right holder, classification and other links for details, URL for the complete document, etc. Besides, a peculiarity of the result page is that one can have various options to narrow down the results, such as year-wise, by classification, or one can prioritize the applications or publications [30, 31].

In addition to simple search, J-PlatPat also offers the option to perform search using keywords or numbers specifically. One can access the "Patent/Utility model" menu and select either "Patent/Utility Model Number Search/OPD" menu for patent search by patent number or "Patent/Utility Model Search" menu for patent search using keywords. In the number search page (refer Fig. 15.10a), one can select "document" type for the published data or "OPD" type for retrieval of application or examination-related information. Other available options are number, number range, number type, country, etc. On the other hand, the keyword search page (refer Fig. 15.10b) provides options for document language selection and document type, for example, domestic documents, foreign documents from different patent offices worldwide, global journal documents like document, science technology terms, etc. Further, various combinations of keywords for additional search items, for example, title, abstract, classification type, etc., can also be used for searching [30, 31].

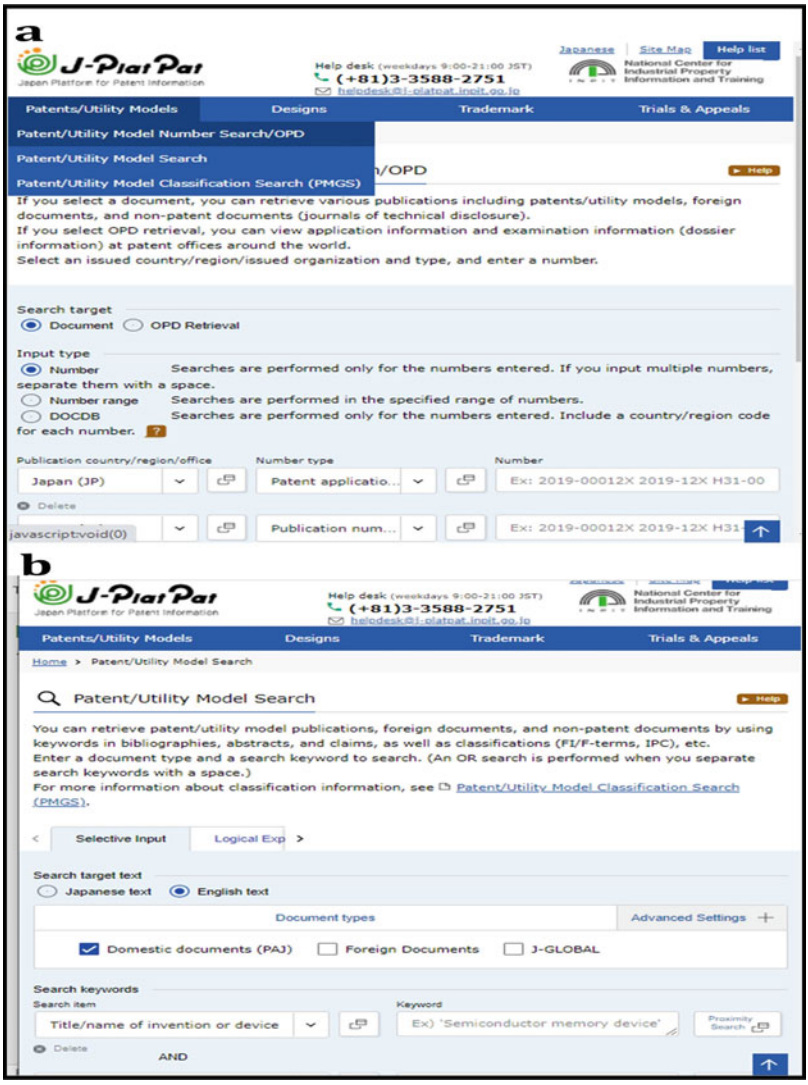


Fig. 15.10 (**a**) Webpage of J-PlatPat showing a dropdown menu of "Patent/Utility model" and "Patent/Utility Model Number Search/OPD" page for performing number search with a number of options (Source of screenshot: https://www.j-platpat.inpit.go.jp/p0000. Accessed 21 Apr 2021); (**b**) "Patent/Utility model search" webpage offering search by using keywords (Source of screenshot: https://www.j-platpat.inpit.go.jp/p0100. Accessed 21 Apr 2021)

15.7.7 Derwent World Patent Index (DWPI)

Clarivate Analytics had developed DWPI to create a single platform for integrating the global patent information in concise abstracts highlighting the nature of the invention and publishing the data in the English language for easy accessibility of searching a specific technology. This is the value added and comprehensive database for effective patent searching. It provides coverage to the databases from almost all the patent offices. DWPI classification system classifies all the technologies irrespective of IPC, USPC, or other types of systems. This unique and consistent system is followed by Clarivate Analytics subject experts for enabling precise invention-based search in a specific field of technology [32].

According to DWPI classification, all the technologies are categorized in only three broad categories of technologies, that is, chemical, engineering, and electrical or electronic engineering. Each category is further classified into 21 different sections, including section A-M for chemical, section P-Q for engineering, and section S-X for electrical or electronic engineering. Classes are further incorporated in each section to classify the technologies based on their specific features and it is denoted by the section code followed by a two-digit number; the example are as follows:

- Code A25 provides the classification as
- Category: Chemical
- Section A: Polymers and plastics
- Class A2: Condensation polymers
- Class A25: Polyurethanes; polyethers

The patent classification can vary country-wise due to national patent laws, and the same patent may have different IPC classifications. However, DWPI provides the benefit of having a single classification system to classify all the technologies worldwide. Unlike IPC, Clarivate Analytics subject experts all over the world consistently use DWPI classification codes. The advantage enables a researcher to combine keywords with a DWPI classification code to make a precise and effective search. For example, “warn” keyword can find patents related to different warning devices; however, the combination of “X22” with the keyword can only precisely find the patents pertaining to automotive warning devices. Another advantage is that DWPI classification is based on the novelty and use of the invention, unlike IPC, based on the chemical structure [33].

15.7.8 Delphion

It is one of the most advanced, comprehensive, and subscription-based online search tools for the patent information and intellectual property professional, a patent agent, an inventor, and a corporate patent information user. Delphion provides a wide range of value-added services and analytical tools that facilitate companies to manage their

intellectual property effectively. Delphion covers the patents from EPO, USPTO, WIPO, JPO, German Patent and Trademark Office, Swiss patent office, DWPI, etc.

After login into the database, one can use Boolean operators like AND, OR, and AND NOT for separating keywords to generate the query for précising the search criterion. Delphion offers multiple search options, for example, one can use a specific name, word, or any other term or phrase in any field and any permutation and combination of terms to conduct the search. Users can also define searching options in front pages, full text, or claims. Multiple fields can be used in advanced searching options, including date range, number, IPC, ECLA, USPC codes, etc. It also offers searching by using patent examiner name, patent agent or firm name, and a patent attorney. Delphion provides an easier search of the non-patent prior art by providing their links at the integrated view of Delphion database. Potential licensees can efficiently use Delphion's IP Listing & License Inquiry Services for searching sellers of intellectual properties. The result page displays the results, including various fields like thumbnails, title, derwent title, abstract, assignee, publication date, filing date, priority date, international patent classification code (IPC)6, score, etc. The user can narrow down the search by customizing and selecting the required fields only or using inbuilt tools. The user-friendly result page allows the user/searcher to perform the search and analyze the search results efficiently. One can also create a work file to save the desired searches for the future. Patent numbers are hyperlinked in a database which facilitates access to the integrated view of the patent. Another advantage of the Delphion is the provision of some inbuilt and user-friendly tools; their description is as follows [34]:

Corporate tree Allow creating the hierarchy of the patent transfer from original to the latest owner

Working with work files Create, annotate, organize, and save the patent records in the work file

Integrated snapshot and features Allow rapid analysis of results in the form of bar charts

Data extract Export the data either from the work files or search results in various formats such as CSV, TXT, and XML.

Citation analysis with analytical tool Allow to visualize and carry out the patent mapping of backward and forward citations

Staying forward Provide automatic updates on saved searches on the regular basis

PDF express For downloading the results in single or multiple or even in bulk quantities (up to 500 numbers in a zip file)

Patent lab for patent mapping and text clustering For efficient result analysis by facilitating the visualization and graphical presentation of the selected fields of result set or work file data

15.7.9 Questel

Orbit.com, run by Questel, is another commercially available patent search database and enables the consumer to perform highly comprehensive searches through various patent databases of the world's major patent offices, for example, WIPO, USPTO, China, France, Germany, Japan, PCT, EPO, G.B. (the United Kingdom), with some commercial services databases like FamPat, DWPI, etc. The FamPat and PlusPat databases provide access to almost 60+ million documents grouped in 40+ million inventions from 90+ patent authorities. Some of the features of orbit.com are as follows [35, 36]:

Express search interface (Orbit express) Offers the query search using Boolean operators or field search through the full-text, assignee/inventor, and number fields only.

Advanced search interface Offers command-line and structured search elements and allows for field searching across the entire range of available fields, including the use of proximity operators.

Semantic search Allows searching the closest result to the text pasted in the search box. Additionally, the cross-language semantic search for English, French, and German can also be possible.

Similarity search interface Allows search for patents from similar patent families regarding ECLA or IPC patent classification, including cited and citing documents.

Family and extended family search interface Permits the identification of patent families related to a particular patent document that may cover the legal status information and citations. The citation data can also be displayed as a graphical citation presentation with priority data.

Citations search interface Permits the identification of forwarding citations through both patent and non-patent literature.

Browse index Assignee/applicant and inventor index viewer.

Patent number wizard Assists in entering a patent number in the desired format or automatic patent number identification tool.

Keyword in-context or KWIC For highlighting the desired or customized keyword in results

ANALYZE tool Identification of top assignees, top classifications, and most cited and citing patents for graphical or pictorial representation of extracted results.

Family citation and PatCitation tools Presentation and analysis of family/citation relationships in the form of graphs.

Questel also offers various services for comprehensive and dedicated search services; the details are as follows:

PatentExaminer An efficient patent portfolio management system that can analyze and organize data.

QPAT A patent search service that keeps up with the requirements of the growing number of patent information users and provides access to worldwide file histories, monitors legal status, strategic analysis of data and stats, and manages your portfolio.

Qweb Comprehensive patent information collections for professionals offering an easy search interface, providing efficient search and analysis functions.

SearchPat Offers various kinds of search options like basic or prior art search, citation search, family/similarity search, inventor search, competitive intelligence search, etc.

Pat Legal Monitor Automatic alert system that notifies the consumer of any update in patent families and legal status by the respective patent office [37].

15.7.10 WIPS Global

The World Intellectual Property Search (WIPS Global) offers full-text coverage of the US, European, and PCT databases. Chinese, Japanese, and Korean patents are available in English translations. It provides access to abstracts of several European patents, whereas for other collections, only bibliographic data is offered. The salient features provided by WIPS Global are as follows [35, 38]:

General search interface Allows the search in two fields; those may be either patent database, date range, document field, type of patent. The query can be effectively constructed using Boolean, proximity, and wildcard operators.

Advanced search interface Offers numerous search fields combined by AND operator.

Step search interface Permits user to combine individual queries using Boolean operators within the query or for connecting the separate queries.

Number search interfaces For retrieving specific document using the number.

Integrated search interface Unlike other interfaces, this peculiar feature allows simultaneous searching through **Search** databases.

Corporate name index Allows performing search by applicant name.

Clustered results Offer clustering of results concerning applicant, IPC class, and application date.

Analysis tool Provides graphical and statistical illustrations and text-based analysis of applicants, citation, date, and classification.

Batch downloading Allows downloading the results in various text-based formats image format.

PM Manager and ThinKlear Offers additional statistical analysis features and report generation tools.

15.7.11 Miscellaneous Sources

In addition to above described sources, many countries provide their database on their webpage where data is restricted to a particular country by their national patent office such as Korean Intellectual property Rights Information Service, State Intellectual Property Office (for People's Republic of China), and so on. Table 15.2 provides the links for websites of various search engines and their databases that can be used to retrieve the patent documents.

Table 15.2 List of different search engines offered by various national authorities and some commercial service providers

Sr. no.	Search engine	Website
1.	China National Intellectual Property Administration	http://ensearch.cnipr.com.cn/sipo_EN/
2.	Delphion	http://www.delphion.com/
3.	Derwent World Patent Index	https://clarivate.com/derwent/solutions/derwent-world-patent-index-dwpi/
4.	Espacenet	https://worldwide.espacenet.com/
5.	Google patents	https://www.google.com/?tbm=pts
6.	Indian Patent Advanced Search System (InPASS)	https://ipindiaservices.gov.in/publicsearch
7.	Japan Platform for Patent Information (J-PlatPat)	https://www.j-platpat.inpit.go.jp/
8.	Korean intellectual Property Rights information services (KIPRIS)	http://eng.kipris.or.kr/enghome/main.jsp
9.	Orbit Search by Questel	https://www.questel.com/business-intelligence-software/orbit-Express/
10.	PatentScope	https://patentscope.wipo.int/search/en/search.jsf
11.	USPTO	https://www.uspto.gov/
12.	WIPS Global	https://inspire.wipo.int/wips-global

15.8 Conclusion

Patents are organized according to the classification codes for indexing those regarding similarities and differences in the field of invention. Patents are made available to the public by the patent offices globally via their search engines. Where and how to search for data to retrieve a patent of interest is a challenging quest. The combination of various keywords by using logic operators and classification codes is generally used for building a search query for patent search. The resulting data can also be concise by applying various filters. Also, some search engines, for example, PatentScope, can depict the results as comprehensive charts and graphs for easy understanding. The analysis of retrieved data serves as a significant guide to check the patentability of an invention, selection of country of interest for the application of patent, keep a close eye on technological developments, survive in the world of competition, and lastly, ensure appropriate financial benefits from the invention.

References

1. Patents. https://www.wipo.int/patents/en/. Accessed 12 Apr 2020
2. Alberts D, Yang C, Fobare-DePonio D, Koubek K, Robins S, Rodgers M et al (2011) Introduction to patent searching-practical experience and requirements for searching the patent

space. In: Current challenges in patent information retrieval. The information retrieval series. Springer, Heidelberg, pp 3–43
3. Marley M (2014) Full-text patent searching on free websites: tools, tips and tricks. Bus Inf Rev 31(4):226–236
4. Intellectual property: protection and enforcement. https://www.wto.org/english/thewto_e/whatis_e/tif_e/agrm7_e.htm. Accessed 16 Apr 2020
5. Alberts D, Yang CB, Fobare-DePonio D, Koubek K, Robins S, Rodgers M et al (2011) Introduction to patent searching. In: Lupu M, Mayer K, Kando N, Trippe AJ (eds) Current challenges in patent information retrieval. The information retrieval series. Springer, Heidelberg, pp 3–43
6. The Patents Act (1970). http://www.ipindia.nic.in/writereaddata/Portal/IPOAct/1_31_1_patent-act-1970-11march2015.pdf. Accessed 12 Apr 2020
7. Clarke NS (2018) The basics of patent searching. World Patent Inf 54:S4–S10
8. Guidelines for examiners. https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_i_1.htm. Accessed 12 Apr 2020
9. Guide to the international patent classification. https://www.wipo.int/edocs/pubdocs/en/wipo_guide_ipc_2020.pdf. Accessed 12 Apr 2020
10. Patent classification. https://www.uspto.gov/patents-application-process/patent-search/classification-standards-and-development. Accessed 13 Apr 2020
11. Overview of the U.S. Patent Classification System (USPC) (2015). https://www.uspto.gov/sites/default/files/patents/resources/classification/overview.pdf. Accessed 10/04/2020
12. Cooperative patent classification. https://worldwide.espacenet.com/classification?locale=en_ep. Accessed 12 Apr 2020
13. Schwander P (2000) An evaluation of patent searching resources: comparing the professional and free online databases. World Patent Inf 22(3):147–165
14. Hantos S (2010) Helping others acquire, license or invest in patents with confidence—a guide for patent searchers to patent due diligence. World Patent Inf 32(3):188–197
15. Smith J (1988) Is 'state of the art' patently ill defined? https://www.latimes.com/archives/la-xpm-1988-06-15-vw-4099-story.html. Accessed 15 Mar 2020
16. Yang YY, Akers L, Yang CB, Klose T, Pavlek S (2010) Enhancing patent landscape analysis with visualization output. World Patent Inf 32(3):203–220
17. van Staveren M (2009) Prior art searching on the Internet: further insights. World Patent Inf 1 (31):54–56
18. Sears CN (2012) Method/system for prior art searching. Google Patents
19. Barford D. Structuring queries; using operators. In: Intellectual property and patent information search strategies. https://www.wipo.int/export/sites/www/tisc/en/ppt/Philippines/structuring_queries.pdf. Accessed 20 Apr 2021
20. Ituku EB. Patent search strategies and techniques. https://www.wipo.int/edocs/mdocs/africa/en/wipo_tiscs_znz_16/wipo_tiscs_znz_16_t_6.pdf. Accessed 20 Apr 2021
21. Rainey MM (2014) Free sources for patent searching: a review. Bus Inf Rev 31(4):216–225
22. Google Patents. https://patents.google.com/. Accessed 10 Apr 2020
23. Tiwari A. EPO adds Espacenet to Its mobile website. https://intellogist.wordpress.com/. Accessed 10 Apr 2020
24. Escapcenet. https://worldwide.espacenet.com/patent/search. Accessed 10 Apr 2020
25. Patentscope. Patentscope. https://www.wipo.int/patentscope/en/news/pctdb/2020/news_0001.html. Accessed 11 Apr 2020
26. Patentscope. https://patentscope.wipo.int/search/en/search.jsf. Accessed 11 Apr 2020
27. Patent full-text database. http://patft.uspto.gov/netahtml/PTO/index.html. Accessed 10 Apr 2020
28. Seven step strategy. https://www.uspto.gov/learning-and-resources/support-centers/patent-and-trademark-resource-centers-ptrc/resources/seven. Accessed 10 Apr 2020
29. Indian Patent Advanced Search System. https://ipindiaservices.gov.in/PublicSearch/. Accessed 11 Apr 2020

30. Searching for patents, utility models, designs, trademarks. https://www.jpo.go.jp/e/faq/yokuaru/searching.html. Accessed 20 Apr 2021
31. Japan platform for patent information operation manual. https://www.inpit.go.jp/content/100868791.pdf. Accessed 20 Apr 2021
32. Derwent World Patents Index. https://clarivate.com/derwent/solutions/derwent-world-patent-index-dwpi/. Accessed 20 Apr 2021
33. Derwent World Patents Index (DWPI) classification. https://clarivate.com/derwent/wp-content/uploads/sites/3/dlm_uploads/2019/08/DWPI-Classification-Guide-2020.pdf. Accessed 20 Apr 2021
34. Raturi MK, Sahoo PK, Tiwari AK (2009) Delphion: a world class patent database—a comprehensive analysis from patent information professional's perspective (November 21, 2009)
35. Guide to technology databases. https://www.wipo.int/edocs/pubdocs/en/patents/434/wipo_pub_l434_11.pdf. Accessed 20 Apr 2021
36. Orbit express. https://www.questel.com/business-intelligence-software/orbit-Express/. Accessed 20 Apr 2021
37. Patent databases by National Science Library. http://nsl.niscair.res.in/Patentdatabases.jsp. Accessed 20 Apr 2021
38. WIPS Global. https://inspire.wipo.int/wips-global. Accessed 20 Apr 2021

Computer Aided Drug Design

16

Bhupinder Kumar, Pooja A. Chawla, and Viney Chawla

Contents

B. Kumar · P. A. Chawla
ISF College of Pharmacy, Moga, Punjab, India

V. Chawla (✉)
University Institute of Pharmaceutical Sciences and Research, Baba Farid University of Health Sciences, Faridkot, Punjab, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_16

Abstract

Various challenges such as cost and time consumption problems associated with traditional drug design and development process have been overcome with in silico approaches known as computer-aided drug design (CADD). It involves the use of various structure-based or ligand-based computational drug designing approaches to invent, modify, and analyse various biologically active molecules or drugs. CADD approaches can speed up the possibilities of identifying molecules with desirable characteristics and features, propel hit-to-lead development, and reduce the chances of failure over the many obstacles of early preclinical studies. The two types of computational drug designing approaches are structure-based drug design (SBDD) and ligand-based drug design (LBDD). The information generated through these approaches provides an idea of the electronic properties (electrostatic potential, polarizability, etc.) required in desired molecule that will influence binding affinity and helps in optimization of known ligands for designing and development of new molecules with improved activity and safety profiles. This chapter focuses on the different computational approaches such as molecular docking, virtual screening, homology modelling, pharmacophore development, and quantitative structure-activity relationship (QSAR) as well as new developments in practical aspects and their role in pharmaceutical applications. This chapter also highlights the clinically approved drugs developed using CADD strategies.

Keywords

Computer-Aided Drug Design · Molecular Modelling · QSAR · Homology modelling · Structure-based drug design · Ligand-based drug design

Chapter Objectives

- Understanding the historical development of CADD
- Understanding the role of CADD in drug discovery
- Understanding the different CADD techniques and their applications
- Understanding the various strategies to design and develop new drug-like molecules.
- Appreciating the role of CADD through case studies

16.1 Introduction

The advent and design of new drugs in early nineteenth century was based upon prototypical molecules which could have a natural or synthetic origin. The classical examples of such an approach include of compounds from various natural origins such as plants, animals, and marine sources. These compounds include several biologically active molecules such as steroidal hormones (cortisone, oestrogen, progesterone, and testosterone), adrenergic drugs (epinephrine), local anaesthetics

(cocaine), opiate analgesics (morphine), and many more. Similarly, prototypical molecules based on synthetic origin include benzodiazepines for various CNS disorders, bisulphonates for osteoporosis, and antipsychotic phenothiazines. However, despite the success of this approach in the early decades, there are practical limitations to its use. An understanding of the structure of receptors involved in the action of drug and the aetiology of the disease can go a long way in designing the new molecules [1]. With the advent of computers and improved algorithms/software coupled with their availability at low cost, there has been a phenomenal in the number of dedicated approaches for new drug development. In line with such an approach, computer-aided drug design (CADD) was conceptualized in 1981 and has many success stories to its credit [2]. It relies on computational methodologies that include mathematical equations such as the one given below:

$$\Phi = f(C) \tag{16.1}$$

where Φ, the physiological action of a molecule, is a function of its chemical constitution C.

CADD includes finding out such relationships of biological activity with chemical structure, searching a library of chemical databases for lead identification, and quick docking of the receptor with the ligand (drug). Docking studies demand information on the three-dimensional (3D) structure of the receptor. It has become possible to predict the 3D structure of the proteins starting with the amino acid sequence. CADD enjoys an edge over high-throughput screening (HTS) since it demands very little compound design or prior information but at the same time can bring out various hit compounds. The primary motive of CADD in process of drug research is to cut short large libraries of possible compounds into smaller groups of possible active moieties, making it easier to optimize lead compounds by improving their biological properties (like affinity and ADMET). The use of clustering as a tool to choose representatives from large libraries is well documented [3].

When compared to HTS, CADD is a more specific approach aimed at generating hits. CADD makes it possible to elucidate the molecular basis for pharmacological action and feasible derivatives and identification of such functional variables that can be improvised to bring about the most suitable drug compound and leading to the ordering of the actives according to their potency. Further, CADD approach has been instrumental in saving time and cost in the search and optimization of potential lead compounds. There are reports of this approach being used at various stages in drug discovery be it target identification, active site validation, ligand design, or identification of interactions of drug candidates with biological targets under consideration [4].

CADD can be either ligand or structure-based. Ligand-based CADD depends on the similarity in chemical structure, and the predictive, quantitative structure-activity relationship (QSAR) models that it develops from the molecules to segregate the known active and inactive molecules. Structure-based CADD relies on the knowledge of the target protein structure in the determination of interaction levels of all derivatives being examined. QSAR modelling reveals an understanding of the effect

of structural factors on biological activity, using various models and the understanding to design derivatives with enhanced and optimal biological profiles. Structure-based CADD is often the preferred choice for soluble proteins that could be crystallized, while ligand-based CADD finds suitability for derivatives with high target binding affinity, devoid of off-target effects, and that could be designed with minimal free energy, favourable drug metabolism, and pharmacokinetic/ADMET properties. CADD offers a better choice where little information about the structure is available such as membrane protein targets [5].

16.2 Structure-Based Drug Design

The drug designing which depends on the information of 3D structure of biomolecule target through x-ray crystallography or NMR spectroscopy is known as structure-based drug design (SBDD) [6, 7]. Sometimes the structure of the target is not known or available. In that case modelling homology of the target with respect to the related protein is carried out. The knowledge of structure of biological target could be utilized to predict the selectivity and binding affinity of the drug with tee specific target through computerized graphics [8]. The aim of using the computer for drug design is to analyse the interactions between the drug and its receptor site and to 'design' molecules that give an optimal fit which ultimately yields a good fit resulting from structural and chemical complementarities to the target receptor. Pharmaceutical industry is widely based upon SBDD tool in designing and optimization of ligands [9–11]. A variety of lead molecules have been developed and introduced to the market by medicinal chemists. The structure-based drug design involves following steps.

1. **Protein Structure Preparation**
   Preparation of target macromolecule structure protein is the beginning of structure-based drug design. The proteins are available at the RCSB Protein Data bank (PDB) [12]. If the structure of target protein is not ascertained, computational methods are explored to determine the structure through homology modelling [13–25].
2. **Binding Site Identification**
   Binding site identification of the target protein helps in determination of protein-ligand interaction, hydrogen bond formation, post docking dynamics, complex free energy determination, etc. that will lead to identify the best pharmacophore of the ligand. X-ray crystallography helps in determination of binding site in the target protein [26, 27]. The information regarding the binding site can be obtained from co-crystallized proteins with their substrates or known inhibitors [28]. If the information regarding the binding site is unavailable, a number of online servers can be used to get the information like MSpocket, metapocket, and DEPTH CASTp to name a few followed by filtration of highly bulky ligands which are unable to approach the binding site during the lead identification process [12, 29–34].

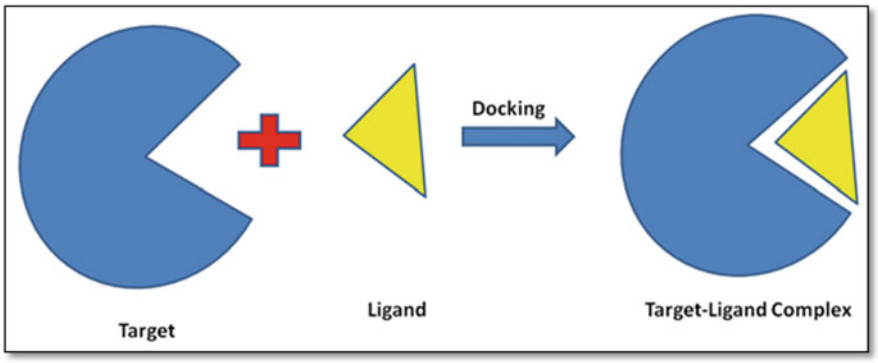


Fig. 16.1 Ligand receptor binding

3. **Ligand Library Preparation**
   The library is prepared by selecting ligands from natural and/or commercial resources, and public repositories [35]. The ligands should be subjected to screening for Lipinski's rule of five followed by ADMET studies and risk parameters before docking studies to get bioactive with requisite pharmacokinetic properties along with safety [36, 37]. The synthetic strategies for the compounds should be designed after carrying out this exercise.
4. **Docking and Scoring Functions**
   Docking is a method which gives an idea about how the ligand is going to bind to the active site of the receptors and also about the extent to which conformational changes can be brought about in the receptor structure when the ligand binds to it and hence the response elicited by the drug can be measured [38]. Active site is the site where a ligand binds to the receptor to produce a particular response (Fig. 16.1). There are around 50 different docking programmes available since the introduction of first docking programme [39].

   To start docking, first of all, generation of 3D structure of the protein target and ligands is carried out. The conformation of ligand affects the determination of binding pose. Usually, it is observed that binding of ligands to macromolecules takes place in low energy conformation but it is not mandatory as a higher energy conformation could be the bioactive conformation [40–43]. Multiple ligand conformations are generated with the help of various software including Corina/Rotate, LigPrep, ConfGen, and MacroModel [44–47]. Energy change calculations are carried out for the estimation of the chances of protein-ligand binding [48] followed by post-docking analysis involving various available scoring functions [49]. The most used scoring functions are force field analysis, empirical scoring function, and knowledge-based scoring function. Actual molecular forces between protein-ligand complex are calculated by force field analysis by optimizing electrostatic forces, van der Waals interactions, and hydrophobicity [50]. Extensive work has been done on systematic improvement [51]. Such functions include CHARMM (Chemistry AT HARward Macromolecular

Mechanics) [52], AMBER (assisted model building and energy refinement) [53], and CVFF (consistent valence force field) [54]. Other force fields with practical importance are modern force fields, quantum mechanics methods, solvation models, etc. [55].

5. **Empirical Scoring Function**
   It is the fastest one among all of the scoring functions. It optimizes empirical terms, like hydrogen bonding energy, van der Waals interaction energies, hydrophobicity, electrostatic energy, entropy, and desolvation energy to calculate the interactions between atoms of binding molecules and change in solvent-accessible surface area [56]. The knowledge-based scoring function utilizes experimentally determined structure geometrical data from databases like PDB [57]. The structural information from protein-ligand complexes is converted into distance-dependent pairwise potentials. There is a huge development in the scoring systems but none of them is perfect in terms of accuracy in hit ranking and general applicability. The scoring functions are associated with advantages and limitations. The results of multiple scoring functions are combined and a consensus proves to be fruitful in scoring and ranking of compounds in getting accurate solutions [58]. More accurate binding affinity of the ligand-receptor complex is predicted through the values from different scoring functions by applying statistical methods [59]. Various tools for consensus scoring functions include CONSENSUS-DOCK [60], MultiScore [61], SeleX-CS [62], GFscore [63], and X-Cscore [64]. SBDD has yielded a number of compounds that have reached clinical trials successfully and even got approvals for marketing. Medicinal chemists are exploiting this tool to discover and design new drug molecules. An important example of SBDD is the discovery of zanamivir, an antiviral drug against influenza A and B infections [65]. Barreca ML. et al. identified novel pyrazolobenzothiazine scaffold as anti-HCV agents with the help of SBDD after rational drug design, organic synthesis, and biological testing [66].

16.2.1 Molecular Modelling

Molecular modelling is defined as the use of theoretical and computational techniques to identify the similarity in behaviour of molecules and their molecular system. In the discovery of new compounds as drug candidates, scientists can visualize the molecule in three dimensions through molecular modelling to perform complex computer simulations on large protein molecules [67]. Molecular modelling has become quite easier due to advent of computers. A large number of human populations across the globe are combating with some dreadful diseases like cancer, AIDS, hepatitis, drug resistance in tuberculosis to name a few. Only a few drugs in the market are available that can combat these diseases. Traditional drug discovery is very expensive, complex, and time-consuming. Molecular modelling treats the diseases at molecular level to reach the root cause of the disease. Here the molecules are modelled based on structural target, that is, receptor and enzyme.

Molecular modelling is generally categorized as direct drug design and indirect drug design [68].

16.2.1.1 Strategies for Molecular Modelling

Direct Drug Designing

It is one of most used approach in drug designing in which the three-dimensional structural features of identified target site of receptor are determined using X-ray crystallography to design optimal compound with high affinity at target site as a lead molecule. In this approach, the geometry of active site of receptor is known while the molecule with optimal structural properties is designed which satisfies some geometry constraints and represents itself as a chemical match for that target site. These designed molecules with specific structural and chemical characteristics can be docked at target site using energy minimization and their binding strength or affinity can be predicted [69].

Indirect Drug Designing

This approach is based on the identification of molecules which bind to the biological target of interest. These molecules can then be aligned on each other to identify a pharmacophore model which describes the minimum structural characteristics required for a molecule in order to bind at the target. In this approach, the active site geometry is unknown and ligands are designed based on earlier reported ligands for that active site of receptor. This approach of drug discovery and development relies on efficient binding at the active site of receptor [69].

16.2.1.2 Molecular Modelling Methods

Molecular modelling methods can be classified as molecular mechanics and quantum mechanics. Molecular modelling methods generate equations for determining total energy of the structures.

Molecular Mechanics (MM)

Molecular mechanics may be defined as a process aimed at finding out the minimum of an empirical potential energy function. As a result, a molecule of idealized geometry is often produced. It acts by attempting to reproduce molecular geometry, molecular energy, and other features through adjustments of bond lengths, bond angles, and torsion angles and aligning them as close as possible to the equilibrium values [70].

MM splits pair wise interaction into

1. Bonded interaction (internal coordination): Atoms which are linked through one to three bonds.
2. Non-bonded interaction: These include electrostatic and Van der Waals interactions.

Normally, the force field equation is expressed as

$$\mathrm{EP}(X) = E_{\text{bonded}} + E_{\text{nonbonded}} \tag{16.2}$$

$$E_{\text{bonded}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} \tag{16.3}$$

$$E_{\text{Nonbonded}} = E_{\text{van der waals}} + E_{\text{electrostatic}} \tag{16.4}$$

Force Fields The MM energy expression usually consists of a simple algebraic equation capable of calculating a compound's energy. Force field refers to a set of equations with their associated constants that are involved in energy expression. There are different aspects of this equation like stretching, bending, torsion, electronic interactions Van der Waals forces, and hydrogen bonding. Some popular force fields are AMBER, CHARMM, and CVFF.

Quantum Mechanics (QM)

QM involves molecular orbital calculations and presents an exhaustive description of a molecule's chemical behaviour. HOMO—highest energy occupied molecular orbital, and LUMO—lowest energy unoccupied molecular orbital, are the frequently used terms. QM uses the basics of particle physics to evaluate structure as a function of electron distribution [71].

$$\text{Schrödinger equation} : H\Psi = E\Psi \tag{16.5}$$

E is energy of the system relative to one in which all atomic particles are separated to infinite distances.

Ψ is wave function that defines the Cartesian and spin coordinates of the atomic particles.

H is Hamiltonian operator that includes kinetic and potential energy terms.

Quantum Mechanical Methods

Ab initio methods These methods are repetitive processes based on SCF (self-consistent field) methods in which there is refinement of initial guess for the value of the coefficients till the attainment of consistent values.

Density functional theory (DFT) The properties of a many-electron system can be determined by using functionals, that is, functions of another function, which in this case is the spatially dependent electron density.

Semiempirical molecular orbital Semiempirical methods are very fast, applicable to large molecules, and may give qualitative accurate results when applied to molecules that are similar to the molecules used for parameterization.

16.2.1.3 Energy Minimization Methods

Energy minimization is the approach to identify the most stable, low energy conformation by continuously changing the configuration of a structure. Alternatively, it can also be defined as identification of a point in the configuration space at which the net force on atom is zero or minimum. In other words, it is identification of the coordinates at which the first derivative of the potential energy function is equal to zero. The potential energy function (or force field) is determined via certain algorithm or a minimizer which shifts the atoms of the molecule to a nearest local minimum. Examples: Steepest Descent, Conjugate Gradient, Newton-Raphson procedure [71].

16.2.2 High-Throughput Screening, Virtual Screening, Docking

High-throughput screening (HTS) is testing of large number of compounds for large number of biological activities. It generally involves automation, robotics, and miniaturization of in vitro tests. HTS is a key approach used for target validation. This process involves the screening of a large number of miscellaneous chemical features to identify 'HITS'. It includes selection of a target and its activity measurement. Currently used targets include cell membrane coupled receptors, ion channels, nuclear receptors, and DNA. After identification of target, a bioassay is developed. Usually, a 96-well microtitre plate is used to carry out the assay. Other microtitre plates having 384, 1536, and 3456 wells are also employed. The assays may be non-radiometric or radiometric. The non-radiometric assay utilizes fluorescence, absorbance, and luminescence spectroscopy whereas radiometric assay include filtration and scintillation proximity assay (SPA) [72]. The HTS helps in the screening of thousands of compounds on repeatable basis. Thus, it helps in identification of lead compound with optimal properties while eliminating the compounds which do not possess measurable activity or properties. HTS is advantageous in the development of drugs at fast rate, and further, it is a cost-effective approach.

16.2.2.1 Virtual Screening

Virtual screening is an approach of CADD in which libraries of small molecules available online are searched for drug discovery. It identifies the structures which possess most binding affinity to a drug target, typically a protein receptor or enzyme [73]. Virtual screening has also been employed for the discovery of bioactive natural products [74]. Virtual screening is advanced molecular modelling methods which discover the new ligands based on the biological structures. It is mainly categorized into two broad categories: ligand-based and structure-based [75, 76].

16.2.2.2 Docking

Docking is an approach to identify the best matching between two molecules. As discussed in Sect. 16.2, it is a method in which the best stable complex of ligand orientation at active site of protein or receptor is identified. Docking is most acclaimed CADD approach used for rational drug designing. It is a structure-based

drug designing approach which helps in the identification of inhibitors for specific target proteins and thus via structural modifications helps in designing of new drugs. The importance of docking is increasing day by day as the more X-ray crystallographic structures are being reported and available for docking. Other than new drug discovery, docking is also an important tool of cellular biology as it helps in identifications of interactions of proteins interacting with themselves or other molecular or cell components.

16.2.2.3 Types of Docking

There are two types of docking:

1. Rigid Docking (Lock and Key method): The docking technique in which the internal geometry of both receptor and ligand fixed (rigid) is entitled as rigid docking.
2. Flexible Docking (Induced Fit): In flexible docking, confirmations of the receptor and the ligand molecules are flexible (Fig. 16.2), as they appear in complex [77].

16.2.3 De Novo Drug Design

De novo means start afresh, from the beginning. It is a structure-based drug design approach in which the 3D structure of target receptor is used in designing of novel or newer molecules complimenting the target site. In this approach, the structure of lead compound in complex with target protein is identified and then structural modifications are performed using molecular modelling tools. Thus, it is a kind of ligand optimization approach. This ligand optimization is performed via identifying the properties or possible interactions at active site of target protein which can play key factor in ligand binding [78–80]. The main requirement of de novo drug design approach is availability of high-resolution structure of target protein along with well-defined binding site. This does not only include the shape constraint of target site but also availability of hypothetical interaction sites such as hydrogen bond acceptor and donor sites, electrostatic, and other non-covalent interaction sites. These defined interactions can reduce the sample size of compounds to large extent as hydrogen

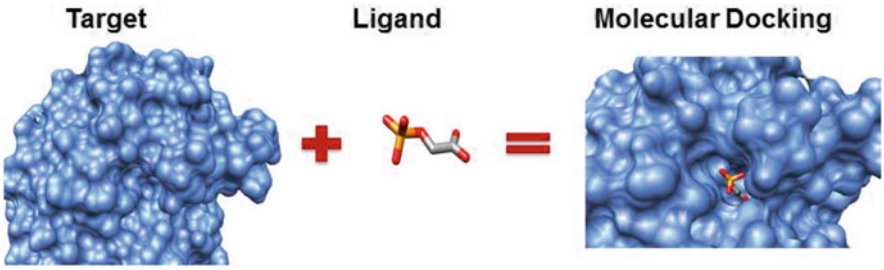


Fig. 16.2 Docking of ligand with target

bonding and other anisotropic interactions are sought to play key role in defining the specific orientation of ligands at active site. The technique uses 3D searching of large databases to identify small molecule fragments with correct size and geometry that can support functional group at favourable orientations. It constitutes an important step aimed at mimicking the closest approximation of the binding site [81, 82].

16.2.4 Homology Modelling

Homology modelling (comparative modelling) is based on determining the three-dimensional structure of a protein from its amino acid sequence. The technique is based on the assumption that in nature the structural conformation of a protein is more conserved than its amino acid sequence and any minute to medium change causes small changes in the 3D structure. If similarity between the target sequence and the template sequence is detected, structural similarity can be assumed. In general, *30% sequence identity* is required to generate a useful model [83]. In fact, this technique is regarded as the most accurate of the computational structure prediction methods [84]. There are four steps in homology modelling: fold assignment, sequence alignment, model building, and model refinement. Different steps are depicted in Fig. 16.3. In the fold assignment, there is identification of template structure (3D proteins) related to the target sequence followed by searching of proteins with known structures from protein data bank (PDB database) with target sequence using sequence similarity search algorithms or threading techniques. After development of a correlation between target protein and a known protein 3D structure, the superimposition or alignment of two proteins is carried out to get optimum correlation between two residues in the target and template sequences. The model building phase involves construction of a model of target protein from the substitution of amino acids in the 3D structure of template protein through deletion and/or introduction of amino acids as per sequence alignment. Alternatively, the two known regions may be connected through generation of a loop. After the construction of a target backbone, energetically favourable conformations can be obtained by additions of side chains. Energy minimization is carried out using force fields. The validation of the constructed model is done by recognizing the important amino acid residues followed by their mutation and their corresponding effect on binding of

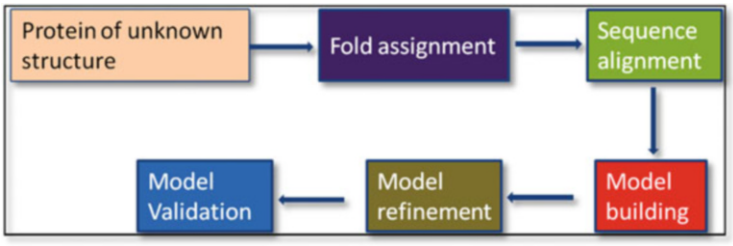


Fig. 16.3 Different steps in homology modelling

model receptor to ligand to note any specific changes [85–93]. Homology modelling has a number of advantages such as locating alpha carbons of key amino acid residues inside the folded protein, to propose structure functional relationships and mutagenesis experiments. These findings related to the positions of conserved regions of the protein binding sites can help identify key interacting amino acid residues, binding pockets, and ultimately in designing of high affinity ligands. Homology models have made remarkable effect on target identification, validation, lead identification, and optimization [94].

16.3 Ligand-Based Drug Design

It is also known as indirect drug designing. In case of absence of reliable target protein structure, ligand-based drug design (LBDD) approach is useful. Principally, LBDD approach is based on the properties of molecules containing similar kind of structures. It is sought that molecules possessing structural similarities tend to have similar properties also [95]. LBDD approach relies upon the ligands which bind to target protein with desired interactions and properties [96]. These ligands of interest are aligned to extract a suitable model with required structural properties for lead molecule which can bind to target protein with desired action. These models aim to retain physiochemical properties with respect to desired interactions required for binding to target while irrelevant interactions are discarded [97]. This in silico approach generally involves the grouping of ligands with chemical similarity known for binding to target protein and extraction of QSAR model which can predict biological activity of novel designed ligands based on this QSAR model.

16.3.1 QSAR and Historical Development of QSAR

Quantitative structure-activity relationship (QSAR) model is a mathematical relationship of structural and chemical characteristics of ligands with their biological properties [98]. The QSAR approach is a process of transferring discovered novel compounds with desired biological properties using chemical instinct and experience to a mathematically calculated and computerized form. The QSAR mathematical equation is given as:

$$\text{Biological Activity} = f(\text{Physicochemical Property}) \tag{16.6}$$

History of QSAR dates back to eighteenth century. In 1868, Crum-Brown and Fraser [99] gave an equation (Eq. 16.1) describing physiological activity (Φ) as a function of the chemical structure (C). This equation is considered as the first equation of QSAR.

In 1869, B.J. Richardson found that narcotic effect of primary alcohols varies in proportion due to their molecular weights. Similarly, in 1893, C. Richet proved that the toxicities of some simple organic compounds are inversely related to their

solubility in water [100]. In addition to this, in 1899 H. Meyer and 1901 E. Overton found the linear correlations between potencies of narcotic compounds and their oil-water partition coefficient (Log P) [101]. L. Hammett in 1930s reported the correlation of electronic properties of various organic acids and bases with their reactivity [102]. In this period, Taft suggested that the substituent effect on the free energy change behaves approximately as a sum of an inductive effect, I, and a resonance effect, R, and gave the first steric parameter termed as E_S [103, 104]. These finding were taken up by Hansch and Fujita and they developed the QSAR paradigm by combining Hammett's electronic constants with hydrophobic constants to give various linear and non-linear Hansch equations [105–107]. These equations are described below. For linear relationship the equation is:

$$\text{Log } 1/C = a\sigma + b\pi + ck \tag{16.7}$$

For non-linear relationship the equation is:

$$\text{Log } 1/C = a \text{ log } P - b(\text{ log } P)^2 + c\sigma + k \tag{16.8}$$

where C = Concentration of compound required for standard activity or response.

Log P = Partition coefficient of compound between octanol and water.

σ = Hammet substituent parameter.

π = Relative hydrophobicity of substituents.

a, b, c, k = Model coefficients.

Apart from this, Free-Wilson approach was given which described the structure-activity relationship studies in a congeneric series. The equation is given as follows:

$$\text{BA} = \sum a_i x_i + u \tag{16.9}$$

where BA = Biological activity of compound.

a_i = Contribution of each structural feature of compound.

x_i = Defines if structural feature is present ($x_i = 1$) or absent ($x_i = 0$).

u = Average contribution of the parent molecule.

There were few limitations with these equations. To overcome these limitations, more sophisticated Fujita-Ban equation was given, which used the logarithm of biological activity and brought the activity parameter in line with other free energy-related terms [108]. This equation was described as:

$$\text{Log BA} = \sum G_i X_i + u \tag{16.10}$$

where Log BA = Logarithm of biological activity of compound.

G_i = Contribution of each structural feature of compound.

X_i = Defines if structural feature is present ($x_i = 1$) or absent ($x_i = 0$).

u = Activity of parent unsubstituted molecule.

Further modifications were done in these activity-based studies were performed time to time by various scientists [109]. The introduction of topological methods was

used to describe relationship between structure of ligands and their biological activity. Based on these, minimum topological difference (MTD) method was provided by Simon. Along with this MTD, the studies on molecular connectivity by Kier and Hall lead the path to development of QSAR/QSPR. In recent developments, Hologram quantitative structure-activity relationship (HQSAR), inverse and binary QSAR are of significant importance.

16.3.2 Two-Dimensional and Three-Dimensional QSAR

16.3.2.1 Two-Dimensional QSAR

After Hansch's seminal works, different methods of QSAR have been developed. Mainly QSAR approaches can be analysed from two facts from which first is the types of structural constraints used to identify molecular structures starting from simple chemical formulas to 3D conformations. Second is the development of mathematical expression to attain the quantitative relationship of structural parameters of molecules with their biological activity.

2D-QSAR is most used category of QSAR using 2D-descriptors. In 2D QSAR methods different classes are described as given below:

1. Free energy models
   Free energy method involves Hansch analysis describing linear free energy relationship (LFER).
2. Mathematical models
   Mathematical models include Free-Wilson analysis and modifications given by Fujita-Ban.
3. Other statistical methods
   In statistical methods, different analysis parameters are included such as discriminant analysis (DA), principle component analysis (PCA), cluster analysis (CA), combine multivariate analysis (CMA), and factor analysis (FA).
4. Pattern recognition
   It is also an important technique and depends upon the similarity of pattern among the molecules and the receptor.
5. Topological methods
   It is based upon graph theory and relies on the fact that the bond connecting the atoms is considered a path undertaken by the atoms for traversing. Topological indices describe the structure in terms of flexibility, ring, and branching.
6. Quantum mechanical methods (as discussed in Sect. 16.2).

Hansch analysis (LFER) It is defined as one of the most successful approach for determinations of interaction ligands with biological target. Hansch Corwin described a linear free energy equation to determine activity of biologically active molecules. It is also termed as linear free energy relationship (LFER) or thermodynamic method. This method describes the various additives, electronic, steric, hydrophobic, and dispersion effects of substituents involved in non-covalent

interactions of ligand and target biomolecule. It is assumed that all these parameters play a key role in efficacy and potency of drug molecule. Hansch analysis is the extended form of Hammet equation from physical organic system to biological targets. According to Hansch analysis action of drug depends upon mainly two processes: the partition coefficient or log P which is important in determining the journey of drug molecule from entry point to target site via passing various membranes, and the interactions of drug molecule with target biomolecule which is mainly dependent on bulk of substituents (steric parameters) and electron density (electronic parameters) of substituents. Based on these parameters, two equations, that is, linear and non-linear equations, are given.

For linear relationship the equation is:

$$\text{Log } 1/C = a \log P + b\sigma + bE_s + d \tag{16.11}$$

For non-linear relationship the equation is:

$$\text{Log } 1/C = a\ (\log\ P)^2 + b(\log\ P) + c\sigma + dE_s + e \tag{16.12}$$

where C = Concentration of compound required for standard activity or response.

Log P = Partition coefficient of compound between octanol and water.

σ = Hammet substituent parameter.

E_s = Electronic parameter.

Log P, σ, and E_s are independent parameters for which values are obtained from experimental tables.

a, b, c, d, and e = Constants for particular biological activity.

Free-Wilson analysis Free-Wilson analysis is a mathematical approach that counts the contribution of number of structural factors to their biological activities. The congeneric equation of Free-Wilson analysis is as follows:

$$\text{BA} = \sum a_i x_i + u \tag{16.13}$$

where BA = Biological activity of compound.

a_i = Contribution of each structural feature of compound.

x_i = Defines if structural feature is present ($x_i = 1$) or absent ($x_i = 0$).

u = Average contribution of the parent molecule.

In this method, indicator x_i is used to indicate if particular structural feature is present or not and if they contribute to biological activity or not. The drawback of this equation is that it involves calculations of large number of parameters which results in loss of degree of freedom statistically. To overcome these limitations, in 1971, a more sophisticated equation was given by Fujita-Ban. A slight modification to their equation brought the activity parameter in line with other free energy-related terms [108]. This equation was described as:

$$\text{Log BA} = \sum G_i X_i + u \tag{16.14}$$

where Log BA = Logarithm of biological activity of compound.
G_i = Contribution of each structural feature of compound.
X_i = Defines if structural feature is present ($x_i = 1$) or absent ($x_i = 0$).
u = Activity of parent unsubstituted molecule.

Statistical Methods

No single method is best suitable to overcome all problems. Thus, statistical methods are most suitable which develops an equation based on correlation between independent variables such as structural features or molecular descriptors and dependents variables such as biological activity. These statistical modelling methods include number of analytical methods such as principle component regression (PCA), discriminant analysis (Linear discriminant analysis and multiple regression analysis), partial least squares (PLS) regression, genetic function approximation (GFA), and genetic partial least squares (G/PLS) techniques. Some of these are discussed below.

Principle component analysis (PCA) It is a method which works via recognizing patterns in data, and expressing these data sets on the basis of their similarities and differences. PCA is one of most useful data compression techniques which works via dropping the number of dimensions without loss of information [110, 111]. Thus, it becomes most useful technique for identifying patterns in data of high dimensions. It represents multivariate data to low-dimensional space variables of data set which are known as principle components. These variables represent the most of data represented by independent variables.

Discrimination analysis It is a statistical method which plays role as foundation in development of QSAR. Most commonly used discrimination analysis is linear discrimination analysis (LDA) over multiple linear regression (MLR). First, LDA was given by Fisher in 1936 [112]. This technique implies to the data having categorized target properties and molecular parameters which are continuous variable [113]. It is most suitable approach also because data is used in sets like active and inactive. LDA tries to reduce the variance within class while variance is increased between classes. Thus, discriminant analysis is only applied in cases where classes are separated symmetrically on basis of parameters.

Partial least square (PLS) PLS was developed by Word et al. and is a generalized form of MLR [114, 115]. It lies somewhere between MLR and principal component analysis. MLR finds the single variable to correlate the predictor with predicted variable while on the other hand PCR counts the greatest amount of variance in the predictors. The PLS regression works on the assumption that the independent variables and the dependent variables can be expected onto a low-dimensional factor

space and there is a linear correlation present between the scores of these two sets. PLS tries to search factors having variance and can achieve correlation also.

Genetic function approximation (GFA) Genetic function approximation technique was given by David Rogers [116]. It develops a QSAR model via combining genetic algorithms to other statistical tools. GFA allows the development of superior quality prediction model and provide some additional information which is not possible in case of standard regression techniques.

16.3.2.2 Three-Dimensional QSAR

3D-QSAR is the advanced technique in QSAR which aims to develop the quantitative correlation between the biological activity and three-dimensional structural properties of molecules. In this, 3D descriptors such as electrostatic and steric factors of molecule structure are calculated with other statistical methods and correlated to their biological activity. 2D-QSAR studies involve use of different techniques such as computational molecular field analysis (CoMFA), comparative molecular indices analysis (CoMSIA), and 3D-pharmacophore generation.

Computational Molecular Field Analysis (CoMFA)

It is a 3D-QSAR technique which utilizes graphics and statistical tools for developing a correlation of molecular structure with biological activities. The main idea behind CoMFA is that the differences observed in biological activities are mainly due to the difference in molecular shape of compound. Thus, to calculate these shape parameters in CoMFA, Lennard–Jones and Coulombic potentials are used to calculate the steric and electrostatic grids respectively. These parameters are sampled at regular intervals in well-defined region. In this process, the compound is selected in such conformation which is biologically active and superimposed with other biologically active molecules according to their interactions at target proteins. After that CoMFA perform the comparison of 3D-strutural parameters (steric and electrostatic) with different probes and generates a quantitative model with information significant for biological activity. CoMFA mainly focuses on alignment of molecules which can result in some errors in calculation of electrostatic and steric parameters.

Comparative Molecular Indices Analysis (CoMSIA)

COMSIA is a modified version of CoMFA with aim to improve some issue with the latter. CoMSIA is ligand-based approach dependent on alignment of ligands and is a linear QSAR method. The working of CoMFA and CoMSIA is almost similar where the similarity of molecular structures is distance-dependent to the atoms creating the molecules and is evaluated at each grid node, but in case of CoMSIA, the grid nodes falling within the molecular volume are also taken into account whereas they are left in case of CoMFA. Unlike CoMFA, SEAL alignment algorithm is used for calculation of steric and electrostatic parameters in case of CoMSIA [117]. In CoMSIA, an energy grid is generated with similar probes positioned at each point grid lattice. Moreover, hydrophobicity term is included in it with purpose of enthalpic component of desolvation. After this, similarity at each grid point is calculated for

electrostatic, steric, and hydrophobic parameters. Based on these descriptors, a correlation model is generated with their biological activity [118].

16.4 List of Software

This section highlights some of the important software used in molecular docking are listed in Table 16.1.

16.5 Applications of CADD

The pivotal role of CADD in the modern discovery of therapeutically important small molecules is well documented. It is possible to speed up various steps such as identification of lead molecules and optimization through CADD. Thus, CADD is a set of useful tools and resources for modelling, storage, analysis, and management of compounds [119]. Therefore, it is no wonder that they find profound applications in different steps of drug discovery, right from designing libraries of small molecules, identification of hits to the optimization of the affinity and selectivity of reference molecules. Digital repositories are a handy tool available with scientists involved in studying significant chemical interactions [120]. It is possible to maximize the yield of true leads thereby minimizing redundancy by optimizing the diversity of a library or its similarity to a particular target [121]. These virtual libraries allow for both sequential as well as parallel selections of suitable motifs on the basis of preferred molecular profiles. A series of tools exist in public domain, with different methods and algorithms, for identification of protein binding sites and molecular functions [122, 123] as well as for designing of molecules possessing favourable physico-chemical properties to be used as drug candidates [124]. Some early success stories of structure-based design include the HIV protease inhibitors Indinavir, Nelfinavir, Ritonavir, and Saquinavir and the carbonic anhydrase inhibitor Dorzolamide [125]. These tools and resources go a long way in improving efficiency of new drug development process and bring down the failure rate during costly late stage clinical trials.

Table 16.1 Important software used to carry out molecular modelling

S. no.	URL	Category of source	Software name
1.	https://www.schrodinger.com/products/glide	Private	About Glide Software
2.	http://autodock.scripps.edu	Private	Autodock Software
3.	https://www.chemcomp.com/	Private	MOE software
4.	http://www.scfbio-iitd.res.in/sanjeevini/sanjeevini.jsp	Government	Sanjeevini Software
5.	https://www.3ds.com	Private	Biovia Discovery Studio Software

16.6 Drugs Developed Using CADD

Generally, the process of drug development from lead molecule takes around average 15 years and costs approximately more than 1 billion dollars. Designing of drugs using computer applications (CADD) reduced the time consumption and cost of drug development. It also reduced the risk factor associated to failure of molecules in preclinical or early clinical studies. Some drug molecules like angiotensin-converting enzyme (ACE) inhibitor captopril, carbonic anhydrase inhibitor dorzolamide, HIV protease inhibitor saquinavir, renin-angiotensin system inhibitor aliskiren, TNF-α converting enzyme inhibitor TMI-005, and Human rhinovirus (HRV) 3C protease inhibitor rupintrivir are discovered or optimized using CADD approach are under various clinical trials or in market which are discussed below.

16.6.1 Captopril (Capoten®, Bristol Myers-Squibb)

Captopril, discovered in 1977 by US based company Bristol Myers-Squibb (BMS), is considered as the first success of structure-based drug designed. The successful designing of this drug was related to the findings that the mechanism of action of enzyme ACE mimics to carboxypeptidase A enzyme except that ACE plays a key role in cleavage of a dipeptide while carboxypeptidase A causes cleavage of single amino acid from the protein [126]. It established the platform for development of captopril. Further, the discovery of carboxypeptidase A inhibitor, L-benzyl succinic acid [1], and ACE inhibitor action of BPP5a (HOOC-Glu-Lys-Trp-Ala-Pro-NH) provided other information required for the development of captopril. Moreover, L-benzyl succinic acid was identified as by-product of carboxypeptidase A inhibitor [127]. Further findings sought that the N-terminal peptide fragments, tetra, tri, and dipeptide fragments (Ala-Pro) retained inhibitory activity of the parent molecule. Thus, formation of two peptide fragments information was used for designing of captopril and L-benzyl succinic acid was used as model entity. It was proposed that succinyl amino acids could act as by-product inhibitors of ACE. Thus, the amino functional group was interchanged with isosteric functional group methylene. Further, using succinyl-proline 2 scaffold (**2**) as lead, SAR studies were performed which guided to the synthesis of captopril (**4**, $IC_{50} = 23$ nM) [128]. In these SAR studies, firstly, the carboxylic functional group of succinic acid was replaced with mercapto moiety which resulted in compound having stronger zinc-coordinating property. During this, a stereospecific *R*-methyl group was merged with succinyl moiety to match naturally present methyl group in the L-Ala residue of Ala-Pro (**3**). These all efforts initiated by Cushman and Ondetti brought the development of Captopril which was approved by FDA on April 6th, 1981, and became frontline drug for treatment of hypertension and heart attack [129–132]. In 1996 on expiration of patent rights of BMS, it became a generic drug (Fig. 16.4).

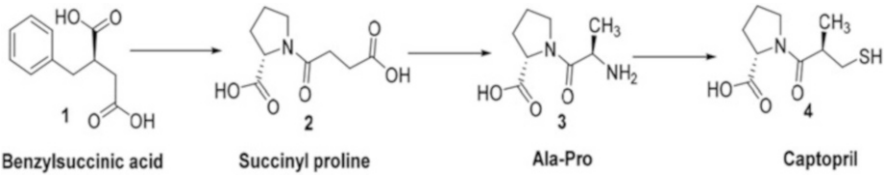


Fig. 16.4 Development of Captopril using structure-based drug design approach

16.6.2 Dorzolamide (Trusopt®, Merck)

Dorzolamide, a carbonic anhydrase (CA) inhibitor, used as antiglaucoma agent is the first example of a drug in market developed from experimentally available X-ray crystal structure of target protein. CA inhibitors are known for inhibiting the conversion of carbon dioxide to bicarbonate which is a key step in active secretion of aqueous humour. Aqueous humour is the fluid which is filled in the anterior and posterior chambers of the eye [133]. The isoenzyme form of CA in these cells is CA II, and the X-ray crystal structural information helped in designing the molecules selectively for this target protein [134]. From crystal structure understanding of CA II, it was observed that its active site is cone shaped and is narrowed down towards catalytic hydrophobic pocket. In this pocket, zinc forms complex with three histidine amino acid residues (His94, His119, and His96) of isoenzyme and forms a tetrahedral geometry [135]. The active site is well organized such that it is hydrophobic on one side while on other side hydrophilic amino acid residues are present which make its active site amphiphilic [136, 137].

As in case of captopril, above discussed approach was used for the structural optimization of a series of carbonic anhydrase inhibitors having suitable hydrophilicity and lipophilicity balance with a target to maximize potency for its use in glaucoma. The development was started from a water-soluble compound MK-927 (**5**) which was reported to penetrate ocular tissue easily in animal models and lowered intraocular pressure (IOP). In further studies, it was observed that it has two isomers (*R, S*) and has difference in affinity for target protein by 100-folds [138]. To understand this difference, both *R*- and *S*- enantiomers were crystallized with CA II and were found to bind to target protein in almost similar manner. In ligand-protein complex structures of both enantiomers, sulfonamide group was found to coordinate with zinc metal ion present at the catalytic site via deprotonated sulfonamide nitrogen. The thiophene ring of both enantiomers was positioned well in between hydrophobic and hydrophilic walls of the active-site cavity [136]. The alkyl amino group was found in less favoured pseudoaxial orientation in both cases. These results suggested almost similar configuration of both enantiomers, although two significant differences were found. First difference was found to be the N-S-C-S dihedral angle which was 150° for *S*-enantiomer while 170° for *R*-enantiomer. Ab initio studies suggested that this twist of 20° of the thiophene ring plays a key role in the reduction affinity of *R*-enantiomer by ΔH of 1 kcal/mol, although preferred angle is of 72°. Second difference found was geometry of the 4-isobutylamino substituent

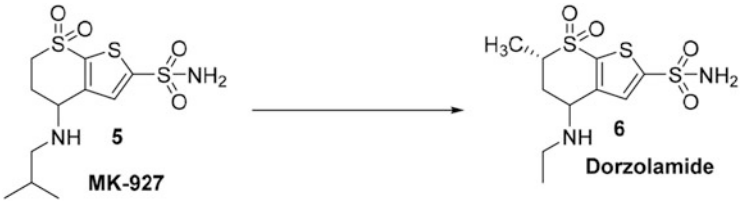


Fig. 16.5 Development of dorzolamide using structure-based drug design approach

present in structure. In *S*-enantiomer this substituent was found in trans form which was found to favour in binding by ΔH of 1 kcal/mol over *Gauche* conformation found in *R*-enantiomer. These two conformational changes were sought to play key role in 100-folds affinity difference for both enantiomers [139]. Further, as isobutylamino group was found to orient in higher energy conformation at the active site of CA II, it required modification. To counter this problem, a methyl group was inserted at sixth position of thienothiopyran ring to eliminate the pseudoequatorial conformation. As lipophilicity was enhanced on introduction of new methyl group, 4-isobutylamino group was replaced by an ethylamino moiety to balance hydrophilicity and hydrophobicity balance which lead to the discovery of dorzolamide (**6**). From all possible conformations of dorzolamide, *S,S*-conformation was found to possess highest affinity for CA II with *Ki* value of 0.37 nM. It was approved by FDA in 1994 for treatment of glaucoma [140–142] (Fig. 16.5).

16.6.3 Saquinavir (Invirase®, Hoffmann-La Roche)

Saquinavir is a HIV protease inhibitor sold under brand name Invirase® by Hoffmann-La Roche. Specifically, it inhibits the human immunodeficiency virus-1 protease (HIV-1 PR) and results in the production of premature and non-infectious viral particles and halts the progress of infection in infected persons [143]. Saquinavir was designed by taking lead from the substrate Phe-Pro which is sought to be cleaved by HIV-1 PR but not by mammalian proteases. This feature of substrate was important in designing selective compound for target protein. Further, the complex of Ro 31-8558 (**7**, peptidic inhibitor) with HIV-1 PR was subjected to X-ray crystallographic studies and found that it binds at target site in expected orientation and represented some modification sites for further development [144, 145]. This lead structure optimization led to development of Ro 31-8959 (**8**) and target drug saquinavir [146]. Ro 31-8959 is a pentapeptide derivative possessing a transition state moiety hydroxylethylamine, decahydro-isoquinolin-3-carbonyl (DIQ), and quinoline groups in place of cleavable peptide bond, at P1' and P3 sites. Saquinavir is reported to inhibit the viral replication in last stage and was found to possess potent activity against both HIV-1 and HIV-2 viruses [147, 148]. In 1995, the FDA approved saquinavir for use in adults in combination with nucleoside analogues [149] (Fig. 16.6).

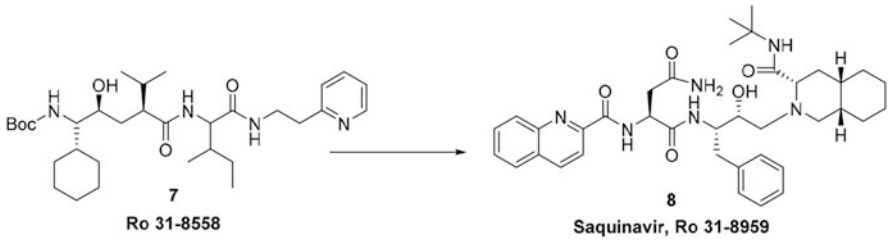


Fig. 16.6 Development of Saquinavir using CADD approach

16.6.4 Aliskiren (Tekturna®, Novartis)

Aliskiren is a renin-angiotensin system (RAS) inhibitor used for treatment of hypertension. Renin-angiotensin system is key system where angiotensin II is synthesized from angiotensinogen via synthesis of angiotensin I. The crystallographic structural data of renin-inhibitor complexes combined with CADD provided important insights into designing novel potent inhibitors with suitable pharmacokinetic properties [150].

First-generation peptide-like renin inhibitor, CGP29287 (**9**) was used as a model in the development of aliskiren due to its structural similarity with natural substrate of the renin system. It led to the development of second-generation peptide-like inhibitor, CGP38560 (**10**), which failed in clinical studies due to low oral absorption and fast excretion [151, 152]. These failures of peptide-like inhibitors suggested that non-peptide inhibitors might be more successful to overcome difficulties like poor oral absorption, poor pharmacokinetic profile, and low specificity. Goschke et al. explored the cavity of renin and used molecular modelling studies to design biologically active compounds having no P1–P4 backbone which was a key structural functionality in first- and second-generation peptide-like inhibitors. It led to use of dipeptide-like hydroxyethylene transition state which mimics the P3–P1 pharmacophore and led to the development of novel compound **11** [153]. During preclinical studies, this lead compound was found to possess high inhibitory potency against purified recombinant human renin at low nanomolar concentrations but was found to be less active in presence of plasma. Thus, further computational molecular modelling approaches were combined with available X-ray crystallographic structural information of enzyme–inhibitor complex for optimization of lead to improve its potency. During these optimization studies, the tertiary butyl and methyl acetoxy substituents present at the aromatic ring of compound **10** were swapped with smaller polar groups $–OCH_3$ and alkylether respectively to give renin inhibitor with activity in low nanomolar range. Further, from X-ray structure of enzyme-**10** complex, S3 subpocket was identified which was found to be unoccupied by peptide-like inhibitors. Exploration of this subpocket and optimizing lead to bind at this site significantly enhanced the binding affinity for renin along with high selectivity over related aspartic peptidases. These structure-based studies concluded that methoxypropoxy side chain is of optimal length for binding and the position of its

distal ether oxygen for forming H-bond with Tyr14 of the S3 subpocket. This leads to journey from lead **11** to Aliskiren (**12**) (SPP-100) via a number of other potent renin inhibitors [154]. In 2007, this small molecule renin inhibitor got FDA approval for use as renin inhibitor [155, 156].

16.6.5 TMI-005 (Apratastat, Wyeth Research)

TMI-005 is a tumour necrosis factor-α converting enzyme (TNF-α convertase, TACE) inhibitor to be used in inflammatory diseases. TNF-α is found to be produced above normal level in inflammatory diseases such as rheumatoid arthritis, multiple sclerosis, diabetes, ulcerative colitis, and Crohn's disease, and in congestive heart failure. Thus, it can play key role as target in these disease conditions [157, 158]. For discovery of TMI-005, early sulfonamide hydroxamate TACE inhibitors (**13** and **14**) were used as lead. The X-ray crystallographic data of these inhibitors crystallized with target enzyme suggested some modifications for increasing potency and selectivity over other protease enzymes. The main difference in binding pocket of TNF-α from other metalloproteases (MMPs) observed was in S1' pocket which resembled with MMP-1 in initial region but rest of region was quite different from MMP-9 and MMP-13 and existed in extended form [159, 160]. Taking lead from these findings, solid combinatorial chemistry approach was utilized for synthesis of library of phenoxyacetylene hydroxamates which provided SAR for their potency and selectivity. Further, optimization for pharmacokinetic properties and enhancing in vivo efficiency led to discovery of TMI-005 (**15**). It reached to phase-II trial against rheumatoid arthritis but due to lack of efficiency this trial was terminated (https://drugs.ncats.io/drug/C6BZ5263BJ) [161, 162] (Figs. 16.7 and 16.8).

16.6.6 Rupintrivir (AG7088, Agouron)

Rupintrivir is a human rhinovirus (HRV) 3C protease inhibitor. This is an enzyme required in replication of viruses in viral cold and upper respiratory infections [163]. This enzyme possesses characteristics of both serine and cysteine protease enzymes which make it unique from other class of enzymes. The earlier reported 3C protease inhibitors were peptide aldehyde-based serine and cysteine protease inhibitors. In these studies, it was observed that carbonyl residue of this aldehyde moiety can bind to cysteine or serine residues of respective protease enzyme in covalent and reversible manner. It was observed the interaction of glutamine amino acid residue present at P1 pocket with aldehyde moiety leads to the cyclization [164, 165]. In structural optimization, to overcome this problem γ-carboxamide moiety was introduced to aldehyde group which solved this problem without any effect on its affinity for S1 subpocket of 3C protease enzyme. It gave a potent type 14 rhinovirus 3C protease inhibitor **16** with *Ki* value of 6 nM [166]. Further, structure-based lead optimization was performed on various 3C protease enzymes in complex with compound **16** and it was observed that residues 69 (Lys or Asn) and

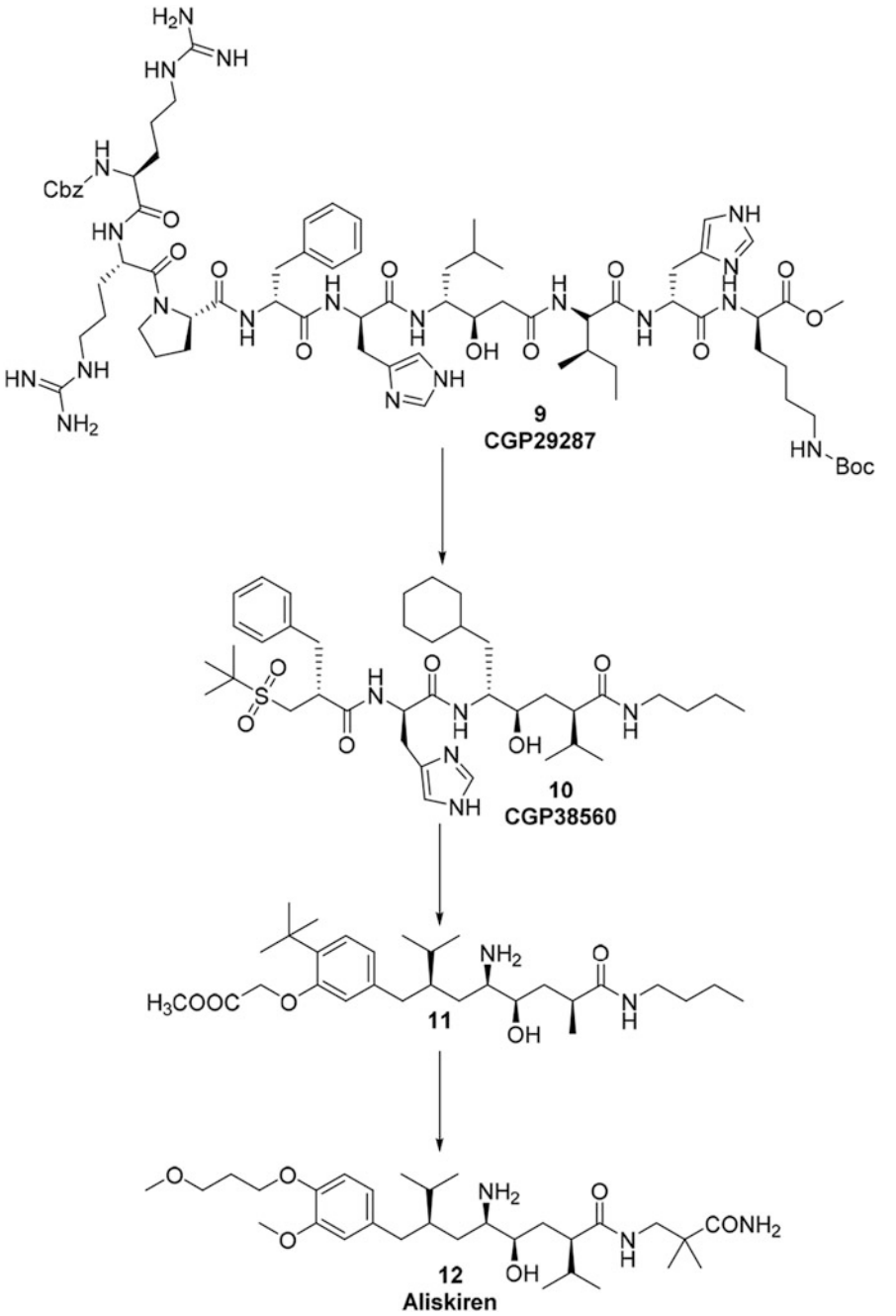


Fig. 16.7 Development of Aliskiren from first-generation peptide-like renin inhibitors

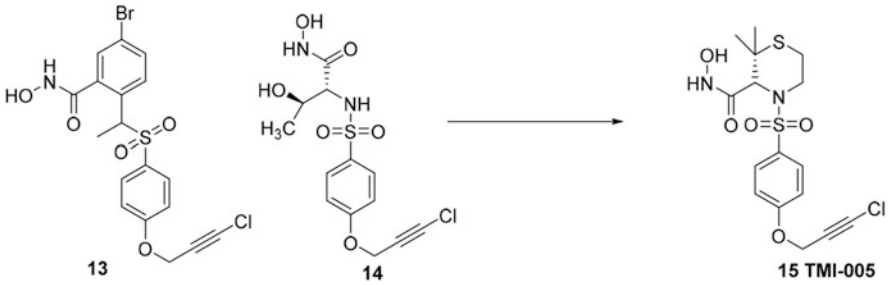


Fig. 16.8 Development of TMI-005 using combinatorial chemistry and CADD techniques

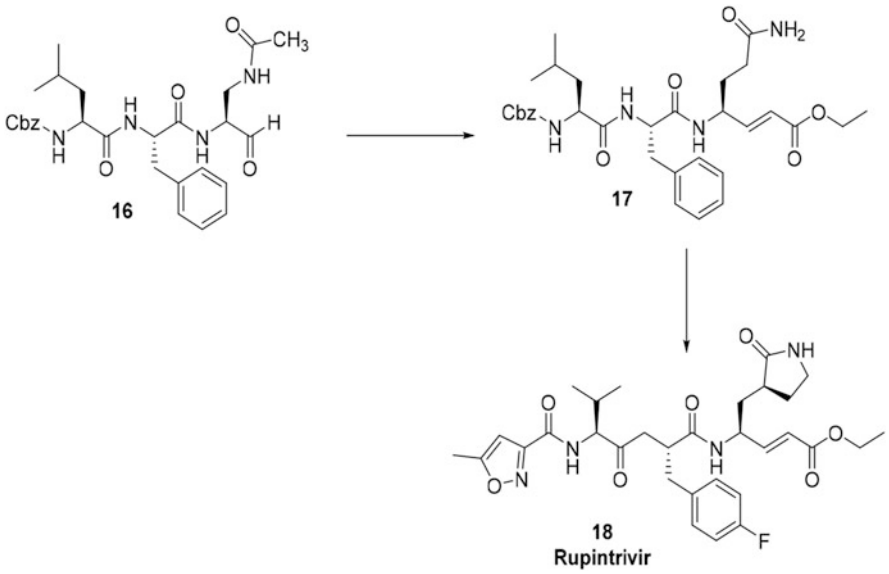


Fig. 16.9 Development of Rupintrivir as an anti-HIV agent

130 (Asn or Thr), depending of type of protease, which are located in S2 specificity pocket are not conserved with their side chains interacting with compound **16** [167, 168]. Further, extensive SAR studies were conducted which resulted in the discovery of trans-α,β-unsaturated esters as the Michael acceptors of choice that initiated the optimization around compound **16.** Taking this lead, SAR studies were carried out around tripeptide derivatives at different sites which led to discovery of Rupintrivir (**18,** AG7088) [169–171]. In initial studies, rupintrivir was found to be successful during phase II/III trials in experimentally infected patients with HRV. But it was unable to alleviate disease severity in patients infected from natural rhinovirus infections which halted the further clinical development of this drug molecule [172] (Fig. 16.9).

Table 16.2 List of some approved drugs developed using CADD tools

S. no.	Name of drug molecule	Therapeutic target/action	Status or year of FDA approval
1.	Indinavir	HIV inhibitor	1996
2.	Ritonavir	HIV inhibitor	1996
3.	Triofiban	Fibrinogen antagonist	1998
4.	Zanamivir	Neuraminidase inhibitor	1999
5.	Oseltamivir	Antagonist of influenza A and B viruses	1999
6.	Raltegravir	HIV inhibitor	2007
7.	Crizotinib	Renal cell carcinoma therapy	2011
8.	Axitinib	Breast cancer therapy	2012
9.	Neratinib maleate	Breast cancer therapy	2017
10.	Luminespib	Anti-cancer therapy	2018
11.	Selinexor	Multiple myeloma therapy	2019
12.	LY-517717	Serine protease inhibitor	Under phase-II studies
13.	Boceprevir	Hepatitis C virus inhibitor	Under phase-III studies
14.	Nolatrexed	For treatment of lung cancer	Under phase-III studies
15.	NVP-AUY922	Heat shock protein 90 inhibitor	Under phase-I studies

Besides this, there are a number of drugs which were discovered using CADD tools and are in market or under various clinical studies [173]. These are described in Table 16.2.

16.7 Conclusion

The success of CADD in bringing many drugs to the market is well documented. It has also reduced the chances of failure rate of development of drug molecules in later stages of clinical trials. A number of tools and resources are freely available for CADD. In fact, it is a suitable alternative and complement to high-throughput screening. However, the choice of software is dependent on the targets of interest and data availability. This chapter has presented a summary of various tools and methods for discovery of new drug candidates. With the advent of newer techniques in the areas of high-throughput screening, bioinformatics, chemical and structural biology, more and more tools are becoming available for a medicinal chemist to further explore and predict the activity of probable motifs, thereby making it possible to exploit the full potential of CADD in future.

16.8 Credible Online Resources for Further Reading

URL	Category of source	What to read
https://www.click2drug.org/	Private (Swiss Institute of Bioinformatics) situated in Switzerland	Click2Drug contains a comprehensive list of CADD software, databases, and web services. These tools are classified according to their application field
https://www.springer.com/gp/book/9789811568145	Private (Computer-Aided Drug Design) book published by Springer	This book provides up-to-date information on bioinformatics tools for the discovery and development of new drug molecules
https://link.springer.com/protocol/10.1007/978-1-4939-7756-7_1	Private (Computer-Aided Drug Design: An oOverview) book chapter published by Springer	This chapter provides an overview on possible approaches to identify active scaffolds (including in silico approximations to approach that task) and computational methods to guide the subsequent optimization process
https://nptel.ac.in/courses/102/106/102106070/	Private	A series of video lectures on drug designing process, molecular and quantum mechanics
https://www.accessengineeringlibrary.com/content/book/9780071701242/chapter/chapter1	Private (Computer-Aided Drug Design and Delivery Systems) published by McGraw-Hill	This book describes the role of computational techniques in drug designing and drug delivery systems
http://pubs.ccmsi.us/pubs/HCC3-2017-2041.pdf	Private	This manuscript defines the role of quantum and molecular mechanics in drug designing along with applications
https://www.google.co.in/books/edition/Drug_Design/EOInhAYMae4C?hl=en	Private (Drug Design Structure- and Ligand-Based Approaches) published by Cambridge University Press	This book discusses the details of structure-based and ligand-based drug designing techniques in details
https://www.tandfonline.com/doi/abs/10.1080/07391102.2014.971429?journalCode=tbsd20	Private	This manuscript describes the role of homology modelling in drug designing and gives interesting examples

References

1. Beale JM, Block JH (eds) (2011) Wilson and Gisvold's textbook of organic, medicinal and pharmaceutical chemistry. Wolters Kluwer and Lippincott Williams and Wilkins. http://repository.fue.edu.eg/xmlui/bitstream/handle/123456789/1520/443.pdf?sequence=1
2. Van Drie JH (2007) Computer-aided drug design: the next 20 years. J Comput Aided Mol Des 21:591–601. https://doi.org/10.1007/s10822-007-9142-y
3. Böcker A, Derksen S, Schmidt E, Teckentrup A, Schneider G (2005) A hierarchical clustering approach for large compound libraries. J Chem Inf Model 45:807–815. https://doi.org/10.1021/ci0500029
4. Kapetanovic IM (2010) Computer-aided drug discovery and development (CADDD): *in silico*-chemico-biological approach. Chem Biol Interact 171:165–176. https://doi.org/10.1016/j.cbi.2006.12.006
5. Kalyaanamoorthy S, Chen YP (2011) Structure-based drug design to augment hit discovery. Drug Discov Today 16:831–839. https://doi.org/10.1016/j.drudis.2011.07.006
6. Reynolds CH, Merz KM, Ringe D (eds) (2010) Drug design: structure- and ligand-based approaches, 1st edn. Cambridge University Press, Cambridge. https://www.google.co.in/books/edition/Drug_Design/R12sAwAAQBAJ?hl=en&gbpv=0
7. Leach AR, Jhoti H (2007) Structure-based drug discovery. Springer, Berlin. https://doi.org/10.1007/1-4020-4407-0. ISBN 978-1-4020-4406-9
8. Mauser H, Guba W (2008) Recent developments in de novo design and scaffold hopping. Curr Opin Drug Discov Devel 11:365–374. https://pubmed.ncbi.nlm.nih.gov/18428090/
9. Lavecchia A, Di Giovanni C (2013) Virtual screening strategies in drug discovery: a critical review. Curr Med Chem 20:2839–2860. https://doi.org/10.2174/09298673113209990001
10. Park H, Chien PN, Ryu SE (2012) Discovery of potent inhibitors of receptor protein tyrosine phosphatase sigma through the structure-based virtual screening. Bioorg Med Chem Lett 22:6333–6337. https://doi.org/10.1016/j.bmcl.2012.08.081
11. Jorgensen WL (2004) The many roles of computation in drug discovery. Science 303:1813–1818. https://doi.org/10.1126/science.1096361
12. Tan KP, Nguyen TB, Patel S et al (2013) Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. Nucleic Acids Res 41(W1):W314–W321. https://doi.org/10.1093/nar/gkt503
13. Gulerez IE, Gehring K (2014) X-ray crystallography and NMR as tools for the study of protein tyrosine phosphatases. Methods 65:175–183. https://doi.org/10.1016/j.ymeth.2013.07.032
14. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM (2018) PDBsum: structural summaries of PDB entries. Protein Sci 27:129–134. https://doi.org/10.1002/pro.3289
15. Svensson F, Bender A, Bailey D (2018) Fragment-based drug discovery of phosphodiesterase inhibitors: miniperspective. J Med Chem 61:1415–1424. https://doi.org/10.1021/acs.jmedchem.7b00404
16. França TC (2015) Homology modeling: an important tool for the drug discovery. J Biomol Struct Dyn 33:1780–1793. https://doi.org/10.1080/07391102.2014.971429
17. Chaudhari R, Heim AJ, Li Z (2015) Improving homology modeling of G-protein coupled receptors through multiple-template derived conserved inter-residue interactions. J Comput Aided Mol Des 29:413–420. https://doi.org/10.1007/s10822-014-9823-2
18. Yadav R, Deepshikha D, Srivastava P (2017) Homology modeling and protein interaction map of CHRNA7 neurogenesis protein. Ann Neurosci 24:173–179. https://doi.org/10.1159/000477155
19. Nim YS, Sun S, Wong KB (2017) Using homology modeling to understand the structural basis of specific interaction of a plant-specific AtSar1a–AtSec23a pair involved in protein ER export. In: Plant protein secretion. Humana Press, New York, pp 59–73. https://doi.org/10.1007/978-1-4939-7262-3_5

20. Jo T, Hou J, Eickholt J, Cheng J (2015) Improving protein fold recognition by deep learning networks. Sci Rep 5:17573. https://doi.org/10.1038/srep17573
21. Jo T, Cheng J (2014) Improving protein fold recognition by random forest. BMC Bioinform 15:S14. https://doi.org/10.1186/1471-2105-15-S11-S14
22. Wang Q, Yan J, Li X (2014) Protein fold recognition based on functional domain composition. Comput Biol Chem 48:71–76. https://doi.org/10.1016/j.compbiolchem.2013.12.001
23. Adhikari B, Cheng J (2017) Improved protein structure reconstruction using secondary structures, contacts at higher distance thresholds, and non-contacts. BMC Bioinform 18:380. https://doi.org/10.1186/s12859-017-1807-5
24. Fischer AW, Anderson DM, Tessmer MH et al (2017) Structure and dynamics of type III secretion effector protein ExoU as determined by SDSL-EPR spectroscopy in conjunction with de novo protein folding. ACS Omega 2:2977–2984. https://doi.org/10.1021/acsomega.7b00349
25. Karakaş M, Woetzel N, Staritzbichler R et al (2012) BCL: fold-de novo prediction of complex and large protein topologies by assembly of secondary structure elements. PLoS One 7: e49240. https://doi.org/10.1371/journal.pone.0049240
26. Zhang Y, Zhang D, Tian H et al (2016) Identification of covalent binding sites targeting cysteines based on computational approaches. Mol Pharm 13:3106–3118. https://doi.org/10.1021/acs.molpharmaceut.6b00302
27. Pan L, Gardner CL, Pagliai FA, Gonzalez CF, Lorca GL (2017) Identification of the tolfenamic acid binding pocket in PrbP from Liberibacter asiaticus. Front Microbiol 8:1591. https://doi.org/10.3389/fmicb.2016.01591
28. Gomez-Gutierrez P, Rubio-Martinez J, Perez JJ (2017) Identification of potential small molecule binding pockets in p38α MAP kinase. J Chem Inf Model 57:2566–2574. https://doi.org/10.1021/acs.jcim.7b00439
29. Zhang Z, Li Y, Lin B, Schroeder M, Huang B (2011) Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics 27:2083–2088. https://doi.org/10.1093/bioinformatics/btr331
30. Zhu H, Pisabarro MT (2011) MSPocket: an orientation-independent algorithm for the detection of ligand binding pockets. Bioinformatics 27:351–358. https://doi.org/10.1093/bioinformatics/btq672
31. Huang B (2009) MetaPocket: a meta approach to improve protein ligand binding site prediction. OMICS 13:325–330. https://doi.org/10.1089/omi.2009.0045
32. Sahu A, Patra PK, Yadav MK, Varma M (2017) Identification and characterization of ErbB4 kinase inhibitors for effective breast cancer therapy. J Recept Signal Transduct 37:470–480. https://doi.org/10.1080/10799893.2016.1342129
33. Wagner JR, Sørensen J, Hensley N et al (2017) POVME 3.0: software for mapping binding pocket flexibility. J Chem Theory Comput 13:4584–1492. https://doi.org/10.1021/acs.jctc.7b00500
34. Stank A, Kokh DB, Horn M et al (2017) TRAPP webserver: predicting protein binding site flexibility and detecting transient binding pockets. Nucleic Acids Res 45(W1):W325–W330. https://doi.org/10.1093/nar/gkx277
35. Kontoyianni M (2017) Docking and virtual screening in drug discovery. In: Proteomics for drug discovery. Humana Press, New York. https://doi.org/10.1007/978-1-4939-7201-2_18
36. Lipinski CA (2004) Lead-and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol 1:337–341. https://doi.org/10.1016/j.ddtec.2004.11.007
37. Wang J, Skolnik S (2009) Recent advances in physicochemical and ADMET profiling in drug discovery. Chem Biodivers 6:1887–1899. https://doi.org/10.1002/cbdv.20090011
38. Shoichet BK, McGovern SL, Wei B, Irwin JJ (2002) Lead discovery using molecular docking. Curr Opin Chem Biol 6:439–446. https://doi.org/10.1016/S1367-5931(02)00339-3
39. Jug G, Anderluh M, Tomašič T (2015) Comparative evaluation of several docking tools for docking small molecule ligands to DC-SIGN. J Mol Model 21:164. https://doi.org/10.1007/s00894-015-2713-2

40. Klebe G (1995) Toward a more efficient handling of conformational flexibility in computer-assisted modelling of drug molecules. Perspect Drug Discov Des 3:85–105. https://doi.org/10.1007/BF02174468
41. Beusen DD, Shands EB (1996) Systematic search strategies in conformational analysis. Drug Discov Today 1:429–437. https://doi.org/10.1016/S1359-6446(96)80009-2
42. Perola E, Charifson PS (2004) Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. J Med Chem 47:2499–2510. https://doi.org/10.1021/jm030563w
43. Vogt AD, Pozzi N, Chen Z, Di Cera E (2014) Essential role of conformational selection in ligand binding. Biophys Chem 186:13–21. https://doi.org/10.1016/j.bpc.2013.09.003
44. Bharatham N, Bharatham K, Shelat AA, Bashford D (2014) Ligand binding mode prediction by docking: mdm2/mdmx inhibitors as a case study. J Chem Inform Model 54:648–659. https://doi.org/10.1021/ci4004656
45. Kumar Mahto M, Yellapu NK, Kilaru RB et al (2014) Molecular designing and in silico evaluation of darunavir derivatives as anticancer agents. Bioinformation 10:221–226. https://doi.org/10.6026/97320630010221
46. Watts KS, Dalal P, Murphy RB et al (2010) ConfGen: a conformational search method for efficient generation of bioactive conformers. J Chem Inform Model 50:534–546. https://doi.org/10.1021/ci100015j
47. Sindhikara D, Spronk SA, Day T et al (2017) Improving accuracy, diversity, and speed with prime macrocycle conformational sampling. J Chem Inform Model 57:1881–1894. https://doi.org/10.1021/acs.jcim.7b00052
48. Huang SY, Grinter SZ, Zou X (2010) Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. Phys Chem Chem Phys 12:12899–12908. https://doi.org/10.1039/C0CP00151A
49. Barradas D, Fernandez-Recio J (2015) A comprehensive analysis of scoring functions for protein-protein docking: PJ-027. Protein Sci 24:250–251
50. Hsieh JH, Yin S, Wang XS et al (2012) Cheminformatics meets molecular mechanics: a combined application of knowledge-based pose scoring and physical force field-based hit scoring functions improves the accuracy of structure-based virtual screening. J Chem Inform Model 52:16–28. https://doi.org/10.1021/ci2002507
51. Dinur U, Hagler AT (1991) New approaches to empirical force fields. Rev Comput Chem 2:99–164. https://www.google.co.in/books/edition/Reviews_in_Computational_Chemistry/qzHcQwQ_T4EC?hl=en&gbpv=0
52. Brooks BR, Bruccoleri RE, Olafson BD et al (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4:187–216. https://doi.org/10.1002/jcc.540040211
53. Weiner SJ, Kollman PA, Case DA et al (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 106:765–784. https://doi.org/10.1021/ja00315a051
54. Hagler AT, Lifson S, Dauber P (1979) Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 2. A benchmark for the objective comparison of alternative force fields. J Am Chem Soc 101:5122–5130. https://doi.org/10.1021/ja00512a002
55. Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed 41:2644–2676. https://doi.org/10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O
56. Li GB, Yang LL, Wang WJ et al (2013) ID-Score: a new empirical scoring function based on a comprehensive set of descriptors related to protein–ligand interactions. J Chem Inform Model 53:592–600. https://doi.org/10.1021/ci300493w
57. Lizunov AY, Gonchar AL, Zaitseva NI, Zosimov VV (2015) Accounting for intraligand interactions in flexible ligand docking with a PMF-based scoring function. J Chem Inform Model 55:2121–2137. https://doi.org/10.1021/acs.jcim.5b00158

58. Park H, Eom JW, Kim YH (2014) Consensus scoring approach to identify the inhibitors of AMP-activated protein kinase α2 with virtual screening. J Chem Inform Model 54:2139–2146. https://doi.org/10.1021/ci500214e
59. Alexander DL, Tropsha A, Winkler DA (2015) Beware of R 2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. J Chem Inform Model 55(7):1316–1322. https://doi.org/10.1021/acs.jcim.5b00206
60. Okamoto M, Masuda Y, Muroya A et al (2010) Evaluation of docking calculations on X-ray structures using CONSENSUS-DOCK. Chem Pharm Bull 58:1655–1657. https://doi.org/10.1248/cpb.58.1655
61. Terp GE, Johansen BN, Christensen IT, Jørgensen FS (2001) A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein− ligand binding affinities. J Med Chem 44:2333–2343. https://doi.org/10.1021/jm001090l
62. Bar-Haim S, Aharon A, Ben-Moshe T et al (2009) SeleX-CS: a new consensus scoring algorithm for hit discovery and lead optimization. J Chem Inform Model 49:623–633. https://doi.org/10.1021/ci800335j
63. Betzi S, Suhre K, Chétrit B et al (2006) GFscore: a general nonlinear consensus scoring function for high-throughput docking. J Chem Inform Model 46:1704–1712. https://doi.org/10.1021/ci0600758
64. Wang R, Lai L, Wang S (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aided Mol Des 16:11–26. https://doi.org/10.1023/a:1016357811882
65. von Itzstein M, Wu WY, Kok GB et al (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. Nature 363:418–423. https://doi.org/10.1038/363418a0
66. Barreca ML, Manfroni G, Leyssen P et al (2013) Structure-based discovery of pyrazolobenzothiazine derivatives as inhibitors of hepatitis C virus replication. J Med Chem 56:2270–2282. https://doi.org/10.1021/jm301643a
67. Perun TJ, Propst CL (eds) (1989) Computer-aided drug design: methods and applications. Marcel Dekker, New York. https://doi.org/10.1177/088391159100600107
68. Cohen NC, Blaney JM, Humblet C et al (1990) Molecular modeling software and methods for medicinal chemistry. J Med Chem 33:883–894. https://doi.org/10.1021/jm00165a001
69. Bultinck P, De Winter H, Langenaeker W, Tollenare JP (eds) (2003) Computational medicinal chemistry for drug discovery. CRC Press. https://www.routledge.com/Computational-Medicinal-Chemistry-for-Drug-Discovery/Bultinck-Winter-Langenaeker-Tollenare/p/book/9780824747749
70. Abraham DJ, Rotella DP Burger's medicinal chemistry, drug discovery, and development, vol 2, Hoboken, p 59. https://www.wiley.com/en-in/Burger%27s+Medicinal+Chemistry%2C+Drug+Discovery%2C+and+Development%2C+8+Volume+Set%2C+Volumes+1+8%2C+7th+Edition-p-9780470278154
71. Abraham DJ, Rotella DP Burger's medicinal chemistry, drug discovery, and development, vol 2, Hoboken, p 62. https://www.wiley.com/en-in/Burger%27s+Medicinal+Chemistry%2C+Drug+Discovery%2C+and+Development%2C+8+Volume+Set%2C+Volumes+1+8%2C+7th+Edition-p-9780470278154
72. Silverman L, Campbell R, Broach JR (1998) New assay technologies for high-throughput screening. Curr Opin Chem Biol 2:397–403. https://doi.org/10.1016/S1367-5931(98)80015-X
73. Rester U (2008) From virtuality to reality-virtual screening in lead discovery and lead optimization: a medicinal chemistry perspective. Curr Opin Drug Disc Dev 11:559–568. https://pubmed.ncbi.nlm.nih.gov/18600572/
74. Rollinger JM, Stuppner H, Langer T (2008) Virtual screening for the discovery of bioactive natural products. In: Natural compounds as drugs, vol I. Birkhäuser, Basel. https://doi.org/10.1007/978-3-7643-8117-2_6

75. Lengauer T, Rarey M (1996) Computational methods for biomolecular docking. Curr Opin Struct Biol 6:402–406. https://doi.org/10.1016/s0959-440x(96)80061-3
76. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3:935–949. https://doi.org/10.1038/nrd1549
77. Bultinck P, De Winter H, Langenaeker W, Tollenare JP (eds) (2003) Computational medicinal chemistry for drug discovery. CRC Press. https://www.google.co.in/books/edition/Computational_Medicinal_Chemistry_for_Dr/gc79DwAAQBAJ?hl=en&gbpv=0
78. Schneider G, Fechner U (2005) Computer-based de novo design of drug-like molecules. Nat Rev Drug Discov 4:649–663. https://doi.org/10.1038/nrd1799
79. Moon JB, Howe WJ (1990) 3D database searching and de novo construction methods in molecular design. Tetrahedron Comput Methodol 3:697–711. https://doi.org/10.1016/0898-5529(90)90168-8
80. Joseph-McCarthy D (1999) Computational approaches to structure-based ligand design. Pharmacol Ther 84:179–191. https://doi.org/10.1016/s0163-7258(99)00031-5
81. Jain SK, Agrawal A (2004) De novo drug design: an overview. Indian J Pharm Sci 66:721–728. https://www.ijpsonline.com/abstract/de-novo-drug-design%2D%2Dan-overview-741.html
82. Foye WO (2008) Foye's principles of medicinal chemistry. Lippincott Williams & Wilkins. https://www.google.co.in/books/edition/Foye_s_Principles_of_Medicinal_Chemistry/R0W1ErpsQpkC?hl=en
83. Krieger E, Nabuurs SB, Vriend G (2003) Homology modeling. Methods Biochem Anal 44:509–524. https://doi.org/10.1002/0471721204.ch25
84. Cavasotto CN, Phatak SS (2009) Homology modeling in drug discovery: current trends and applications. Drug Disc Today 14:676–683. https://doi.org/10.1016/j.drudis.2009.04.006
85. Rose PW, Prlić A, Altunkaya A et al (2016) The RCSB protein data bank: integrative view of protein, gene and 3D structural information. Nucleic Acids Res 45:D271–D281. https://doi.org/10.1093/nar/gkw1000
86. Rost B, Sander C (1996) Bridging the protein sequence-structure gap by structure predictions. Annu Rev Biophys Biomol Struct 25:113–136. https://doi.org/10.1146/annurev.bb.25.060196.000553
87. Dalton JA, Jackson RM (2007) An evaluation of automated homology modelling methods at low target–template sequence similarity. Bioinformatics 23:1901–1908. https://doi.org/10.1093/bioinformatics/btm262
88. Kuntal BK, Aparoy P, Reddanna P (2010) EasyModeller: a graphical interface to MODELLER. BMC Res Notes 3:81–88. https://doi.org/10.1186/1756-0500-3-226
89. Kmiecik S, Gront D, Kolinski M et al (2016) Coarse-grained protein models and their applications. Chem Rev 116:7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163
90. Werner T, Morris MB, Dastmalchi S, Church WB (2012) Structural modelling and dynamics of proteins for insights into drug interactions. Adv Drug Deliv Rev 64:323–343. https://doi.org/10.1016/j.addr.2011.11.011
91. Samudrala R, Moult J (1998) Determinants of side chain conformational preferences in protein structures. Protein Eng 11:991–997. https://doi.org/10.1093/protein/11.11.991
92. Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. Proteins 47:393–402. https://doi.org/10.1002/prot.10104
93. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410. https://doi.org/10.1093/nar/gkm290
94. Koseki Y, Aoki S (2014) Computational medicinal chemistry for rational drug design: identification of novel chemical structures with potential anti-tuberculosis activity. Curr Top Med Chem 14:176–188. https://doi.org/10.2174/1568026613666131113155042

95. Kaushik AC, Kumar A, Bharadwaj S et al (2018) Ligand-based approach for in-silico drug designing. In: Bioinformatics techniques for drug discovery. Springer, Cham, pp 11–19. https://doi.org/10.1007/978-3-319-75732-2
96. Johnson MA, Maggiora GM (1990) Concepts and applications of molecular similarity. Wiley
97. Mestres J, Martín-Couce L, Gregori-Puigjané E et al (2006) Ligand-based approach to in silico pharmacology: nuclear receptor profiling. J Chem Inform Model 46:2725–2736. https://doi.org/10.1021/ci600300k
98. Gramatica P (2007) Principles of QSAR models validation: internal and external. QSAR Combi Sci 26:694–701. https://doi.org/10.1002/qsar.200610151
99. Crum-Brown A, Fraser T (1868) On the connection between chemical constitution and physiological action. Part I. On the physiological action of salts of the ammonium bases derived from strychnia, bucia, thebaia, codeia, morphia and nicotia. Part II. On the physiological action of the ammonium bases derived from atropia and conia. Trans R Soc Edinburgh 25: 693–739
100. Richet C (1893) On the relationship between the toxicity and the physical properties of substances. Compt Rendus Seances Soc Biol 9:775–776
101. Rekker RF (1992) The history of drug research: from Overton to Hansch. QSAR 11:195–199. https://doi.org/10.1002/qsar.19920110214
102. Hammett LP (1940) Physical organic chemistry. McGraw-Hill Book Company, New York. https://doi.org/10.1002/jps.3030290733
103. Taft RW Jr, Lewis IC (1959) Evaluation of resonance effects on reactivity by application of the linear inductive energy relationship. V. Concerning a σR scale of resonance effects[1,2]. J Am Chem Soc 81:5343–5352. https://doi.org/10.1021/ja01529a025
104. Taft RW (1956) Separation of polar, steric and resonance effects in reactivity. In: Steric effects in organic chemistry. pp 556–675
105. Hansch C, Maloney PP, Fujita T et al (1962) Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. Nature 194:178–180. https://doi.org/10.1038/194178b0
106. Fujita T, Iwasa J, Hansch C (1964) A new substituent constant, π, derived from partition coefficients. J Am Chem Soc 86:5175–5180. https://doi.org/10.1021/ja01077a028
107. Hansch C, Fujita T (1964) p-σ-π Analysis. A method for the correlation of biological activity and chemical structure. J Am Chem Soc 86:1616–1626. https://doi.org/10.1021/ja01062a035
108. Fujita T, Ban T (1971) Structure-activity relation. 3. Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. J Med Chem 14:148–152. https://doi.org/10.1021/jm00284a016
109. Klopman G, Li JY, Wang S et al (1994) Computer automated log P calculations based on an extended group contribution approach. J Chem Inform Comput Sci 34:752–781. https://doi.org/10.1021/ci00020a009
110. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemom Intell Lab Syst 2:37–52. https://doi.org/10.1016/0169-7439(87)80084-9
111. Geladi P, Isaksson H, Lindqvist L et al (1989) Principal component analysis of multivariate images. Chemom Intell Lab Syst 5:209–220. https://doi.org/10.1016/0169-7439(89)80049-8
112. Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Eugenics 7: 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x
113. Pirhadi S, Shiri F, Ghasemi JB (2015) Multivariate statistical analysis methods in QSAR. RSC Adv 5:104635–104665. https://doi.org/10.1039/C5RA10729F
114. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 58:109–130. https://doi.org/10.1016/S0169-7439(01)00155-1
115. Geladi P, Kowalski BR (1986) An example of 2-block predictive partial least-squares regression with simulated data. Anal Chim Acta 185:19–32. https://doi.org/10.1016/0003-2670(86)80029-0

116. Rogers D, Hopfinger AJ (1994) Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. J Chem Inform Comput Sci 34:854–866. https://doi.org/10.1021/ci00020a020
117. Roy K, Kar S, Das RN (2015) Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment. Academic, Elsevier. https://www.google.co.in/books/edition/Understanding_the_Basics_of_QSAR_for_App/bkFOBQAAQBAJ?hl=en&gbpv=0
118. Chackalamannil S, Rotella D, Ward S (2017) Comprehensive medicinal chemistry III. Elsevier. https://www.google.co.in/books/edition/Comprehensive_Medicinal_Chemistry_III/iZHUDQAAQBAJ?hl=en&gbpv=0
119. Song CM, Lim SJ, Tong JC (2009) Recent advances in computer-aided drug design. Brief Bioinform 10:579–591. https://doi.org/10.1093/bib/bbp023
120. Rigden DJ, Fernández-Suárez XM, Galperin MY (2016) The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection. Nucleic Acids Res 44: D1–D6. https://doi.org/10.1093/nar/gkv1356
121. Brown RD, Hassan M, Waldman M (2000) Combinatorial library design for diversity, cost efficiency, and drug-like character libraries. J Mol Graph Model 4:537. https://doi.org/10.1016/S1093-3263(00)00072-3
122. Mizuguchi K (2004) Fold recognition for drug discovery. Drug Discov Today: Target 3:18–23. https://doi.org/10.1016/S1741-8372(04)02392-8
123. Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253:164–170. https://doi.org/10.1126/science.1853201
124. Lyne PD (2002) Structure-based virtual screening: an overview. Drug Discov Today 7:1047–1055. https://doi.org/10.1016/S1359-6446(02)02483-2
125. Johnson MA, Maggiora GM (1990) Concepts and applications of molecular similarity. Wiley. https://www.wiley.com/en-us/Concepts+and+Applications+of+Molecular+Similarity-p-9780471621751
126. Cushman DW, Cheung HS, Sabo EF et al (1977) Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids. Biochemistry 16:5484–5491. https://doi.org/10.1021/bi00644a014
127. Byers LD, Wolfenden R (1973) Binding of the by-product analog benzylsuccinic acid by carboxypeptidase A. Biochemist 12:2070–2078. https://doi.org/10.1021/bi00735a008
128. Ondetti MA, Rubin B, Cushman DW (1977) Design of specific inhibitors of angiotensin-converting enzyme: new class of orally active antihypertensive agents. Science 196:441–444. https://doi.org/10.1126/science.191908
129. Harris C, Smith GH (1981) Captopril (Capoten®, ER Squibb & Sons). Drug Intell Clin Pharm 15:932–939. https://doi.org/10.1177/106002808101501203
130. Maxwell RA, Eckhardt SB (1990) Captopril. Drug discovery. Humana Press, Totowa. https://doi.org/10.1007/978-1-4612-0469-5_2
131. Ahmadi R (2012) Computational study of chemical properties of Captopril drug and the connected form to Fullerene (C60) as a medicine nano carrier. J Phys Theor Chem 9:185–190. http://jptc.srbiau.ac.ir/article_2569.html
132. Surabhi S, Singh BK (2018) Computer aided drug design: an overview. J Drug Deliv Ther 8: 504–509. https://doi.org/10.22270/jddt.v8i5.1894
133. Maren TH (1987) Carbonic anhydrase: general perspective and advances in glaucoma research. Drug Dev Res 10:255–276. https://doi.org/10.1002/ddr.430100407
134. Baldwin JJ, Ponticello GS, Anderson PS et al (1989) Thienothiopyran-2-sulfonamides: novel topically active carbonic anhydrase inhibitors for the treatment of glaucoma. J Med Chem 32: 2510–2513. https://doi.org/10.1021/jm00132a003
135. Eriksson AE, Jones TA, Liljas A (1988) Refined structure of human carbonic anhydrase II at 2.0 Å resolution. Proteins: Struct Funct Bioinform 4:274–282. https://doi.org/10.1002/prot.340040406

136. Coleman JE (1975) Chemical reactions of sulfonamides with carbonic anhydrase. Annu Rev Pharmacol 15:221–242. https://doi.org/10.1146/annurev.pa.15.040175.001253
137. Christianson DW (1991) Structural biology of zinc. Adv Protein Chem 42:281–355. https://doi.org/10.1016/S0065-3233(08)60538-0. Academic, Elsevier
138. Maren TH (1995) The development of topical carbonic anhydrase inhibitors. J Glaucoma 4: 49–62. https://journals.lww.com/glaucomajournal/Citation/1995/02000/The_Development_of_Topical_Carbonic_Anhydrase.12.aspx
139. Talele TT, Khedkar SA, Rigby AC (2010) Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. Curr Top Med Chem 10:127–141. https://doi.org/10.2174/156802610790232251
140. Balfour JA, Wilde MI (1997) Dorzolamide. Drugs Aging 10:384–403. https://doi.org/10.2165/00002512-199710050-00006
141. Wilkerson M, Cyrlin M, Lippa EA et al (1993) Four-week safety and efficacy study of dorzolamide, a novel, active topical carbonic anhydrase inhibitor. Arch Ophthalmol 111: 1343–1350. https://doi.org/10.1001/archopht.1993.01090100051026
142. Leelananda SP, Lindert S (2016) Computational methods in drug discovery. Beilstein J Org Chem 12:2694–2718. https://doi.org/10.3762/bjoc.12.267. eCollection 2016
143. Kohl NE, Emini EA, Schleif WA et al (1988) Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci U S A 85:4686–4690. https://doi.org/10.1073/pnas.85.13.4686
144. McQuade TJ, Tomasselli AG, Liu L et al (1990) A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. Science 247:454–456. https://doi.org/10.1126/science.2405486
145. Vacca JP, Condra JH (1997) Clinically effective HIV-1 protease inhibitors. Drug Discov Today 2:261–272. https://doi.org/10.1016/S1359-6446(97)01053-2
146. Roberts NA, Martin JA, Kinchington D et al (1990) Rational design of peptide-based HIV proteinase inhibitors. Science 248:358–361. https://doi.org/10.1126/science.2183354
147. Graves BJ, Hatada MH, Miller JK et al (1991) The three-dimensional x-ray crystal structure of HIV-1 protease complexed with a hydroxyethylene inhibitor. In: Structure and function of the aspartic proteinases. Springer, Boston. https://doi.org/10.1007/978-1-4684-6012-4_58
148. Krohn A, Redshaw S, Ritchie JC et al (1991) Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere. J Med Chem 34:3340–3342. https://doi.org/10.1021/jm00115a028
149. Noble S, Faulds D (1996) Saquinavir. Drugs 52:93–112. https://doi.org/10.2165/00003495-199652010-00007
150. Azizi M, Guyene TT, Chatellier G et al (1994) Blood pressure effects of acute intravenous renin or oral angiotensin converting enzyme inhibition in essential hypertension. J Hypertens 12:419–428. https://europepmc.org/article/med/8064166
151. Wolfenden R (1972) Analog approaches to the structure of the transition state in enzyme reactions. Acc Chem Res 5:10–18. https://doi.org/10.1021/ar50049a002
152. Rahuel J, Priestle JP, Grütter MG (1991) The crystal structures of recombinant glycosylated human renin alone and in complex with a transition state analog inhibitor. J Struct Biol 107: 227–236. https://doi.org/10.1016/1047-8477(91)90048-2
153. Göschke R, Stutz S, Rasetti V et al (2007) Novel 2, 7-dialkyl-substituted 5 (S)-amino-4 (S)-hydroxy-8-phenyl-octanecarboxamide transition state peptidomimetics are potent and orally active inhibitors of human renin. J Med Chem 50:4818–4831. https://doi.org/10.1021/jm070314y
154. Rahuel J, Rasetti V, Maibaum J et al (2000) Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. Chem Biol 7:493–504. https://doi.org/10.1016/S1074-5521(00)00134-4
155. Frampton JE, Curran MP (2007) Aliskiren. Drugs 67:1767–1792. https://doi.org/10.2165/00003495-200767120-00008

156. Brown MJ (2008) Aliskiren. Circulation 118:773–784. https://doi.org/10.1161/CIRCULATIONAHA.108.787630
157. Black RA, Rauch CT, Kozlosky CJ et al (1997) A metalloproteinase disintegrin that releases tumour-necrosis factor-α from cells. Nature 385:729–733. https://doi.org/10.1038/385729a0
158. Vassalli P (1992) The pathophysiology of tumor necrosis factors. Annu Rev Immunol 10:411–452. https://doi.org/10.1146/annurev.iy.10.040192.002211
159. Maskos K, Fernandez-Catalan C, Huber R et al (1998) Crystal structure of the catalytic domain of human tumor necrosis factor-α-converting enzyme. Proc Natl Acad Sci U S A 95:3408–3412. https://doi.org/10.1073/pnas.95.7.3408
160. Levin JI, Chen JM, Cheung K et al (2003) Acetylenic TACE inhibitors. Part 1. SAR of the acyclic sulfonamide hydroxamates. Bioorg Med Chem Lett 13:2799–2803. https://doi.org/10.1016/S0960-894X(03)00514-6
161. Levin JI, Chen JM, Laakso LM et al (2006) Acetylenic TACE inhibitors. Part 3: thiomorpholine sulfonamide hydroxamates. Bioorg Med Chem Lett 6:1605–1609. https://doi.org/10.1016/j.bmcl.2005.12.020
162. Fleischman R, Durez P, Genovese M et al (2006) Lack of efficacy with 3 oral dose levels of TMI-005 (Apratastat), in subjects with active rheumatoid arthritis on a background of methotrexate-A phase 2 double-blind, placebo-controlled, parallel, randomized study. Ann Rheum Dis 65:339. http://hdl.handle.net/2078.1/59841
163. Orr DC, Long AC, Kay J et al (1989) Hydrolysis of a series of synthetic peptide substrates by the human rhinovirus 14 3C proteinase, cloned and expressed in Escherichia coli. J Gen Virol 70:2931–2942. https://doi.org/10.1099/0022-1317-70-11-2931
164. Shepherd TA, Cox GA, McKinney E et al (1996) Small peptidic aldehyde inhibitors of human rhinovirus 3C protease. Bioorg Med Chem Lett 6:2893–2896. https://doi.org/10.1016/S0960-894X(96)00537-9
165. Kaldor SW, Hammond M, Dressman BA et al (1995) Glutamine-derived aldehydes for the inhibition of human rhinovirus 3C protease. Bioorg Med Chem Lett 5:2021–2026. https://doi.org/10.1016/0960-894X(95)00345-T
166. Webber SE, Okano K, Little TL et al (1998) Tripeptide aldehyde inhibitors of human rhinovirus 3C protease: design, synthesis, biological evaluation, and cocrystal structure solution of P1 glutamine isosteric replacements. J Med Chem 41:2786–2805. https://doi.org/10.1021/jm980071x
167. Matthews DA, Dragovich PS, Webber SE et al (1999) Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. Proc Natl Acad Sci U S A 96:11000–11007. https://doi.org/10.1073/pnas.96.20.11000
168. Dragovich PS, Webber SE, Babine RE et al (1998) Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 2. Peptide structure− activity studies. J Med Chem 41:2819–2834. https://doi.org/10.1021/jm9800696
169. Dragovich PS, Webber SE, Babine RE et al (1998) Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 1. Michael acceptor structure-activity studies. J Med Chem 41:2806–2818. https://doi.org/10.1021/jm9800696

170. Dragovich PS, Prins TJ, Zhou R et al (2002) Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 6. Structure−activity studies of orally bioavailable, 2-pyridone-containing peptidomimetics. J Med Chem 45: 1607–1623. https://doi.org/10.1021/jm010469k
171. Dragovich PS, Prins TJ, Zhou R et al (1999) Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 4. Incorporation of P1 lactam moieties as L-glutamine replacements. J Med Chem 42:1213–1224. https://doi.org/10.1021/jm9805384
172. Patick AK (2006) Rhinovirus chemotherapy. Antivir Res 71:391–396. https://doi.org/10.1016/j.antiviral.2006.03.011
173. Berdigaliyev N, Aljofan M (2020) An overview of drug discovery and development. Future Med Chem 12:939–947. https://doi.org/10.4155/fmc-2019-0307

Quantitative Structure-Property Relationship (QSPR) Modeling Applications in Formulation Development

17

Pankaj Wadhwa and Amit Mittal

Contents

Abstract

Quantitative structure-property relationships (QSPRs) are generally represented as mathematical model which are developed by correlating the physicochemical properties/biological activity of compounds to their chemical structures. It differs from quantitative structure-activity relationships (QSARs) in the terms of their dependent variable, biological activity (QSAR) vs. biophysicochemical property (QSPR). It plays an important role in the area of drug discovery. It also provides a model which states about features of molecule through various molecular descriptors such as conformational, electronic, quantum mechanical, spatial, topological, thermodynamic, and many more. For generating successful development of QSPR model, various different steps like selection of data set and extraction of structural/empirical descriptors, variable selection, model construction, and validation evaluation should be followed. Developed QSPR models are beneficial in predicting the biological activities of untested/new molecules,

P. Wadhwa (✉) · A. Mittal
Department of Pharmaceutical Chemistry, School of Pharmaceutical Sciences, Lovely Professional University, Phagwara, Punjab, India
e-mail: Pankaj.23400@lpu.co.in; amit.13145@lpu.co.in

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_17

alternative of animal experiments, thus reducing animal use, refinement of synthetic targets, assumption of ADME and toxicity studies, etc. In this chapter, we focus on the various 2D/3D descriptors, methods, and their applications in formulation development.

Keywords

Aqueous solubility · Formulation development · Molecular descriptors · Partition coefficient · Physical properties · QSPR

Chapter Objectives

- Illustrate the available in silico modeling tools related to QSPR modeling.
- Introduce various descriptors for designing molecules with desired physicochemical properties to achieve better pharmaceutical formulations.
- Understand the approach for development of mathematical models to correlate the biological activity and descriptive parameters related to the molecular structure.
- To understand the established QSPR models with suitable representative examples.

17.1 Introduction

In the twentieth century, a variety of quantitative structure-activity relationship (QSAR) approaches have been developed progressively and used as a precious predictive tool for industrial and environmental chemistry and also for designing a vast range of pharmaceuticals and agrochemicals [1]. The term quantitative structure-property relationship (QSPR) acronymous is used when a property is modeled. QSARs are based on the theory that the geometric, steric, and electronic properties of a molecule are responsible for their biological activity [2]. In other words, it can be stated as a QSAR/QSPR equation which is generated on the basis of different numerical descriptor(s), and which is further used to assess the biological activity of a new molecules or chemicals [3]. The mathematical representation of Eq. 17.1 can be given as:

$$y = Xb + c \tag{17.1}$$

Here in Eq. 17.1, X describes a set of predictor variables (X) with a predicted variable (y) by means of a regression vector (b) [4].

The initial step in building the QSPR models is calculating different types of molecular descriptors or a set of known variables that stand for variation in the constitutional, structural, topological, or geometrical properties of the molecules by a number [5]. These descriptors are not based on experimental properties and also do not have any relationship with other descriptors. The basic requirements for optimal descriptors include that those should be simple and can be applied to local structure and must generalize to "higher" descriptors [6]. Aside from that, they should have

the ability of structural interpretation, be able to differentiate among isomers, and have a good correlation with at least one property [7].

Nowadays, a vast range (more than 2000) of descriptors are being used in these studies which can be classified into different categories such as 0D, 1D, 2D, 3D, and 4D [8]. The simplest form of molecule demonstration is the chemical formula that is not able to give any information about their atom connections, and this is generally known as 0D molecular descriptors [9]. The various structural fragments of molecules are considered as 1D molecular descriptors which plays a main role in similarity/diversity analysis and virtual screening [10]. Others such as fragment counts, H-bond acceptor/donor, and polar surface area also belong to this section. Next, topological indices are usually explained as 2D descriptors that is mainly topological representation of molecular structures (i.e., molecular graph, kappa, and chi index) [9]. However, these are not able to provide any detailed information about spatial distribution of the atoms [11]. Next, 3D descriptor, usually known as geometrical descriptor, includes weighted holistic invariant molecular (WHIM) descriptor analysis/autocorrelation or Morse along with other surface properties [12]. The 3D-QSAR techniques are further categorized into two sections which include alignment-based methods like comparative molecular field analysis (CoMFA), comparative molecular similarity indices analysis (CoMSIA), self-organizing molecular field analysis (SOMFA), receptor surface analysis (RSA), and molecular shape analysis (MSA) and alignment-independent methods like comparative molecular moment analysis (CoMMA), Vol Surf, Compass [13], comparative spectral analysis (CoSA), and grid-independent descriptors (GRIND) [14]. The most recent one is different types of 4D-3D coordinates and conformations [15]. This confirms that as the numbers of chemicals are growing daily, the interest on identifying the role of their molecular properties or research on their molecular descriptor is also rapid enhancing [16].

17.2 Development of QSPR Models

A large dataset of similar class along with their molecular properties is required to cover a large diversity space. Initially, a dataset of accumulated chemical structures along with interested experimental property data for each compound is required. The structures representation can be done as 2-D hydrogen-suppressed graph [17]. Suitable software to handle the molecules (such as SMILES, SDF in 2D, or optimized 3D structure) along with their calculated descriptors is highly preferred [18]. A crucial step for QSPR study is the generation of molecular descriptors [19], which usually represents the complete molecular structure or any structural fragment, as shown in Fig. 17.1. A descriptor package for calculating various types of descriptors is required.

Understanding of feature selection or a statistical process for removing invariant or correlated descriptors is necessary. The use of more numbers of calculated or inappropriate descriptors leads to overfitting and poor performance of developed model [20]. The relevant descriptors selection can be achieved by using different

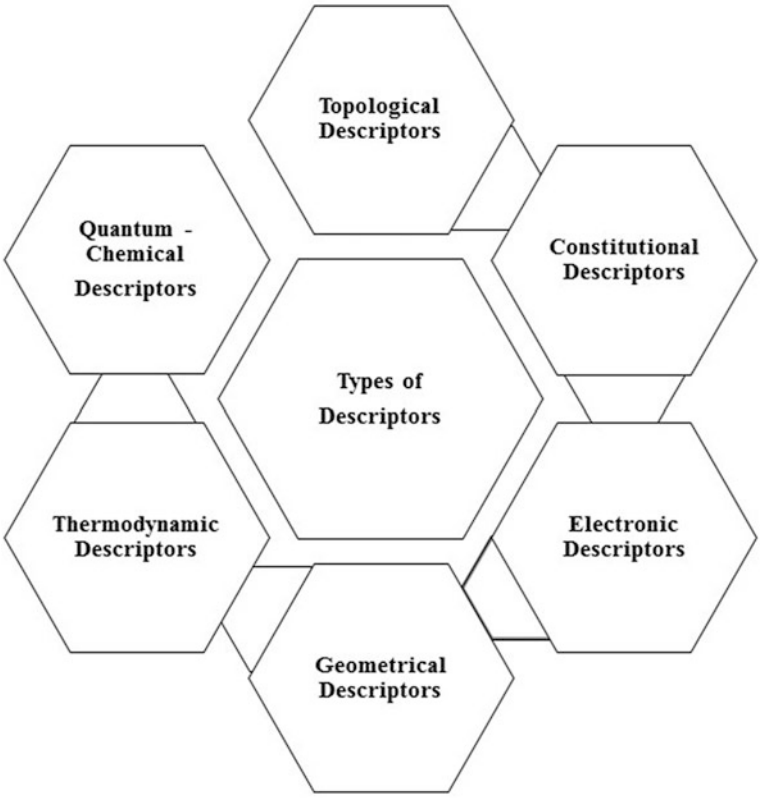


Fig. 17.1 Representation of various types of QSPR descriptors

methods such as genetic algorithm, principal component analysis, partial least square regression, etc. The mostly used method unsupervised forward selection (UFS) works by removing irrelevant variables from selected data matrix [21].

The molecules will be classified into three sets such as a training (70%), validation (30%), and an additional external training or validation set which is not used in either method [22]. In few cases, the validation set is also known as testing set or vice versa. If molecules are not enough, then cross-validation step has to be carried out. After this, regression or classification methods should be applied [23].

On the basis of these discussed steps, a block diagram representing the main component of QSPR studies is shown in Fig. 17.2.

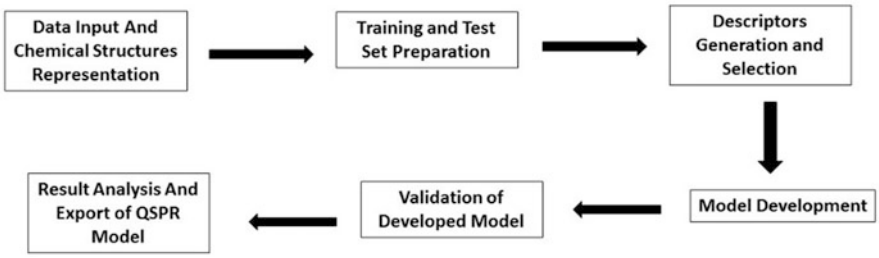


Fig. 17.2 Representation of various QSPR modeling components

17.3 Various Tools for Performing QSPR/QSAR Studies

Earlier, the reliability of quantitative structure-property relationship (QSPR) models was often difficult due to less availability of controlling tools. But nowadays, the bioinformatics field has emerged as a powerful discipline because of available databases and other software tools which are playing crucial role in scientific research. The easy accessibility of DNA sequences as well as protein structures (for example, Gen Bank and Protein Data Bank databases) are also one of the major reasons behind the development of numerous methods, tools, and resources for large-scale data analysis [24]. Here we have summarized a list of available various types of databases for chemical compounds as well as their calculated descriptors such as ZINC, NIC, and MOLE db in Table 17.1. The web links for drawing chemical structures in 2D and 3D forms are also provided here. These listed tools will help in 2D drawing and further creates optimized 3D structures using MMFF atom type assignment and force-field optimization [25]. Next, for quantitative representations of physical, chemical, or topological characteristics of molecules, various available tools for calculating different types of descriptors along different software list for virtual libraries designing are also listed in this table. QSAR models are one of best alternative to animal testing because of reduced time for experimentation and minimized costs [26]. The available tools for developing QSPR models using statistical parameters are also incorporated.

17.4 Applications of QSPR Modeling in Formulation Development

Chemicals represent an obligatory part of humankind due to their vast ranges of applications in daily routine. The applications of this present technique are enormous, and by employing QSPR models, it is easy to predict activity/property/toxicity of various pharmaceutical agents [27]. The obtained nature of chemicals gives abundant results for a considerable range of chemicals such as analytical reagents, agro/pharmaceutical chemicals, etc. [28]. In this section, we discuss

Table 17.1 List of available online tools for conducting QSPR studies

Name	Description	Website links
ZINC Database	Commercially available chemical compounds for virtual screening	https://zinc.docking.org/
NCI Database	Provides thousands of chemical structures	https://ccr.cancer.gov/chemical-biology-laboratory
The MOLE db	Freely available online molecular descriptor database	https://michem.unimib.it/mole-db-molecular-descriptors-data-base/
MOE	Online available tool for calculating around 300 topological, physical properties, and structural descriptor	https://www.chemcomp.com/
ACD/ ChemSketch	A tool for drawing chemical structures for different types of organics, organometallics, polymer derivatives	https://www.acdlabs.com/resources/freeware/chemsketch/index.php
BioPPSy	An open-source Platform for QSAR/QSPR analysis	https://sourceforge.net/projects/bioppsy/
Biovia Draw	Mainly used for chemical drawing application that provides support for IUPAC names, canonical SMILES strings, InChI strings	https://www.3ds.com/products-services/biovia/disciplines/
CORINA	Useful tool for converting 2D structures into 3D structures	https://www.mn-am.com/online_demos/corina_demo
Chem Des	An integrated web-based platform for molecular descriptor and fingerprint computation	http://www.scbdd.com/chemdes/
E-Dragon	It is helpful in calculating around 1600 molecular descriptors that are usually divided into different 20 logical blocks	http://www.vcclab.org/lab/edragon/
Chemistry Development Kit (CDK)	It is one kind of collected libraries for providing structural and cheminformatics details	https://cdk.github.io/
JChem for Excel	For generating virtual libraries and 2D descriptors	https://chemaxon.com/products/jchem-for-office
VCCLAB PLS	Partial least squares regression	http://www.vcclab.org/lab/pls/conf_f.html

about the important role of QSPR study in sector of pharmaceutical formulations. Electrolytes are the major component of our pharmaceutical formulations, and they exist in various forms such as weak acids/bases, etc. At a wide range of pH, they show different level of apparent/buffer solubility, which is usually a function of its intrinsic solubility, pH value, and its pKa [29]. Apart from this, biomimetic properties of drug compound are essential for assessing their plasma protein binding [30].

In the drug development process, the toxicological and pharmacokinetic properties of compounds are major factors. Physicochemical properties are examined, such as aqueous solubility, acid dissociation constant pKa and lipophilicity,

and many more, helps us to overcome bioavailability-related problems [31]. In many reports, it was suggested that many atom type/group counts (ATC), BCUT 3D indices [32], and electrotopological index categories of molecular descriptors are widely used in solubility prediction [33]. These BCUT metrics are extensions of parameters usually called Burden parameters, originally developed by Burden. These parameters are mainly based upon combination of the atomic number for each atom and a description of the nominal bond-type for adjacent and nonadjacent atoms. The BCUT metrics expand the number and type of atomic features that can be considered and also provide a greater variety of proximity measures and weighting schemes [34].

17.4.1 Aqueous Solubility

A solubility measurement is a critical for investigating either the thermodynamic or the kinetic solubility of the pharmaceutical compounds via practical ways [35]. It regulates two things mainly, viz. dissolution rate and the maximum amount traveling to GI (gastrointestinal) fluid [36]. By using the QSPR methods, we can easily predict the aqueous solubility. The aqueous solubility of different pharmaceuticals provides an idea about their solubility in the intestinal fluids and involvement in bioavailability [37]. Optimum solubility is required for absorption of drugs at their site and also useful for in vivo testing [38]. The major thing that should be considered is what type of descriptors controls aqueous solubility of compounds. To answer this question, we should have idea about various molecular interactions (for example, intra-/intermolecular hydrogen bonding) occurring with solvents to completely dissolve the compound [39]. Different descriptors such as polar surface area (PSA), molecular weights and polarizability [40], partition coefficient (clogP), energy required for crystal packing, solvation [41], followed by energy required for cavity formation within solvent, topological indices are accountable to maximum extent [42].

Louis Hamette first correlated electronic functions of organic acid and bases with their equilibrium constant and reactivity [43]. Following that, a number of mathematics models that interrelate structure and have been established. QSPR interrelates structural or property descriptors of the compounds with their chemical or biological activities. The general expectation of QSPR model is that molecular structure is culpable for the observed behavior of the compound. The physiochemical descriptors like hydrophobicity, topology, electronic properties, and steric effects are determined empirically or more recently by computational method [44]. Multiple structural and other descriptors may represent ligand. Therefore, selection of key descriptor is a crucial step in any QSPR study. Identification of patterns (predictive fingerprints and combination of features) is another important step that correlates with activity [45].

Various types of clustering approaches, such as hierarchical divisive clustering, hierarchical agglomerative clustering, and others, have been employed in the pharmaceutical sector to solve various drug development-related challenges [44]. Dror along with his colleagues have reviewed the application of clustering techniques and genetic algorithms toward predicting molecular interactions [46]. Khan et al. have

discussed about their role for feature selection in successful development of QSAR model [47]. Similarly, another method known as consensus k-nearest-neighbor (kNN) QSAR has been developed toward predicting estrogenic activity. The main concept behind this approach is that the activity can be anticipated by averaging activities over k-nearest neighbors. These multiple models are used for making use of different sets of descriptors followed by consensus prediction [48].

In drug design, neural networks have been widely used to solve many problems. Winkler et al. have studied about the application part of neural networks in a variety of QSAR problems [49]. Their main agenda was to explore the role of neural networks in prediction of the physiochemical, toxicological, and pharmacokinetic parameters. Polanski and their group members have used self-organized maps (SOM) for studying molecular diversity and engaged in drug development [50]. In particular, SOM-based method of comparative molecular surface analysis (CoMSA) has been presented in detail. In another different study, the least squares support vector machine (LSSVM) technique was employed to screen calcium channel antagonists using QSPR approach [51]. Furthermore, evolutionary QSAR techniques that employed genetic algorithms are being developed for docking and related studies. Nicolotti et al. have reported the use of multiple objective genetic QSAR to study neuronal nicotinic acetylcholine ion channel receptors (nAChRs) [52]. Other research used a genetic algorithm to predict receptor ligand binding affinities [53] and to classify compounds based on their structure-activity connection (SAR) [54]. Particle swarm optimization (PSO) is another most important developing technique which is mainly used in QSPR study for optimum biomarker selection [55]. In the present study, few strategies which were used for development of QSPR for analyzing solubility along with their detailed algorithms and descriptors are discussed. Formulations having optimum solubility characteristics exhibit improved metabolism and excretion processes, and this further leads to minimize the chances of drug accumulation and other associated effects. Fullerenes are one of the significant elements of many nanotechnologies. On the basis of chemical structure of various organic solvents, the solubility prediction of fullerene C_{60} was carried out. The developed model has shown reproducible results with correlation coefficient value around 0.81 for training set after Monte-Carlo optimization [56]. Later, researchers calculated molecular descriptors associated to charge distribution, quantum mechanical energy field, topological, and geometrical properties for C_{60} and C_{70} fullerene derivatives. The finest model was developed by multiple linear regression technique with correlation coefficient (R^2) values of 0.801 and 0.792 for training and test set, respectively [57]. Their results revealed various types of independent factors like electron–electron and nuclear–nuclear repulsion energy, electro-nuclear attraction energy, rotational–vibrational energies, polar interactions and, reactivity effects solubility of the fullerenes.

A linear QSPR model for predicating aqueous solubility (log S_w) of 150 organic compounds by the means of stepwise and multiple linear regression (MLR) was developed [58]. Their results showed that solubility was mainly dependent on three properties of molecules, viz. octanol/water partition coefficient (log P), molecular

volume (MV), and hydrogen bond-forming ability (HB) with square correlation coefficient (R^2) value of 0.9954.

Palmer and his group members reported that random forest regression (RF) is the best way to predict solubility of almost 988 organic compounds as compared to available other methods such as artificial neural networks (ANN), partial least squares (PLS) regression, and support vector machines (SVM) [59]. Their developed model via RF approach has shown correlation coefficient (R^2) and root-mean-square-error (RMSE) values of 0.89 and 0.69, respectively.

The dissolving power of solvents is an important aspect in pharmaceutical industry. A QSPR study of 1228 solvents was performed by the ways of genetic algorithm-based multivariate linear regression (GA-MLR) and generalized regression neural network (GRNN) methods [60]. The first one was used to get the optimum descriptors as an input for the second one. The GRNN method was found to be more accurate with correlation coefficient (R^2) value of 0.98 as compared to GA-MLR which has shown 0.821 R^2 value. Another similar study was performed via using scores of extended connectivity fingerprint as descriptors and partial least squares (PLS) method to build a robust QSPR model [61]. Their results concluded that calculated descriptors were of high performance with correlation coefficient of 0.83 for around 1300 different chemical compounds. Duchowicz and Castro have performed QSPR study to predict aqueous solubility of 145 diverse drug-like organic compounds [62]. Their developed QSPR model by linear regression method comprised of three DRAGON molecular descriptors such as X1sol, RDF060u, and MLOGP. The correlation coefficient was found around 0.871 indicating the variation of the experimental solubilities in a satisfactory level and also explaining characterization of both training and test in a proper way. The same group further reported that by calculating novel descriptors related to bioavailability and other kind of descriptors such as constitutional, electronic, or topological, we can correlate the molecular structures with their solubility [63].

Shayanfar et al. reported that QSPR model developed for predicting solubility of 220 drug candidates with average absolute error (AAE) and mean percentage deviation (MPD) methods comprises two descriptors: excess molar refraction and partition coefficient [64]. Their results revealed that the proposed QSPR model was more accurate than general solubility equation (GSE) and the linear solvation energy relationship (LSER) models. Two QSPR models for predicting aqueous solubility of polychlorinated dibenzo-p-dioxins and polychlorinated dibenzofurans (PCDD/Fs) analogs with the help of linear artificial neural network (L-ANN) in combination with stepwise regression method were developed [65]. Validation of both developed models was carried out via leave-one-out cross-validation and split-sample technique, respectively. Their results revealed that both models have shown difference in predictive ability due to difference in input descriptors.

In the same year, a robust model was developed by calculating topological descriptors and using random forest and multiple linear regression methods for more than two thousand five hundred molecules [66]. The developed models were further validated and have shown its correlation coefficient 0.99 (for RF) and 0.85 (for MLR method), respectively. In pharmaceutical formulations, phase segregation

is a vital step which depends upon solubility. A model for this purpose was developed by using rank-based ant system feature selection (RBAS–FS) for more than diverse eighteen hundred organic and inorganic molecules. The obtained optimized model has shown accuracy with correlation coefficient value of 0.75 [67]. This ant-based rank system is mainly one kind of probabilistic technique that is used for finding optimal paths that further leads to construct a regression tree model. The developed ant-based algorithm represents each ant as a regression tree, while pheromone trails are emulated by a separate binary reference tree that represents the topological union of all the "ant" trees encountered in the course of the simulation [68].

Many perfluorinated chemicals (PFCs) have issues with their stability and biodegradation. To counteract this problem, a QSPR model was developed to predict their solubility, vapor pressure, and critical micelle concentration (CMC) [69]. Their results concluded that the developed model was able to assume water solubility and vapor pressure of PFCs in a better way and that information was much useful in understanding their environmental behavior.

It was stated that for the distribution of organic contaminants in the environment, two factors like solubility and partition coefficient are important. On this basis, a QSPR model for 134 halogenated methyl-phenyl ethers individually for each factor using Gaussian 98 program was generated [70]. The developed model comprises of three variables such as energy in the form of lowest unoccupied molecular orbital (ELUMO), atomic partial charge in (q^+), and quadrupole moment. The validation step was performed by leave-one-out (LOO) methods, and their correlation coefficient values were 0.992 and 0.970.

In another study, the relationship between structures of surfactants and their effect on water–oil interfacial tension was analyzed by developing a QSPR model [71]. The selection of almost 24 compounds was done using principal component analysis approach, and their critical micelle concentrations were used to calculate their interfacial tensions. The geometrical optimization was carried out and different quantum chemical and structural descriptors were calculated. A best model having three descriptors was developed using genetic algorithm (GA-MLR) method and that was found to predict the interfacial tension. Earlier it has been discussed about three types of interactions such as lattice energy required for dissociation, solvent−solvent interactions, and the solvation energy that mainly affects the aqueous solubility of compounds. Using these fundamentals, Salahinejad and their group members have developed a model for more than four thousand five hundred molecules [72]. Two approaches such as multiple linear regressions with expectation maximization (MLREM) and a Bayesian regularized artificial neural network with a Laplacian prior (BRANNLP) method were employed for predicting a model. Their results showed that the model of BRANNLP technique was most suitable with squared correlation coefficients of 0.90.

A QSPR model for assuming solubility of anticancer nano drugs was developed through genetic function approximation (GFA) algorithm [73]. The quantum and other types of descriptors were calculated by Ab initio, density functional theories methods and Gaussian 09 program. The calibration step was performed by partial

least squares (PLS) regression and found that descriptors such as quantum chemical, WHIM, GETAWAY, H–GETAWAY, R–GETAWAY, Geometrical, 3D–MoRSE play an important role in controlling the solubility.

Another group have used MODESLAB software to determine the molecular descriptors for various pharmaceutical organic compounds [74]. Almost 24 mathematical models were developed through the use of IBM SPSS Statistics (data reduction) and Build QSAR computational system to predict the aqueous solubility in respect to various pH ranges. The correlation coefficient values were observed in the range of 0.7–0.8.

17.4.2 Hydrophobicity and Partition Coefficients

Recently, a QSPR technique was developed based on a machine learning software to predict drug solubility in binary solvent systems [75]. Various types of structural features such as molar refractivity, McGowan volume, and topological surface area were used and their refine selection was carried out by the use of genetic algorithm. Their results have shown piecewise linear relationship which was further validated using a Levenberg–Marquardt training algorithm. It is well reported that hydrophobicity plays a crucial function in deciding drug's absorption, binding, diffusion, and partitioning into other parts of body. A QSAR model for establishing relationship between binding (KD) and hydrophobicity (ClogP) was developed for a set of drugs [76]. Their study results showed that both strength of binding and the loss of water because of binding increase with drug hydrophobicity. The hydrophobicity which is usually known as molecular lipophilicity, abbreviated as log *P*, is an important characteristic in drug discovery. This parameter commonly represents lipophilicity of organic compounds and has a strong impact on their pharmacokinetic properties. Earlier it was measured using known experimental methods, but recently available computational methods are very much helpful in its determination. The number of reports published for an online prediction of this important parameter remains limited. Here we will understand about various developed QSPR models to predict log *P* and its significant relationship with pharmacodynamic and pharmacokinetic properties of pharmaceutical formulations. It affects mainly bioavailability, permeability, and toxicity of pharmaceutical compounds [77].

Nowadays, prediction of ADMET properties using QSPR models is one of the inexpensive and faster way. Using this phenomenon, Talevi et al. have generated both linear and nonlinear models for estimating percentage of human intestinal absorption (HIA). The best developed QSPR model has shown correlation coefficient value of 0.659 and is represented below as Eq. 17.2 [78].

$$\log_{10}(\mathrm{HIA}+10) = 1.801\ (\pm 0.1) - 0.100\ (\pm 0.02) * \mathrm{BEHm1} + 1.639\ (\pm 0:4) * \mathrm{RNCG} - 0.139\ (\pm 0.04) * \mathrm{n-COOH} + 0.118\ (\pm 0.01) * \mathrm{MLOGP} \tag{17.2}$$

First 3D QSPR model was developed using various kinds of structural descriptors such as molecular electrostatic potential (MEP), local ionization energy (IE_L), local electron affinity (EA_L), hardness, polarizability, and field normal to the surface (FN) [79]. The calculation was done on the basis of AM1 semi-empirical molecular orbital theory, and the validated model has shown significant performance. A similar study to predict corneal permeability for fluoroquinolones was also performed by another group of scientists [80]. Their generated QSPR model has shown that permeability coefficient (represented as log PC) was dependent upon two descriptors, viz. partition coefficients (log *P*) and dissociation constant (denoted as log *D*), with R^2 value of more than 0.9. Arabinoxylan (AX) is a kind of biopolymer which is usually isolated from ispaghula seeds. The release profile of various drugs complexed with AX polymer was assumed on the basis of QSPR methods using a multiple linear regression (MLR) and neural networks, respectively [81]. The validation step was carried out by leave-one-out cross and y-scrambling techniques. Their results concluded that release profile is dependent upon various kinds of descriptors, for example, softness, lipophilicity, unsaturation, polarization, and topological as well as geometrical performance of the molecules.

Liposomes are one of best novel drug delivery systems, and around ten or more approved liposomal drugs are currently marketed. In many reports, it is suggested that the effective concentration of liposomal drugs can be achieved by remote/active loading of liposomes. The estimation of this parameter can be done by QSPR approach using various machine learning methods [82]. The study results of renowned group have shown that two types of highly specific and reliable results providing models were generated. The reported approach is helpful in selection of new candidates for liposome remote loading, helping the society in formulation experiments designing.

Many types of volatile organic compounds (VOCs) are the part of household things and may have adverse effects on human health. To overcome this problem, Palomba et al. have developed a QSPR model to predict blood-to-liver partition coefficients (log P_{liver}) of VOCs [83]. The selection of different molecular descriptors have been carried out using machine learning methods, and both decision trees and neural networks are employed as a part of regression step. The obtained results were highly specific and accurate as compared to previous developed models.

A QSPR model was generated for predicting n-octanol/water partition coefficients of almost 195 substituted aromatic drugs. The selection of descriptors and model development was carried out by using genetic algorithm and multiple linear regressions (GA/MLR) techniques [84]. The validated model has shown correlation coefficient value of 0.9433. For the treatment of skin-related disorders, human skin permeability of drugs matters. Baba et al. have compiled the data sets of

compound with their experimentally calculated values of permeability coefficients [85]. After selection of suitable descriptors, QSPR models were developed through use of machine learning techniques such as support vector regression (SVR) and random forest (RF), respectively. Their study results showed that results of nonlinear SVR method was more significant with correlation coefficient value of 0.91 as compared to RF.

Recently, QSPR model for sulfa drugs was developed using genetic algorithm-multiple linear regressions (GA-MLR) and genetic algorithm-artificial neural network (GAANN)-based methods [86]. Various descriptors were calculated using Dragon software, and R^2 values of the MLR and ANN models were found to be 0.312 and 0.9869, respectively. In drug delivery system, various kinds of pH-responsive copolymer micelles are commonly employed as carriers. Two QSPR models for predicting net cumulative drug release percentage (E_n) were developed at pH 7.4 and 5.0 of two different time-wise stages [87]. Their study results have showed that LUMO energy enhanced drug release in acidic condition. Apart from this, hydropathy property of micelle and polarity were other factors influencing the drug release. Another QSPR model was generated to investigate the binding way of almost 28 drugs with patiromer (a nonabsorbed, potassium binding drug approved for the treatment of hyperkalemia) at different pH range and ionic conditions using stepwise linear regression method [88]. Their study results revealed that four descriptors mainly such as surface area, ionization potential, electron affinity, and lipophilicity were part of model with correlation coefficient value of 0.7. Encapsulating the drug material in polymer nanoparticles leads to improvement in their pharmacokinetic and pharmacodynamic properties. Mainly two types of polymers such as (poly-[D,L-lactic acid] or PLA) and (poly-(ethylene glycol) or PEG) are used for this purpose. In 2017, a study for analyzing the drug affinity along with their efficient drug loading for PLA-PEG nanoparticles using in silico tools (for example, molecular dynamics and Monte-Carlo techniques) was conducted [89]. Their study results revealed that calculated adsorption energies and log *P* have significant correlation with experimental drug-loading values.

17.5 Conclusion

QSPR is one of most advanced technique in drug discovery and design. Their advanced application helps in development of a more robust QSPR model. The development of QSPR models intended to assist scientists with design ideas and understanding the behavior of various chemical components in successful delivery of the drugs in silico. The accessibility of molecular descriptors and available machine learning methods represent the utility of QSPRs in drug discovery research. The application of various QSPR modeling methods is discussed here in detail. We believe that this study helps the pharmaceutical community to explore more aspects of QSPR in pharmaceutical sector.

References

1. Gramatica P (2007) Principles of QSAR models validation: internal and external. QSAR Comb Sci 26(5):694–701. https://doi.org/10.1002/qsar.200610151
2. Dearden JC (2017) The history and development of quantitative structure-activity relationships (QSARs): addendum. Int J Quant Str Pro Rel 2(2):36–46. https://doi.org/10.4018/IJQSPR.2017070104
3. Buchwald P, Bodor N (2002) Computer-aided drug design: the role of quantitative structure–property, structure–activity and structure–metabolism relationships (QSPR, QSAR, QSMR). Drugs Future 27(6):577–588. https://doi.org/10.1358/dof.2002.027.06.856934
4. Roy K, Das RN, Kar S (2015) A primer on QSAR/QSPR modeling: fundamental concepts. Springer, London
5. Karelson M, Lobanov VS, Katritzky A (1996) Quantum-chemical descriptors in QSAR/QSPR studies. Chem Rev 96:1027–1044. https://doi.org/10.1021/cr950202r
6. Khan AU (2016) Descriptors and their selection methods in QSAR analysis: paradigm for drug design. Drug Discov Today 21(8):1291–1302. https://doi.org/10.1016/j.drudis.2016.06.013
7. Katritzky AR, Gordeeva EV (1993) Traditional topological indexes vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research. J Chem Inf Comput Sci 33 (6):835–857. https://doi.org/10.1021/ci00016a005
8. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32(7):1466–1474. https://doi.org/10.1002/jcc.21707
9. Ponce YM (2004) Total and local (atom and atom type) molecular quadratic indices: significance interpretation, comparison to other molecular descriptors, and QSPR/QSAR applications. Bioorg Med Chem 12(24):6351–6369. https://doi.org/10.1016/j.bmc.2004.09.034
10. Kuz'min VE, Artemenko AG, Polischuk PG, Muratov EN, Hromov AI, Liahovskiy AV, Andronati SA, Makan SY (2005) Hierarchic system of QSAR models (1D–4D) on the base of simplex representation of molecular structure. J Mol Model 11(6):457–467. https://doi.org/10.1007/s00894-005-0237-x
11. Oprea TI (2002) On the information content of 2D and 3D descriptors for QSAR. J Braz Chem Soc 13(6):811–815. https://doi.org/10.1590/S0103-50532002000600013
12. Todeschini R, Gramatica P (1997) The WHIM theory: new 3D molecular descriptors for QSAR in environmental modelling. SAR QSAR Environ Res 7(1–4):89–115. https://doi.org/10.1080/10629369708039126
13. Consonni V, Todeschini R, Pavan M (2002) Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. J Chem Inf Comput Sci 42(3):682–692. https://doi.org/10.1021/ci015504a
14. Klein CT, Kaiser D, Ecker G (2004) Topological distance based 3D descriptors for use in QSAR and diversity analysis. J Chem Inf Comput Sci 44(1):200–209. https://doi.org/10.1021/ci0256236
15. Senese CL, Duca J, Pan D, Hopfinger AJ, Tseng YJ (2004) 4D-fingerprints, universal QSAR and QSPR descriptors. J Chem Inf Comput Sci 44(5):1526–1539. https://doi.org/10.1021/ci049898s
16. Sun H (2004) A universal molecular descriptor system for prediction of logP, logS, logBB, and absorption. J Chem Inf Comput Sci 44(2):748–757. https://doi.org/10.1021/ci030304f
17. Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci 22(1):69–77. https://doi.org/10.1002/qsar.200390007
18. Bajorath J (2001) Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. J Chem Inf Comput Sci 41(2):233–245. https://doi.org/10.1021/ci0001482
19. Tropsha A, Weifan Z (2001) Identification of the descriptor pharmacophores using variable selection QSAR applications to database mining. Curr Pharm Des 7(7):599–612. https://doi.org/10.2174/1381612013397834

20. Singh KP, Gupta S (2014) Nano-QSAR modeling for predicting biological activity of diverse nanomaterials. RSC Adv 4(26):13215–13230. https://doi.org/10.1039/C4RA01274G
21. Whitley DC, Ford MG, Livingstone DJ (2000) Unsupervised forward selection: a method for eliminating redundant variables. J Chem Inf Comput Sci 40(5):1160–1168. https://doi.org/10.1021/ci000384c
22. Gramatica P, Gramatica P (2013) On the development and validation of QSAR models. In: Computational toxicology. Humana Press, Totowa, pp 499–526. https://doi.org/10.1007/978-1-62703-059-5_21
23. Wold S, Eriksson L, Clementi S (1995) Statistical validation of QSAR results. In: Chemometric methods in molecular design. pp 309–338. https://doi.org/10.1002/9783527615452.ch5
24. Zhulin IB (2015) Databases for microbiologists. J Bacteriol 197(15):2458–2467. https://doi.org/10.1128/JB.00330-15
25. Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. Mol Inform 29(6–7):476–488. https://doi.org/10.1002/minf.201000061
26. Madden JC, Enoch SJ, Paini A, Cronin MT (2020) A review of in silico tools as alternatives to animal testing: principles, resources and applications. Altern Lab Anim 48(4):146–172. https://doi.org/10.1177/0261192920965977
27. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M et al (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57(12):4977–5010. https://doi.org/10.1021/jm4004285
28. Ojha Lokendra K, Rachana S, Rani BM (2013) Modern drug design with advancement in QSAR: a review. Int J Res Biosci 2:1–2. http://www.ijrbs.in/index.php/ijrbs/article/view/58
29. Bhattachar SN, Deschenes LA, Wesley JA (2006) Solubility: it's not just for physical chemists. Drug Discov Today 11(21–22):1012–1018. https://doi.org/10.1016/j.drudis.2006.09.002
30. Tsopelas F, Giaginis C, Tsantili-Kakoulidou A (2017) Lipophilicity and biomimetic properties to support drug discovery. Expert Opin Drug Discov 12(9):885–896. https://doi.org/10.1080/17460441.2017.1344210
31. Butina D (2004) Performance of Kier-Hall E-state descriptors in quantitative structure activity relationship (QSAR) studies of multifunctional molecules. Molecules 9(12):1004–1009. https://doi.org/10.3390/91201004
32. Klopman G, Zhu H (2001) Estimation of the aqueous solubility of organic molecules by the group contribution approach. J Chem Inf Comput Sci 41(2):439–445. https://doi.org/10.1021/ci000152d
33. Manallack DT, Tehan BG, Gancia E, Hudson BD, Ford MG, Livingstone DJ, Whitley DC, Pitt WR (2003) A consensus neural network-based technique for discriminating soluble and poorly soluble compounds. J Chem Inf Comput Sci 43(2):674–679. https://doi.org/10.1021/ci0202741
34. Gao H (2001) Application of BCUT metrics and genetic algorithm in binary QSAR analysis. J Chem Inf Comput Sci 41(2):402–407. https://doi.org/10.1021/ci000306p
35. Savjani KT, Gajjar AK, Savjani JK (2012) Drug solubility: importance and enhancement techniques. ISRN Pharm 2012. https://doi.org/10.5402/2012/195727
36. Krishnaiah YS (2010) Pharmaceutical technologies for enhancing oral bioavailability of poorly soluble drugs. J Bioequiv Availab 2(2):28–36. https://doi.org/10.4172/jbb.1000027
37. Jouyban A (2009) Handbook of solubility data for pharmaceuticals. CRC Press. https://doi.org/10.1201/9781439804889
38. Hörter D, Dressman JB (2001) Influence of physicochemical properties on dissolution of drugs in the gastrointestinal tract. Adv Drug Deliv Rev 46(1–3):75–87. https://doi.org/10.1016/S0169-409X(00)00130-7
39. Hewitt M, Cronin MT, Enoch SJ, Madden JC, Roberts DW, Dearden JC (2009) In silico prediction of aqueous solubility: the solubility challenge. J Chem Inf Model 49(11):2572–2587. https://doi.org/10.1021/ci900286s
40. Bergström CA, Norinder U, Luthman K, Artursson P (2002) Experimental and computational screening models for prediction of aqueous drug solubility. Pharm Res 19(2):182–188. https://doi.org/10.1023/A:1014224900524

41. Delaney JS (2005) Predicting aqueous solubility from structure. Drug Discov Today 10 (4):289–295. https://doi.org/10.1016/S1359-6446(04)03365-3
42. Wang J, Hou T (2011) Recent advances on aqueous solubility prediction. Comb Chem High Throughput Screen 14(5):328–338. https://doi.org/10.2174/138620711795508331
43. Hammett LP (1940) Physical organic chemistry. https://doi.org/10.1021/ed017p551.3
44. Kitchen DB, Stahura FL, Bajorath J (2004) Computational techniques for diversity analysis and compound classification. Mini-Rev Med Chem 4(10):1029–1039. https://doi.org/10.2174/1389557043402982
45. Koch MA, Waldmann H (2005) Protein structure similarity clustering and natural product structure as guiding principles in drug discovery. Drug Discov Today 10(7):471–483. https://doi.org/10.1016/s1359-6446(05)03419-7
46. Dror O, Shulman-Peleg A, Nussinov R, Wolfson HJ (2004) Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. Curr Med Chem 11(1):71–90. https://doi.org/10.2174/0929867043456287
47. Khan PM, Roy K (2018) Current approaches for choosing feature selection and learning algorithms in quantitative structure–activity relationships (QSAR). Expert Opin Drug Discov 13(12):1075–1089. https://doi.org/10.1080/17460441.2018.1542428
48. Asikainen AH, Ruuskanen J, Tuppurainen KA (2004) Consensus kNN QSAR: a versatile method for predicting the estrogenic activity of organic compounds in silico. A comparative study with five estrogen receptors and a large, diverse set of ligands. Environ Sci Technol 38 (24):6724–6729. https://doi.org/10.1021/es049665h
49. Winkler DA (2004) Neural networks as robust tools in drug lead discovery and development. Mol Biotechnol 27(2):139–167. https://doi.org/10.1385/mb:27:2:139
50. Polanski J, Gieleciak R (2003) The comparative molecular surface analysis (CoMSA) with modified uniformative variable elimination-PLS (UVE-PLS) method: application to the steroids binding the aromatase enzyme. J Chem Inf Comput Sci 43(2):656–666. https://doi.org/10.1021/ci020038q
51. Yao X, Liu H, Zhang R, Liu M, Hu Z, Panaye A, Doucet JP, Fan B (2005) QSAR and classification study of 1, 4-dihydropyridine calcium channel antagonists based on least squares support vector machines. Mol Pharm 2(5):348–356. https://doi.org/10.1021/mp050027v
52. Nicolotti O, Altomare C, Pellegrini-Calace M, Carotti A (2004) Neuronal nicotinic acetylcholine receptor agonists: pharmacophores, evolutionary QSAR and 3D-QSAR models. Curr Top Med Chem 4(3):335–360. https://doi.org/10.2174/1568026043451384
53. Deng W, Breneman C, Embrechts MJ (2004) Predicting protein–ligand binding affinities using novel geometrical descriptors and machine-learning methods. J Chem Inf Comput Sci 44 (2):699–703. https://doi.org/10.1021/ci034246
54. Lü Q, Wu H, Yu R, Shen G (2004) The lifetime of CFC substitutes studied by a network trained with chaotic mapping modified genetic algorithm and DFT calculations. SAR QSAR Environ Res 15(4):279–292. https://doi.org/10.1080/10629360410001724923
55. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of ICNN'95-international conference on neural networks, vol 4, IEEE, pp 1942–1948. https://doi.org/10.1109/ICNN.1995.488968
56. Toropov AA, Leszczynska D, Leszczynski J (2007) QSPR study on solubility of fullerene C60 in organic solvents using optimal descriptors calculated with SMILES. Chem Phys Lett 441 (1–3):119–122. https://doi.org/10.1016/j.cplett.2007.04.094
57. Pourbasheer E, Aalizadeh R, Ardabili JS, Ganjali MR (2015) QSPR study on solubility of some fullerenes derivatives using the genetic algorithms—multiple linear regression. J Mol Liquids 204:162–169. https://doi.org/10.1016/j.molliq.2015.01.028
58. Ghasemi J, Saaidpour S (2007) QSPR prediction of aqueous solubility of drug-like organic compounds. Chem Pharm Bull 55(4):669–674
59. Palmer DS, O'Boyle NM, Glen RC, Mitchell JB (2007) Random forest models to predict aqueous solubility. J Chem Inf Model 47(1):150–158. https://doi.org/10.1021/ci060164k

60. Gharagheizi F (2008) QSPR studies for solubility parameter by means of genetic algorithm-based multivariate linear regression and generalized regression neural network. QSAR Comb Sci 27(2):165–170. https://doi.org/10.1002/qsar.200630159
61. Zhou D, Alelyunas Y, Liu R (2008) Scores of extended connectivity fingerprint as descriptors in QSPR study of melting point and aqueous solubility. J Chem Inf Model 48(5):981–987. https://doi.org/10.1021/ci800024c
62. Duchowicz PR, Talevi A, Bruno-Blanch LE, Castro EA (2008) New QSPR study for the prediction of aqueous solubility of drug-like compounds. Bioorg Med Chem 16 (17):7944–7955. https://doi.org/10.1016/j.bmc.2008.07.067
63. Duchowicz PR, Castro EA (2009) QSPR studies on aqueous solubilities of drug-like compounds. Int J Mol Sci 10(6):2558–2577. https://doi.org/10.3390/ijms10062558
64. Shayanfar A, Fakhree MA, Jouyban A (2010) A simple QSPR model to predict aqueous solubility of drugs. J Drug Deliv Sci Technol 20(6):467–476. https://doi.org/10.1016/S1773-2247(10)50080-7
65. Long J (2011) QSPR studies on the aqueous solubility of selected PCDD/FS by using artificial neural network combined with principal component analysis. In: 2011 5th international conference on bioinformatics and biomedical engineering, IEEE, pp 1–4. https://doi.org/10.1109/icbbe.2011.5781211
66. Kovdienko NA, Polishchuk PG, Muratov EN, Artemenko AG, Kuz'min VE, Gorb L, Hill F, Leszczynski J (2010) Application of random forest and multiple linear regression techniques to QSPR prediction of an aqueous solubility for military compounds. Mol Inform 29(5):394–406. https://doi.org/10.1002/minf.201000001
67. Bagheri M, Golbraikh A (2012) Rank-based ant system method for non-linear QSPR analysis: QSPR studies of the solubility parameter. SAR QSAR Environ Res 23(1–2):59–86. https://doi.org/10.1080/1062936x.2011.623356
68. Izrailev S, Agrafiotis D (2001) A novel method for building regression tree models for QSAR based on artificial ant colony systems. J Chem Inf Comput Sci 41(1):176–180. https://doi.org/10.1021/ci000336s
69. Bhhatarai B, Gramatica P (2011) Prediction of aqueous solubility, vapor pressure and critical micelle concentration for aquatic partitioning of perfluorinated chemicals. Environ Sci Technol 45(19):8120–8128. https://doi.org/10.1021/es101181g
70. Zeng XL, Wang HJ, Wang Y (2012) QSPR models of n-octanol/water partition coefficients and aqueous solubility of halogenated methyl-phenyl ethers by DFT method. Chemosphere 86 (6):619–625. https://doi.org/10.1016/j.chemosphere.2011.10.051
71. Fallah Fini M, Riahi S, Bahramian A (2012) Experimental and QSPR studies on the effect of ionic surfactants on n-decane–water interfacial tension. J Surfactant Deterg 15(4):477–484. https://doi.org/10.1007/s11743-012-1330-7
72. Salahinejad M, Le TC, Winkler DA (2013) Aqueous solubility prediction: do crystal lattice interactions help? Mol Pharm 10(7):2757–2766. https://doi.org/10.1021/mp4001958
73. Heidari A, Esposito J, Caissutti A (2019) Symbiodinolide time–resolved absorption and resonance FT–IR and Raman biospectroscopy and density functional theory (DFT) investigation of vibronic–mode coupling structure in vibrational spectra analysis. Clin Case Stud Rep 2 (3):1–4. https://doi.org/10.15761/CCSR.1000130
74. Cabrera Y, Polo J, Machín L, Torres L (2019) Modeling aqueous solubility of drug like organic compounds in strongly acidic medium. https://doi.org/10.3390/mol2net-05-06268
75. Chinta S, Rengaswamy R (2019) Machine learning derived quantitative structure property relationship (QSPR) to predict drug solubility in binary solvent systems. Ind Eng Chem Res 58(8):3082–3092. https://doi.org/10.1021/acs.iecr.8b04584
76. Al-Husseini JK, Stanton NJ, Selassie CR, Johal MS (2019) The binding of drug molecules to serum albumin: the effect of drug hydrophobicity on binding strength and protein desolvation. Langmuir 35(52):1705460. https://doi.org/10.1021/acs.langmuir.9b02318

77. Kujawski J, Popielarska H, Myka A, Drabińska B, Bernard MK (2012) The log P parameter as a molecular descriptor in the computer-aided drug design—an overview. Comput Methods Sci Technol 18(2):81–88. https://doi.org/10.12921/cmst.2012.18.02.81-88
78. Talevi A, Goodarzi M, Ortiz EV, Duchowicz PR, Bellera CL, Pesce G, Castro EA, Bruno-Blanch LE (2011) Prediction of drug intestinal absorption by new linear and non-linear QSPR. Eur J Med Chem 46(1):218–228. https://doi.org/10.1016/j.ejmech.2010.11.005
79. Kramer C, Beck B, Clark T (2010) A surface-integral model for log P_{OW}. J Chem Inf Model 50 (3):429–436. https://doi.org/10.1021/ci900431f
80. Sharma C (2011) Development of novel in silico model to predict corneal permeability for congeneric drugs: a QSPR approach. J Biomed Biotechnol 2011. https://doi.org/10.1155/2011/483869
81. Akbar J, Iqbal MS, Chaudhary MT, Yasin T, Massey S (2012) A QSPR study of drug release from an arabinoxylan using ab initio optimization and neural networks. Carbohydr Polym 88 (4):1348–1357. https://doi.org/10.1016/j.carbpol.2012.02.016
82. Cern A, Golbraikh A, Sedykh A, Tropsha A, Barenholz Y, Goldblum A (2012) Quantitative structure-property relationship modeling of remote liposome loading of drugs. J Control Release 160(2):147–157. https://doi.org/10.1016/j.jconrel.2011.11.029
83. Palomba D, Martínez MJ, Ponzoni I, Díaz MF, Vazquez GE, Soto AJ (2012) QSPR models for predicting log p liver values for volatile organic compounds combining statistical methods and domain knowledge. Molecules 17(12):14937–14953. https://doi.org/10.3390/molecules171214937
84. Bahmani A, Saaidpour S, Rostami A (2017) A simple, robust and efficient computational method for n-octanol/water partition coefficients of substituted aromatic drugs. Sci Rep 7 (1):1–4. https://doi.org/10.1038/s41598-017-05964-z
85. Baba H, Takahara JI, Mamitsuka H (2015) In silico predictions of human skin permeability using nonlinear quantitative structure–property relationship models. Pharm Res 32 (7):2360–2371. https://doi.org/10.1007/s11095-015-1629-y
86. Dadfar E, Shafiei F, Isfahani TM (2020) Structural relationship study of octanol-water partition coefficient of some sulfa drugs using GA-MLR and GA-ANN methods. Curr Comput Aided Drug Des 16(3):207–221. https://doi.org/10.2174/1573409915666190301124714
87. Zhang R, Wen LY, Wu WS, Yuan XZ, Zhang LJ (2019) Quantitative structure-property relationship for pH-triggered drug release performance of acid-responsive four/six-arms star polymeric micelles. Pharm Res 36(1):1. https://doi.org/10.1007/s11095-018-2549-4
88. Brew CT, Blake JF, Mistry A, Liu F, Carreno D, Madsen D, Mu Y, Mayo M, Stahl W, Matthews D, Maclean D (2018) Use of QSPR modeling to characterize in vitro binding of drugs to a gut-restricted polymer. Pharm Res 35(4):1. https://doi.org/10.1007/s11095-018-2356-y
89. Meunier M, Goupil A, Lienard P (2017) Predicting drug loading in PLA-PEG nanoparticles. Int J Pharm 526(1–2):157–166. https://doi.org/10.1016/j.ijpharm.2017.04.043

Modelling Approaches for Studies of Drug-Polymer Interactions in Drug Delivery Systems

18

Mire Zloh and Nuno Martinho

Contents

Abstract

Medicines are often mixtures comprising active pharmaceutical ingredients (APIs) and excipients, components that usually bulk up formulation or stabilize APIs. The need for therapeutic potential enhancements of APIs with poor

M. Zloh (✉)
Faculty of Pharmacy Novi Sad, University Business Academy in Novi Sad, Novi Sad, Serbia

UCL School of Pharmacy, University College London, London, UK
e-mail: mire.zloh@faculty-pharmacy.com

N. Martinho
iBB—Institute for Bioengineering and Biosciences and Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_18

solubility and permeability or for their targeted delivery requires more complex formulations. Polymers belong to a class of pharmaceutical excipients that are often used as API (drug) carriers. The advancement in the design of polymers and synthesis methods allows the control of drug delivery either via improving drug dissolution, aiding the drug diffusion, or degradation of the carrier matrix at the site of action. These mechanisms depend on the structures of formulation components and importantly on their intermolecular interactions. The structure elucidation of polymers is often impeded by their nature and lack of homogeneity, while the experimental evaluation of intermolecular interactions provides limited information. That information at the atomistic level can be obtained by using computational chemistry. The focus of this chapter is on the in silico approaches to generate three-dimensional models of drugs and polymers and evaluate interactions between them as a basis for the rational design of pharmaceutical formulations.

Keywords

Nanomedicine · Drug delivery · Polymers · Dendrimers · Molecular modelling

Chapter Objectives

- Grasp the drug discovery and development process with the inclusion of the design of the drug delivery systems
- Realize key factors contribute to the development of formulations
- Understand in silico methods that can provide insights into interactions of drugs with the components of drug delivery systems
- Appreciate different software packages and their applications in studying intermolecular interactions
- Understand the basis of molecular docking and molecular dynamics and know where to find additional resources
- Gain information on a basic protocol for generation, simulation and analysis of a drug delivery system with a therapeutic molecule
- Understand the importance of solvent in the design of stimuli-responsive polymers for drug encapsulation and release
- Know the differences in behaviour of linear, hyperbranched and dendrimer polymers

18.1 Introduction

Pharmaceutical research and industry are focused on bringing health to the world. Health is defined as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity” by the World Health Organization [1]. Human organisms and other living beings have complex biology that depends on the interplay of many factors, including their composition and interactions between components. Furthermore, beyond the intermolecular interactions between

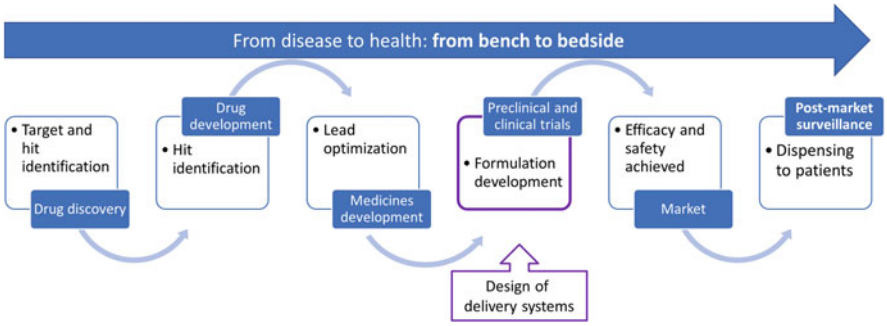


Fig. 18.1 An overview of the drug discovery and development process

the building blocks of life, organisms are exposed to the environment that in turn influences the biochemical pathways essential for maintaining their normal state and function. The knowledge about this interplay is on the increase by the study of systems biology and stored in relevant databases, such as Reactome knowledgebase [2], Kyoto Encyclopedia of Genes and Genomes (KEGG) [3] and Human Annotated and Predicted Protein Interactions (HAPPI) [4], among others.

Diseases are generally associated with abnormal states of parts of a body or deviation from the normal functions that are often manifested with certain signs and symptoms. Often diseases are a consequence of very complicated interactions between genetic, epigenetic and external chemical agents that may disrupt biological pathways. Understanding the molecular processes that are associated with disease-modifying pathways can guide the development of novel therapeutic agents [5, 6]. Such knowledge can support medicine development (Fig. 18.1) by revealing which biomacromolecules are potential targets for the intervention by a ligand (small molecule or another biomacromolecule). Ligands generally interact with a target biomacromolecule and influence disease-modifying pathways. These molecules could inhibit or activate pathways in question and, depending on the nature of the disease, have the potential to alleviate symptoms of the disease and restore the normal state of the organisms.

The discovery of bioactive ligands (hits) is the first step in the development of drugs and medicines. The iterative changes of the molecular structure of the hit are often required to obtain a drug candidate with adequate solubility and stability, as well as absorption, distribution, metabolism, excretion (ADME) and toxicity properties. The final stages of medicine development include mixing the drug with different components into a form that would be suitable for its optimal or acceptable delivery to the targeted biomacromolecule. Formulated forms of a drug, also known as an active pharmaceutical ingredient (API), are then submitted to clinical trials before entering the market and used in clinics.

Components of the medicine formulation often termed as excipients have various functions that either are responsible for the physical appearance of the medicine, enable drug administration, ensure the stability of medicines or improve ADME

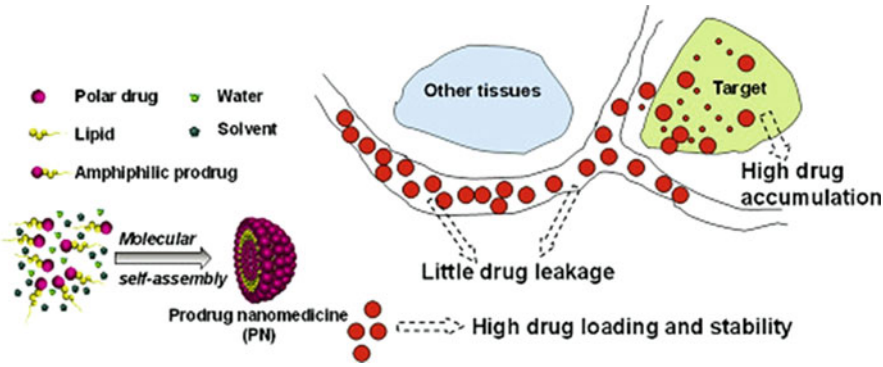


Fig. 18.2 An example of a drug delivery system, from assembly to targeted delivery. (Reproduced from [14])

properties of the API. The development of formulations to a final product to be used in the clinic uses a combination of experimental evaluations and quality-by-design approaches to optimize the properties of medicines [7]. It is a given that the function of the excipients in the formulation production and API delivery depends on the structure and molecular properties of both API and excipients, and consequently the interplay of various intermolecular interactions between them. Moreover, these are the interactions that determine the function of the excipient in different formulations and the type of formulation to be used, such as tablets [8], injectables [9, 10], creams for topical delivery [11] and eye drops [12]. Most of the drug formulations are multicomponent systems that aim to provide a balanced effect on the availability and activity of API (drug) molecules.

Such a variety of excipient functions used in different physical states of formulations requires a diversity of chemical structures of molecules that belong to different chemical classes. The advancement of the science of formulation development enabled the development of drug delivery systems with additional functions such as a prolonged release of APIs or their targeted delivery to a specific tissue or protein of interest. Controlled and targeted drug release functions are often achieved by the use of polymeric systems [13] to form a drug-polymer complex. Drug molecules could be potentially on the surface or inside the cavities of folded polymer chains. The formation of these complexes and the subsequent release of the drug at the desired site/tissue (Fig. 18.2) depend on various intramolecular and intermolecular interactions formed between polymers and the API and/or surrounding body fluids [14].

These interactions can be evaluated using various experimental techniques, including nuclear magnetic resonance (NMR) spectroscopy [15, 16], X-ray crystallography [17–19], infrared (IR) spectroscopy [20], small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) [21]. However, such studies are not always feasible due to various constraints. A plethora of computational tools has also been developed that enables the study of these systems in silico to overcome

experimental limitations or to use the results for predictive purposes. The prediction of interactions between biopolymers and different small molecules including drugs and natural products is reviewed elsewhere [22–24]. It has been shown that the insights into drug-polymer interactions can be obtained using molecular dynamics (MD) simulations [25]. The advancement of the computational methods combined with hardware development opened opportunities to study larger molecular systems including polymer-based formulations.

The aim of this chapter is to provide an overview and progress of the in silico predictions of drug-polymer interactions illustrated by selected examples. Furthermore, the use of different computational tools to model, at the atomistic level, the three-dimensional (3D) structures of small molecule drugs, various polymers and their complexes will be explored with the focus on selected open-source software and free-for-academics software.

18.2 Polymers as Drug Carriers

Polymers, substances that are based on macromolecules, have repeating units named monomers that are covalently linked to form a chain. Monomers are usually small molecules with two or more linking points, leading to either linear polymers (LP), cross-linked polymers (CLP) or hyperbranched polymers (HBP) that include dendrimers as a special case. These molecules usually have high molecular weights and can have either a narrow (monodisperse) or wide (polydisperse) molecular weight range. Additionally, the use of only one type of monomer results in the production of homopolymers, while the use of two or more types of monomers results in heteropolymers.

The diversity of monomers that can be used to produce polymers leads to almost endless possibilities in generating substances with desired properties. There has been a growing interest in formulation development to overcome the chemical and physical shortcomings associated with drugs. In particular, the use of polymers for drug delivery systems is on the rise. Drug encapsulation by polymers is a major advantage to increase availability and targetability in drug delivery, a process that can be shifted towards more versatile polymers. The nature, physico-chemical properties and available functional groups in monomers drive intermolecular interactions within the polymer chain and influence the 3D arrangement of atoms in space (polymer folding). However, final 3D structures of polymers depend on the presence and nature of solvents, salts, ions and other small molecules that could be part of either formulation or body fluids. These factors are behind the main mechanisms of action of polymers responsible for controlled drug release, namely diffusion, osmotic effects and polymer erosion [13]. Although the effects of these mechanisms on the drug release can often be measured in vitro, it still remains challenging to predict them without inputs from computational chemistry.

Importantly, the pharmaceutical industry is concerned with polymers that are suitable for biomedical applications which require non-toxic and often biodegradable materials [26, 27]. Such polymers could either be thermo-responsive [28] or

respond to other stimuli in vivo [29, 30]. It was demonstrated in many clinical examples that polymers can modulate the release of drugs, prevent their degradation and effluxes and aid active targeting to the desired tissue, thus overcoming the limitations of the drug alone. Therefore, the proper choice of drug formulation via drug delivery systems is of utmost importance and understanding their molecular interactions with drugs is crucial to enhance their performance. However, our understanding of the interactions between polymers and drugs is still limited, particularly the intricacies that allow design of novel biocompatible polymers with physicochemical properties compatible with the API. As a result, the emerging knowledge of polymers synthesis combined with molecular modelling to understand their interactions with drugs is going to be a next major step in developing new medicines.

18.2.1 Linear Polymers

Linear polymers are long chains of monomers linearly connected to each other with potentially random 3D arrangement in space. As a result of the intramolecular interactions and nature of monomers, these molecules may form flexible structures that could form linear, loop and coil segments. Such polymer chains exhibit surfaces for interactions with neighbouring polymer chains (aggregation) and potentially forming cavities that may accommodate other small molecules (encapsulation). Both phenomena depend on the balance between attraction and repulsion interactions, such as van der Waals, ionic and dipole interactions. It is long recognized that the choice of polymers for biomedical applications has to consider chemical, interfacial, biological and mechanical properties, and workflow for polymer selection was proposed [31].

Often, natural polymers, such as hyaluronic acid, chitosan, alginate, gelatin and cellulose, are considered in the development of the drug delivery systems [32–34]. Although these polymers based on carbohydrates and protein building blocks have many advantages, there is a need for the use of synthetic polymers with greater control of their composition and tunable properties [35, 36]. Some of the common monomers or precursors used in the generation of such polymers are shown in Fig. 18.3.

18.2.2 Hyperbranched Polymers

Dendritic polymers such as dendrimers and hyperbranched polymers are highly three-dimensional branched molecular entities. As these polymers grow in size, a large number of intramolecular cavities are often formed that facilitate encapsulation of small molecules. Dendrimers, contrary to linear polymers and randomly branched polymers that commonly attain a multitude of random conformations, generally result in a “defined” characteristic globular conformation that can be utilized for drug delivery purposes. Moreover, they display large surface areas as well as

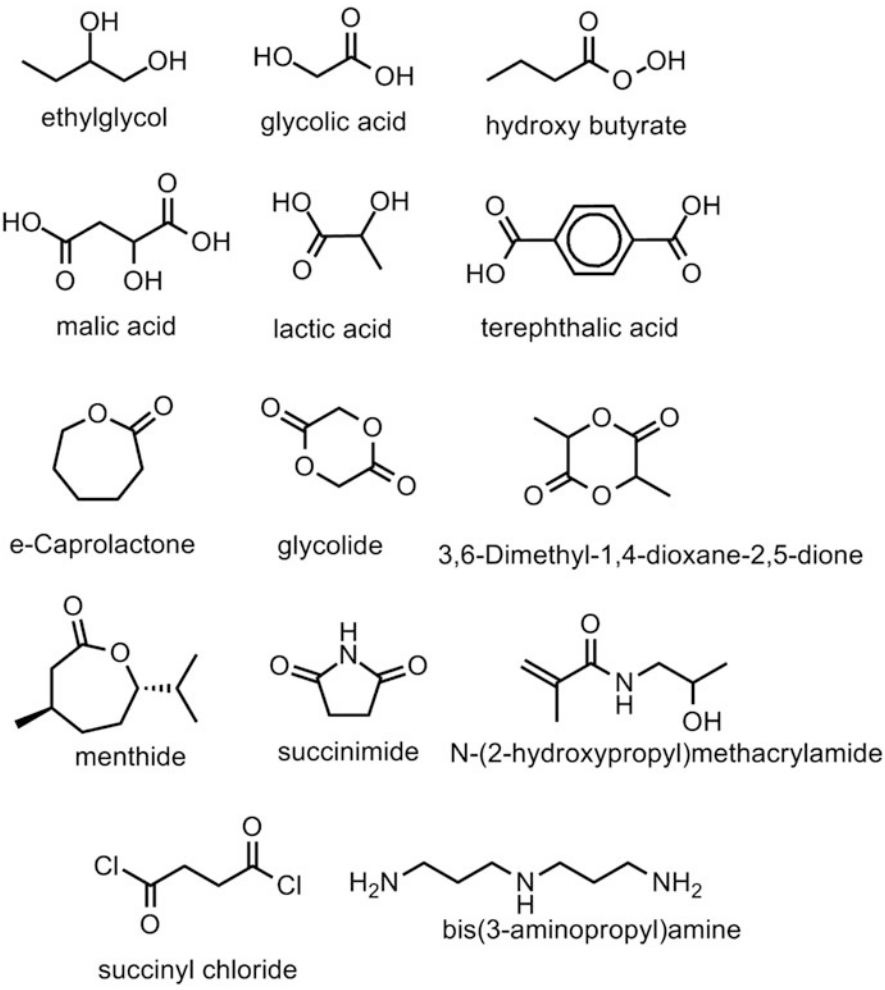


Fig. 18.3 Chemical structures of some monomers and precursors used in the generation of biodegradable polymers

multivalent terminal groups that lead to advantageous properties such as increased solubility and biocompatibility.

Chemically, both dendrimers and hyperbranched polymers have analogous properties. Dendrimers are uniformly monodispersed polymers with predictable and more controlled behaviour, which is highly desirable for pharmaceutical applications. A range of dendrimers generated from different building blocks (Fig. 18.4) is considered for biomedical applications. However, due to their precise nature, it may be costly to synthesize and purify them. On the other hand, hyperbranched polymers are irregular molecules with some comprising overgrown branches while others shorter branches. This results in a mixture of structurally

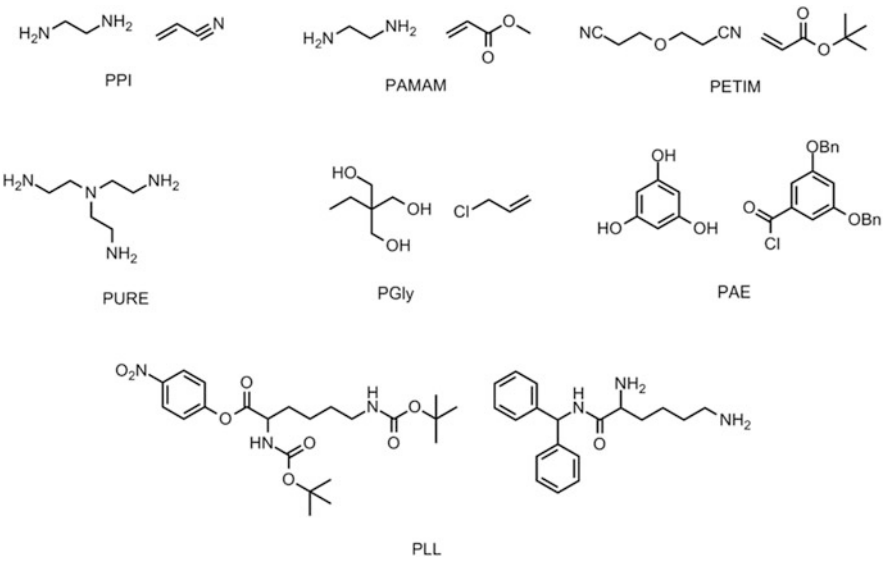


Fig. 18.4 Chemical structures of some monomers used as dendrimer building blocks with their corresponding reagents (PPI: Ethylenediamine; Acetonitrile; PAMAM: Ethylenediamine; Methyl Acrylate; PETIM: Bis(2-cyanoethyl) ether; tert-Butyl acrylate; PURE: Tris(2-aminoethylamine);-PGly: 2-Ethyl-2-(hydroxymethyl)-1,3-propanediol, 3-chloroprop-1-ene; PAE: Phloroglucinol 3,5-bis(benzyloxy)benzoyl chloride; PLL: Boc-L-Lys(Boc)-OPh(4-NO2); 2,6-diamino-*N*-[(2*S*)-1-phenylpropan-2-yl]hexanamide)

similar polymers with different molecular weights. Even though it is not easy to predict their conformations, it is known that they still form randomly distributed cavities. Since hyperbranched polymers can be prepared through one-step procedures, it is thus much easier to use for large-scale applications.

The versatility of these polymers to be used in drug delivery systems results for the possibility to tailor their branching and terminal groups for different goals and applications. Such modifications of their architecture can optimize their capability to encapsulate both hydrophilic and hydrophobic small molecules in their interior. Therefore, the variation of their chemical composition results in cavities with different polar/apolar properties as observed for poly(propyleneimine) (PPI) vs poly(amidoamine) PAMAM dendrimers [37]. This also means that polymers can be designed as unimolecular micelle-like polymers with amphiphilic characteristics that do not rely on critical micellar concentration to act as solubilizers [38]. The incorporation of drugs into these cavities by the formation of host-guest interactions can thus be exploited for drug delivery systems but can also be employed for purifications purposes [39, 40].

Host-guest interactions can also be established at the surface of dendritic polymers. However, the driving interactions at the surface greatly affect their loading efficiency and stability of drugs as well as the solubilization and aggregation of the polymer itself. Moreover, these interactions can not only greatly be affected by the

structural diversity of the polymers but also the environment where solvent and other molecules compete for the same type of interactions. For instance, even though the binding at the surface is generally driven by nonspecific interactions such as electrostatic, hydrophobic, coordination and hydrogen bonding, some specificity of binding can be achieved by the shape complementarity and charge of the binding molecules. This can be observed for the PPI generation 5 dendrimers, which showed greater affinity for Rose Bengal and Erythrosin when compared to other dyes. It was also shown that the pH changes affected both the loading and kinetics of these molecules [41]. Similarly, the incorporation of 5-fluorouracil into the PAMAM generation 5 was higher for amine-terminated dendrimers (70 molecules) in comparison to hydroxyl-terminated ones (14 molecules) due to the ability of the drug to protonate the terminal groups [42]. Alternatively, drugs can be covalently attached at the surface groups and act as prodrugs. This has been shown to be advantageous as it can protect the drug from degradation and cellular detoxification with the possibility to improve their activities [43, 44].

Despite the success of dendritic polymers in encapsulating small molecules, there is still a need for better understanding of the interactions between polymers and drugs. Understanding these interactions becomes increasingly difficult when multiple molecules can bind simultaneously, and one promotes the binding of the other [45]. Moreover, the same interactions that lead to host-guest complexation need to be tuned to control the release of drugs out of the polymer in response to a stimulus. Furthermore, because dendritic polymers do not have specific binding sites, the exchange of their guest molecules with other molecules in their surroundings can create toxicity problems by picking up various biomolecules or by chelation of important ions. Even though there have been considerable efforts to increase the selectivity for recognition of specific molecules, there are still opportunities for considerable improvements in the tailored design of polymers for drug delivery. For this reason, there has been a growing role in using computational modelling approaches to understand the interactions of dendritic polymers at the atomic level and their interaction with small molecules to design novel polymers.

18.3 Molecular Modelling of Drugs and Polymers

Molecular modelling is a discipline that uses computer-based methods to mimic the behaviour of molecules and predict their properties including reactivity. These methods can use the structure of molecules as a detailed 3D representation of its atoms in space, which consequently can be used in further in silico estimation of the relevant properties and interactions with surrounding molecules. Although the initial 3D structure of molecules can be obtained experimentally by using X-ray crystallography, NMR and electron microscopy, the number of molecular species with unknown 3D structures is far greater. Molecular modelling techniques can be employed to generate structures for such known compounds as well as for the design of novel molecular species, including small molecules and polymers. This chapter does not aim to provide a comprehensive introduction to various approaches used in

molecular modelling. The comprehensive introduction to modelling and computational chemistry can be found elsewhere [46].

Briefly, the most accurate prediction of the structure of molecules can be achieved by considering the positions of nuclei and electron distribution in molecular orbitals. These high-level theory methods, including quantum mechanics methods such as ab initio and density functional theory approaches, are mostly suitable for small molecules studies. However, these methods are generally not suitable to study drug formulation systems due to the size of the system that includes many components of large molecular weight such as polymers. Furthermore, it is often beneficial to obtain information on dynamic processes within drug delivery systems and/or their interactions with target proteins or biological membrane which further increases the size and complexity of the system.

The development of empirical methods that can model systems containing several hundred thousand atoms enabled all-atom simulation studies of drug delivery systems in the presence of solvent molecules and other components that could be found in the surrounding environment. The size of the systems to be modelled can be increased by using coarse-grained methods, where several atoms can be represented as a single particle; however, the level of details on intermolecular interactions is decreased. Therefore, the method of choice is molecular mechanics which considers atoms (nuclei and electrons) as a particle with each atom type being defined by its specific radius and a charge. It is considered that the system is frozen (kinetic energy is neglected), while the potential energy of the whole system (E_{pot}) can be separated into bonded (E_{bonded}) and nonbonded energies ($E_{nonbonded}$) (Eq. 18.1). The bonded energy often comprises four components (Eq. 18.2), while nonbonded generally has up to three components (Eq. 18.3).

$$E_{pot} = E_{bonded} + E_{nonbonded} \tag{18.1}$$

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral} + E_{improper} \tag{18.2}$$

$$E_{nonbonded} = E_{elec} + E_{vdW} + E_{H\text{-}bond} \tag{18.3}$$

Atoms of one element can be assigned different atom types (i.e. properties) depending on which functional group they are found in. The covalent bonds are represented by springs with a reference bond length and binding force constant that depends on the atom types in the bond. Three additional geometrical parameters are defined, namely bond angle, torsional angle (dihedral angle between four atoms connected by three sequential bonds) and improper angle (dihedral angle between four non-consecutive atoms), with their respective reference values for angles, force constants and other parameters that need to be defined for each type of connectivity. The non-bonded interactions include energies that arise from Coulomb interactions (E_{elec}), van der Waals forces (E_{vdW}) and hydrogen bonding ($E_{H\text{-}bond}$). These are distance-dependent energy terms between different atom types found in molecular systems. All parameters for the above energy terms are derived from experimental

data and stored in a set of parameters and equations named force field. There are several different force fields developed for atomistic molecular dynamics simulations that have been used for polymers including but not limited to CHARMM [47], AMBER [48], CVFF [49], Dreiding [50], GROMOS [51] and OPLS [52]. Most of these force fields are being implemented in different computational chemistry software packages used for simulations of polymers drug systems. A suitable combination of the force field and minimization algorithms can be used to predict a low energy 3D conformation of a molecule or polymer.

Furthermore, the integration of Newton's equation of the motion can be used to calculate the forces exerted on each atom in the structure at each time step and predict their accelerations and positions in the next time step. These calculations termed molecular dynamics simulations result in a trajectory that provides information on the position of each molecule over a time period. The trajectory analyses can provide information on the dynamic behaviour of a molecule and can describe the change of the molecular properties with time. One of the added values of molecular dynamics simulation in these molecular systems is the information of the influence of solvent molecules and other components such as APIs, buffers, salts and other excipients on the intermolecular interactions of polymers.

For further introduction into molecular mechanics with simplified theoretical basis behind molecular dynamics see [53, 54]. An extensive introduction in polymer simulations using various levels of theory with a proposed roadmap is provided elsewhere [55], which covers best practices, key challenges, selection of computational methods and analyses. A simplified outline of the molecular modelling process of the drug delivery system study is presented in Fig. 18.5.

18.3.1 Structure Generation

Initial 3D structures of different molecules to be used in simulations can be found in databases of small molecules [56] and proteins [57], or derived from experimental data. However, for many small molecules and polymers, such information is not available and in silico generation of 3D structures for those must be employed. In principle, this can be achieved with any chemical drawing program in conjunction with adequate molecular modelling packages, although there are considerations and limitations for each type of molecule.

18.3.2 Small Molecules

The most often used approach is to generate a simplified molecular-input line-entry system (SMILES) or International Chemical Identifier (InChi) string for a selected molecule from a 2D structure. This can be achieved using a chemistry drawing package like JChemPaint [58] or Ketcher [59]. Additionally, SMILES strings for many compounds can be found in online databases, that is PubChem [60], DrugBank [61], etc. These SMILES strings can be converted into 3D structures

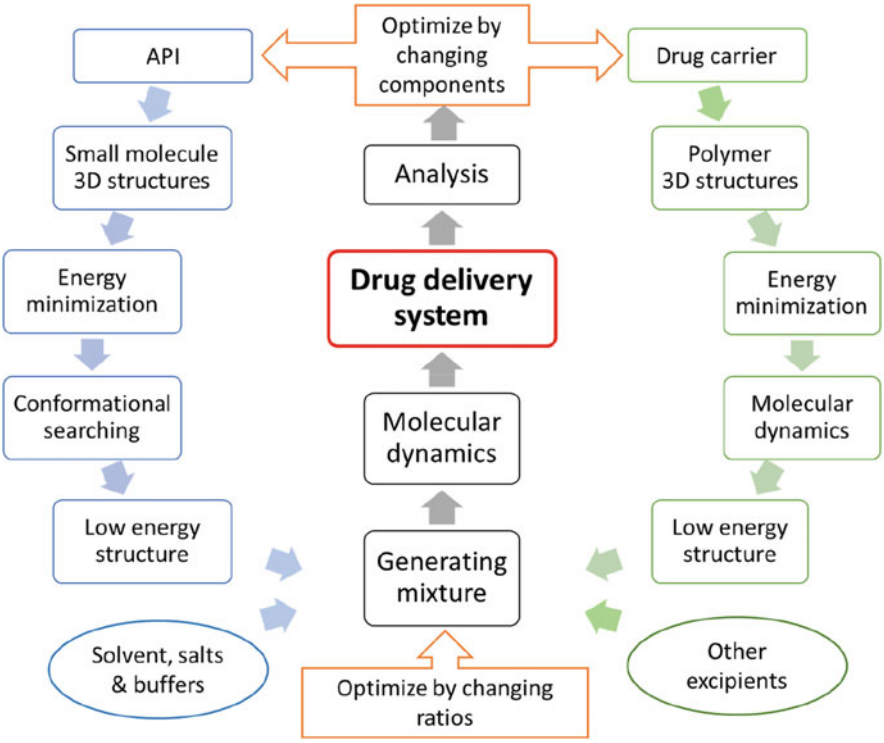


Fig. 18.5 A general outline of the suggested process for drug delivery system design

using software such as Avogadro [62], Jmol [63], OpenBabel [64] and VegaZZ [65]. This approach is preferred to drawing structures in a 3D editor due to complexity of the visualization of larger molecules and possible errors in selecting correct atoms when making bonds between them. Most of these packages optimize the 3D structure that often results in a local energy minimum, while Avogadro and VegaZZ can be used for conformational searching to obtain a near-global minimum structure. Such structures can be used for the prediction of molecular properties and as input files for further molecular dynamics and/or docking studies.

Box 18.1 List of Structure Drawing Software and Molecular Editors

JChemPaint, version 3.3-1210, https://jchempaint.github.io/. Available for Linux, Windows and as a java applet. It requires JAVA to run. First launch in 2000. Available for free and it is open source. This is chemistry drawing software and allows drawing 2D structures of drugs and polymers [58].

Avogadro, version 1.90.0, https://avogadro.cc/ Available for Linux, macOS, and Windows. First launch 2008. Available for free. It is a cross-

(continued)

Box 18.1 (continued)
platform that can be used to design small molecules and polymers. Allows visualization, minimization and creation of input files for other software. It is also molecular editor with built-in molecular mechanics tool for minimization and conformational searching that can provide optimized structures in a local and near-global minimum [62]. It can be used for the preparation of structures for molecular dynamics simulations [66, 67].

VegaZZ, version 3.1.2.39, https://www.ddl.unimi.it/cms/index.php?Software_projects:VEGA_ZZ. Its GUI runs only on the Windows platform, while command line version is available for Linux. First launch 2000. VegaZZ is free for non-commercial use [68]. It is a complete molecular modelling package for editing and visualizing molecules, with the interface to several packages for molecular mechanics (AMMP and NAMD) and semi-empirical chemistry software (MOPAC), as well as built-in graphical user interface (GUI) scripts for molecular docking using Vina, PLANTS and Escher NG. It has a range of features for molecular properties calculations and analyses of small molecules, polymers and proteins [69, 70].

Ambertools, version 20, https://ambermd.org/AmberTools.php. Available for Linux and Windows compiled using Cygwin environment. First launch 1997. Available for free. Ambertools is a toolkit that provides programs to facilitate force-field parameter development in addition to other functionalities [71].

18.3.3 Linear Polymers

The repeating nature of polymers and their sizes impede generating 3D structures linear chains via drawing 2D structures by connecting a chain of monomers. There are only a few free-for-academics software packages that can be used for building polymers. The most comprehensive polymer builder is a web-based platform CHARMM-GUI [72], which allows building polymers from 60 different monomers, providing files compatible with the CHARMM force field. Additionally, it is possible to generate solutions of these polymers in ten different solvents. Polymatic [73] is a command line package that uses Perl scripts to generate polymers as input files for LAMMPS molecular dynamics package [74]. Currently, most of the polymer builders with GUI are part of the commercial software platforms, such as Materials Studio (Biovia), Materials Science (Schrodinger), QuantumATK (Synopsys) and AMS2019 (Software for Chemistry & Materials B.V.).

18.3.4 Hyperbranched Polymers and Dendrimers

On the other hand, the task of building branched polymers is very cumbersome and limited. This is especially true for larger molecules where mistakes drawing the 2D

structures can be easily made. To automate this process several tools have been available to build dendrimers such as Dendrimer Building Toolkit (DBT) [48], Starmaker [75], Dendrimer Builder (for peptide dendrimers) [76] and other in-house tools. There are reports of tools that allow both the generation of dendrimers and hyperbranched polymers but have not yet been shared with the public domain [77, 78]. It is worth noting that some of the methods used to build dendrimers, that is DBT, can be used to generate structures of linear polymers if the branching parameter is set to 1. On the other hand, contrary to dendrimers and some linear polymers that can be represented as single monodisperse entities, hyperbranched polymers can be a mixture of thousands of different topologically similar polymers, and therefore appropriate random representations of these polymers need to be selected. This includes measuring the degree of branching, polymerization, polydispersity and how compact they are [78].

One of the key problems in the generation of dendritic polymers is the possibility of steric overlaps of branches that create a biased local energy minimum. The steric overlap in such conformations cannot often be resolved using energy minimization or molecular dynamics simulations, in such cases, energy barriers can be overcome using simulated annealing simulations [77].

Another constraint of these tools is the applicability of available force fields while building polymers, including dendrimers. There are no dedicated force fields developed for dendrimers and hyperbranched polymers that are at the same time applicable for small molecules and biopolymers. This is expected due to their complex nature and diverse chemistry, for which some parameters may not be available in the widely used software packages. Therefore, missing parameters must be proposed and validated as exemplified by poly(ethylene oxide) studies [79]. Nevertheless, it is possible to assign missing parameters for new atom types by comparing to those in available force fields. Modelling and simulations of dendrimers and hyperbranched polymers using this approach have been shown to accurately represent these types of polymers [80, 81]. Additionally, tools such as those within Ambertools [71] or websites such as LigParGen [82] can be used to obtain missing force field parameters for atoms in polymers.

The major expected advantage of modelling dendritic polymers is the virtual exploration of both the chemical space and molecular designs. In these simulations variables such as size, branching architecture and terminal groups can be tested on how they affect the overall properties and how these correlate with activity. For example, modelling of PPI and PPI-PAMAM dendrimers allowed determination of the parts responsible for their loading capacity of a small molecule. This in turn enabled the design of a better dendrimer construct with lower toxicity [83]. It is expected that modelling of dendrimers could also guide their synthesis by providing reliable predictions of retrosynthesis processes prior to further in vitro experiments [84].

18.3.5 Molecular Dynamics Simulation of Polymers

Computational simulations are now a crucial tool for drug development and have been widely adopted for material design, particularly for biomedical applications. Ab initio methods remain the most accurate method to identify low energy structures for small molecules without prior information, but its use in polymers has been limited due to the demanding computational cost. For this reason, classical all-atom molecular dynamic simulations have been the cornerstone of polymer modelling and simulation, especially to understand their behaviours for biomedical applications. In the case of dendritic polymers, however, molecular dynamics simulations have been crucial to probe their dynamics in solution such as branch distribution, water penetration, number of cavities and the backfolding of the terminal groups. Generally, these are properties too difficult to examine experimentally due to their high complexity and repetitive nature. To this end, several studies using Brownian Dynamics, Monte Carlo and molecular dynamics simulations have been carried out to characterize polymers in solution [80, 81, 85, 86]. Even though these methods can be applied in a certain way to any type of polymer, the study of linear polymers will be limited as they generally demonstrate random conformations, and therefore their interactions with small molecules are more related to bulk properties. As expected, several factors impact the behaviour of dendritic polymers in an aqueous solution including their chemistry but also the solvent (type and presence of salts). Notably, molecular dynamics of dendrimers using common force fields have been able to corroborate in vitro observations of dendrimer scaffolding transition upon variations of pH which result in different internal distributions of the branches [84]. This conformational analysis is difficult to probe experimentally and has remarkable consequences as the change in shape or configuration assumed by these polymers will determine the polarity nature of both the surface and cavities as well as their interaction with the solvent or targets (Fig. 18.6).

The understanding of polymer behaviour is also fundamental to design better drug delivery systems. A good example was the study of the disassemble response of an amphiphilic dendritic polymer upon interaction with its targeted protein [87]. In these simulations, it was possible to assess where the targeting moiety needed to be attached to remain exposed at the surface and how this impacted the structure of polymers. This type of approach has also been commonly used to design other dendrimers, such as triazine dendrimers with bio-labile linkages to ensure the proper release of paclitaxel [88].

Box 18.2 List of Molecular Docking and Molecular Dynamics Simulation Software

AutodockVina, version 1.1.12, http://vina.scripps.edu/index.html. Available for Linux, Windows and macOS. First launch 2009. It requires graphical user interface Autodock Tools. Vina is available for free. Autodock tools is

(continued)

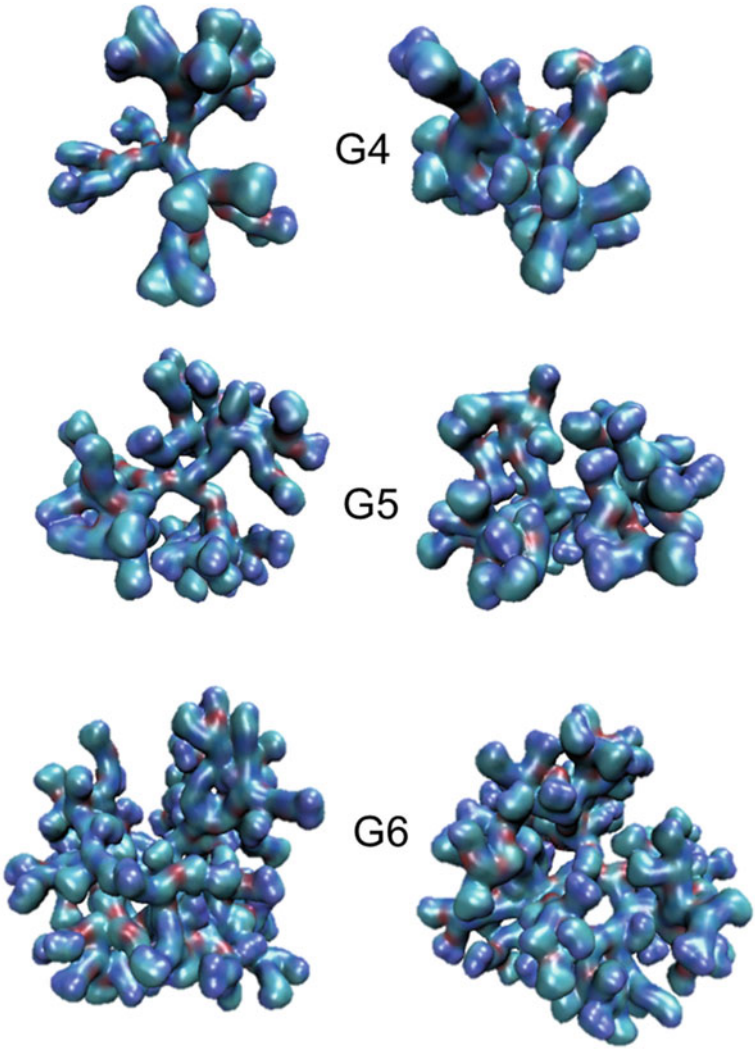


Fig. 18.6 Surface representation of example conformations of different generations' dendrimer (cyan—neutral surface; blue—positively charged; and red—negatively charged surfaces)

Box 18.2 (continued)
free for academic use. VegaZZ can also be used as GUI. Vina is a molecular docking software that can be used for evaluation of binding affinities of ligands to various biopolymers and polymers [70, 89, 90].

(continued)

Box 18.2 (continued)

NAMD, version 2.1,4, https://www.ks.uiuc.edu/Research/namd/. Available for Linux, Windows and macOS. First launch in 1995. Available for free for all. NAMD software is used for molecular dynamics simulations of different systems [91]. The default force field with this software is CHARMM, and missing parameters for polymers can be automatically assigned using CHARMM-GUI [72]. CHARMM-GUI, http://charmm-gui.org/?doc=input/polymer, has a polymer builder functionality. This program is used in conjunction with VMD, https://www.ks.uiuc.edu/Research/vmd/, a molecular graphics software [92]. VMD and NAMD that provide tools for setting up and running simulations and analysis of results stored as trajectories of various systems [6, 93].

Desmond, version 2020-1, https://www.deshawresearch.com/. Available for Linux, integrated with Maestro graphical user interface. First launch 2005. Free for academics [94]. The default force field with the free version is OPLS-2005, but other force fields can be used too. Desmond can be used for simulation of proteins, membranes, polymers and drugs [95, 96].

Packmol, version 20.2, https://github.com/m3g/packmol. First launch 2008. Available as a source code and requires Fortran 90 compiler. Available for free. Packmol can be used for building multi-component molecular system [97]. Specified number of each component, including proteins, drugs, polymers and salts, is positioned randomly within a box of specified size [96].

Another area where molecular dynamics of dendritic polymers has provided invaluable knowledge on their possible toxicity mechanism has been simulations with lipid bilayers. Simulations have shown that some dendrimers at certain generations can interact with lipid membranes and create holes in them. Such predictive power can therefore be leveraged to a significant advantage for the design of novel polymers devoid of these mechanisms of toxicity. The detailed analysis of molecular dynamic simulations is also important to predict the in vivo performance of dendritic polymers including the strength of host-guest complexation as well as the availability of targeting groups or exposure of labile molecules.

18.4 In Silico Evaluation of Intermolecular Interactions

18.4.1 Molecular Docking

In silico evaluation of intermolecular interactions provides information on the energetics associated with the driving forces that lead to the host-guest complexation. This can provide free energy predictions of the complexation process and binding affinity of drugs to target molecules. For this, molecular docking has been a popular choice used to study host-guest interactions in various systems, initially

between proteins and small molecules [98], then between small molecules and polymers, dendrimers and other targets (e.g. DNA and solid interfaces). The binding affinity is estimated using scoring functions, using parametrized energy terms that describe nonbonded interactions between the host and guest molecules. The integral part of the docking process is the generation of different orientations of the guest with respect to the host (poses), and for each predicted pose, a scoring value is calculated. Such evaluation provides information on how well the guest molecule fits into the binding pocket of the host and how strong favourable interactions are. This allows not only relative ranking of the poses for one host and guest pair but also the comparison to relative binding affinities between other hosts and guest. Commonly, several top poses are analysed to gain insights on the intermolecular interactions that potentially can be formed. A recent summary of progress in the development of the molecular docking can be found elsewhere [99].

Although most of the docking software has been developed with the focus on the protein-ligand interactions by parametrizing scoring functions against experimental data available for protein complexes, these are being used to study interactions between polymers and small molecules as well. Such use does not generally result in reliable results when the chemical nature of monomers is considerably different than amino acids. Furthermore, studying binding affinity is a difficult task due to the absence of the solvent effects on conformational changes, including entropy and enthalpy compensations, and therefore it makes rationalization of polymer-ligand association a complex task. Moreover, the flexible nature of polymers and the fact that multiple molecules can bind simultaneously influence the conformation of the dendrimer where cavities are filled to accommodate these molecules. Therefore, it is not surprising that docking studies can show contradictory results in predicting the affinity and loading capacity of a dendrimer [100].

Nevertheless, docking has been commonly used for the evaluation of polymer-drug interactions. Particularly, molecular docking may be used as a preliminary filtering tool for the selection of a polymer for development of the drug delivery system. A set of 3D structures of different chitosan-based polymers generated by molecular dynamics simulations was used to evaluate their affinity to curcumin. The polymer with the highest affinity for the ligand was chosen as a material for the delivery of novel nanoparticles [101]. Similarly, molecular docking was used to evaluate the affinity of curcumin and insulin for two different polymers, chitin and chitosan. The polymers were represented as chains of ten monomer units and were energy minimized before setting them as targets for docking. Most favourable complexes were used as a starting point for molecular dynamics simulations using Desmond software to evaluate dynamics nature of intermolecular interactions. The predicted properties of different complexes were in a good agreement with in vitro observations [102].

Additionally, docking can be used to inform polymer modifications and optimize cavity size to accommodate a selected drug and their intermolecular interactions. Cavities can also be measured using a spherical probe to roll around the van der Waals spheres of the atoms. However, these do not account for the favourable interactions that may occur between polymers and drugs that result in an adaptation

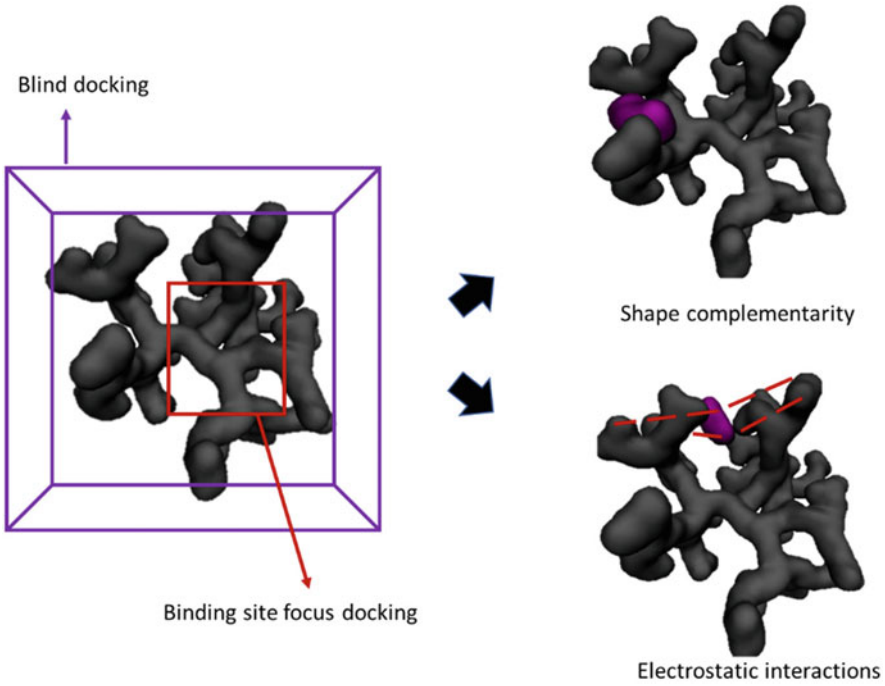


Fig. 18.7 Strategies for docking drugs against a dendrimer as an example system. Blind docking uses the whole surface of the polymer as a target for the small molecule binding (purple box), while the preselected cavity can be chosen for focused docking (red box). The docking results provide information on the shape complementarity as well as electrostatic interactions

of the polymer to the small molecule. Another interesting application of docking was to verify if the linked agonist to a PAMAM G3 could simultaneously occupy both subunits of the receptor [103].

Molecular docking can also provide some information between shape complementarity of the ligand and the polymer cavities. Thus, docking can be useful to (1) determine how many drug molecules can bind to polymers and (2) identify putative binding site locations on large molecules. For the former, blind docking, the surface of the whole polymers is explored as a target for ligand binding, while the latter is usually focused on the interaction of the ligand with a specified binding surface or cavity (Fig. 18.7). As polymers are generally flexible, it has to be considered that their surface and cavity geometry and properties change with time. One way to overcome issues related to target flexibility is the use of several conformations extracted from the molecular dynamics trajectory of the polymer to be considered as a target for docking.

Despite docking alone not always being enough to study binding affinity and loading, it can provide valuable information that can be used in combination with molecular dynamics simulations or to set up molecular dynamics simulations of the polymer-drug complexes.

18.4.2 Molecular Dynamics Simulations

Molecular dynamics simulations of polymers have been a versatile approach to study host-guest interaction mechanisms. Even though there are still some limitations associated with the simulation of a large number of atoms including the solvent molecules, simulations, in general, have been shown to correlate with in vitro observations. In particular, simulations with a ligand free to diffuse in solution have contributed to a better understanding of the host-guest interactions governing the absorption capabilities and loading capacity. The interactions between polymers and small molecules are highly dependent on many other interactions with water molecules, ions and other polymer copies, all of which may be synergistic or competitive and play a crucial role in both encapsulation and release [104].

For example, all-atom molecular dynamics simulations of PAMAM dendrimers with ibuprofen placed in its proximity have shown that ibuprofen could penetrate closer to the dendrimer core than water [105]. This is critical as ibuprofen is poorly soluble in water. Moreover, these simulations showed that depending on the pH ibuprofen was evenly distributed throughout the dendrimers or formed clusters which were consistent with experimental solubility data. The self-association of drug molecules at the surface is expected to significantly alter the surface properties of the polymers and thus their solubility. The pH of the solution has a significant effect on the dendrimer charges, as shown by molecular dynamic simulations of PAMAM generation 4 with different terminal groups (NH_2 and acetyl) that resulted in pH-dependent conformational changes. These, along with the effects of terminal groups, determine the loadings of drug molecules and their distribution within the dendrimer interior (Fig. 18.8) [106].

Similar results were obtained for phenanthrene molecules that penetrated into the interior of PAMAM of generation 5 at neutral pH (open structure), whereas at high pH (compact structure) most ligands remained located near the surface. Simulations of PAMAM of generation 4 with methotrexate predicted loading capacity similar to that observed in vitro [100]. Under acidic conditions, 30 methotrexate molecules were able to associate with the dendrimer, whereas at basic pH only one molecule was involved in the interaction. In fact, simulations of different dendrimers with small molecules have been good at differentiating and estimating the number of molecules, their location and the forces involved in these interactions. The main uses of this approach were to characterize the mass distribution of both dendrimer, water and small molecules as well as the spatial arrangement of the small molecules with respect to the polymer.

In general drug encapsulation efficiency has been a characteristic of the effect of solvent, size and architecture of the polymers where for example a smaller degree of flexibility can lead to restriction of motion and thus lower degree of association. Among the forces driving complexation, the small molecules themselves can enhance their association in a cooperative effect [40].

Alternatively, drugs can also be attached at the surface of the polymer. In this particular case, simulations can provide information regarding their location in different environments and how this correlates with their activity. If the APIs need

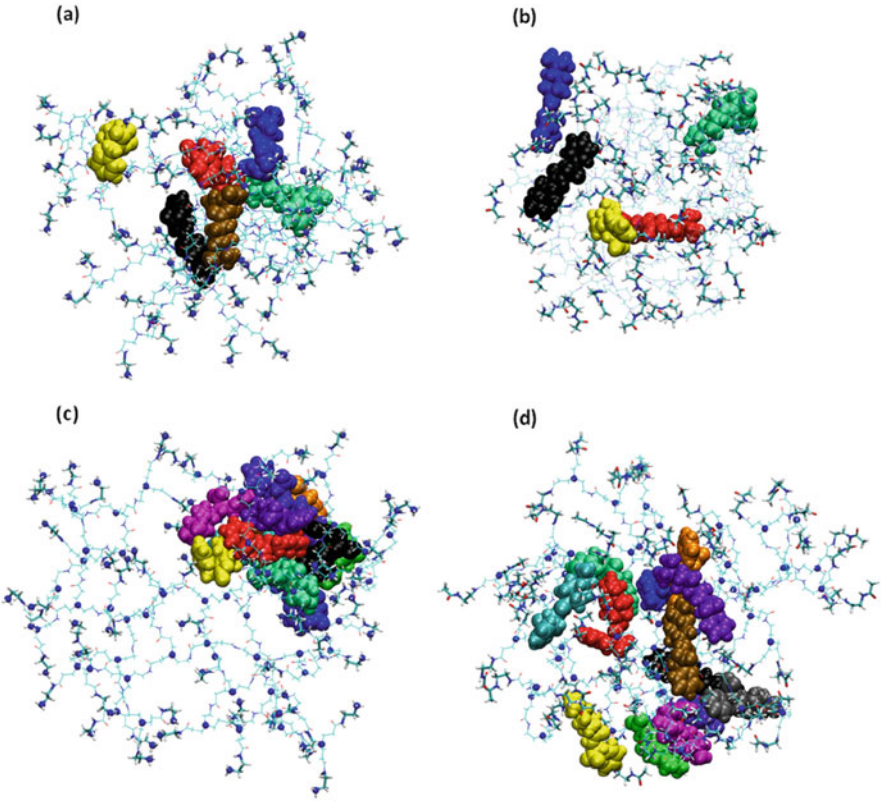


Fig. 18.8 Instantaneous snapshots of the complex after 30 ns simulation run (**a**) G4 PAMAM (NH_2)–Ntg (6 molecules) at neutral pH, (**b**) G4 PAMAM(Ac)–Ntg (5 molecules) at neutral pH, (**c**) G4 PAMAM(NH_2)–Ntg (12 molecules) at low pH, and (**d**) G4 PAMAM(NH_2)–Ntg (13 molecules) at low pH. Bound Ntg drug molecules are shown as space filled representation. Reproduced from [106]

to be protected from degradation, then promoting the API entry into the polymer interior may be important. On the other hand, if the small molecules have a function of a targeting moiety, its exposure to the solvent is of utmost importance. Finally, its presence at the surface will also change the properties of the polymer itself including its solubility and interaction with the lipid membrane and this can be studied using molecular dynamics simulations. In fact, coarse-grained simulations of PAMAM generation 4 linked with paclitaxel showed that the interaction mechanism and penetration were dependent on pH, distribution and concentration of the conjugate [107].

As previously mentioned, the combination of molecular docking followed by molecular dynamics simulations has also allowed a better insight into the energetic basis of molecular binding, loading and encapsulation of small organic molecules, nucleic acids and peptides by modified dendrimers at surface groups. Such an

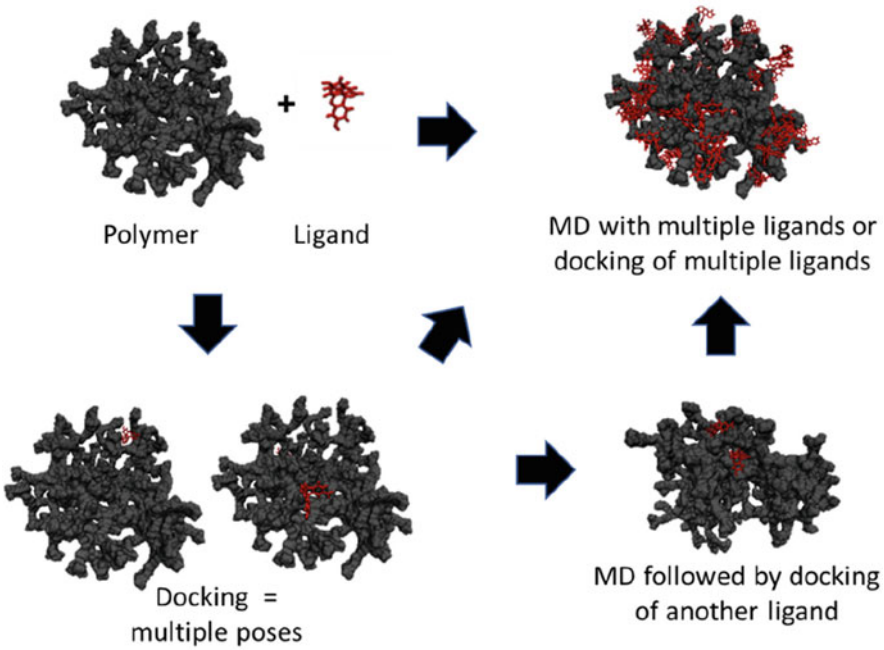


Fig. 18.9 Schematic representation of possible utilization of docking results for building molecular systems needed for molecular dynamics simulations

example was the use of docking with molecular dynamic simulations to study the interactions of PAMAM-(5-Fluorouracil) conjugates at different generations against two proteins and the possible mechanisms of binding interactions and compare it to the free molecule [108].

The preparation of the molecular system for molecular dynamics simulations is one of the key steps in the studies of drug delivery systems. In one of the approaches, molecular dynamics simulations studies could utilize the results of docking experiments, mainly to guide establishing the initial positions of the ligands with respect to the polymer. In this case, drugs can be placed in a predetermined position with favourable interactions within the interior of the dendrimer (Fig. 18.9). The final generated complex can be used as a starting structure for generating molecular systems for the simulation by addition of the solvent molecules, ions to minimize the overall charge of the system, buffer and salts to mimic relevant in vitro or in vivo environments. These additional components are randomly positioned around polymer drugs complex and completely solvated systems can be submitted to molecular dynamics simulation protocols that are dependent on the software used.

Another approach to developing starting points for the molecular dynamics simulation is based on the random positioning of all components within a box of a specified size. The ratio of the number of components should correspond to the desired molar ratio to mimic experimental conditions. The random position of the

components may lead to the system stuck in local minimum energy leading to poor sampling. To improve sampling of possible arrangements of small molecules in respect to the polymer, simulated heating simulations can be employed by heating and cooling of the molecular system followed by the selection of the final frame conformation for further molecular dynamics simulation. It is possible to repeat the simulated annealing protocol several times to obtain a different starting point for molecular dynamics simulations. Simulations should be submitted to a protocol relevant to the software used for molecular dynamics simulations.

Other considerations include the direct calculation of changes in free energy for some complexes. However, this is still a difficult task due to the intrinsic flexibility and natural change in conformation upon binding. Recently, Potential Mean Force (PMF) has been used to compute the energy between small molecules and dendrimer along a reaction path. This allows binding/unbinding calculation of the free energy profile to be accelerated. In this way, drug release can be simulated by PMF calculations as the free energy barrier that takes to unbind the small molecule. It was observed for example that the energy barrier of insoluble ligands from PAMAM of generation 5 was higher compared to soluble ligands. Similar to the evaluation of binding and identification of the various energetic contributions, these simulations were successfully applied between peptide dendrimers of different generations and 5-fluorouracil. Results showed weak interactions involving all parts via hydrogen bonding of the dendrimer and were highly affected by solvent effects [104]. Nevertheless, it was still able to differentiate between charged and neutral dendrimers in a similar manner to simulations of amine- and acetyl-terminated PAMAM generation 3 with nicotinic acid at different pH [109]. Moreover, this technique allows probing the behaviour at parts of the dendrimer that would be otherwise difficult to explore and give indications of how it can be improved.

In general, the calculations have been successfully able to reproduce experimental stoichiometry of host-guest binding and which forces lead to the binding affinity. Using Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) approach to study the binding of dexamethasone-21-phosphate to PAMAM generation 5, it was possible to study the favourable binding region of dendrimer and the complex driving forces behind the stabilization of host-guest complex that could not be attributed to a single type but rather an overall contribution [110]. Similarly, the use of Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) approach was able to differentiate the better encapsulation properties of phenyl butazone for PAMAM generation 3 in comparison to PPI generation 3 dendrimers [37].

In addition to small molecules, dendrimers can be used for the development of delivery systems for bigger therapeutics such as proteins and DNA. In particular, a variety of simulations that include docking, steered molecular dynamics and molecular dynamics have been carried to understand how dendrimers bind to nucleic acids. Generally, driven by electrostatic forces, DNA dendrimers can form strong complexes that result in wrapping distortions of the nucleic acids to neutralize their charge. This compact binding also results in aggregation of these complexes, but this can be desirable to protect the nucleic acids from degradation and increase their entry into cells.

18.5 Conclusions and Future Prospects

Overall, computational studies have generally supported the rationalization of experimental observations for the drug delivery systems by providing information on their size, surface and compositions. These features are shown to be key to design and develop novel drug carriers. The progress of understanding on how these polymers work together using molecular modelling tools will undoubtedly allow improving the efficacy of these small molecule carriers and support the development of novel drugs. Despite the current challenges, the growing interest in the use of polymers for biomedical applications will require further development of computational tools to guide the polymer design for specific applications. The development of drug delivery systems requires greater integration of available tools into an integrated platform that combines QbD approaches, molecular modelling and analysis of results of experiments. This platform should be community supported by building a database of up-to-date experimental information on drug delivery systems, polymer structures, properties and their intermolecular interactions, as well as their 3D models and predicted molecular properties. Such resources would promote the drug delivery systems development into a big data problem allowing the use of deep learning approaches in the design of novel polymers for specific applications.

18.6 Credible Online Resources for Further Reading

URL	Category of source	What to read
https://www.ddl.unimi.it/manual/pages/tu_build.htm	Institution	Tutorial for building small molecules including drug molecules
http://www.ks.uiuc.edu/Training/Tutorials/namd-index.html	Institution	Tutorials to learn how to set up a basic molecular dynamic simulation using NAMD and VMD as a graphical users interface
https://github.com/supramolecular-toolkit/stk	Institution	Software to build complex molecules
http://autodock.scripps.edu/faqs-help/tutorial	Institution	Tutorials to learn molecular docking using AutoDock
https://www.cambridgemedchemconsulting.com/resources/	Institution	Learn more about medicinal chemistry and drug discovery in general
http://cgbind.chem.ox.ac.uk/	Institution	Build molecular cages and add a substrate to its interior
https://www.mdanalysis.org/	Institution	Learn how to analyse a molecular dynamics simulation
https://openforcefield.org/	Institution	Learn more about an open initiative to develop force fields
http://cyclo-lib.mduse.com/	Institution	A database of molecular dynamics simulations of cyclodextrins
https://sites.google.com/site/dendrimerlibrary/home	Institution	Database containing 3D models of dendrimers

References

1. WHO (2010) Basic documents: forty-ninth edition (including amendments adopted up to 31 May 2019). World Health Organization, Geneva. https://apps.who.int/gb/bd/PDF/bd48/basic-documents-48th-edition-en.pdf. Accessed 5 Mar 2021
2. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R et al (2015) The Reactome pathway knowledgebase. Nucleic Acids Res 44(D1):D481–D4D7. https://doi.org/10.1093/nar/gkv1351
3. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30. https://doi.org/10.1093/nar/28.1.27
4. Chen JY, Pandey R, Nguyen TM (2017) HAPPI-2: a comprehensive and high-quality map of human annotated and predicted protein interactions. BMC Genomics 18(1):182. https://doi.org/10.1186/s12864-017-3512-1
5. Hao T, Wang Q, Zhao L, Wu D, Wang E, Sun J (2018) Analyzing of molecular networks for human diseases and drug discovery. Curr Top Med Chem 18(12):1007–1014. https://doi.org/10.2174/1568026618666180813143408
6. Barraza LF, Jiménez VA, Alderete JB (2017) Association of methotrexate with native and PEGylated PAMAM-G4 dendrimers: effect of the PEGylation degree on the drug-loading capacity and preferential binding sites. J Phys Chem B 121(1):4–12. https://doi.org/10.1021/acs.jpcb.6b08882
7. Badawy SIF, Narang AS, LaMarche KR, Subramanian GA, Varia SA, Lin J et al (2019) Chapter 20—Integrated application of quality-by-design principles to drug product and its control strategy development. In: Narang AS, Badawy SIF (eds) Handbook of pharmaceutical wet granulation. Academic, pp 665–702
8. Aleeva G, Zhuravleva M, Khafiz'yanova RK (2009) The role of excipients in determining the pharmaceutical and therapeutic properties of medicinal agents. Pharm Chem J 43(4):230–234. https://doi.org/10.1007/s11094-009-0265-4
9. Nema S, Washkuhn RJ, Brendel RJ (1997) Excipients and their use in injectable products. PDA J Pharm Sci Technol 51(4):166–171. https://doi.org/10.5731/pdajpst.2011.00634
10. Rayaprolu BM, Strawser JJ, Anyarambhatla G (2018) Excipients in parenteral formulations: selection considerations and effective utilization with small molecules and biologics. Drug Dev Ind Pharm 44(10):1565–1571. https://doi.org/10.1080/03639045.2018.1483392
11. Garg T, Rath G, Goyal AK (2015) Comprehensive review on additives of topical dosage forms for drug delivery. Drug Deliv 22(8):969–987. https://doi.org/10.3109/10717544.2013.879355
12. Baranowski P, Karolewicz B, Gajda M, Pluta J (2014) Ophthalmic drug dosage forms: characterisation and research methods. Sci World J 2014:861904. https://doi.org/10.1155/2014/861904
13. Harrison K (2007) 2—Introduction to polymeric drug delivery systems. In: Jenkins M (ed) Biomedical polymers. Woodhead Publishing, pp 33–56
14. Li C, Wang J, Wang Y, Gao H, Wei G, Huang Y et al (2019) Recent progress in drug delivery. Acta Pharm Sin B 9(6):1145–1162. https://doi.org/10.1016/j.apsb.2019.08.003
15. Wyss DF, Zartler ER (2020) NMR studies of protein–small molecule interactions for drug discovery. In: Renaud JP (ed) Structural biology in drug discovery. Wiley, pp 325–345
16. Nitsche C, Otting G (2018) NMR studies of ligand binding. Curr Opin Struct Biol 48:16–22. https://doi.org/10.1016/j.sbi.2017.09.001
17. Maveyraud L, Mourey L (2020) Protein X-ray crystallography and drug discovery. Molecules 25(5):1030. https://doi.org/10.3390/molecules25051030
18. Aitipamula S, Vangala VR (2017) X-ray crystallography and its role in understanding the physicochemical properties of pharmaceutical cocrystals. J Indian Inst Sci 97(2):227–243. https://doi.org/10.1007/s41745-017-0026-4
19. Maniruzzaman M, Snowden MJ, Bradely MS, Douroumis D (2015) Studies of intermolecular interactions in solid dispersions using advanced surface chemical analysis. RSC Adv 5(91):74212–74219. https://doi.org/10.1039/C5RA13176F

20. Mignolet A, Derenne A, Smolina M, Wood BR, Goormaghtigh E (2016) FTIR spectral signature of anticancer drugs. Can drug mode of action be identified? Biochim Biophys Acta Proteins Proteom 1864(1):85–101. https://doi.org/10.1016/j.bbapap.2015.08.010
21. Di Cola E, Grillo I, Ristori S (2016) Small angle X-ray and neutron scattering: powerful tools for studying the structure of drug-loaded liposomes. Pharmaceutics 8(2):10. https://doi.org/10.3390/pharmaceutics8020010
22. Moumbock AFA, Li J, Mishra P, Gao M, Günther S (2019) Current computational methods for predicting protein interactions of natural products. Comput Struct Biotechnol J 17:1367–1376. https://doi.org/10.1016/j.csbj.2019.08.008
23. Zhou L, Li Z, Yang J, Tian G, Liu F, Wen H et al (2019) Revealing drug-target interactions with computational models and algorithms. Molecules 24(9):1714. https://doi.org/10.3390/molecules24091714
24. Chen X, Yan CC, Zhang X, Zhang X, Dai F, Yin J et al (2015) Drug–target interaction prediction: databases, web servers and computational models. Brief Bioinform 17(4):696–712. https://doi.org/10.1093/bib/bbv066
25. Ahmad S, Johnston BF, Mackay SP, Schatzlein AG, Gellert P, Sengupta D et al (2010) In silico modelling of drug–polymer interactions for pharmaceutical formulations. J R Soc Interface 7(Suppl 4):S423–SS33. https://doi.org/10.1098/rsif.2010.0190.focus
26. Kulshrestha AS, Mahapatro A (2008) Polymers for biomedical applications. In: ACS Symp Ser. 977: ACS Publications. https://doi.org/10.1021/bk-2008-0977.ch001
27. Song R, Murphy M, Li C, Ting K, Soo C, Zheng Z (2018) Current development of biodegradable polymeric materials for biomedical applications. Drug Des Devel Ther 12:3117. https://doi.org/10.2147/DDDT.S165440
28. Sarwan T, Kumar P, Choonara YE, Pillay V (2020) Hybrid thermo-responsive polymer systems and their biomedical applications. Front Mater 7:73. https://doi.org/10.3389/fmats.2020.00073
29. Mane SR, Sathyan A, Shunmugam R (2020) Biomedical applications of pH-responsive amphiphilic polymer nanoassemblies. ACS Appl Nano Mater 3(3):2104–2117. https://doi.org/10.1021/acsanm.0c00410
30. Sepulveda AF, Borges R, Marchi J, de Araujo DR (2020) Biomedical applications of stimuli-responsive hydrogels. In: Green nanoparticles. Springer, pp 1–20. https://doi.org/10.1007/978-3-030-39246-8_1
31. Angelova N, Hunkeler D (1999) Rationalizing the design of polymeric biomaterials. Trends Biotechnol 17(10):409–421. https://doi.org/10.1016/S0167-7799(99)01356-6
32. Wang B, Wang S, Zhang Q, Deng Y, Li X, Peng L et al (2019) Recent advances in polymer-based drug delivery systems for local anesthetics. Acta Biomater 96:55–67. https://doi.org/10.1016/j.actbio.2019.05.044
33. Sabra S, Abdelmoneem M, Abdelwakil M, Mabrouk MT, Anwar D, Mohamed R et al (2017) Self-assembled nanocarriers based on amphiphilic natural polymers for anti-cancer drug delivery applications. Curr Pharm Des 23(35):5213–5229. https://doi.org/10.2174/1381612823666170526111029
34. Tong X, Pan W, Su T, Zhang M, Dong W, Qi X (2020) Recent advances in natural polymer-based drug delivery systems. React Funct Polym 148:104501. https://doi.org/10.1016/j.reactfunctpolym.2020.104501
35. Neuse EW (2008) Synthetic polymers as drug-delivery vehicles in medicine. Met Based Drugs 2008:469518. https://doi.org/10.1155/2008/469518
36. Pillai O, Panchagnula R (2001) Polymers in drug delivery. Curr Opin Chem Biol 5(4):447–451. https://doi.org/10.1016/S1367-5931(00)00227-1
37. Jain V, Maingi V, Maiti PK, Bharatam PV (2013) Molecular dynamics simulations of PPI dendrimer–drug complexes. Soft Matter 9(28):6482–6496. https://doi.org/10.1039/C3SM50434D

38. Moorefield CN, Newkome GR (2003) Unimolecular micelles: supramolecular use of dendritic constructs to create versatile molecular containers. C R Chim 6(8):715–724. https://doi.org/10.1016/j.crci.2003.04.002
39. Guo D, Muhammad N, Lou C, Shou D, Zhu Y (2019) Synthesis of dendrimer functionalized adsorbents for rapid removal of glyphosate from aqueous solution. New J Chem 43(1):121–129. https://doi.org/10.1039/C8NJ04433C
40. DeFever RS, Geitner NK, Bhattacharya P, Ding F, Ke PC, Sarupria S (2015) PAMAM dendrimers and graphene: materials for removing aromatic contaminants from water. Environ Sci Technol 49(7):4490–4497. https://doi.org/10.1021/es505518r
41. Chen S, Yu Q, Li L, Boozer CL, Homola J, Yee SS et al (2002) Detecting the adsorption of dye molecules in homogeneous poly(propylene imine) dendrimer monolayers by surface plasmon resonance sensor. J Am Chem Soc 124(13):3395–3401. https://doi.org/10.1021/ja0170201
42. Buczkowski A, Olesinski T, Zbicinska E, Urbaniak P, Palecz B (2015) Spectroscopic and calorimetric studies of formation of the supramolecular complexes of PAMAM G5-NH2 and G5-OH dendrimers with 5-fluorouracil in aqueous solution. Int J Pharm 490(1):102–111. https://doi.org/10.1016/j.ijpharm.2015.05.033
43. Kolhe P, Khandare J, Pillai O, Kannan S, Lieh-Lai M, Kannan R (2004) Hyperbranched polymer-drug conjugates with high drug payload for enhanced cellular delivery. Pharm Res 21(12):2185–2195. https://doi.org/10.1007/s11095-004-7670-x
44. da Silva Santos S, Igne Ferreira E, Giarolla J (2016) Dendrimer prodrugs. Molecules 21(6):686. https://doi.org/10.3390/molecules21060686
45. Liu C, Gao C, Yan D (2006) Synergistic supramolecular encapsulation of amphiphilic hyperbranched polymer to dyes. Macromolecules 39(23):8102–8111. https://doi.org/10.1021/ma0608065
46. Hinchliffe A (2008) Molecular modelling for beginners, 2nd edn. Wiley
47. Barata TS, Shaunak S, Teo I, Zloh M, Brocchini S (2011) Structural studies of biologically active glycosylated polyamidoamine (PAMAM) dendrimers. J Mol Model 17(8):2051–2060. https://doi.org/10.1007/s00894-010-0907-1
48. Maingi V, Jain V, Bharatam PV, Maiti PK (2012) Dendrimer building toolkit: model building and characterization of various dendrimer architectures. J Comput Chem 33(25):1997–2011. https://doi.org/10.1002/jcc.23018
49. Quintana A, Raczka E, Piehler L, Lee I, Myc A, Majoros I et al (2002) Design and function of a dendrimer-based therapeutic nanodevice targeted to tumor cells through the folate receptor. Pharm Res 19(9):1310–1316. https://doi.org/10.1023/a:1020398624602
50. Liu Y, Bryantsev VS, Diallo MS, Goddard Iii WA (2009) PAMAM dendrimers undergo pH responsive conformational changes without swelling. J Am Chem Soc 131(8):2798–2799. https://doi.org/10.1021/ja8100227
51. Javor S, Reymond J-L (2009) Molecular dynamics and docking studies of single site esterase peptide dendrimers. J Org Chem 74(10):3665–3674. https://doi.org/10.1021/jo802743c
52. Roberts BP, Krippner GY, Scanlon MJ, Chalmers DK (2009) Molecular dynamics of variegated polyamide dendrimers. Macromolecules 42(7):2784–2794. https://doi.org/10.1021/ma8021579
53. Genheden S, Reymer A, Saenz-Méndez P, Eriksson LA (2018) Chapter 1: Computational chemistry and molecular modelling basics. In: Computational tools for chemical biology. chemical biology. The Royal Society of Chemistry, pp 1–38
54. Vanommeslaeghe K, Guvench O, MacKerell AD Jr (2014) Molecular mechanics. Curr Pharm Des 20(20):3281–3292. https://doi.org/10.2174/13816128113199990600
55. Gartner TE, Jayaraman A (2019) Modeling and simulations of polymers: a roadmap. Macromolecules 52(3):755–786. https://doi.org/10.1021/acs.macromol.8b01836
56. Groom CR, Bruno IJ, Lightfoot MP, Ward SC (2016) The Cambridge structural database. Acta Crystallogr B Struct Sci Cryst Eng Mater 72(2):171–179. https://doi.org/10.1107/S2052520616003954

57. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K et al (2002) The protein data bank. Acta Crystallogr Sect D Biol Crystallogr 58(6):899–907. https://doi.org/10.1107/s0907444902003451
58. Krause S, Willighagen E, Steinbeck C (2000) JChemPaint-using the collaborative forces of the internet to develop a free editor for 2D chemical structures. Molecules 5(1):93–98. https://doi.org/10.3390/50100093
59. Karulin B, Kozhevnikov M (2011) Ketcher: web-based chemical structure editor. J Cheminform 3(1):1. https://doi.org/10.1186/1758-2946-3-S1-P3
60. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A et al (2016) PubChem substance and compound databases. Nucleic Acids Res 44(D1):D1202–D1D13. https://doi.org/10.1093/nar/gkv951
61. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D et al (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 36(Suppl 1):D901–D9D6. https://doi.org/10.1093/nar/gkm958
62. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 4(1):17. https://doi.org/10.1186/1758-2946-4-17
63. Jmol.Jmol, an open-source Java viewer for chemical structures in 3D (2013) Jmol web page: http://www.jmol.org/. Accessed 5 Mar 2021
64. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. J Cheminform 3(1):33. https://doi.org/10.1186/1758-2946-3-33
65. Pedretti A, Villa L, Vistoli G (2002) VEGA: a versatile program to convert, handle and visualize molecular structure on Windows-based PCs. J Mol Graph Model 21(1):47–49. https://doi.org/10.1016/s1093-3263(02)00123-7
66. Maleki R, Khoshoei A, Ghasemy E, Rashidi A (2020) Molecular insight into the smart functionalized TMC-Fullerene nanocarrier in the pH-responsive adsorption and release of anti-cancer drugs. J Mol Grap Mod 100:107660. https://doi.org/10.1016/j.jmgm.2020.107660
67. Moradi S, Taran M, Mohajeri P, Sadrjavadi K, Sarrami F, Karton A et al (2018) Study of dual encapsulation possibility of hydrophobic and hydrophilic drugs into a nanocarrier based on bio-polymer coated graphene oxide using density functional theory, molecular dynamics simulation and experimental methods. J Mol Liq 262:204–217. https://doi.org/10.1016/j.molliq.2018.04.089
68. Pedretti A, Villa L, Vistoli G (2004) VEGA—an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. J Comput Aided Mol Des 18(3):167–173. https://doi.org/10.1023/B:JCAM.0000035186.90683.f2
69. Barmpalexis P, Karagianni A, Kachrimanis K (2018) Molecular simulations for amorphous drug formulation: polymeric matrix properties relevant to hot-melt extrusion. Eur J Pharm Sci 119:259–267. https://doi.org/10.1016/j.ejps.2018.04.035
70. Palazi E, Karavas E, Barmpalexis P, Kostoglou M, Nanaki S, Christodoulou E et al (2018) Melt extrusion process for adjusting drug release of poorly water soluble drug felodipine using different polymer matrices. Eur J Pharm Sci 114:332–345. https://doi.org/10.1016/j.ejps.2018.01.004
71. Duke R, Giese T, Gohlke H, Goetz A (2016) AmberTools 20; University of California, San Francisco: San Francsico, CA. 2020. https://ambermd.org/AmberTools.php
72. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 29(11):1859–1865. https://doi.org/10.1002/jcc.20945
73. Abbott LJ, Hart KE, Colina CM (2013) Polymatic: a generalized simulated polymerization algorithm for amorphous polymers. Theor Chem Accounts 132(3):1334. https://doi.org/10.1007/s00214-013-1334-z
74. Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117(1):1–19. https://doi.org/10.1006/jcph.1995.1039

75. Charlmers D, Roberts B (2011) Silico—a perl molecular modeling toolkit. http://silico.sourceforge.net. Accessed 5 Mar 2021
76. Reymond J-L (2020) Dendrimer builder. https://dendrimerbuilder.gdb.tools/. Accessed 5 Mar 2021
77. Martinho N, Silva LC, Florindo HF, Brocchini S, Barata T, Zloh M (2017) Practical computational toolkits for dendrimers and dendrons structure design. J Comput Aided Mol Des 31(9):817–827. https://doi.org/10.1007/s10822-017-0041-6
78. Yu C, Ma L, Li S, Tan H, Zhou Y, Yan D (2016) HBP Builder: a tool to generate hyperbranched polymers and hyperbranched multi-arm copolymers for coarse-grained and fully atomistic molecular simulations. Sci Rep 6:26264. https://doi.org/10.1038/srep26264
79. Borodin O, Smith GD, Douglas R (2003) Force field development and MD simulations of poly (ethylene oxide)/LiBF4 polymer electrolytes. J Phys Chem B 107(28):6824–6837. https://doi.org/10.1021/jp027539z
80. Martinho N, Florindo H, Silva L, Brocchini S, Zloh M, Barata T (2014) Molecular modeling to study dendrimers for biomedical applications. Molecules 19(12):20424–20467. https://doi.org/10.3390/molecules191220424
81. Zloh M, Barata TS (2020) An update on the use of molecular modeling in dendrimers design for biomedical applications: are we using its full potential? Expert Opin Drug Discov 15(9):1015–1024. https://doi.org/10.1080/17460441.2020.1769597
82. Dodda LS, Cabeza de Vaca I, Tirado-Rives J, Jorgensen WL (2017) LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. Nucleic Acids Res 45(W1): W331–W3W6. https://doi.org/10.1093/nar/gkx312
83. Kavyani S, Amjad-Iranagh S, Dadvar M, Modarress H (2016) Hybrid dendrimers of PPI (core)–PAMAM(shell): a molecular dynamics simulation study. J Phys Chem B 120(36):9564–9575. https://doi.org/10.1021/acs.jpcb.6b05142
84. Martinho N, Silva LC, Florindo HF, Brocchini S, Zloh M, Barata TS (2017) Rational design of novel, fluorescent, tagged glutamic acid dendrimers with different terminal groups and in silico analysis of their properties. Int J Nanomed 12:7053. https://doi.org/10.2147/IJN.S135475
85. García EJ, Hasse H (2019) Studying equilibria of polymers in solution by direct molecular dynamics simulations: poly(N-isopropylacrylamide) in water as a test case. Eur Phys J Spec Top 227(14):1547–1558. https://doi.org/10.1140/epjst/e2018-800171-y
86. Rukmani SJ, Kupgan G, Anstine DM, Colina CM (2019) A molecular dynamics study of water-soluble polymers: analysis of force fields from atomistic simulations. Mol Simul 45(4–5):310–321. https://doi.org/10.1080/08927022.2018.1531401
87. Amado Torres D, Garzoni M, Subrahmanyam AV, Pavan GM, Thayumanavan S (2014) Protein-triggered supramolecular disassembly: insights based on variations in ligand location in amphiphilic dendrons. J Am Chem Soc 136(14):5385–5399. https://doi.org/10.1021/ja500634u
88. Lim J, Lo S-T, Hill S, Pavan GM, Sun X, Simanek EE (2012) Antitumor activity and molecular dynamics simulations of paclitaxel-laden triazine dendrimers. Mol Pharm 9(3):404–412. https://doi.org/10.1021/mp2005017
89. Nguyen NT, Nguyen TH, Pham TNH, Huy NT, Bay MV, Pham MQ et al (2019) Autodock vina adopts more accurate binding poses but autodock4 forms better binding affinity. J Chem Inf Model 60(1):204–211. https://doi.org/10.1021/acs.jcim.9b00778
90. Hameed HA, Khan S, Shahid M, Ullah R, Bari A, Ali SS et al (2020) Engineering of naproxen loaded polymer hybrid enteric microspheres for modified release tablets: development, characterization, in silico modelling and in vivo evaluation. Drug Des Devel Ther 14:27–41. https://doi.org/10.2147/DDDT.S232111
91. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E et al (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802. https://doi.org/10.1002/jcc.20289
92. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(1):33–38. https://doi.org/10.1016/0263-7855(96)00018-5

93. Khezri A, Karimi A, Yazdian F, Jokar M, Mofradnia SR, Rashedi H et al (2018) Molecular dynamic of curcumin/chitosan interaction using a computational molecular approach: emphasis on biofilm reduction. Int J Biol Macromol 114:972–978. https://doi.org/10.1016/j.ijbiomac.2018.03.100
94. Bowers KJ, Chow DE, Xu H, Dror RO, Eastwood MP, Gregersen BA et al (eds) (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. In: SC'06: proceedings of the 2006 ACM/IEEE conference on supercomputing, IEEE. https://doi.org/10.1109/SC.2006.54
95. Avila-Salas F, Pereira A, Rojas MA, Saavedra-Torres M, Montecinos R, Bonardd S et al (2017) An experimental and theoretical comparative study of the entrapment and release of dexamethasone from micellar and vesicular aggregates of PAMAM-PCL dendrimers. Eur Polym J 93:507–520. https://doi.org/10.1016/j.eurpolymj.2017.06.023
96. Avila-Salas F, Rodriguez Nuñez YA, Marican A, Castro RI, Villaseñor J, Santos LS et al (2018) Rational development of a novel hydrogel as a pH-sensitive controlled release system for nifedipine. Polymers 10(7):806. https://doi.org/10.3390/polym10070806
97. Martínez L, Andrade R, Birgin EG, Martínez JM (2009) PACKMOL: a package for building initial configurations for molecular dynamics simulations. J Comput Chem 30(13):2157–2164. https://doi.org/10.1002/jcc.21224
98. Meng X-Y, Zhang H-X, Mezei M, Cui M (2011) Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des 7(2):146–157. https://doi.org/10.2174/157340911795677602
99. Fan J, Fu A, Zhang L (2019) Progress in molecular docking. Quant Biol 7(2):83–89. https://doi.org/10.1007/s40484-019-0172-y
100. Bello M, Rodríguez-Fonseca RA (2019) Complexation of methotrexate via ligand diffusion molecular dynamic simulations under neutral, basic, and acidic conditions. J Mol Graph Mod 93:107443. https://doi.org/10.1016/j.jmgm.2019.107443
101. Yadav P, Bandyopadhyay A, Chakraborty A, Sarkar K (2018) Enhancement of anticancer activity and drug delivery of chitosan-curcumin nanoparticle via molecular docking and simulation analysis. Carbohydr Polym 182:188–198. https://doi.org/10.1016/j.carbpol.2017.10.102
102. Dhanasekaran S, Rameshthangam P, Venkatesan S, Singh SK, Vijayan SR (2018) In vitro and in silico studies of chitin and chitosan based nanocarriers for curcumin and insulin delivery. J Pol Environ 26(10):4095–4113. https://doi.org/10.1007/s10924-018-1282-8
103. Ivanov AA, Jacobson KA (2008) Molecular modeling of a PAMAM-CGS21680 dendrimer bound to an A2A adenosine receptor homodimer. Bioorg Med Chem Lett 18(15):4312–4315. https://doi.org/10.1016/j.bmcl.2008.06.087
104. De Luca S, Seal P, Ouyang D, Parekh HS, Kannam SK, Smith SC (2016) Dynamical interactions of 5-fluorouracil drug with dendritic peptide vectors: the impact of dendrimer generation, charge, counterions, and structured water. J Phys Chem B 120(25):5732–5743. https://doi.org/10.1021/acs.jpcb.6b00533
105. Tanis I, Karatasos K (2009) Association of a weakly acidic anti-inflammatory drug (ibuprofen) with a poly(amidoamine) dendrimer as studied by molecular dynamics simulations. J Phys Chem B 113(31):10984–10993. https://doi.org/10.1021/jp9039176
106. Jain V, Maiti PK, Bharatam PV (2016) Atomic level insights into realistic molecular models of dendrimer-drug complexes through MD simulations. J Chem Phys 145(12):124902. https://doi.org/10.1063/1.4962582

107. He X, Lin M, Lu T, Qu Z, Xu F (2015) Molecular analysis of interactions between a PAMAM dendrimer-paclitaxel conjugate and a biomembrane. Phys Chem Chem Phys 17(44):29507–29517. https://doi.org/10.1039/c5cp02242h
108. Rengaraj A, Subbiah B, Haldorai Y, Yesudhas D, Yun HJ, Kwon S et al (2017) PAMAM/5-fluorouracil drug conjugate for targeting E6 and E7 oncoproteins in cervical cancer: a combined experimental/in silico approach. RSC Adv 7(9):5046–5054. https://doi.org/10.1039/C6RA26511A
109. Caballero J, Poblete H, Navarro C, Alzate-Morales JH (2013) Association of nicotinic acid with a poly(amidoamine) dendrimer studied by molecular dynamics simulations. J Mol Graph Mod 39:71–78. https://doi.org/10.1016/j.jmgm.2012.11.003
110. Vergara-Jaque A, Comer J, Monsalve L, González-Nilo FD, Sandoval C (2013) Computationally efficient methodology for atomic-level characterization of dendrimer–drug complexes: a comparison of amine- and acetyl-terminated PAMAM. J Phys Chem B 117(22):6801–6813. https://doi.org/10.1021/jp4000363

Computers in Pharmaceutical Analysis

19

Mukesh Maithani, Viney Chawla, and Pooja A. Chawla

Contents

Abstract

Automation of analytical techniques becomes a necessity both in research and pharmaceuticals manufacturing especially when a large number of analyses have to be carried out as rationally and reliably as possible. With the evolution of technology, there is a simultaneous increase in the levels of quality, safety, and

M. Maithani
Multidisciplinary Research Unit, Veer Chandra Singh Garhwali Government Institute of Medical Science and Research, Srinagar, India

V. Chawla
Department of Pharmaceutics, University Institute of Pharmaceutical Sciences and Research, Baba Farid University of Health Sciences, Faridkot, Punjab, India

P. A. Chawla (✉)
Department of Pharmaceutical Chemistry and Analysis, ISF College of Pharmacy, Moga, India

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_19

reliability. Additionally, the revolution of the use of computers in pharmaceutical analysis provided by the development of flow analysis concepts and process analysis strategies offer a link between modern instrumentation and social or technological problems. Automation of computer in analysis as well as analytical methodology provides an opportunity to the pharmaceutical industry in its attempts to use risk management and try scientifically designed manufacturing processes. Such attempts often lead to a better understanding of the product and thereby promote quality assurance. With an aim to reduce the increasing costs for product development and to overcome the regulatory hurdles toward invention and creativity, the Federal regulatory agency of the USA, that is, FDA, is promoting automation, and computers are an integral part of achieving this objective. This chapter summarizes current state of automation and computer-aided analysis, computer-assisted analysis of drug delivery systems, different chromatographic data systems, use of computer-/software-assisted analytical method development, role of analytical QbD as well as its application in analytical process, and importance of nanoparticle tracking analysis.

Keywords

Computer-assisted analysis · Analytical QbD · Method development · Nanoparticle tracking analysis · Dissolution software

Chapter Objectives

- Understanding the importance of computers in pharmaceutical analysis
- Understanding of chromatographic data systems
- Understanding of the software used in quality by design
- Understanding of the software in dissolution testing
- Insight into nanoparticles tracking analysis

19.1 Introduction

In today's world, computers are an integral part of everyone's life, and it is difficult to imagine a world without human-computer interface. No wonder medical profession and its associated activities are also seeing an upsurge in the use of computers. In its journey from lab to the market, a drug molecule has its brush up with computer at different stages that include search by chemists in structural databases, use of computer algorithms for synthesis of molecules, prediction of interaction of these molecules with target proteins, prediction of pharmacokinetics of a candidate drug, prediction of side effects, computer-based simulations in choosing a potential molecule, analyzing results of preclinical studies, and use of bioinformatics tools aimed at discovery of drugs. Computers can also be used for statistical modeling in pharmaceutical R&D and for assessment of involved risks in various stages. It is also possible to collect and analyze the data generated through clinical studies and present it in a scientific manner using computers. Production and distribution records

can also be computerized. Hospital setup has also seen the increased use of computers in the recent past, and telemedicine is an upcoming field in the era of COVID-19. It is possible for a computer to easily retrieve the enormous data stored in it thereby saving time, money, and number of personnel employed [1–3]. Whereas the traditional methods of finding new drugs relied upon trial and error, the novel techniques often based on some level of computer interface have successfully curtailed the expenses and time invested in bringing out a novel drug candidate to market with added advantages and tailored properties [4, 5].

Computers in pharmaceutical analysis are chiefly used for data storage and processing of data to and search for information. With advancement, computer technology and software used in the instruments employed in pharmaceutical industry carry out the analysis, data storage, and processing followed by its interpretation leading to an overall improvement of the process. These latest software carry out all necessary process required for analytical purpose. These programs possess in-built libraries which prove to be useful in search of data related to different chemical entities. If a mass spectrum of any unknown chemical is obtained, then command can be given to the program for detection of similar mass spectrum from the in-built library that ultimately helps to detect unknown chemical entity [6, 7]. Commercial software packages are available for interpretation of IR spectra. Wiley's Sadtler database, a well-known database, contains 26,400 IR spectra of commercial and standard substances. Furthermore identity, potency, and purity are very important characteristics of the drug substance to obtain quality products with the anticipated biological response. Identity, potency, and purity of the products are tested by different sophisticated analytical instruments in pharmaceutical industries. Some of the commonly used instruments in pharmaceutical analysis are microscopes, spectroscopic instruments (UV visible spectrophotometer, IR, mass spectrometer, NMR, fluorimeter, etc.), chromatographic instruments and hyphenated or tandem instruments (high-performance liquid chromatography, high-performance thin layer chromatography, gas chromatography, LC-MS, GC-MS, LC-MS-MS, etc.), and much more. In addition, process analytical technology (PAT) is a system for analyzing, controlling, and designing production by timely measurements of the quality of critical raw materials and performance characteristics, the materials being processed, and their processes, with the aim of achieving the quality of the final product. Application of PAT in the pharmaceutical companies is to analyze and capture the observed data. It is also used to establish identity of the product and confirm its quality so that poor quality products do not reach the market. The difficulties in the development of formulations of newly discovered drugs mainly stem from the fact that research and development is a time-consuming and expensive process in pharmaceutical industries. Artificial intelligence, statistically organized experiments, and other computational technologies can be used to improve both the process development and the formulation. The concurrent development and research of pharmaceutical products and processes allows for the implementation of quality in accordance with the design concept supported by regulators all over the world. The computer-aided applications in the field of pharmaceutical technology cover the chemometric methods with emphasis of their application in process control,

interpretation and fundamentals of experimental design application, neural computing (decision trees, artificial neural networks and fuzzy logic, self-organizing maps, genetic algorithms, and evolutionary computing), and computer-aided biopharmaceutical characterization along with application of computational fluid dynamics in pharmaceutical technology. All of these procedures are critical tools for ensuring the innate quality of pharmaceutical processes and products from the beginning of their development to the selection of the finest [8, 9].

Recently, automation of analytical instruments (such as auto samplers, control of instruments using microprocessors, databases, and data treatment) has accelerated the advancement of analytical chemistry techniques in the analysis of pharmaceuticals. Automation has long been used to solve a variety of industrial and laboratory issues. In the pharmaceutical analysis, these automatic analytical systems are typically based on chromatography, a technique used for separating and analyzing various types of compounds [10]. Similar methods, such as fast sampling instruments, pneumatic tube for conveying samples to a lab, and atomic emission spectrometers, are currently in use in metallurgical operations. Automatic systems for analysis in the pharmaceutical laboratory frequently involve continuous flow (especially in the drug industry) or flow injection analysis. Advances in software and computing technology have also resulted in the creation of ever-growing data collections that must be organized, stored, and evaluated. As a result, analytical chemists dealing with these large data sets must be knowledgeable in statistics and chemometrics, as well as sampling, apparatus, and interpretation. The usage of laboratory robots is also on the rise, prompting fears that robotics may suffocate this historically hands-on discipline and result in job losses. However, because analytical chemists' cognitive skills are required to render the data useful, the vast data created by current procedures and instruments really represent an opportunity for them [11].

19.2 Automation and Computer-Aided Analysis

Pharmaceutical analysis owes its beginnings to classical analytical techniques. It also emphasizes the need to have a sound hold on advanced analytical instruments; scientist/analyst must have good grasp over the basics of classical techniques. Today the horizons of analytical studies have expanded at a tremendous pace. The expectation levels from the analytical scientist have been raised in terms of ultralow detection limits, speed of analysis, and commitment on the reported results due to increasing safety concerns, development of new pharmaceutical products, and consumer demands. Pharmaceutical companies demand immediate results on products manufactured in process flows or assembly lines. These judgments have far-reaching implications on commercial use and quality of manufactured goods. Application of these decisions is possible using online process monitoring and advanced offline or online instrumental analysis techniques. Classical techniques that undoubtedly deliver high accuracy results do not cope up when it comes to high-throughput analysis [12, 13]. In addition to the speed of analysis, the automated

instrumental methods also guarantee the absence of personal prejudices and errors. Instrumental methods of analysis have exploited several basic characteristics of matter like absorbance of light, electrical conductivity, viscosity, phase transitions, presence of isotopes, molecular vibrations, selective adsorption of certain species, etc. to develop well-defined analytical streams such as spectroscopy, chromatography, thermal analysis, and electrometric methods. The limitless options available have helped the scientist/analyst gain insight into arrangement of atoms and molecules in solids (both crystalline and amorphous), geological specimens, and biological samples. Instrumental methods have also helped significantly to understand and develop nano-materials, bio-fuels, semiconductor devices, and clinical diagnostics. Rapid advances in optics, automation, and microelectronics have further added to advancements in instrumental techniques and their field of application. However, once again it is stressed that the scientist/analyst must constantly update their analytical skills, update the basic concepts, and avoid the pitfalls of becoming an ordinary technical operator of analytical tools [14].

The use of various control methods for operating applications and equipment with minimal human interaction is referred to as automation. The usage of automation in pharmaceutical analysis enables to perform many tests by analytical instruments with least human intervention. Automation in instruments/equipment has advantages that laboratories can easily handle high-throughput samples with minimum involvement of manpower. Additionally, automation reduces the danger of findings' fluctuation and errors that might occur during manual analysis. The incorporation of integrated computer software and hardware into analytical methodologies has made pharmaceutical analysis easier, as it allows for autonomous data processing and control [15]. The use of an automated analyzer has numerous advantages, including reduced workload, reduced time spent per sample analysis, increased number of tests completed in less time, use of minute amounts of sample, reduced chances of human errors, high accuracy and reproducibility, and controlled and minimized use of chemicals and reagents, resulting in cost savings per test as well as environmental friendliness. Figure 19.1 depicts many types of laboratory computers and their diverse applications. In a subject that is evolving and improving at a fast pace in terms of digitalization of instruments and computers, it is extremely difficult to look into the future and foresee the kind of applications that may be on the horizon. Some tendencies, on the other hand, appear to be very visible right now. One of them is that digital devices and computers are becoming much less expensive, faster, more powerful, and more reliable in general. At the same time, the cost of producing software does not go down. As a result, sophisticated hardware devices are more likely to handle experimental control and simple data handling and data processing functions than programmable general-purpose computers [16, 17]. Furthermore, future development will most likely shift the focus of online digital computer from passive to active purposes. The two fundamental distinguishing qualities of laboratory computer applications are described by the terms "passive" and "active." A passive application means that there isn't much experimenting going on. An active application means that the computer was involved in some way in the experiment control. There are several subcategories within the broad definition of

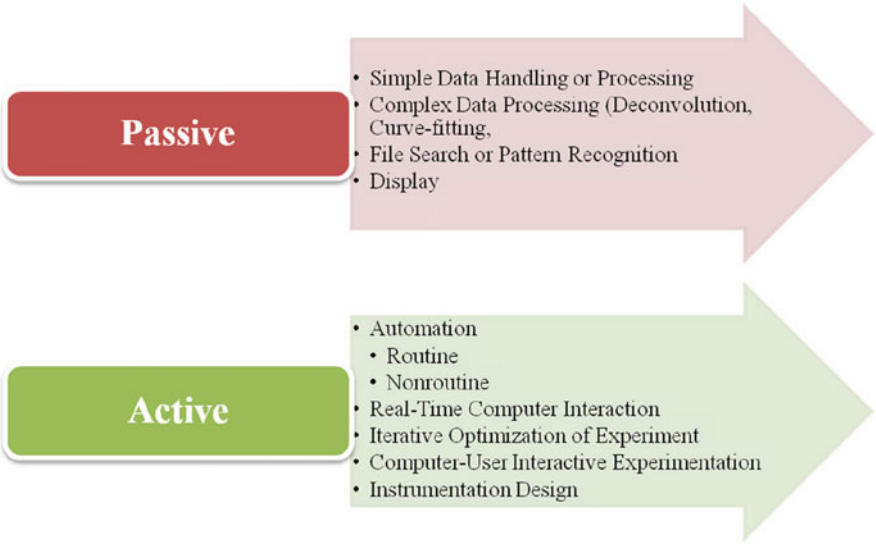


Fig. 19.1 Different types of laboratory computers with their applications

passive application. In general, these applications depict circumstances in which the digital computer is used primarily as a data-logging device. These applications merely make use of the high-speed data capture, large data storage capacity, and quick data handling or data processing capabilities of an online computer. Furthermore, depending on their online data-logging properties, these applications are categorized into four subcategories: Simple data handling and processing, complicated data processing, file search or pattern recognition, and raw or linked data display are all examples. In addition, some applications will include components from more than one of the subcategories indicated above. Applications that fall into the active category are those in which the digital computer is used to regulate and/or manipulate experimental parameters in some way. This control can be preprogrammed (basic automation), or it can be executed in real time with computer decision-making. Automation (routine and non-routine), interaction with computer on real-time basis, iterative optimization of experiments, experimentation based on interaction between computer and user, and designs of instruments are some of the subcategories [18, 19].

19.3 Computer-Assisted Analysis of Drug Delivery Systems

The progress of pharmaceutical analysis has been substantially expedited by the increasing sophistication of computers and software, particularly in the study of medicines from biological tissues such as plasma. Rapid modeling of processes and rapid solution of complicated techniques is now allowed by computer software

programs very efficiently. Tedious and problematic calculations are simplified by computers, and it also gives you more time to experiment with new data analysis and modeling techniques. In addition, computer software is frequently used for the graphical representation of data, development of experimental designs, data manipulation, projection or prediction of drug action, statistical data treatment, and pharmacokinetic model simulation [20, 21]. Particularly, interesting results in the treatment of cardiovascular disease, autoimmune syndromes, cancer, and infectious diseases are yielded by recombinant monoclonal antibodies (mAb) [22]. These big macromolecules have a number of benefits for patients, including excellent efficacy, specificity, a broad therapeutic range, and few side effects. Furthermore, the patents on the oldest approved biopharmaceuticals expire, generic biologics or "biosimilars" are entering the market. Regulatory organizations such as the FDA and EMA demand a full drug characterization owing to the fact that variations in the manufacturing of biopharmaceuticals are often larger than that of conventional pharmaceuticals. A variety of orthogonal approaches are frequently required to completely describe these complex samples. Spectrophotometry, electrophoresis, chromatography, and mass spectrometry are the most essential techniques [23, 24]. Another complex analytical project is proteome analysis, sometimes known as "proteomics." The proteome is dynamic in nature, as it is shaped by the genome, the environment, and the history of the cell. Analytical criteria include high sensitivity, high resolution, and high throughput. As a result, liquid chromatography is regarded as a critical tool in proteomic research. Proteins can be digested or separated before being separated and identified using mass spectrometry (MS), or they can be separated first and then fragmented in the MS. These two proteomics methods are referred to as "bottom-up" and "top-down," respectively [25, 26].

Computer technology has improved significantly since Hounsfield introduced the first CT (computed tomography) scanner prototype in 1969, allowing for details in reconstructed pictures with pixel sizes down to the nanometer range. Microcomputer tomography is now employed in a variety of fields, including biology, palaeontology, geology, material science, the automobile industry, and medicine, among others. During Quality by Design investigations, pharmaceutical corporations have used this technique to analyze and improve their production processes and medicine development (QbD). Micro CT, in fact, allows for a direct evaluation of the impact of physical parameters specified during pharmaceutical formulation (cracks, surface defects, or internal failures). Medications are currently developed and delivered directly in medical devices via physical, mechanical, or thermal methods. There could be defects or technical concerns with these mechanisms. CT systems are useful for analyzing and studying the internal and external structures of objects without having to touch them when making non-destructive 3D measurements [27–29].

Micro CT is a sort of 3D imaging that enables for virtual reconstruction of objects with micrometric pixel sizes. A sample rotates with relation to a stationary X-ray source and detector, or the X-ray source and detector move around the sample, generating X-rays toward the sample. The X-rays are attenuated as they pass through the sample due to the length traveled in the absorbent material (thickness), as well as

the material's composition and density (i.e., the attenuation coefficient). Through analyte, the detector monitors the intensity of the transmitted X-rays. With the image of the X-ray shadow, variable signal intensity levels offer a gray scale that depicts the sample and its attributes [30]. The micro-CT device's monitor displays this high-resolution X-ray shadow image in real time. X-ray transmission images are obtained at numerous discrete angular increments, similar to a map of the relative atomic density of analyte. Finally, the rebuilt 2D radiographs are stacked and gathered at the same time. Then it provides the calculation of complete 3D map of the sample for further processing. Simultaneously, the image analysis program examines 2D cross sections and 3D objects. In most investigations, the first stage is 3D visualization. In the majority of investigations, 3D visualization is used as a preliminary stage. Semi-transparent representations can provide a quick 3D glimpse of the edges of the work item [31, 32].

There are currently a variety of computers to choose from. PCs can be used alone or in conjunction with local networks (LANs) that share various application software packages and are widely used for written reports, documentation, and archiving. An operating system (OS) is a collection of programs that manages resources and allows algorithms (well-defined rules or methods for solving a problem in a finite number of steps) to be executed on a computer. Commonly used operating systems include Windows, macOS, and, more recently, LINUX. PC users can access remote information from numerous websites that offer a variety of free or commercial programs used in pharmaceutical analysis via dial-up, ISDN, DSL, Cable TV internet connection, satellite internet connection, or wireless internet connection. A computer package, often known as software, is a set of instructions written in computer language. This program is required for the computer to function. The software's language must be supported by the computer operating system. In the past, computer users had to be proficient in computer programming and have at least one computer language under their belt, such as Pascal, C, or Basic, python, java, java script, C++, ruby/ruby on rails, and so on. Because of the numerous commercial and noncommercial pharmacokinetic apps and spreadsheets, such as Excel, just basic computer programming is necessary for several pharmacokinetics applications. Pharmacokinetic software consists of computer programs that are meant to compute and solve pharmacokinetic problems quickly. Many computer programs suit the needs of users, such as fitting drug concentration–time data to a series of pharmacokinetic models and selecting the one that statistically best describes the data, fitting data into a pharmacokinetic or pharmacodynamic model defined by the user, and many others [33–36].

19.4 Chromatographic Data Systems

Chromatography data systems (CDS) are data-collection technologies that interface and/or integrate with chromatography equipment in a laboratory. CDS is a method for obtaining data from chromatography equipment, processing that data, storing it in a database, and interacting with other laboratory informatics systems for exporting

and importing files and data. It automates data collecting, letting analysts to run jobs without manually managing databases, decreasing human error, and ensuring that data are accurate. Furthermore, a sophisticated CDS ensures that laboratory operations and practices comply with the stringent regulatory criteria that govern data integrity. Currently, analytical laboratories must be able to provide clues to the data relevant to the audit in compliance with the increasing regulations. Attempting to manually gather and report data without a robust system, on the other hand, can soon become error-prone and difficult, necessitating the usage of advanced software tools in modern labs [37, 38]. In the 1970s, the first attempts to use electronics to automate the examination of analytical data were made. To output the results, these analysis instruments used microprocessor-based integrators and a printer plotter. Because those early systems had little memory, they couldn't store more than one chromatograph at a time. In the mid-1970s, this became less of a concern for large labs with larger resources, as costlier centralized data systems were implemented, allowing more data exchange and storage possibilities [39]. In 1980, entrepreneur and Hewlett-Packard prodigy Dave Nelson identified the prospects of using a personal computer in analytical chemistry, and he formed Nelson Analytical Inc. with Harmon Brown. They created the first CDS personal computer program in the same year, followed by Turbochrom, the first CDS system for Microsoft Windows. This innovation quickly spread beyond analytical chemistry labs to other fields like environmental, forensic, and pharmaceutical sciences. The use of two chromatography minicomputers, Perkin-Elmer's Laboratory Information Management Systems (LIMS) 2000 CLAS and Hewlett-3350 Packard's LAS Lab Automation System, increased simultaneously due to their ability to acquire and process data from up to 32 or more chromatographs at the same time. In the 1990s, networks of CDSs, particularly those installed on personal computers, became possible because of the increasingly affordable high-performance PCs and tighter networking standards. CDSs could build up a methodology and analytically run information in the late 1990s, control some instruments, receive injection data, process the data in different ways, save the data, and send it to other systems like a LIMS. CDS functions were being upgraded by 2008, thanks to advances in liquid chromatographs (LC) and gas chromatographs (GC). Faster data collection, increased separation, and higher resolutions and sensitivities were all possibilities with the new high-speed LC and GC devices. While these next-generation devices would provide additional processing power to chromatography labs, it would also need suppliers to improve CDSs, particularly the sample rates of analog-to-digital converters. Some vendors predicted that data capture sample rates of 100–300 Hz would be required to keep up with the new wave of faster chromatographic instruments at the time. Due to the expansion of pharmaceutical corporations in the Far East, South America, and Latin America, additional concerns about scalability and remote access were becoming more important [40, 41].

19.4.1 Data Integrity

Data integrity is very crucial in pharmaceutical industries because regulatory authorities are increasingly scrutinizing the authentication and accuracy of laboratory results through audit. Pharmaceutical companies want to ensure complete traceability of every analyst from beginning to end in order to comply with regulatory audits, which necessitated software that not only logged any changes done by the analyst/user but also allowed for quick retrieval of relevant audit trails for any workflow. This was made possible by the integrated CDS, which also allows the analyst/user to visualize the changes and roll back to previous versions if necessary [42]. These advancements have made it easier to demonstrate regulatory compliance by streamlining report preparation and speeding up the data evaluation process. Analysts' ability to monitor their own audit trails is critical for modern analytical labs. As part of the review process, it is now important to study audit records; thus analysts and users must become experts in order to comprehend these audit data. As a result, it is critical that CDS has the capacity to make this procedure as straightforward as feasible [43–45].

19.4.2 Workflow Automation

Every month, analytical labs perform hundreds of injections, necessitating the confidence that procedures are running well. The previous CDS nature has led to a considerable waste of time in the manual configuration of the repeated sequences. With integrated CDS analysts, on the other hand, can obtain high-quality data right away, with user-defined procedures like re-injections and sample dilutions being performed automatically. In past workflows, data processing consumed a significant amount of time, but the pharmaceutical industry is now more efficient thanks to the employment of new computer algorithms. All calculations during pharmaceutical analysis are performed within the software, resulting in consolidated data and the elimination of the need for separate packages. This has resulted in more reliable data and a significant decrease in human error [46–48].

19.5 Computer-/Software-Assisted Analytical Method Development

Method development (MD) in chromatography is the search for the best chromatographic operating conditions (type of mobile and stationary phase, ionic strength, temperature, pH, gradient steepness, and so on) for separating a mixture of samples into its parts. Because of the considerable risk of peak overlap and the sensitive dependence of the retention time on the applied chromatographic parameters, MD is a time-consuming and exhausting procedure (takes up to many weeks). MD requires the analyst's knowledge and experience, as well as a number of trial-and-error methods [49, 50]. Computer-assisted MD has the potential to

drastically speed up the MD process if appropriate retention models are available. Peptides and proteins, which are more sensitive to changes in solvent strength than small molecules and hence require very lengthy and shallow gradients, may benefit from the time savings (on-off retention mechanism). As a result, creating long gradients without retention modeling is time-consuming and tedious. The optimization and scoping phases of MD are the most important (scouting and screening). During the optimization phase, very accurate retention modeling is required to identify the appropriate separation settings (errors as low as 1–2% in retention time). However, depending on the column stationary phase and chromatographic technology, prediction errors of up to 10% may be accepted during the scoping phase. Quantitative structure retention relationships (QSRR), which could replace first exploratory trials with a prediction based merely on the molecule's structure, are promising for speeding up the scoping phase. Different software are commercialized such as empirical models (Chromgenius/ACD, ChromSword) and employing empirical models (Osiris, DryLab). There are many more software designed for various chromatographic instruments, for example GC-SOS by The 4S Co; Malcom by Schlumberger; Chemstation, EZchrom Elite, and Open Lab CDS by Agilent; Chromeleon and Atlas CDS by Thermo Scientific; Empower and MassLynx by Waters; OpenChrome by Philip Wenig; Clarity by DataApex; Mass Frontier by High Chem; Lab Solutions by Shimadzu; Power Chrom by eDAQ; ChromStar 7 and PrepCon 5 by SCPA; Unicorn 7.0 by GE Life Scienses; Trilution LC by Gilson; ChromNAV by Jasco; CHROMuLAN by PiKRON; ChromaTOF by LECO; Peaksel by Elsci; etc.

Furthermore, ChromSwordAuto delivers an advanced MD tool and completed 40–50 injections overnight. The analyst/scientist can select the best and optimum chromatograms in the morning for further analysis of pharmaceuticals. In addition, "Drylab" is the best widespread and successful software in the present era. Other optimization algorithms that have been reported in the literature were created in-house utilizing Microsoft Excel or Matlab software [51–55]. Compliance can boost the robustness of a chromatographic method while also speeding up the MD process with a logical approach to using these software programs. Weiyong and Henrik developed a strategic MD by QbD involving three steps: screening of multiple column/mobile phase, use of various organic modifiers for further mobile phase optimization, and use of Plackett-Burman (PB) design for optimization of multiple-factor method for assay/impurity testing of pharmaceuticals using LC. This procedure significantly decreased the number of runs required for method development, and after the method had achieved a reasonable separation, PB experimental designs were used to fine-tune it. In a QbD with design of experiments (DoE) method, Peter et al. employed UPLC technology in tandem with specialized software to separate vancomycin impurities [56]. Since Lawrence's demonstration, the method has been shown to be far superior to gradient HPLC, with the routine use of a sub-2-pm ACQUITY UPLC Column with QbD methodology being able to separate 26 contaminants in vancomycin, compared to only 13 impurities using the old method [57]. Finally, using computer-assisted MD, the solvent consumption was minimized by decreasing the number of tests necessary, which can be considered a

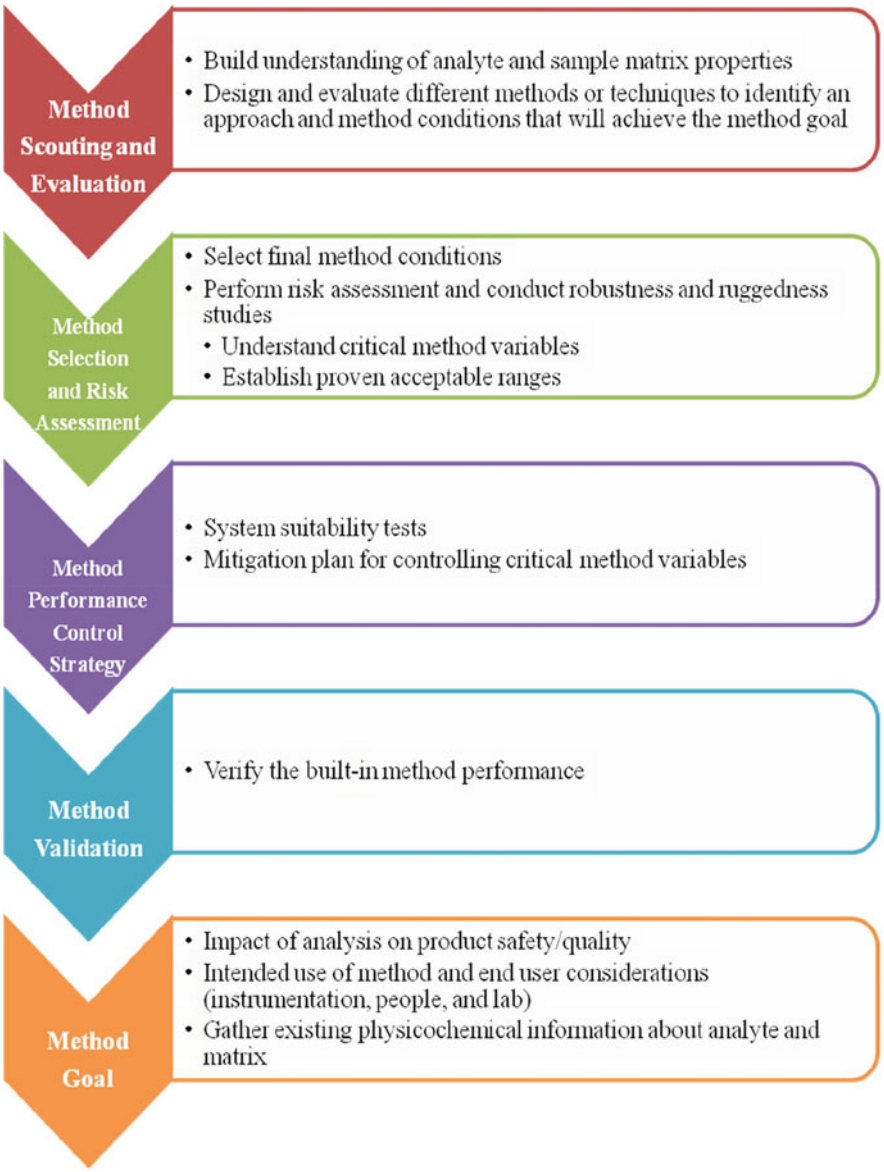


Fig. 19.2 Common approach of method development strategy (MDS) for QbD analytical methods

green chromatography technique [58, 59]. Figure 19.2 depicts a common method development strategy (MDS) approach for QbD analytical methods. Although, to avoid any misunderstanding, method validation is viewed as a separate step following the MDS process [60].

19.6 Analytical QbD

Dr. JM Juran, a quality pioneer, was the first to invent the concept of QbD. Dr. Juran thought that quality must be designed/planned into a product, and most of the critical situations and issues related to quality can be attributed to this approach. According to Woodcock, a high-quality drug substance is free of contamination and offers the consumer with the therapeutic benefit promised on the product label. For manufacturing and regulation, the US Food and Drug Administration (USFDA) strongly encourages a risk-based strategy based on QbD principles in drug product development. The USFDA recognizes that simply increasing testing would not result in an improvement in quality, hence the focus on QbD. Furthermore, QbD stresses product and process understanding through design efforts and precise scientific concepts in order to reach stated goals and objectives [61–64]. The applications of QbD in pharmaceutical analysis are given in Table 19.1 [65]. The QbD is a regulatory strategy that encourages systematic product development, beginning with established objectives and stressing the product, as well as the control and understanding of the process, all based on strong science and quality risk management. This paradigm's response surface methodology (RSM) and DoE modules are critical for generating a design space for formulation input variables. RSM and DoEs can be used to fit linear or quadratic functions to response surfaces, yielding mathematical models of output responses as input variables. The expected performance of the formulation will also be determined by the optimal selection of these input variables. The International Conference on Harmonization (ICH), often known

Table 19.1 Applications of QbD in pharmaceutical analysis (Reproduced from [65])

S. no	Implementation (stage wise)	Details
1	Measurement of the target	Decide what to measure and when to measure it. Develop measurement criteria based on product's quality target product profiles (QTPP) and CQAs and define analytical target profile (ATP)
2	Choose a technique	Choose the right analytical technique for the ATP test you want to do. Define the performance requirements for the method
3	Assessment of the risk	Examine the risks of method input variables, sample variance, and environmental factors. Failure mode effective analysis (FMEA) and other risk assessment tools can be employed
4	Validation and development of methods	To comprehend method robustness and ruggedness, DoE examines probable multivariate interactions and defines the method operable design region (MODR)
5	Methodology of control	To meet ATP, define the control space and system suitability; meet method performance criteria
6	Continual development	Analysts proactively identify and rectify out-of-trend method performance by monitoring method performance that remains compliant with ATP criteria. Update with new analytical and process technology

Fig. 19.3 The QbD approach in analytical process

as Harmonization for Better Health, has produced four guidelines: ICH Q8(R2), Q9, Q10, and Q11, as well as implementation documents. In such guidelines and documents ICH provides a common outline of QbD application for manufacturing and development of drug substances as well drug products. The key fundamentals of the QbD framework were defined and described particularly in ICH Q8 (R2) (Revised in 2009) [66].

Since its approval by the USFDA, QbD has become a major paradigm in the pharmaceutical industry. The concept of QbD can be stretched to analytical methods. The current state of QbD is to develop analytical methods to illustrate wider issues and the general state of the sector, before focusing the discussion on the analytical aspects of QbD. MDS approach is used for applications of analytical QbD to different types of techniques including genotoxic impurity analysis, spectroscopic determinations of identity and color, HPLC method development, tablet dissolution testing, and water content determinations. The QbD approach in analytical process is presented in Fig. 19.3 [67–70].

19.6.1 Applications of QbD Analytical Process

To obtain optimal method performance, QbD is utilized to describe the method employed in a complete examination and analysis of the alternative technique. The

chosen approach is next evaluated for risks using risk assessment methods, and its ruggedness and robustness are tested. These studies aid in gaining a better understanding of method performance, making adjustments, and developing a risk management control approach so that the method works as intended following validation [71, 72]. Some crucial QbD applications in pharmaceutical analysis especially chromatographic techniques are as follows:

19.6.1.1 In Validation and Development of Ultrahigh Performance Liquid Chromatography (UHPLC)

The purpose of developing a reversed-phase UHPLC method was to analyze total benzalkonium chloride present as a preservative in different compositions. The various HPLC settings, such as mobile phases and gradient parameter, were optimized using a QbD technique with the help of Fusion AE®. An ACE Excel 2 C18-AR column was employed. The aqueous mobile phase was ammonium phosphate buffer (pH 3.3; 10 mM), the organic mobile phase was methanol/acetonitrile (85/15, v/v), and UV detections were performed at 214 nm. The method was found to be precise and accurate, with linearity between 0.025 and 0.075 mg/mL and a recovery of 99–103% at linear concentration ranges, and it took less than 2 min to separate the compound's primary homologues (C12 and C14) [73–75].

19.6.1.2 In Hydrophilic Interaction Liquid Chromatography (HILIC) Development

HILIC analytical method using gradient elution with analytical quality by design (AQbD) approach was employed and validated for olanzapine and its seven related substances by Tumpa et al. [76]. The critical process parameters (CPPs) were defined as the duration of the linear gradient, the initial content of the aqueous phase, and the temperature, while the critical quality attributes were recognized as the separation criterion of critical pairs of substances (CQAs). Models constructed using Rechtschaffen design were used to describe the interdependence between CQAs and CPPs, and the optimal/best conditions were chosen utilizing the design space gained. QbD ideas were used for the development and manufacturing of a lyophilized protein product by Jameel and Khan [77], who concluded that BD comprises quality target profiles, risk evaluations, screening and optimization studies, scale-up studies, and controlled techniques.

19.6.1.3 In Chromatography Column Screening

Connie et al. evaluated reputed companies of column manufacturers which were very commonly used and manufactured. The analytical columns, as well as the assessment criteria and experimental design details, have been described [78]. A systematic approach against predetermined performance criteria of QbD method development was adopted for the assessment of seven different RP-HPLC columns [79]. The generated data is of great help to scientists/analysts as it would help them to develop rugged and robust methods that could be used in QbD. Liu et al. [80] advocated that QbD approaches have been also used in the selection of the best UPLC columns.

19.6.1.4 In the Development of HPLC Methods for Drug Products/Substances

Monks et al. introduced a new approach to HPLC method development based on QbD principles, using a column database and computer modeling software to evaluate four essential factors such as aqueous eluent pH, stationary phase, gradient time, and temperature [81]. Aydar [82] used components of QbD to develop and optimize an analytical method for protamine sulfate, which was followed by a study of interaction and quadratic effects on responses utilizing response surface methodology in BB design. The four peptide peaks of protamine sulfate were observed with a tailing factor of 1.02–1.45 and a peak resolution of 1.99–3.61 when the QbD method was applied to optimal conditions.

19.6.1.5 Response Surface Methodology (RSM)

In modern well-furnished laboratories, the computer-controlled intricate instruments have wiped out the common problems of chemists, where the practical experiments generally used to cause the major limitation in obtaining relevant information. RSM is a tool that was introduced in the early 1950s by Box and Wilson. It is a commonly used mathematical and statistical method for modeling and analyzing a process in which the response of interest is influenced by a number of variables, with the goal of optimizing the response. It is a crucial engineering tool used in the development of processes through experimental design. It helps the pharmaceutical industry in a variety of ways, including product design, process development, quality, production engineering, and operations, as well as formulation design and intermediate and final products. The optimization procedure involves systematic formulation design to minimize the number of trials, analysis of the response surfaces in order to facilitate understanding of the effect of independent factors, obtaining appropriate formulation with the achievement of target goals and also with the acceptable component region as process control condition in practical preparation [82].

19.6.1.6 RSM's Fundamentals and Theoretical Aspects

The most crucial part of RSM is the DoE. The goal of the DoE is to choose the most appropriate points where the response should be thoroughly analyzed. The process's mathematical model is primarily based on DoE. As a result, the experiment design chosen has a significant impact on the validity of the response surface construction. The RSM's benefits can be stated as detecting the interaction between independent variables, mathematically modeling the system, and saving time and money by reducing the number of trials [82].

19.6.1.7 Application of RSM

Because of its advantages over traditional one-variable-at-a-time optimization, such as the ability to generate large amounts of data from a small number of experiments and the ability to evaluate the interaction effect between the variables on the response, RSM is widely used in the optimization of analytical procedures today. To use this method for experimental optimization, it is customary to first choose an experimental design, then fit a suitable mathematical function, evaluate the quality

and accuracy of the fitted model, and finally make predictions based on the experimental data. Because of its inefficiency with greater numbers of variables, three-level factorial designs are rarely used, and their application has been limited to the optimization of two variables. In contrast, the Box–Behnken and Doehlert designs present more efficient matrices and have experienced an increase in the number of published articles in recent years. Until now, the application of desirability as a function in multiple response optimization has been restricted to the chromatographic field, its related techniques, and electrochemical approaches. Its concepts, on the other hand, can be applied to the construction of procedures utilizing various analytical techniques that necessitate the simultaneous search for optimal circumstances for a set of replies. As an alternative to traditional modeling, an adaptive learning technique combining neural networks and experimental design can be utilized to describe a dependent connection. This technique has demonstrated to be more accurate in data learning and prediction than the standard RSM. Furthermore, designing and optimizing pharmaceutical formulation with an adequate analysis of dissolution rate in a shorter time period and with a minimal number of trials is an essential issue encountered in the development of a sustained release dosage form. A computer-based polynomial equation and an Artificial Neutral Network (ANN) are commonly employed to solve this problem [83].

19.7 Software for Dissolution/Drug Release Analysis

Testing of drug dissolution is used during drug development for many dosage forms including capsules and tablets in both the early and late stages. Initially dissolution test in drug development helps scientists/researchers to find the best formulation of the oral dosage form to adapt to the in vitro behavior. Later, dissolution profiles may be used to establish an in vivo/in vitro correlation (IVIVC) which can reduce the requirement of costly bioequivalence studies. In the final stage of drug development, the testing of dissolution is used for quality control, that is, to test batch-to-batch consistency, stability, and to detect manufacturing defects which might lead to an entire lot rejection [84, 85]. Dissolution tests are listed as in vitro performance test in the United State Pharmacopeia (USP), and it is mandatory to perform as per guidelines of USFDA. In the development of pharmaceutical formulation, early determination of dissolution investigation and behavior on solubility for the influence of formulation factors is very crucial and important. Using in silico techniques in the stages of drug discovery to estimate drug dissolution can eliminate many trial-and-error experiments and save time and money. Analysis of dissolution data for characterization and quantification of drug release from a pharmaceutical dosage form is performed by either mathematical models or statistically comparing dissolution profiles. Most commercial statistical software in pharmaceutical R&D are used for evaluation of the pharmacokinetic parameters. Researchers have designed numerous programs (DDDPlus, KinetDS, DDSolver, etc.) to eliminate errors of calculation which reduce calculation time. It allows dissolution data model up to 40 built-in dissolution patterns. In addition, these software allow a similarity analysis

to be performed using well-established profile comparison approaches. These programs provide an efficient report of data analysis to summarize the data of dissolution analysis [86–90].

DDSolver is a menu-driven add-in tool for Microsoft Excel calculations in Visual Basic programs. Calculating with Excel has a number of advantages over alternative tools; the most notable of which is its ease of use. Excel software is familiar to most analysts and scientists due to its considerable versatility and accessibility. When you start Excel after installing the program, a drop-down option called DDSolver will appear in the menu bar. Users can select any form from the drop-down menu and insert dissolution data by choosing the appropriate cell range in the spreadsheet. For each module, DDSolver provides a number of customizable options, including chart output, maximum number and convergence of iterations of the non-linear algorithm optimization, number of decimal places in the calculated results, initial parameter estimates, and report generation in Microsoft Word [91, 92].

KinetDS software is primarily used to describe the cumulative dissolution curve using a set of equations or a simple curve fitting equation. Alternative curves generated from various data sources could be assessed if the dependent-variable range is between 0 and 100. The Higuchi, Hixson-Crowell, zero-to-third-order kinetic models, Korsmeyer-Peppas, Weibull (two- and three-parameter), Michaelis-Menten, and Hill equations were selected as the most popular empirical and mechanistic models used to report drug dissolution curves [93]. For each accessible dataset, the values of dissolution efficiency (DE) and mean dissolution time (MDT) are automatically generated at the end of the report file. (1) Coefficient of determination (r2) and empirical coefficient of determination (r2 emp); (2) Akaike information criterion (AIC); (3) Bayesian information criterion or Schwarz criterion (BIC); and (4) root-mean-squared error (RMSE). The standard error (SE) of correlation coefficients, which is also presented in percent of the relevant parameter value, is an additional diagnostic tool for the linear regression function (relative SE of the correlation coefficient). Linear and nonlinear regression are used to set model parameters (NLR). So far, all of the models in KinetDS 3.0 have been found to be linearizable. As a result, NLR and linear regression may be used to fit them. In Version 4.0, only nonlinear regression-specific models will be introduced. Because of their mathematical simplicity, kinetic models from zero to third order rarely require NLR. The zero-order model is specifically excluded from the NLR procedure because of its linearity. As a result, the user selects the option "employ nonlinear regression (NLR)" without regard for the models, opting instead for the projected model performance enhancement. The final report will incorporate these enhancements if the software notices them. Furthermore, the software will always be mentioned in the final report for reference in all sets of coefficients derived from the NLR and linear regression. In this case, checking the "Use nonlinear regression (NLR)" option as a means of improving the model is often recommended. This option does, however, come at a cost in terms of processing resources. Because NLR is based on an iterative technique (basic method), its convergence can take a long time in some situations [94–96].

DDDPlus is used in many pharmaceutical industries to study dissolution pattern and disintegration of active ingredients and dosage form. Formulation scientists utilize this software as an advanced computer program to simulate in vitro dissolution and disintegration of formulations, active pharmaceutical ingredients (API), and excipients under a variety of experimental settings. DDDPlus provides for analytical and formulation analyst/scientist in vitro mechanistic dissolution software. Scientist can simulate the in vitro dissolution of formulation excipients and API with the help of DDDPlus under various experimental conditions in few seconds. A single calibration experiment is usually required in the development of new API, after which DDDPlus predicts how changes in formulation or experimental parameters would alter rate dissolution. Because this software provides exact disintegration and dissolving rate information, there is no need to rely on the traditional 'cut and try' method for finalizing formulation design. DDDPlus also lets you choose from five mathematical models and five dosage forms to show how a single chemical dissolves [97]. The mathematical models used for in vitro dissolution simulations describe the effect of the following parameters on dissolution:

- The formulation ingredients' physicochemical qualities, such as solubility, pKa, density, and diffusion coefficient.
- Manufacturing properties for immediate-release dosage formulations.
- Particle size distribution for each ingredient in the recipe.
- Each experimental apparatus has a different flow pattern and fluid velocity.
- Interactions between the active substance and the formulation excipients.
- Solubility and dissolution/precipitation are pH-dependent in microclimates.
- Micelle dissolution in media is aided by the presence of surfactants.
- Automatically calculates fluid velocity based on instrument speed and equipment type.

19.8 Nanoparticle Tracking Analysis (NTA)

The combination of the human eye's outstanding ability to differentiate particle outlines, a touch screen, and a stylus pen may give advantages over both manual image analysis (accuracy) and image analysis programs (detailed size and shape information). The diffusion coefficient is calculated for each individual particle by tracking their path. It is a high-resolution analytic technique that employs diffusion and Brownian motion (or the intensity of light scattering) to detect tiny differences between two portions or populations. The NTA image analysis tool collects a series of video files of the particles under observation (typically 30–60 s in length), then recognizes and tracks the center of each particle frame by frame. The average distance traveled by each particle in the x and y directions is then calculated using image analysis software. The particle diffusion coefficient may be computed using this value, and the hydrodynamic diameter can be derived using the Stokes-Einstein equation. Furthermore, the NTA image analysis program tracks and sizes a large number of particles at the same time. The results can be viewed as frequency size

distributions or exported to a spreadsheet. These size distributions are coupled with additional scatter intensity data to create 3D graphs that highlight distinct populations. Each sample's video clip is recorded and saved for later examination. The "movie" is taken and stored at 30 frames per second in NTA. It is then presented using the NTA software to illustrate each particle's individual "track." The user then chooses the final output, which can range from a simple particle distribution curve to a complicated three-dimensional display that allows various materials with very similar particle sizes to be distinguished easily. NanoTrackerTM 2 software is one of the best examples of NTA image analysis software. It is a research-grade inverted optical microscope-based optical tweezers platform for sensitive manipulation, force, and tracking investigations. The NanoTrackerTM 2 allows the user to control, manipulate, and examine samples in real time with nanoscale precision and resolution, trapping and tracking particles from several meters down to 30 nm. The NanoTrackerTM technology allows for precise and repeatable studies of particle/cell interactions. The device provides exact information about single molecule mechanics and can be used to determine mechanical properties on single molecules such as adhesion, elasticity, and stiffness [98].

The growing popularity of protein-based treatments necessitates the development of simple and reliable analytical methods for characterizing protein compositions. Particle sizes in the 1–10 mm range have been measured using analytical approaches such as flow cytometry, electrical sensing zone, and light obscuration. Furthermore, because of the potential for adverse effects on drug safety and efficacy, increased emphasis has been made on expanding particle characterization to smaller particle sizes in the submicrometer range [99]. Dynamic light scattering (DLS) is a popular method for particle sizing in this size range due to its high throughput and ease of usage. Only monodisperse samples within a concentration range for a specific particle size are correct using this method. NTA, a new method for measuring and characterizing particle-size distributions in the submicrometer range, has been shown to overcome DLS's limitations. A desirable feature of nanoparticle tracking is the ability to visualize and image scattered light from individual particle solutions using laser illumination. Unlike the DLS, which simply reflects the average particle size of the studied sample, the NTA is not an aggregate measure. When comparing DLS with NTA, it is clear that NTA can solve particle sizes for polydisperse mixtures of PS particles with a size diameter ratio of 2:110, whereas DLS can only solve particle sizes with a diameter ratio of 5:1. DLS is unable to differentiate or detect individual particles, which has an impact on particle size and count. Laser illumination of particles is used in all NTA instrumentation that allows for the imaging of scattered light from a particle [100–103]. Following the Brownian motion of these distinct and observable particles over a time, series of micrographs can be used to compute the mean squared displacement of the particles. This method can be used to estimate individual diffusion coefficients. The Stokes-Einstein equation is used to calculate individual particle sizes from the diffusion coefficient, assuming spherical particles. The range's maximum limit is set by the requirement for adequate particle separation, which allows for lengthy tracks from clearly distinguishable particles. On the other hand, altering the camera sensitivity to

decrease or increase the number of tracked particles visible can possibly alter the ratio of visible particles per unit volume to the number of tracked particles per milliliter volume [104–107]. The quantity of particles detected is determined by the instrument's configuration, particle characteristics, and particle placement in the focused laser beam path. These parameters have an impact on particle illumination, which in turn has an impact on the captured pictures needed to determine particle size. NTA can calculate particle concentration by counting the number of particles observed in the sample volume. This characteristic has received little attention because the initial focus on NTA was on determining its practicality for particle size and dispersion. A qualified analyst/operator is required to decide whether the imaged particles are suitable for tracking to estimate particle size for a common conclusion of NTA investigations. Furthermore, because software and equipment are upgraded or altered on a daily basis, analyst/operator training is critical. In order to find appropriate particle size, instrument optimization may not provide reliable particle concentration. Because they offer the right particle size over a large range of attribute values, some image qualities, such as image intensity, individual particle illumination, or the imaged area of scattered light, may be presumed to be appropriate for particle sizing. When reporting particle concentration, however, caution should be exercised because these image properties can alter it. It is impossible to tell if particle solutions are adequately sized and monitored for concentration without a clearer grasp of the relationship between image characteristics and particle count [108, 109].

19.9 Conclusions

Recently, the use of computers has expanded dramatically in complete pharmaceutical organizations around the world to access various details and obtain information. This technology is also groundbreaking in the field of pharmaceutical analysis, since it may be utilized to improve analytical processes while lowering erroneous results. The pharmaceutical analyst employed computers to search a scientific database. It also serves as a well-established strategy for method development, particularly liquid chromatographic method development, in order to save time and money when developing methods for a variety of compounds using various analytical techniques. Furthermore, QbD is anticipated to continue to grow within pharmaceutical businesses, given positive results have already been achieved through its use. The most prevalent advantages are methods that are more robust and resistant, and hence can withstand the rigors of long-term use by manufacturing laboratories and quality control with a low risk of failure. Furthermore, computers are used to store and manage obtained analytical data, as well as to forecast chromatographic characteristics using a variety of applications. Taking into account all of these advantages, such software will undoubtedly be used extensively in the near future, not only in academics but also in the pharmaceutical sector, to develop rapid methods for various analytical techniques. At last, we conclude that analysis of pharmaceuticals products from discovery of drug to market is inevitable without computer and software.

19.10 Credible Online Resources for Further Reading

URL	Category of source	What to read
https://www.simulations-plus.com/software/dddplus/	Private company situated in Lancaster, California, United States	Applications of DDDPlus software Different types of DDDPlus software Services of DDDPlus software
https://www.malvernpanalytical.com/en/products/technology/light-scattering/nanoparticle-tracking-analysis	Private company situated in Enigma Business Park, Grovewood Road, Malvern WR14 1XZ, United Kingdom	About Nanosight NTA software Visualize and measure particle size and concentration Concept of NTA software
https://www.sartorius.com/en/products/process-analytical-technology/data-analytics-software/doe-software	Private company situated in Göttingen, Germany	Design of experiments software Role of QbD software Importance of QbD software
http://www.dewinterimaging.net/pharmaceutical-analysis-software.htm	Private company situated in West Patel Nagar, New Delhi, India	Different types of software used in the pharmaceutical analysis Applications of different software Selection of software in pharmaceutical analysis
https://www.chromatographyonline.com/view/chromatography-data-systems-perspectives-principles-and-trends	Private company situated in LCGC North America	Perspectives of chromatography data systems Principles of chromatography data systems Recent trends in chromatography data systems
https://www.astrazeneca.com/what-science-can-do/topics/next-generation-therapeutics/developing-the-next-generation-of-drug-delivery-technologies.html	Private company situated in Cambridge, United Kingdom	Novel types of drug delivery systems Types of drug delivery systems Usage of software in analysis
https://www.acdlabs.com/products/com_iden/meth_dev/autochrom/	Private company situated in Toronto, Canada	Automate method development Free consultation and demo of software Ensure an optimal method development strategy that Follows QbD principles
https://www.mastercontrol.com/quality/qms/quality-by-design/	Private company situated in Salt Lake City, UT	Streamline QbD processes with MasterControl's integrated software solutions QbD videos MasterControl QMS Toolkit: A quality manager's guide

References

1. Bruno A, Costantino G, Sartori L, Radi M (2019) The insilico drug discovery toolbox: applications in lead discovery and optimization. Curr Med Chem 26:3838–3873. https://doi.org/10.2174/0929867324666171107101035
2. Maia EH, Assis LC, de Oliveira TA, da Silva AM (2020) Taranto AG Structure-based virtual screening: from classical to artificial intelligence. Front Chem 8:343. https://doi.org/10.3389/fchem.2019.00343
3. Ekins S, Wang B (2006) In: Ekins S (ed) Computer applications in pharmaceutical research and development. Wiley
4. Rahman MM, Karim MR, Ahsan MQ, Khalipha ABR, Chowdhury MR, Saifuzzaman M (2012) Use of computer in drug design and drug discovery: a review. Int J Pharm Life Sci 1:1–21. https://doi.org/10.3329/ijpls.v1i2.12955
5. Tang Y, Zhu W, Chen K, Jiang H (2006) New technologies in computer-aided drug design: toward target identification and new chemical entity discovery. Drug Disc Today Technol 3:307–313. https://doi.org/10.1016/j.ddtec.2006.09.004
6. Geldenhuys WJ, Gaasch KE, Watson M, Allen DD, der Schyf CJV (2006) Optimizing the use of open-source software applications in drug discovery. Drug Disc Today 11:127–132. https://doi.org/10.1016/S1359-6446(05)03692-5
7. Song CM, Lim SJ, Tong JC (2009) Recent advances in computer-aided drug design. Brief Bioinform 10:579–591. https://doi.org/10.1093/bib/bbp023
8. Shah P, Mistry J, Reche PA, Gatherer D, Flower DR (2018) In silico design of Mycobacterium tuberculosis epitope ensemble vaccines. Mol Immunol 97:56–62. https://doi.org/10.1016/j.molimm.2018.03.007
9. Mali AS, Jagtap M, Karekar P, Maruška A (2016) A brief review on process analytical technology (PAT). Int J Curr Pharm Res 8:10–15. https://doi.org/10.1016/j.molimm.2018.03.007
10. Blundell TL (1996) Structure-based drug design. Nature 384:23–26. https://doi.org/10.1038/384023a0
11. Zöldhegyi A, Rieger HJ, Molnár I, Fekhretdinova L (2018) Automated UHPLC separation of 10 pharmaceutical compounds using software-modeling. J Pharm Biomed Anal 156:379–388. https://doi.org/10.1016/j.jpba.2018.03.039
12. Dick V, Sinz C, Mittlböck M, Kittler H, Tschandl P (2019) Accuracy of computer-aided diagnosis of melanoma: a meta-analysis. JAMA Dermatol 155:1291–1299. https://doi.org/10.1001/jamadermatol.2019.137
13. Bakan E, Bayraktutan Z, KilicBaygutalp N, Gul MA, Umudum FZ, Bakan N (2018) Evaluation of the analytical performances of Cobas 6500 and Sysmex UN series automated urinalysis systems with manual microscopic particle counting. Biochem Med 28:329–339. https://doi.org/10.11613/BM.2018.020712
14. Perone SP (1971) Computer applications in the chemistry laboratory. Survey. Anal Chem 43:1288–1299. https://doi.org/10.1021/ac60304a036
15. Podlogar BL, Muegge I, Brice LJ (2001) Computational methods to estimate drug development parameters. Curr Opin Drug Discov Devel 4:102–109. PMID: 11727315
16. Shelmerdine SC, Singh M, Norman W, Jones R, Sebire NJ, Arthurs OJ (2019) Automated data extraction and report analysis in computer-aided radiology audit: practice implications from post-mortem paediatric imaging. Clin Radiol 74:733e11–733e18. https://doi.org/10.1016/j.crad.2019.04.021
17. Lee AC, Harris JL, Khanna KK, Hong JH (2019) A comprehensive review on current advances in peptide drug development and design. Int J Mol Sci 20:2383. https://doi.org/10.3390/ijms20102383
18. Engel KM, Grunewald S, Schiller J, Paasch U (2019) Automated semen analysis by SQA Vision® versus the manual approach—a prospective double-blind study. Andrologia 51:e13149. https://doi.org/10.1111/and.13149

19. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL (2012) Quantifying the chemical beauty of drugs. Nat Chem 4:90–98. https://doi.org/10.1038/nchem.1243
20. Garofalo M, Grazioso G, Cavalli A, Sgrignani J (2020) How computational chemistry and drug delivery techniques can support the development of new anticancer drugs. Molecules 25: 1756. https://doi.org/10.3390/molecules25071756
21. Yang X, Wang Y, Byrne R, Schneider G, Yang S (2019) Concepts of artificial intelligence for computer-assisted drug discovery. Chem Rev 119:10520–10594. https://doi.org/10.1021/acs.chemrev.8b00728
22. Fekete S, Veuthey JL, Beck A, Guillarme D (2016) Hydrophobic interaction chromatography for the characterization of monoclonal antibodies and related products. J Pharm Biomed Anal 130:3–18. https://doi.org/10.1016/j.jpba.2016.04.004
23. Pinzi L, Rastelli G (2019) Molecular docking: shifting paradigms in drug discovery. Int J Mol Sci 20:4331. https://doi.org/10.3390/ijms20184331
24. Gillings N, Todde S, Behe M, Decristoforo C, Elsinga P, Ferrari V et al (2020) EANM guideline on the validation of analytical methods for radiopharmaceuticals. EJNMMI Radiopharm Chem 5(1):7. https://doi.org/10.1186/s41181-019-0086-z
25. Fekete S, Guillarme D (2012) Reversed-phase liquid chromatography for the analysis of therapeutic proteins and recombinant monoclonal antibodies. LCGC Europe 25:540–550. https://doi.org/10.1039/C6AN01520D
26. Graves PR, Haystead TA (2002) Molecular biologist's guide to proteomics. Microbiol Mol Biol Rev 66:39–63. https://doi.org/10.1128/MMBR.66.1.39-63.2002
27. Ambrose J (2006) Radiologists who co-developed computer tomography and performed the first scan. Br Med J 332:977. https://doi.org/10.1016/j.jpba.2015.01.045
28. Hindelang F, Zurbach R, Roggo Y (2015) Microcomputer tomography for medical device and pharmaceutical packaging analysis. J Pharm Biomed Anal 108:38–48. https://doi.org/10.1016/j.jpba.2015.01.045
29. Stock SR (2012) Trends in the micro-and nano Computed Tomography 2010-2012. In: Developments in X-ray tomography VIII, vol 8506. Int Society Optics and Photonics, p 850602. https://doi.org/10.1117/12.930157
30. Schambach SJ, Bag S, Schilling L (2010) Application of micro-CT in small animal imaging. Methods 50:2–13. https://doi.org/10.1016/j.ymeth.2009.08.007
31. Baker DR, Mancini L, Polacci M (2012) An introduction to the application of X-ray microtomography to the three-dimensional study of igneous rocks. Lithos 148:262–276. https://doi.org/10.1016/j.lithos.2012.06.008
32. Naik NN, Jupe AC, Stock SR, Wilkinson AP, Lee PL, Kurtis KE (2006) Sulfate attack monitored by microCT and EDXRD: influence of cement type, water-to-cement ratio, and aggregate. Cem Concr Res 36:144–159. https://doi.org/10.1016/j.cemconres.2005.06.004
33. Salvo L, Cloetens P, Maire E, Zabler S, Blandin JJ, Buffiere J et al (2003) X-ray micro-tomography an attractive characterisation technique in materials science. Nucl Instrum Methods Phys Res B 200:273–286. https://doi.org/10.1016/S0168-583X(02)01689-0
34. De Chiffre L, Carmignato S, Kruth JP, Schmitt R, Weckenmann A (2014) Industrial applications of computed tomography. CIRP Ann 63:655–677. https://doi.org/10.1016/j.cirp.2014.05.011
35. Lee SH, Bajracharya R, Min JY, Han J, Park BJ, Han H (2020) Strategic approaches for colon targeted drug delivery: an overview of recent advancements. Pharmaceutics 12:68. https://doi.org/10.3390/pharmaceutics12010068
36. Acharya C, Coop A, Polli JE, MacKerell AD (2011) Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Curr Comput Aided Drug Des 7(1):10–22. https://doi.org/10.2174/157340911793743547
37. Muehlwald S, Meyburg N, Rohn S, Buchner N (2020) Comparing a two-dimensional liquid chromatography with a quick, easy, cheap, effective, rugged, and safe protocol-based liquid chromatography method for matrix removal in pesticide analysis using time-of-flight mass spectrometry. J Chromatogr A 3:461153. https://doi.org/10.1016/j.chroma.2019.461153

38. Davis JM, Giddings JC (1983) Statistical theory of component overlap in multicomponent chromatograms. Anal Chem 55:418–424. https://doi.org/10.1021/ac00254a003
39. Tyrkkö E, Pelander A, Ojanperä I (2012) Prediction of liquid chromatographic retention for differentiation of structural isomers. Anal Chim Acta 720:142–148. https://doi.org/10.1016/j.aca.2012.01.024
40. Goga-Remont S, Heinisch S, Rocca JL (2000) Use of optimization software to determine rugged analysis conditions in high-performance liquid chromatography. J Chromatogr A 868: 13–29. https://doi.org/10.1016/S0021-9673(99)01203-0
41. Wang X, Yang B, Sun H, Zhang A (2012) Pattern recognition approaches and computational systems tools for ultra-performance liquid chromatography–mass spectrometry-based comprehensive metabolomic profiling and pathways analysis of biological data sets. Anal Chem 84: 428–439. https://doi.org/10.1021/ac202828r
42. Rogers CA, Ahearn JD, Bartlett MG (2020) Data integrity in the pharmaceutical industry: analysis of inspections and warning letters issued by the bioresearch monitoring program between fiscal years 2007–2018. Ther Innov Regul Sci:1–1. https://doi.org/10.1007/s43441-020-00129-z
43. Khin NA, Francis G, Mulinde J, Grandinetti C, Skeete R, Yu B et al (2020) Data integrity in global clinical trials: discussions from joint US Food and Drug Administration and UK medicines and healthcare products regulatory agency good clinical practice workshop. Clin Pharmacol Ther 108:949–963. https://doi.org/10.1002/cpt.1794
44. Shafiei N, De Montardy R, Rivera-Martinez E (2015) Data integrity—a study of current regulatory thinking and action. PDA J Pharm Sci Technol 69:762–770. https://doi.org/10.5731/pdajpst.2015.01082
45. Steinwandter V, Herwig C (2019) Provable data integrity in the pharmaceutical industry based on version control systems and the blockchain. PDA J Pharm Sci Technol 73:373–390. https://doi.org/10.5731/pdajpst.2018.009407
46. Achey TS, McEwen CL, Hamm MW (2019) Implementation of a workflow system with electronic verification for preparation of oral syringes. Am J Health Syst Pharm 76:S28–S33. https://doi.org/10.1093/ajhp/zxy019
47. Spjuth O, Bongcam-Rudloff E, Hernández GC, Forer L, Giovacchini M, Guimera RV et al (2015) Experiences with workflows for automating data-intensive bioinformatics. Biol Direct 10:1–2. https://doi.org/10.1186/s13062-015-0071-8
48. Velghe S, Deprez S, Stove CP (2019) Fully automated therapeutic drug monitoring of antiepileptic drugs making use of dried blood spots. J Chromatogr A 1601:95–103. https://doi.org/10.1016/j.chroma.2019.06.022
49. Tyteca E, Veuthey JL, Desmet G, Guillarme D, Fekete S (2016) Computer assisted liquid chromatographic method development for the separation of therapeutic proteins. Analyst 141: 5488–5501. https://doi.org/10.1039/C6AN01520D
50. Shi Y, Xiang R, Horváth C, Wilkins JA (2004) The role of liquid chromatography in proteomics. J Chromatogr A 1053:27–36. https://doi.org/10.1016/j.chroma.2004.07.044
51. Van Heukelem L, Thomas CS (2001) Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. J Chromatogr A 910:31–49. https://doi.org/10.1016/S0378-4347(00)00603-4
52. Płotka J, Tobiszewski M, Sulej AM, Kupska M, Górecki T, Namieśnik J (2013) Green chromatography. J Chromatogr A 1307:1–19. https://doi.org/10.1016/j.chroma.2013.07.099
53. Molnar I (2002) Computerized design of separation strategies by reversed-phase liquid chromatography: development of DryLab software. J Chromatogr A 965:175–194. https://doi.org/10.1016/S0021-9673(02)00731-8
54. Torres-Lapasió JR, García-Álvarez-Coque MC (2006) Levels in the interpretive optimisation of selectivity in high-performance liquid chromatography: a magical mystery tour. J Chromatogr A 1120:308–321. https://doi.org/10.1016/j.chroma.2006.03.008
55. Tyteca E, Liekens A, Clicq D, Fanigliulo A, Debrus B, Rudaz S et al (2012) Predictive elution window stretching and shifting as a generic search strategy for automated method

development for liquid chromatography. Anal Chem 84:7823–7830. https://doi.org/10.1021/ac301331g
56. Peter GA, Warren BP, Dana Y. https://www.waters.com/nextgen/es/es/library/application-notes/2010/qbd-design-experiments-approach-development-of-chromatographic-method-for-separation-impurities-vancomycin.html. Accessed 23 Mar 2021
57. Lawrence XY (2008) Pharmaceutical quality by design: product and process development, understanding, and control. Pharm Res 25:781–791. https://doi.org/10.1007/s11095-007-9511-1
58. De Beer M, Lynen F, Chen K, Ferguson P, Hanna-Brown M, Sandra P (2010) Stationary-phase optimized selectivity liquid chromatography: development of a linear gradient prediction algorithm. Anal Chem 82:1733–1743. https://doi.org/10.1021/ac902287v
59. Wang X, Stoll DR, Schellinger AP, Carr PW (2006) Peak capacity optimization of peptide separations in reversed-phase gradient elution chromatography: fixed column format. Anal Chem 78:3406–3416. https://doi.org/10.1021/ac0600149
60. Parr MK, Schmidt AH (2018) Life cycle management of analytical methods. J Pharm Biomed Anal 147:506–517. https://doi.org/10.1016/j.jpba.2017.06.020
61. de Sousa J, Holt D, Butterworth PA (2018) Analytical method design, development, and lifecycle management in pharmaceutical quality by design: a practical approach. In: Schlindwein WS, Gibson M (eds) Pharmaceutical quality by design: a practical approach, 1st edn. Wiley, New York, pp 257–279. https://doi.org/10.1002/9781118895238
62. Vogt FG, Kord AS (2011) Development of quality-by-design analytical methods. J Pharm Sci 100:797–812. https://doi.org/10.1002/jps.22325
63. Kochling J, Bridgewater J, Naji R (2010) Introducing a science-based quality by design concept to analytical methods development. In: Pharmaceutical stability testing to support global markets. Springer, New York, pp 169–179. https://doi.org/10.1007/978-1-4419-0889-6_22
64. Mhatre R, Rathore AS (2009) Quality by design: an overview of the basic concepts. In: Rathore AS, Mhatre R (eds) Quality by design for biopharmaceuticals: principles and case studies, 1st edn. Wiley, New York, pp 1–7. https://doi.org/10.1002/9780470466315.ch1
65. Peraman R, Bhadraya K, Padmanabha Reddy Y (2015) Analytical quality by design: a tool for regulatory flexibility and robust analytics. Int J Anal Chem:1–10. https://doi.org/10.1155/2015/868727
66. International Conference on Harmonization of technical requirements for registration of pharmaceuticals for human use, ICH harmonized tripartite guideline, Draft Step 4. Pharmaceutical Development Q8(R1) (2008) https://database.ich.org/sites/default/files/Q8_R2_Guideline.pdf. Accessed 23 Mar 2021
67. International Conference on harmonization of technical requirements for registration of pharmaceuticals for human use, ICH harmonized tripartite guideline. Pharmaceutical Development Q8(R2) (2009) https://database.ich.org/sites/default/files/Q8_R2_Guideline.pdf. Accessed 23 Mar 2021
68. International Conference on harmonization of technical requirements for registration of pharmaceuticals for human use, ICH harmonized tripartite guideline. Quality Risk Management Q9 (2005) https://database.ich.org/sites/default/files/Q9_Guideline.pdf Accessed 23 Mar 2021
69. International Conference on harmonization of technical requirements for registration of pharmaceuticals for human use, ICH harmonized tripartite guideline (2008) Pharmaceutical Quality System Q10. https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human_en.pdf Accessed 23 Mar 2021
70. Mishra V, Thakur S, Patil A, Shukla A (2018) Quality by design (QbD) approaches in current pharmaceutical set-up. Expert Opin Drug Deliv 15:737–758. https://doi.org/10.1080/17425247.2018.1504768

71. Sun M, Liu DQ, Kord AS (2010) A systematic method development strategy for determination of pharmaceutical genotoxic impurities. Org Process Res Dev 14:977–985. https://doi.org/10.1021/op100089p
72. Hanna-Brown M, Borman PJ, Bale S, Szucs R, Roberts J, Jones C (2010) Development of chromatographic methods in the era of quality by design for analytical methods. Sep Sci 2:12–19. https://doi.org/10.1016/j.jscs.2012.12.001
73. Mallik R, Raman S, Liang X, Grobin AW, Choudhury D (2015) Development and validation of a rapid ultra-high performance liquid chromatography method for the assay of benzalkonium chloride using a quality-by-design approach. J Chromatogr A 1413:22–32. https://doi.org/10.1016/j.chroma.2015.08.010
74. Gavin PF, Olsen BA (2008) A quality by design approach to impurity method development for atomoxetine hydrochloride (LY139603). J Pharm Biomed Anal 46:431–441. https://doi.org/10.1016/j.jpba.2007.10.037
75. Li Y, Liu DQ, Yang S, Sudini R, McGuire MA, Bhanushali DS et al (2010) Analytical control of process impurities in Pazopanib hydrochloride by impurity fate mapping. J Pharm Biomed Anal 52:493–507. https://doi.org/10.1016/j.jpba.2010.01.043
76. Tumpa A, Stajić A, Jančić-Stojanović B, Medenica M (2017) Quality by design in the development of hydrophilic interaction liquid chromatography method with gradient elution for the analysis of olanzapine. J Pharm Biomed Anal 134:18–26. https://doi.org/10.1016/j.jpba.2016.11.010
77. Jameel F, Khan M (2009) Quality-by-design as applied to the development and manufacturing of a lyophilized protein product. Am Pharm Rev 12:20–24
78. Ye C, Terfloth G, Li Y, Kord A (2009) A systematic stability evaluation of analytical RP-HPLC columns. J Pharm Biomed Anal 50:426–431. https://doi.org/10.1016/j.jpba.2009.05.028
79. Kormány R, Molnár I, Rieger HJ (2013) Exploring better column selectivity choices in ultra-high performance liquid chromatography using Quality by Design principles. J Pharm Biomed Anal 80:79–88. https://doi.org/10.1016/j.jpba.2013.02.028
80. Liu DQ, Chen TK, McGuire MA, Kord AS (2009) Analytical control of genotoxic impurities in the pazopanib hydrochloride manufacturing process. J Pharm Biomed Anal 50:144–150. https://doi.org/10.1016/j.jpba.2009.04.002
81. Monks KE, Rieger HJ, Molnár I (2011) Expanding the term "Design Space" in high performance liquid chromatography (I). J Pharm Biomed Anal 56:874–879. https://doi.org/10.1016/j.jpba.2011.04.015
82. Aydar AY (2018) Utilization of response surface methodology in optimization of extraction of plant materials. Statistical approaches with emphasis on design of experiments applied to chemical processes, vol 7. pp 157–169. https://doi.org/10.5772/intechopen.73690
83. Bezerra MA, Santelli RE, Oliveira EP, Villar LS, Escaleira LA (2008) Response surface methodology (RSM) as a tool for optimization in analytical chemistry. Talanta 76:965–977. https://doi.org/10.1016/j.talanta.2008.05.019
84. Jamkhande PG, Ghante MH, Ajgunde BR (2017) Software based approaches for drug designing and development: a systematic review on commonly used software and its applications. Bull Fac Pharm Cairo Univ 55:203–210. https://doi.org/10.1016/j.bfopcu.2017.10.001
85. Milanowski B, Hejduk A, Bawiec MA, Jakubowska E, Urbańska A, Wiśniewska A et al (2020) Biorelevant in vitro release testing and in vivo study of extended-release niacin hydrophilic matrix tablets. AAPS PharmSciTech 21:83. https://doi.org/10.1208/s12249-019-1600-z
86. Espíndola B, Bortolon FF, Pinto JM, Pezzini BR, Stulzer HK (2018) New approach for the application of USP apparatus 3 in dissolution tests: case studies of three antihypertensive immediate-release tablets. AAPS PharmSciTech 19:2866–2874. https://doi.org/10.1208/s12249-018-1086-0

87. Li J, Li LB, Nessah N, Huang Y, Hidalgo C, Owen A et al (2019) Simultaneous analysis of dissolution and permeation profiles of nanosized and microsized formulations of indomethacin using the in vitro dissolution absorption system. J Pharm Sci 108:2334–2340. https://doi.org/10.1016/j.xphs.2019.01.032
88. Feng X, Zidan A, Kamal NS, Xu X, Sun D, Walenga R et al (2020) Assessing drug release from manipulated abuse deterrent formulations. AAPS PharmSciTech 21:1–1. https://doi.org/10.1208/s12249-019-1595-5
89. Costa P, Lobo JM (2001) Modeling and comparison of dissolution profiles. Eur J Pharm Sci 13:123–133. https://doi.org/10.1016/S0928-0987(01)00095-1
90. Hofsäss MA, Dressman J (2020) Suitability of the z-factor for dissolution simulation of solid oral dosage forms: potential pitfalls and refinements. J Pharm Sci 109:2735–2745. https://doi.org/10.1016/j.xphs.2019.05.019
91. Zuo J, Gao Y, Bou-Chacra N, Löbenberg R (2014) Evaluation of the DDSolver software applications. Biomed Res Int 2014:204925. https://doi.org/10.1155/2014/204925
92. Zhang Y, Huo M, Zhou J, Zou A, Li W, Yao C et al (2010) DDSolver: an add-in program for modeling and comparison of drug dissolution profiles. AAPS J 12:263–271. https://doi.org/10.1208/s12248-010-9185-1
93. Mendyk A, Jachowicz R, Fijorek K, Dorozynski P, Kulinowski P, Polak S (2012) KinetDS: an open source software for dissolution test data analysis. Dissolut Technol 19:6–11. https://doi.org/10.14227/DT190112P6
94. Belew S, Suleman S, Duguma M, Teshome H, Wynendaele E, Duchateau L et al (2020) Development of a dissolution method for lumefantrine and artemether in immediate release fixed dose artemether/lumefantrine tablets. Malar J 19:1–2. https://doi.org/10.1186/s12936-020-03209-5
95. Nair AB, Al-Dhubiab BE, Shah J, Jacob S, Saraiya V, Attimarad M et al (2020) Mucoadhesive buccal film of almotriptan improved therapeutic delivery in rabbit model. Saudi Pharm J 28:201–209. https://doi.org/10.1016/j.jsps.2019.11.022
96. Guilherme VA, Ribeiro LN, Alcântara AC, Castro SR, da Silva GH, da Silva CG et al (2019) Improved efficacy of naproxen-loaded NLC for temporomandibular joint administration. Sci Rep 9:1–1. https://doi.org/10.1038/s41598-019-47486-w
97. Almukainzi M, Okumu A, Wei H, Löbenberg R (2015) Simulation of in vitro dissolution behavior using DDDPlus™. AAPS Pharm Sci Technol 16:217–221. https://doi.org/10.1208/s12249-014-0241-5
98. Hole P, Sillence K, Hannell C, Maguire CM, Roesslein M, Suarez G et al (2013) Interlaboratory comparison of size measurements on nanoparticles using nanoparticle tracking analysis (NTA). J Nanopart Res 15:2101. https://doi.org/10.1007/s11051-013-2101-8
99. Probst C, Zeng Y, Zhu RR (2017) Characterization of protein particles in therapeutic formulations using imaging flow cytometry. J Pharm Sci 106:1952–1960. https://doi.org/10.1016/j.xphs.2017.04.034
100. Panchal J, Kotarek J, Marszal E, Topp EM (2014) Analyzing subvisible particles in protein drug products: a comparison of dynamic light scattering (DLS) and resonant mass measurement (RMM). AAPS J 16:440–451. https://doi.org/10.1208/s12248-014-9579-6
101. Xin Y, Yin M, Zhao L, Meng F, Luo L (2017) Recent progress on nanoparticle-based drug delivery systems for cancer therapy. Cancer Biol Med 14:228–241. https://doi.org/10.20892/j.issn.2095-3941.2017.0052
102. Zhang Z, Zhuang L, Lin Y, Yan M, Lv J, Li X, Lin H, Zhu P, Lin Q, Xu Y (2020) Novel drug delivery system based on hollow mesoporous magnetic nanoparticles for head and neck cancers—targeted therapy in vitro and in vivo. Am J Cancer Res 10:350–364. PMID: 32064172
103. Varenne F, Makky A, Gaucher-Delmas M, Violleau F, Vauthier C (2016) Multimodal dispersion of nanoparticles: a comprehensive evaluation of size distribution with 9 size measurement methods. Pharm Res 33:1220–1234. https://doi.org/10.1007/s11095-016-1867-7

104. Filipe V, Hawe A, Jiskoot W (2010) Critical evaluation of Nanoparticle Tracking Analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates. Pharm Res 27:796–810. https://doi.org/10.1007/s11095-010-0073-2
105. Defante AP, Vreeland WN, Benkstein KD, Ripple DC (2018) Using image attributes to assure accurate particle size and count using nanoparticle tracking analysis. J Pharm Sci 107:1383–1391. https://doi.org/10.1016/j.xphs.2017.12.016
106. Tian X, Nejadnik MR, Baunsgaard D (2016) A comprehensive evaluation of nanoparticle tracking analysis (NanoSight) for characterization of proteinaceous submicron particles. J Pharm Sci 05:3366–3375. https://doi.org/10.1016/j.xphs.2016.08.009
107. Bai K, Barnett GV, Kar SR, Das TK (2017) Interference from proteins and surfactants on particle size distributions measured by nanoparticle tracking analysis (NTA). Pharm Res 34:800–808. https://doi.org/10.1007/s11095-017-2109-3
108. Kousaka Y, Endo Y, Ichitsubo H, Alonso M (1996) Orientation-specific dynamic shape factors for doublets and triplets of spheres in the transition regime. Aerosol Sci Technol 24:36–44. https://doi.org/10.1080/02786829608965350
109. Fung J, Manoharan VN (2013) Holographic measurements of anisotropic three-dimensional diffusion of colloidal clusters. Phys Rev E 88:020302. https://doi.org/10.1103/PhysRevE.88.020302

Telemedicine

20

Mehdi Chamani, Parsa Khoshkhat, and Farid Abedin Dorkoosh

Contents

Abstract

Nowadays, there is a growing demand for alternative ways to access medical care to solve healthcare problems. Electronic information and telecommunication technologies, for instance, telemedicine, could be a solution to address these problems. This chapter explores telemedicine's background and definition, followed by some of its drawbacks and issues. Finally, telemedicine's future

M. Chamani · P. Khoshkhat
Faculty of Pharmacy, Tehran University of Medical Sciences, Tehran, Iran

F. A. Dorkoosh (✉)
Department of Pharmaceutics, Faculty of Pharmacy, Tehran University of Medical Sciences, Tehran, Iran

Medical Biomaterial Research Center (MBRC), Tehran University of Medical Sciences, Tehran, Iran
e-mail: dorkoosk@tums.ac.ir

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_20

perspective as a new frontier in providing a quality health service will be discussed.

Keywords

Telemedicine · Telehealth · Health service · Telecommunication technologies · COVID-19 · Medical urgency · Increasing global health

Chapter Objectives

- Present a general definition of telemedicine and differences with similar terms
- Look back into telemedicine evolving history
- Introduce the services that a modern telemedicine system can offer
- Explain the Telepharmacy concept and how it can be practiced successfully through modern technologies such as artificial intelligence
- Illustrate how health emergencies such as the COVID pandemic these days can be dealt with by telemedicine technology
- Provide a future image of telemedicine use and its capabilities

20.1 Introduction

Telemedicine was first introduced in the 1970s, which meant healing at a distance [1]. It was defined in 2018 as a service delivery in which the patient experiences remote interaction using Information and Communications Technology (ICT) instead of the traditional face-to-face method. According to a 2007 study, no exact scientific definition of telemedicine has been provided [2]. This book chapter delves into telemedicine's history and origin, which will provide different definitions for this technology and its exact background. Then we will go to the services and usage of this technology, whose services are divided into four categories, each of which we will be examined thoroughly.

The next subject is about health emergencies in today's world and utilizing telemedicine as a potential tool to overcome the challenges involved. In emergencies, the solutions provided by this technology can be essential and save time and money. One of the current services provided for it in this section is helping prevent the spread of COVID-19. In this chapter, the effect of telemedicine to prevent the spread of COVID-19 will be discussed. Worldwide medical emergencies help assess and control health issues. Telemedicine helps reduce side effects and increase health management [3, 4]. One of the significant medical problems right now is the unrestricted spread of COVID-19. This technology can also work well in epidemic diseases and play an essential role in better managing the crisis and spreading it by providing private medical information and remote advice to patients [5, 6]. Telemedicine has played an essential role in treatment in today's world, but it cannot take the place of traditional medicine. All service providers need to remember that awareness of people and patients is one of the most critical matters. Finally, after

reviewing all the cases, we will mention telemedicine's future and goals and how it will be used shortly.

20.2 Telemedicine Definition

Technology breakthroughs are impacting almost every aspect of science categories we understand today in positive and negative ways. Meanwhile, healthcare systems are not an exception. The intersection of technology and healthcare systems gives birth to new areas such as telemedicine and telehealth. Telemedicine and telehealth are used most of the time interchangeably, and there is no distinct barrier between them in terms of definition. Achieving a more in-depth insight into telemedicine and telehealth requires declaring their definitions. Up to this moment, there are numerous definitions provided for telemedicine. Telemedicine is defined by using telecommunications technology to diagnose medically, monitor, and execute therapeutic goals when the physical distance detaches the patient and the physician [7]. In different words, WHO defines it as: "the delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for the diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of health care providers, all in the interests of advancing the health of individuals and their communities" [8].

The many definitions feature that telemedicine is an open and steadily growing science, as it consolidates new advancements in technology and responds to the altering health demands and conditions of societies. According to the WHO, telehealth involves using telecommunications and virtual technology to deliver health care outside of traditional healthcare facilities, for example, a virtual home health care, where patients such as the chronically ill or the elderly may receive guidance in specific procedures while remaining at home [8]. It would be a little confusing when it comes to merging technology and medicine. Commonly, inappropriate language instead of technical terms is used by policymakers, vendors, and advocacy groups. As mentioned earlier, telemedicine and telehealth terms can be used interchangeably, but their minor differences can distinguish them. Telehealth term compasses a wide range of technologies and services to provide care to the patient and generally improves the healthcare delivery system. Telehealth is distinct from telemedicine; because it refers to a broader range of services than telemedicine. As telemedicine specifically refers to clinical services, telehealth can also provide nonclinical distance services such as provider training, office meetings, and continuing education for physicians [9–11]. Miller suggests that telehealth refers to clinical and nonclinical applications in education, administration, and research, while telemedicine is often reserved for clinical, patient care applications [12]. Figure 20.1 helps to develop a better understanding of telemedicine and telehealth relations. However, for this chapter, telemedicine and telehealth are synonymous and used interchangeably.

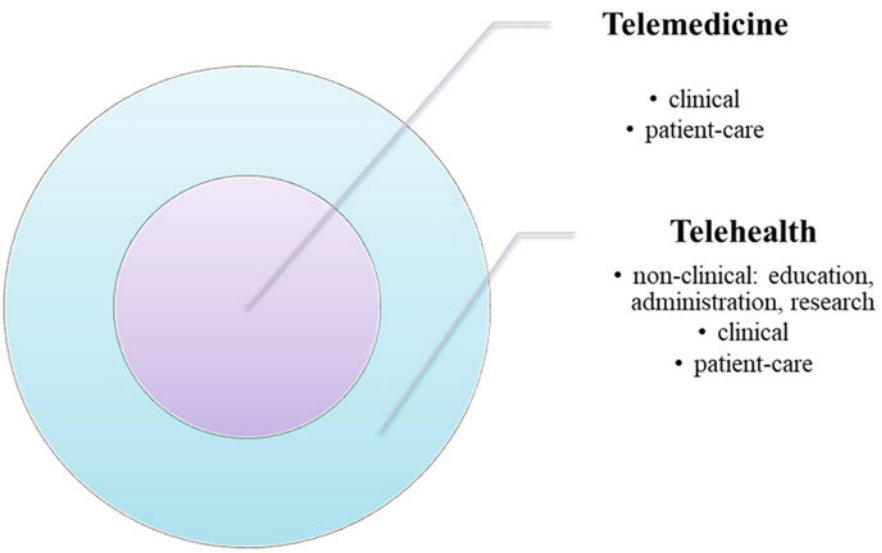


Fig. 20.1 Abstract model of the telemedicine-telehealth relationship

20.3 Telemedicine History

The principle of conducting medical tests and assessments using telecommunications networks is not new. A look at every telemedicine article or book will show many different interpretations of the method's history and different dates for the same events. Telemedicine history can be confusing because it has not been invented as a well-defined discipline with advanced equipment and protocols. In 1844, telemedicine was first invented as the telegraph in its early form, and in 1876 the telephone made it easier for patients to seek physicians conveniently and made it easier for physicians to be consulted by their counterparts [13]. American Civil War was when the telegraph was used to deliver lists of casualties and order medical supplies. After telegraphy, telephones were created and became the primary means of remote healthcare contact [14, 15]. The telegraph was soon substituted by the telephone as a general means of connection, but it survived for a long time because of Australia's vast distances [16]. The subsequent development of widespread significance was at the end of the nineteenth century when radio communication became possible, which was done initially by Morse code and later by voice. Steps were taken to transfer heart and lung sounds to a qualified specialist who could evaluate the organs' state. However, inadequate transmission systems made the attempts a failure [17]. To be more structured in terms of telemedicine's history, we will discuss historical milestones in telemedicine's evolution one by one.

In 1906, the father of electrocardiography (Einthoven) made the first ECG transmission on telephone lines. Radios were used to support ships throughout the

1920s. It merely links doctors at shore stations in medical emergency circumstances to offer medical advice to clinics on ships at sea. The next breakthrough was to change a standard psychiatric health service into telepsychiatry. In 1955 Nebraska Psychiatric Institute used a closed-circuit television service and began developing it into a two-way link with Norfolk State Hospital with a 112 miles distance. This development occurred in 1964 and continued to expand in 1971 [18].

Connections have been used for learning and counseling sessions between advisors and physicians. This experiment is one of the first many examples of telepsychiatry. One earlier example of television connecting physicians and patients was the Massachusetts General Hospital/Logan International Airport Medical Center, founded in 1967. It used a two-way audio-visual microwave circuit, which allowed nurses to provide treatment to passengers and airport personnel 24 h a day, complemented by medical experts using an audio-visual connection. The results of 1000 episodes were recorded in an early report of this form of health care's practical delivery [17]. Paramedics were affiliated with hospitals in distant towns or villages throughout rural Alaskan and Canadian villages. The National Aeronautics and Space Administration (NASA) played a crucial role in telemedicine's development stages. The telemedicine initiative of NASA began when people started to be in space in the early 1960s. NASA is a leader in research and developments in telemedicine. The rising difficulty of space operations since the first days of suborbital flight has led to the changes applied to telemedicine. The NASA team has begun tracking astronaut physiological parameters (telemetry) and cabin and exterior conditions as telemedicine applications. In the 1970s, the first set of commercial communication satellites (ATS-6 satellites) was operational, and telemedicine opportunities were immediately noticeable by physicians. The scene of the primary operation was once more North America. The most known examples are the Alaskan Satellite Biomedical Demonstration Program for Village Wellbeing strengthened and many Canadian programs servicing remote areas using NASA's Hermes satellite [19, 20]. Telemedicine for emergency purposes was demonstrated during the aftermath of earthquakes in Mexico. While other communication systems were out of service, satellite technology provided first responders access to information and supported those working directly with survivors [21]. During this era, satellite systems' importance cannot be overlooked; however, telemedicine's technology and communication systems were costly. Since the mid-1990s, we have seen several medical specialties invade telemedicine, and the overall number of appointments has grown.

In the United States, Allen and Grigsby report that within 1998 over 40,000 telecommunications consultations took place in over 35 different specialties [22]. About 70% of the episodes use interactive footage. The remainder is based on pre-recorded or not considered as video technologies. The number of telecommunication consultations is inevitably limited compared with traditional face-to-face consultations. The statistics exclude Teleradiology, which, with more than 250,000 consultations only conducted in the United States in 1997, remains the single most common (and pre-recorded) application [23].

The last number represents Teleradiology's long-term status as one of the few telemedicine specialties which have become established healthcare services. As a result, the US Medicare healthcare system provides Teleradiology with total

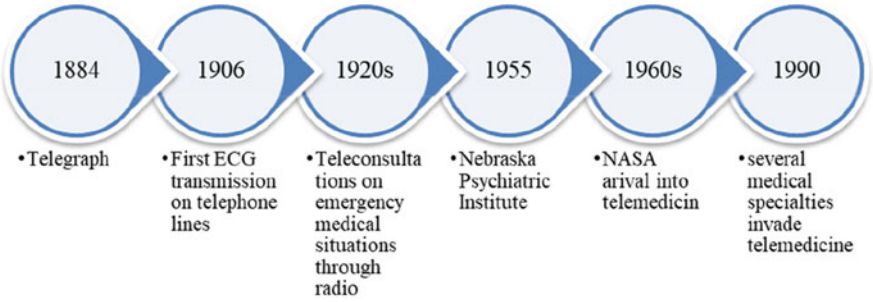


Fig. 20.2 Timetable of telemedicine evolution

reimbursements. Since 1994, not only as a source of data (tele-education) but as a medium of contact, the Internet's general influence has become apparent. Audio and video streaming increases the possibility of remote connectivity through the Superhighway. Considering the broad-ranged problems like security issues can be fixed [24]. At last, an overview of certain important moments in telemedicine evolving process schemed in Fig. 20.2.

20.4 Telemedicine Services

Telemedicine provides a range of health programs that promote well-being and treat illnesses and disorders and personal health services. Telemedicine was initially intended to make healthcare facilities more available to rural areas with minimal health service infrastructure. It was later developed to enhance patient care quality by teaching and improving medical practitioners' decision-making processes in rural areas. More recently, it has been suggested to increase the quality of health care, as it encourages resources that are geographically distant and reshapable to leverage health resources to be organized. Fortunately, for most specialties, telemedicine programs are currently available.

This growth in the availability of multiple telemedicine systems is due to increasing digitalization and evolving technologies. There is no such unique classification of telemedicine services because of the diversity of its applications. However, from the writers' point of view, the following classification can be viewed as a broad and detailed telemedicine services categories:

1. Teleconsultation
   1.1. Store and forward
   1.2. Real-time video consultations
2. Remote patient monitoring (RPM)
3. Telesurgery
4. Telepharmacy
   4.1 Modern ICTs and software in Telepharmacy
   4.2 Artificial intelligence and Telepharmacy

20.4.1 Teleconsultation

Teleconsultation comprises synchronous (real-time video consultations) and asynchronous (store and forward). The problem of geographic separation in therapy is solved by teleconsultation. "Remote consultation" is the higher term for teleconsulting in MeSH (Medical Subject Headings). Its definition is mentioned as "Consultation by remote telecommunications, generally for diagnosis or treatment of a patient at a site remote from the patient or primary physician" [25, 26]. This technology's mechanism is that two functions of storage and transfer images to the medical center and video conferencing are carried out when a patient is referred.

20.4.1.1 Store and Forward

Various materials, such as laboratory papers and images, demographic information, and medical history, are stored in this technology electronically and then sent to the treatment facility [27]. The most important applications of the platforms are ophthalmology, radiology, and pathology. Center for Connected Health Policy states that it includes X-rays, MRIs, medical pictures, and video evidence to be checked by the physician and forwarded by the patient. Video calling can lead to a face-to-face examination if needed. Five advantages to this technology:

- Care will be provided without the patient.
- Waiting for shifts, especially in areas with a shortage of medical staff, is reduced.
- Specialists can provide primary care, regardless of the patient's location.
- Doctors can choose the right patients.
- This technology can remove linguistic and cultural limitations.

This technology provides valuable services such as providing the best remote medical services to patients and even on-the-go consumers, and doctors can respond to stored information at any time, depending on their expertise [28].

20.4.1.2 Real-Time Video Consultations

This technology, also known as live video, allows the patient and physician to communicate and talk simultaneously with video call technology. In addition to medical services, psychological and counseling services can also be provided in this way [17]. In this technology, video devices can be video conferencing units, side cameras, and video films. Monitors such as computer monitors, LED and LCD TVs, and projectors can be used for limited access by patients or prisoners. Telemedicine video conferencing uses video and communication technology to transmit patient information between multiple centers. The doctor examines various instruments, such as stethoscopes in ophthalmology. The information collected by telemedicine is sent to the doctor simultaneously to treat the disease.

20.4.2 Remote Patient Monitoring (RPM)

It is a way to provide medical services to the patient by collecting and expressing it differently from traditional and home methods. This method reduces access for further improvement and lower costs for communication. RPM can be very useful in administering treatments such as hemodialysis [29]. This technology's main features, such as remote monitoring and analysis of physiological factors, make it possible to diagnose the disease's severity quickly. Also, other patients will not waste their time and money on hospitalization. RPM has a range of applications that boost its performance, but components need to be supplied to make these applications doable. It should be equipped with integrated wireless sensors for physiological information collection and view.

There is a need to store information in the patient's site that connects it to the central data bank or the healthcare providers. A robust data repository center should be established to store the collected data from the patient's site. Finally, RPM requires intelligent software that offers care advice by reviewing available patient medical reports [30]. Depending on the nature of the illness and its factors, and even the type of sensors and data storage, various configurations of the mentioned components may be planned for this technology. Although this technology is designed to be used faster and better by the general public, specific barriers can stop it from achieving this goal. This technology success is dependent mainly on the patient's satisfaction and motivation, and without patient commitment, it would not be practicable. Another limitation of using RPM technology properly could be its high cost for both the patient and the service provider. Security and data privacy in using this technology is essential and necessary due to an online base. The Internet is the backbone of the RPM process. Therefore, this system will have almost nothing to say in places where it is not feasible to access the Internet and wireless networking [31].

20.4.3 Telesurgery

The technology is an entry-level surgical system that links the surgeon to unique robots using the wireless network while the patient is outside of the doctor's accessibility. Telesurgery addresses issues such as the absence of a surgeon, long geographic distance unavailability of the patient, and tremendous potential costs [32, 33]. This procedure is also for both the patient's benefit and the surgeon's professional accuracy and wellness. Many advances have been made in surgery since the first operation took place in 2001. The ZEUS robotic system was one of the main challenges for the surgical team in New York (Intuitive Surgical, Sunnyvale, CA, USA). The surgery was performed in a hospital in Strasbourg. This successful laparoscopic cholecystectomy procedure was done on a woman patient and lasted 2 h, but the patient was not well recovered [34]. This technology has an issue with optimum time management; messages exchanged on both sides are delayed due to the high demand for data transfer servers. This delay can cut negatively on surgery

efficiency. However, optimistically, this limitation would fade away because of the fast pace in technological and network progress. Health providers must achieve experiences with teleoperations' technologies and devices to decrease unnecessary time and money consumption [35].

20.4.4 Telepharmacy

It is a form of pharmaceutical care in which pharmacists and patients communicate in a remote location using information technology (IT) and communication facilities. Telepharmacy is very important in today's world, where the pharmacy is an integral part of public health. In addition to distributing medications, pharmacists offer other services such as diet counseling and pharmacovigilance. Pharmacies often face the problem of unequal distribution across the country. Besides, with the declining number of pharmacists, this problem is increasing [36, 37]. This technology includes drug selection, distribution, counseling, and clinical services. Its features include the significant participation in the pharmaceutical service of regions that suffer from economic or regional problems, and patients can receive services without a referral. It also has disadvantages, such as difficulties in evaluating drugs and increasing the risk to patient information security and accuracy [36, 38]. Figure 20.3 represents Telepharmacy practice steps. Telepharmacy is now available in Spain, Denmark, Egypt, France, Canada, Italy, Scotland, and Germany. The aging of the population elevates the need for medication. According to the WHO, the number of pharmacists is declining, and this technology's need is felt [39]. European institutions estimated a shortage of one million doctors by 2020, 10% of whom were pharmacists. These statistics are disappointing, and technology solutions should be considered. The International Federation of Pharmaceuticals (FIP) recognizes the need to use up-to-date technologies to combat efficient human resources shortages. Utilizing Telepharmacy can help improve cooperation between the public and private sectors

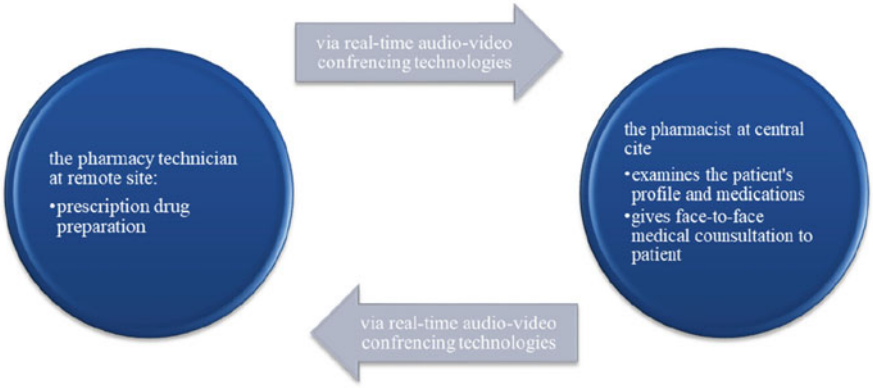


Fig. 20.3 Telepharmacy practice steps

and scientific institutions and universities to achieve better results in healthcare delivery [39, 40].

20.4.4.1 Modern ICTs and Software in Telepharmacy

The broad ICT definition covers all communication devices, including the Internet, cellular networks, mobile phones, laptops, apps, middleware, teleconferencing, social networking, and the use and provision of information and communication technologies. It may help to solve the issue of health providers' lack of accessibility. Various studies and examples are demonstrating that Telepharmacy practice through ICT is practical and enjoyable for people.

A wide-scale project was initiated in 2003 by the Community Health Association of Spokane. In order to support pharmacists' online visits, Webcams are located in participating pharmacies as part of the trial. The efficiency assessment found that more than 70% of surveyed were very pleased with the extra service offered [41]. Besides, Telepharmacy can be used as a management method to minimize pharmaceutical services costs within the hospital industry [42]. Another point to consider is that medication delivery and interdisciplinary collaboration between doctors and pharmacists are simultaneously guaranteed through ICT [43].

Recently, broad use of the Internet and technology has contributed to the growth of a wide variety of Telepharmacy networks in the future. Amazon's arrival in the pharmaceutical sector as a new distributor in and of itself demonstrates Telepharmacy's high potential [40]. Also, it is essential to use the Internet's potentials to build a unique online software network between customers, pharmacies, and hospitals. For instance, Updox and zoom software packages for health care are the successful ones that create real-time pharmacist-patient contact (either visually or by audio). These software programs needed regular updates by the developer and regular upgrades of hardware systems before cloud-based technology. With the help of Cloud-based technology, there is no longer hardware necessity in any location that needs pharmacist like in the past. Cloud spaces can hold and store much patient information and make its management more convenient [44, 45]. Telepharm is one of these cloud-based software programs in which the pharmacy management system (PMS) receives patients' prescriptions, and all prescription dispensing steps are recorded until the final step. Finally, these processes are evaluated by the pharmacist, and the prescription is delivered to the patients. So, without the pharmacist's physical presence, all the medicine supply steps are done correctly. This system will inevitably decrease workflow, increase patient Outcomes, and make pharmaceutical services more cost-effective.

20.4.4.2 Artificial Intelligence and Telepharmacy

Artificial intelligence (AI) is an IT field that highlights intelligent machines functioning like people. AI is still developing, and in many industries, it has proved that it is instrumental. The pharmaceutical industry has used AI in many phases of pharmaceutical products in recent years. The pharmaceutical industry uses AI to make decisions effectively and reliably for research and development, 3D drug printing, and complicated operational details [46, 47]. Scientists seek to level up patient safety

by using state-of-the-art AI techniques. Johns Hopkins, a health official, said more than 250,000 deaths a year were due to a medical mistake while examining the medical death rate in the United States for 8 years [48]. Human mistakes can cause medication errors, but most occur from an incorrect backup system whose aim is to detect errors.

For example, MedEye is a full-service medical company that screens medications for accuracy by visual recognition and machine learning tools and at the same time connected to the hospital information system all by using advanced AI technologies. By avoiding drug overdose, AI also plays a role in patient safety. PerceptiMed is another example of a pharmaceutical company that uses AI technologies to manage and analyze prescription supplies. In other words, it determines the correct treatment supply and each drug's accurate dosage in real time. It is a channel through which the Telepharmacy concept balances workloads while ensuring the effectiveness of healthcare facilities and pharmacies.

20.5 Telemedicine in Emergencies

As we all know, medical emergencies might happen widespread in the developed world today. Finding comprehensive and new solutions is mandatory because of the increasing growth of communities and the difficulty of handling widespread and global situations [49, 50]. An immediate and effective response to health emergencies plays a paramount role in mitigating the destructive and fatal consequences. All health and medical emergencies should be managed in a timely and very efficient manner at a higher level than typical situations. This management is highly dependent on the accessibility of healthcare specialists. Health providers should evaluate significant health changes for the impacted group mentally and physically and offer necessary health risk prognostic information to the population and patients. These massive scales of medical assessments and information providing in public medical emergencies need a modern solution. Therefore the necessity of telemedicine practice as a capable solution is strongly felt [51].

A prominent example of a modern life health emergency is the prevalence of COVID-19 as a pandemic disease. The prevalence of COVID-19 showed that classical methods of treatment are not responsive. In pandemic diseases, the prevalence is so widespread and dangerous that there is a shortage of treatment space in hospitals worldwide, so it can also be handled in public areas such as hotels, sports centers, and at home providing emergency facilities [52]. Telemedicine helps patients to support themselves and improve global health. EHR (electronic health record) is required for each country; it explains the continuity of services and improves the epidemiological surveillance of sick people. In these situations, telemedicine methods are performing telemedicine at home, clinical video conferencing, and store and forward type [53, 54]. Before the outbreak of COVID-19, this technology was not very common, but today, due to the upcoming new challenges, it can be used in the prevention and treatment; for example, using phone calls, video conferencing, and answering online questionnaires can serve to provide people with

more information and assistance [55]. Utilizing telemedicine can be beneficial in handling this disease in several ways. It reduces unnecessary hospital visits while there is a heavy workload on hospital staff and prevents the spread of infections through people who would have been hospitalized for no serious reason. Centers that have been fully occupied and could not accept more COVID-19 patients can provide their health services through this technology. There are many examples of beneficial and successful telemedicine practices. In a short period, Cleveland Clinic, Oregon Health Sciences University (OHSU), Intermountain Health Care, Medical University of South Carolina (MUSC), Rush University Medical Center, University of Washington (UW), and NYU Langone have received valuable services from telemedicine to manage the Coronavirus pandemic. At last, considering all the details and facts mentioned above, the potential benefits of telemedicine practice in health emergencies can be concluded in following points:

- Reduce possible infections when people visit the hospital
- Increasing public awareness to comply with health protocols
- Providing services such as a virtual visit that invites people to be more relaxed and calm
- Reduce treatment costs and labor needs

20.6 The Future of Telemedicine

The broad adoption of telehealth in the future is now recognized as a certainty. However, health professionals should not fear one more technological trend. The conventional on-site hospital visit was never meant to be replaced by telemedicine. The overall goal is to help serve our patients and enhance organizational quality. Alternatively, telemedicine should be used as a therapeutic method on the clinician's work ground. In the future, however, telemedicine practices may lose patients to the competitors who offer this alternative. The development process of this technology in the world of medicine and health depends on various factors. Similar to every other aspect of technology to grow, the financial elements are the most evident ones. Insurance and reimbursement policies are essential factors in health economics. The telemedicine repayment mechanism is a challenge that is the responsibility of communities and their economic management. The general budget allocated in Canada, Australia, the United Kingdom, New Zealand, and the Veterans Management Organization has led to this technology's large-scale development [56, 57]. Another factor that can lead to its excellent development in medicine soon is the overall growth of technology itself, which is the backbone of telemedicine and its applications [58].

Having all been mentioned, one and the most crucial element in the telemedicine growing pathway is people's aspect. As we know, human culture and behavior, and attitudes in today's world are influential in all aspects, including telemedicine [59]. Patients' behavior and perception of technology can shape this health technology future because they are the ones who receive the most contact and treatment

information. The following can be a sign of the popularity of this technology among patients [60–62]:

- Increasing the use of the Internet for the advancement of medical science
- Increased dissatisfaction with the traditional healthcare system
- Patients' accepted satisfaction with telemedicine
- More involvement in healthcare decision-making

Patients who gravitate more and more into using telemedicine can also play an essential role in developing technology. The doctors and nurses in healthcare centers, such as hospitals, communicate with each other and other medical centers. This need must be met by technology, which telemedicine does well [63, 64]. Various medical service providers must make information and news available to the public before an illness or outbreak occurs. The basis of these organizations' progress is that, on the one hand, they reduce the costs of treatment, and on the other hand, they increase the quality [65]. Technology support for health organizations is done by following ICT. Organizations such as the American Association of Retired Persons (AARP), the Institute of Medicine (IOM), the National Health Service (NHS), the World Health Organization (WHO), and the World Bank enjoy this support [66]. With the progress in industrialized countries, medical science's role is to facilitate treatment, which will lead to the growth of the scientific and health level of the people of the world. On the other hand, technology using existing devices such as mobile phones and the Internet can help better inform people in telemedicine [67]. As an example of oversight, the federal government has placed telemedicine oversight on its people to make the technology more efficient and follow traditional medicine rules [68].

20.7 Conclusion

Telemedicine attempts to replicate the patient-to-doctor relationship in a realistic, productive manner. As technology becomes user-friendly, people become more comfortable with technology, and as smartphones become widely available, telemedicine will continue to show rapid development [69]. Thus, the telemedicine benefits and, at the same time, its challenges became more evident with the increased public participation in the use of telemedicine systems than ever. Telemedicine has many advantages to overcome several of the health system's problems. First, access to health facilities will be more accessible, and patients could be better off treatment. The convenience and well-being of patients will eventually raise and reduce medical and therapeutic costs. Like other advanced innovations, telemedicine systems contribute to ease and well-being for consumers.

On the other hand, the release and use of telemedicine are universal. Another essential advantage of this technology could be improving the global health condition in different communities and countries [70, 71]. Other advantages of this technology include having an integrated and modifiable system to advance medical

science goals because every data is stored in different patients and diseases [72]. Therefore, it is possible to predict the increase in the world scientific level. Telemedicine expansion in today's life, despite many advances, has been limited by problems such as a lack of trust in technology, the challenge of using data transfer over the Internet, and a lack of experience for the general public. Similarly, technological knowledge and convenience through smartphones and healthcare applications are also significant obstacles to telemedicine. Today with the extensive usage of telemedicine, particularly in this COVID-pandemic period, one of the critical concerns of this technology that must be explored is telemedicine's legal complexities.

The challenges associated with the development of telemedicine have not yet been eliminated from their regulatory obstacles. Eventually, as technology progresses and has a growing effect on people, the telemedicine role is becoming increasingly significant day by day. Its existence as a permanent and impressive medical services component is unavoidable in the not too distant future.

20.8 Credible Online Resources for Further Reading

URL	Category of source	What to read or refer?
https://medium.com/@gloriaimodia/how-telepharmacy-can-increase-access-to-pharmacists-a7286f6af45f	Private company	Tutorial for building small molecules including drug molecules
https://blog.telepharm.com/4-technology-trends-in-the-pharmacy-industry	Private company	Technology trends in the pharmacy industry
https://medium.datadriveninvestor.com/ai-enabled-robotic-delivery-and-telemedicine-for-meds-to-beds-programs-drives-adherence-reduces-bcef4b178e16	Private company	AI-enabled robotic delivery and telemedicine for Meds-to-Beds programs drives adherence, reduces pathogen transmission potential
https://www.americantelemed.org/resource_categories/practice-guidelines/	Institution	American Telemedicine Association Practice Guidelines
https://www.everydayhealth.com/healthy-living/the-best-telemedicine-apps/	Government	The best telemedicine apps to use during the Coronavirus pandemic
https://www.virtualmedstaff.com/	Private company	How this telemedicine company connects hospitals, clinicians and patients to deliver high-quality healthcare whenever and wherever it's needed

References

1. Strehle E, Shabde N (2006) One hundred years of telemedicine: does this new technology have a place in paediatrics? Arch Dis Child 91(12):956–959. https://doi.org/10.1136/adc.2006.099622
2. Sood SP, Negash S, Mbarika VW, Kifle M, Prakash N (2007) Differences in public and private sector adoption of telemedicine: Indian case study for sectoral adoption. Stud Health Technol Inform 130:257
3. Bashshur RL (1995) On the definition and evaluation of telemedicine. Telemed J 1(1):19–30. https://doi.org/10.1089/tmj.1.1995.1.19
4. Hjelm N (2005) Benefits and drawbacks of telemedicine. J Telemed Telecare 11(2):60–70. https://doi.org/10.1258/1357633053499886
5. Hollander JE, Carr BG (2020) Virtually perfect? Telemedicine for Covid-19. N Engl J Med 382(18):1679–1681. https://doi.org/10.1056/NEJMp2003539
6. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J et al (2020) Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 26(5):309–313. https://doi.org/10.1177/1357633X20916567
7. Shaw DK (2009) Overview of telehealth and its application to cardiopulmonary physical therapy. Cardiopulm Phys Ther J 20(2):13. https://doi.org/10.1097/01823246-200920020-00003
8. World Health Organization (ed) (2010) Telemedicine: opportunities and developments in member states: report on the second Global survey on eHealth [Internet]. World Health Organization, Geneva, 93 p. (Global Observatory for eHealth Series). http://www.who.int/goe/publications/goe_telemedicine_2010.pdf. Accessed 15 June 2020
9. Dorsey ER, Topol EJ (2016) State of telehealth. N Engl J Med 375(2):154–161. https://doi.org/10.1056/NEJMc1610233#SA1
10. Stowe S, Harding S (2010) Telecare, telehealth and telemedicine. Eur Geriatr Med 1(3):193–197. https://doi.org/10.1016/j.eurger.2010.04.002
11. Tuckson RV, Edmunds M, Hodgkins ML (2017) Telehealth. N Engl J Med 377(16):1585–1592. https://doi.org/10.1056/NEJMsr1503323
12. Miller EA (2007) Solving the disjuncture between research and practice: telehealth trends in the 21st century. Health Policy 82(2):133–141. https://doi.org/10.1016/j.healthpol.2006.09.011
13. Cipolat C, Geiges M (2003) The history of telemedicine. Curr Probl Dermatol 32:6–11. https://doi.org/10.1159/000067346
14. Craig J, Patterson V (2005) Introduction to the practice of telemedicine. J Telemed Telecare 11(1):3–9. https://doi.org/10.1177/1357633x0501100102
15. Zundel KM (1996) Telemedicine: history, applications, and impact on librarianship. Bull Med Libr Assoc 84(1):71–79
16. Eikelboom RH (2012) The telegraph and the beginnings of telemedicine in Australia. Stud Health Technol Inform 182:67–72. https://doi.org/10.3233/978-1-61499-152-6-67
17. Wootton R (2006) Realtime telemedicine. J Telemed Telecare 12(7):328–336. https://doi.org/10.1258/135763306778682387
18. Benschoter RA (1971) CCTV—pioneering Nebraska Medical Center. Educational Broadcasting
19. Patricoski C (2004) Alaska telemedicine: growth through collaboration. Int J Circumpolar Health 63(4):365–386. https://doi.org/10.3402/ijch.v63i4.17755
20. Stanberry B (2006) Legal and ethical aspects of telemedicine. J Telemed Telecare 12(4):166–175. https://doi.org/10.1258/135763306777488825
21. Nicogossian AE, Pober DF, Roy SA (2001) Evolution of telemedicine in the space program and earth applications. Telemed J E Health 7(1):1–15. https://doi.org/10.1089/153056201300093813
22. Allen A, Grigsby B (1998) 5th annual program survey—part 2. Consultation activity in 35 specialties. Telemed Today 6(5):18–19

23. Coiera E (2002) Essentials of telemedicine and telecare. BMJ 324(7345):7
24. Zollo SA, Kienzle MG, Henshaw Z, Crist LG, Wakefield DS (1999) Tele-education in a telemedicine environment: implications for rural health care and academic medical centers. J Med Syst 23(2):107–122. https://doi.org/10.1023/A:1020589219289
25. Bove AA, Homko CJ, Santamore WP, Kashem M, Kerper M, Elliott DJ (2013) Managing hypertension in urban underserved subjects using telemedicine a clinical trial. Am Heart J 165(4):615–621. https://doi.org/10.1016/j.ahj.2013.01.004
26. Deldar K, Bahaadinbeigy K, Tara SM (2016) Teleconsultation and clinical decision making: a systematic review. Acta Inform Med 24(4):286. https://doi.org/10.5455/aim.2016.24.286-292
27. Patterson V, Hoque F, Vassallo D, Roberts MF, Swinfen P, Swinfen R (2001) Store-and-forward teleneurology in developing countries. J Telemed Telecare 7(1 Suppl):52–53. https://doi.org/10.1258/1357633011936714
28. Bonnardot L, Rainis R (2009) Store-and-forward telemedicine for doctors working in remote areas. J Telemed Telecare 15(1):1–6. https://doi.org/10.1258/jtt.2008.008004
29. Malasinghe LP, Ramzan N, Dahal K (2019) Remote patient monitoring: a comprehensive study. J Ambient Intell Humaniz Comput 10(1):57–76. https://doi.org/10.1007/s12652-017-0598-x
30. Syed L, Jabeen S, Manimala S, Elsayed HA (2019) Data science algorithms and techniques for smart healthcare using IoT and big data analytics. In: Smart techniques for a smarter planet. Springer, pp 211–241
31. Vegesna A, Tran M, Angelaccio M, Arcona S (2017) Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 23(1):3–17. https://doi.org/10.1089/tmj.2016.0051
32. Haidegger T, Sándor J, Benyó Z (2011) Surgery in space: the future of robotic telesurgery. Surg Endosc 25(3):681–690. https://doi.org/10.1007/s00464-010-1243-3
33. Hung AJ, Chen J, Shah A, Gill IS (2018) Telementoring and telesurgery for minimally invasive procedures. J Urol 199(2):355–369. https://doi.org/10.1016/j.juro.2017.06.082
34. Marescaux J, Leroy J, Gagner M, Rubino F, Mutter D, Vix M et al (2001) Transatlantic robot-assisted telesurgery. Nature 413(6854):379–380. https://doi.org/10.1038/35096636
35. Choi PJ, Oskouian RJ, Tubbs RS (2018) Telesurgery: past, present, and future. Cureus 10(5). https://doi.org/10.7759/cureus.2716
36. Poudel A, Nissen LM (2016) Telepharmacy: a pharmacist's perspective on the clinical benefits and challenges. Integr Pharm Res Pract 5:75–82. https://doi.org/10.2147/iprp.S101685
37. Win AZ (2017) Telepharmacy: time to pick up the line. Res Soc Adm Pharm 13(4):882–883. https://doi.org/10.1016/j.sapharm.2015.06.002
38. Kimber MB, Peterson GM (2006) Telepharmacy enabling technology to provide quality pharmacy services in rural and remote communities. J Pharm Pract 36(2):128–133. https://doi.org/10.1002/j.2055-2335.2006.tb00588.x
39. Baldoni S, Amenta F, Ricci G (2019) Telepharmacy services: present status and future perspectives: a review. Medicina 55(7):327. https://doi.org/10.3390/medicina55070327
40. Le T, Toscani M, Colaizzi J (2020) Telepharmacy: a new paradigm for our profession. J Pharm Pract 33(2):176–182. https://doi.org/10.1177/0897190018791060
41. Clifton GD, Byer H, Heaton K, Haberman DJ, Gill H (2003) Provision of pharmacy services to underserved populations via remote dispensing and two-way videoconferencing. Am J Health Syst Pharm 60(24):2577–2582. https://doi.org/10.1093/ajhp/60.24.2577
42. Boon AD (2007) Telepharmacy at a critical access hospital. Am J Health Syst Pharm 64(3):242–244. https://doi.org/10.2146/ajhp060392
43. Lam AY, Rose D (2009) Telepharmacy services in an urban community health clinic system. J Am Pharm Assoc (2003) 49(5):652–659. https://doi.org/10.1331/JAPhA.2009.08128
44. Sankaranarayanan J, Murante LJ, Moffett LM (2014) A retrospective evaluation of remote pharmacist interventions in a telepharmacy service model using a conceptual framework. Telemed J E Health 20(10):893–901. https://doi.org/10.1089/tmj.2013.0362
45. Soman A (2011) Cloud-based solutions for healthcare IT. CRC Press

46. Damiati SA (2020) Digital pharmaceutical sciences. AAPS PharmSciTech 21(6):206. https://doi.org/10.1208/s12249-020-01747-4
47. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H (2019) Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 9(4):e1312. https://doi.org/10.1002/widm.1312
48. Makary MA, Daniel M (2016) Medical error the third leading cause of death in the US. BMJ 353. https://doi.org/10.1136/bmj.i2139
49. Doarn CR, Merrell RC (2014) Telemedicine and e-health in disaster response. Telemed J E Health 20(7):605–606. https://doi.org/10.1089/tmj.2014.9983
50. Rockwell KL, Gilroy AS (2020) Incorporating telemedicine as part of COVID-19 outbreak response systems. Am J Manag Care 26(4):147–148. https://doi.org/10.37765/ajmc.2020.42784
51. Latifi R, Tilley EH (2014) Telemedicine for disaster management: can it transform chaos into an organized, structured care from the distance? Am J Disaster Med 9(1):25–37. https://doi.org/10.5055/ajdm.2014.0139
52. Bashshur RL, Shannon GW, Smith BR, Alverson DC, Antoniotti N, Barsan WG et al (2014) The empirical foundations of telemedicine interventions for chronic disease management. Telemed J E Health 20(9):769–800. https://doi.org/10.1089/tmj.2014.9981
53. Moazzami B, Razavi-Khorasani N, Moghadam AD, Farokhi E, Rezaei N (2020) COVID-19 and telemedicine: immediate action required for maintaining healthcare providers well-being. J Clin Virol 126:104345. https://doi.org/10.1016/j.jcv.2020.104345
54. Xiong W, Bair A, Sandrock C, Wang S, Siddiqui J, Hupert N (2012) Implementing telemedicine in medical emergency response: concept of operation for a regional telemedicine hub. J Med Syst 36(3):1651–1660. https://doi.org/10.1007/s10916-010-9626-5
55. Portnoy J, Waller M, Elliott T (2020) Telemedicine in the era of COVID-19. J Allergy Clin Immunol Pract 8(5):1489–1491. https://doi.org/10.3329/birdem.v10i0.50971
56. Dávalos ME, French MT, Burdick AE, Simmons SC (2009) Economic evaluation of telemedicine: review of the literature and research guidelines for benefit–cost analysis. Telemed J E Health 15(10):933–948. https://doi.org/10.1089/tmj.2009.0067
57. Ruckdäschel S, Reiher M, Rohrbacher R et al (2006) The role of health economics in telemedicine. Dis Manag Health Outcomes 14:3–7. https://doi.org/10.2165/00115677-200614001-00003
58. Burney SA, Mahmood N, Abbas Z (2010) Information and communication technology in healthcare management systems: prospects for developing countries. Int J Comput Appl 4(2):27–32. https://doi.org/10.5120/801-1138
59. Heinzelmann PJ, Lugn NE, Kvedar JC (2005) Telemedicine in the future. J Telemed Telecare 11(8):384–390. https://doi.org/10.1177/1357633X0501100802
60. Frey KA, Bratton RL (2002) Role of telemedicine in the health care deliver system. J Am Board Fam Pract 15(2):170–171
61. Nesbitt TS, Marcin JP, Daschbach MM, Cole SL (2005) Perceptions of local health care quality in 7 rural communities with telemedicine. J Rural Health 21(1):79–85. https://doi.org/10.1111/j.1748-0361.2005.tb00066.x
62. Williams AM, Bhatti UF, Alam HB, Nikolian VC (2018) The role of telemedicine in postoperative care. Mhealth 4. https://doi.org/10.21037/mhealth.2018.04.03
63. Hibbert D, Mair FS, May CR, Boland A, O'Connor J, Capewell S et al (2004) Health professionals' responses to the introduction of a home telehealth service. J Telemed Telecare 10(4):226–230. https://doi.org/10.1258/1357633041424386
64. Spetz J, Given R (2003) The future of the nurse shortage: will wage increases close the gap? Health Aff (Millwood) 22(6):199–206. https://doi.org/10.1377/hlthaff.22.6.199
65. Lamothe L, Fortin J-P, Labbe F, Gagnon M-P, Messikh D (2006) Impacts of telehomecare on patients, providers, and organizations. Telemed J E Health 12(3):363–369. https://doi.org/10.1089/tmj.2006.12.363

66. Hailey D, Ohinmaa A, Roine R (2004) Study quality and evidence of benefit in recent assessments of telemedicine. J Telemed Telecare 10(6):318–324. https://doi.org/10.1258/1357633042602053
67. Evans RG, McGrail KM, Morgan SG, Barer ML, Hertzman C (2001) Apocalypseno: population aging and the future of health care systems. Can J Aging 20(S1):160–191. https://doi.org/10.1017/S0714980800015282
68. Saliba V, Legido-Quigley H, Hallik R, Aaviksoo A, Car J, McKee M (2012) Telemedicine across borders: a systematic review of factors that hinder or support implementation. Int J Med Inform 81(12):793–809. https://doi.org/10.1016/j.ijmedinf.2012.08.003
69. Lupton D, Maslen S (2017) Telemedicine and the senses: a review. Sociol Health Illn 39(8):1557–1571. https://doi.org/10.1111/14679566.12617
70. Combi C, Pozzani G, Pozzi G (2016) Telemedicine for developing countries: a survey and some design issues. Appl Clin Inform 7(4):1025. https://doi.org/10.4338/ACI-2016-06-R-0089
71. Ohannessian R, Duong TA, Odone A (2020) Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 6(2):e18810. https://doi.org/10.2196/18810
72. Kuznetsov D (2016) Development of telemedicine system. MATEC Web of Conferences: EDP Sciences, p 01074

Bioinformatics in Drug Design and Delivery 21

Rishav Kumar, Meenakshi Gupta, and Maryam Sarwat

Contents

R. Kumar · M. Gupta · M. Sarwat (✉)
Amity Institute of Pharmacy, Amity University, Noida, Uttar Pradesh, India
e-mail: msarwat@amity.edu

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_21

Abstract

Bioinformatics is the one-stop solution to the problems in drug designing like large capital investments, human resources, technological expertise, regulations, and deadlines. Even after spending a large amount of money and effort, sometimes a drug fails to make up to the market. Genome sequence achieved by bioinformatics is very important because it helps scientists to find out the genes more easily and quickly. Structural biology determines the 3-D structure of biomolecules associated with cancer progression and development. Bioinformatics has extensive use of the genomic technique in diagnosis and management for pathogenic diseases like fungal, bacterial, and viral infections. Pathogen identification, strain typing, resistome analysis, and identification of virulence factors are some of the genomic approaches. Bioinformatics is used to overcome drug resistance by locating the antibiotic resistance genes present on a plasmid from the entire genome sequences of pathogens. It has therefore the revolutionary solution for the worldwide problem of antibiotic resistance. Its capacity to model the living cell and protein docking has enabled effective drug strategies to find the solution. Using computational biology and homology modeling biological annotations can be analyzed. This chapter follows the different methods regarding the role of bioinformatics in the current drug design and delivery system.

Keywords

Genomics · Proteomics · Ribosome profiling · Personalized medicine · Transcryptomics · Phylogenetics

Chapter Objectives

- To understand the fundamentals of bioinformatics and its application
- Basic understanding of protein structures and their visualization
- To increase the concepts of genome mapping, pharmacogenomics, and ribosome profiling
- Applications of next-generation sequencing
- Applications of bioinformatics in Personalized medication

21.1 Introduction

Computation, mathematics, and statistics are the quantitative tools that are requisite for modern research in medical science. These tools help to record, analyze, and integrate the raw information to mysteries of the perplexing molecular world of biology. It took Gregor Mendel a long time to find out the fundamentals of genetics. Imagine what else he could have done with the help of fast, accurate, and highly specialized features of bioinformatics. The study of molecular biology, genome, chromosomal structure, computational and mathematical tools are now an integral part of research today [1]. The quantitative analysis of information related to biological molecules with the help of a computer is called bioinformatics [2]. It is an interface between the biological and computer sciences. This technology is used to retrieve, manipulate, store, and distribute information from macromolecules like protein, DNA, and RNA [3]. Computers are used to process this information because the analyses are mathematically complex and highly repetitive. For example, obligation of file allocation table (FAT) during the formatting of computer hard drive at a high level and then consequent loading of an operating system (OS).

The computational tools and databases and then using these in better understanding of the living world. The development of the tool encompasses software for sequencing, functional and structural analysis, and curating the biological databases. Bioinformatics is a vibrant and constantly evolving field. Cheminformatics, systemics biology, biostatistics, computational languages, and database management systems are some of the fields under the umbrella [4].

The data analysis in bioinformatics steps includes collection of data, check for quality, and cleaning, processing, modeling, visualization, and reporting. These steps mentioned are done in a linear sequential fashion; however we get to the previous step and then repeat the steps with different parameters. Generally, data analysis is done by going through the same process over and over again through the combination of the following:

(a) One dataset can be used to answer the query of other related questions such as a single dataset can be used to obtain information on phylogenetic as well as genomic analysis.
(b) It deals with later found data quality issues.
(c) To inclusion of new data sets going to be analyzed.

Bioinformatics mining is carried out using the database for annotation, visualization, and integrated discovery. More advanced bioinformatics analysis can also be performed using the IPA (Ingenuity Pathway Analysis) software. These tools and software are employed for better understanding of the living cell, and its functions at the molecular level. All the cellular functions are performed by the proteins which are synthesized in the cell through a process called central dogma from RNA and DNA, respectively, and the sequence decides the capabilities [5]. Having the data for a protein structure provides a deeper state of understanding of how a protein works,

which can help us to create assumptions about how to affect it, control it, or modify it. Hence, taking an approach from the structural point of view and using sequence is proved to be useful. For example, recognizing a protein's structure could allow the designing of site-directed mutations with the intention of changing function. Or you could envisage molecules that bind to a protein.

Bioinformatics is not only limited to the subjects like molecular biology and genomics but also influences biotechnology and other biomedical science. It has other applications in drug design for higher potency and lower side effects, forensic DNA analysis, and agriculture technology, not only in drug design but also in drug development. Bioinformatics tools are very effective in the prediction, analysis, and interpretation of clinical and preclinical findings [6]. It helps spot leads for synthetic drugs. Bioinformatics with genomics has revolutionized the healthcare system for the development of personalized medicine. Early diagnosis, effective treatment, and potent mutation harms can be accessed using the patient's genome. Plant genome database and gene expression profile can help to get higher yields and developing disease-resistance crops.

21.2 Software Used in Bioinformatics

The increased dependence of scientists on advanced software and computational methods for research purposes paves the way for developing newer and user-friendly tools. Multiple software (Table 21.1) have already been developed and have been utilized by the scientific community. These software help researchers in interpreting and analyzing biological data. Clinical data management and statistical analysis can also be used in the drug development process by these software. Along with this, these software cover a wide range of functions such as genome analysis, prediction of binding models, and analysis of protein structures.

With the understanding of bioinformatics, we must also realize its limitations so that it does not add to the overexpectation for the output. Bioinformatic tools are completely based on the extraction of information on the cellular and micromolecular levels. Incompleteness of the raw data and flawed analysis may give wrong outputs. Moreover, the predictions made by bioinformatics are not proof of concepts in a certain way. The quality of prediction depends on the class of data and the configuration of algorithms used [7]. In case of some incorrect sequence and wrong annotations, the results may be misleading. This is the reason we need to look forward to the realistic perspective as well.

21.3 Structural Bioinformatics

Structural bioinformatics aided in lead optimization and identification of pre-defined target in which they have well-recognized roles; they can now contribute to lead discovery, exploiting high-throughput methods of structure fortitude that provide powerful approaches to a screening of fragment binding. It also includes the

Table 21.1 Available online software for protein and genome-based data prediction

S. no.	Name of software	Application	URL
1.	ChemGenome 2.0	Ab initio gene prediction software, which find genes in prokaryotic genomes in all six reading frames	http://www.scfbio-iitd.res.in/chemgenome/chemgenomenew.jsp
2.	3DStructural Superimposition (3D-SS)	Web-based interactive computing server, principally designed to support researchers, to lay over two or several three-dimensional protein structures	http://cluster.physics.iisc.ernet.in/3dss/
3.	Sanjeevini	A complete drug design software	http://www.scfbio-iitd.res.in/sanjeevini/sanjeevini.jsp
4.	StrGen	Captures the sequence and/or secondary structure information as input and presents the file with Ramachandran values for helix, sheet and loop dihedrals	http://www.scfbio-iitd.res.in/software/proteomics/p2pdb.jsp
5.	Protein Structure Analysis Package (PSAP)	A suite based on web for protein structure analysis	http://iris.physics.iisc.ernet.in/psap/
6.	Bhageerath	Protein Structure Prediction Software	http://www.scfbio-iitd.res.in/bhageerath/index.jsp
7.	ActiveSite prediction	Active Site Prediction of Protein server computes the cavities in a given protein	http://www.scfbio-iitd.res.in/dock/ActiveSite.jsp
8.	Biomolecules Segment Display Device (BSDD)	Shared web-based display tool; created to search for and visualize a user-defined motif or fragment among the protein structures which are available in the Protein Data Bank (PDB)	http://iris.physics.iisc.ernet.in/bsdd/

structural study of the molecules like proteins, amino acids, formation of peptide linkages, dihedral angles, stabilizing forces involved, secondary, tertiary structures, and protein structure database because these determine the functions of the basic structural units in a living cell.

21.3.1 Protein Structure Basis

Proteins are the working modules in a living cell and most of the cellular functions. Protein structure study is vital to understand functions and comportment for the specific proteins. Fundamentally, proteins are polypeptides that are formed by the attachment of amino acids together with peptide bonds [8]; some of the amino acids have been discussed in Table 21.2.

Rotational angles describe the folding of polypeptide around the primary chain bonds like Φ and ψ angles. The confirmation of the preferred proteins regulates the

Table 21.2 Some basic amino acids categorized by their common side chain qualities

S. no.	Group of amino acids	Names of amino acids	Three- and one-letter code	Main functions of amino acids
1.	Non-polar and small	Alanine Glycine Proline	Ala, A Gly, G Pro, P	Non-reactive in a chemical reaction: Pro and Gly interrupt regular secondary structures
2.	Polar and small	Cysteine Threonine Serine	Cys, C Thr, T Ser, S	Providing as post-translational modification sites and contributing in active sites of enzyme or binding metal
3.	Polar and large	Asparagine Glutamine	Asn, N Gln, Q	Hydrogen bonding, participation in enzyme active sites
4.	Polar (basic) and large	Lysine Arginine Histidine	Lys, K Arg, R His, H	Observed on the surface of globular protein; His participate in enzyme catalysis
5.	Polar (acidic) and large	Aspartate Glutamate	Asp, DGlu, E	Observed on the surface of globular protein which provides salt bridges
6.	Non-polar and large (aliphatic)	Leucine Isoleucine Valine Methionine	Leu, L Ile, I Val, V Met, M	Take part in hydrophobic reactions
7.	Non-polar and large (aromatic)	Tryptophan Phenylalanine Tyrosine	Trp, W Phe, F Tyr, Y	Offering sites for aromatic packing interactions

degree of rotation. Ramachandran plot specifies the permissible Φ and ψ angles in the protein. Primary, secondary, tertiary, and quaternary are the different levels of protein formation, in which the primary structure is the sequence of amino acid residues, secondary is the formation of α-helices and β-sheets by repeated main-chain conformation, tertiary structure is the 3-dimensional conformation of polypeptide chains, and quaternary structure is a very complex arrangement of multiple polypeptide chains [9]. These structures of the protein are stabilized by certain fundamental forces like electrostatic interactions, van der Waals forces, and hydrogen bonds including sulfur bonds at some positions. The structure of proteins varies largely from integral membrane proteins to soluble globular proteins. These structures are then ascertained by techniques like NMR spectroscopy and X-ray crystallography [10]. These complimentary methods have both advantages and disadvantages. Finally, these solved structures are stored in Protein Data Banks (PDB), through which the details in these structures can be described via appropriate format. But the original format available in PDB has restricted capacity and sometimes it is not easy to be interpreted by computer software [11]. Thankfully, we have developed formats like molecular modeling database (MMDBand macromolecular crystallographic information file (mmCIF)) which can retrieve the information separately and these are highly parsable by software like ADIT validation server, checkCIF, CIFIO, cod-tools, CYCLOPS, encipher, and PLATON [12].

21.3.2 Protein Structure Visualization, Comparison, and Classification

After the protein structure is resolved, cartesian coordinates are used to create the 3-D view of the protein structure. Pre visual-software era, physical models were used to represent the structure using balls and rods. But we have developed highly specified computer graphic programs to visualize, compare, and analyze the complicated 3-dimensional structures which developed a deeper understanding of proteins [13]. Many visualization programs are now available for this purpose. For simple structure viewing and complex structure imaging, we have now developed lightweight web-based programs and stand-alone programs, respectively. The comparative study of protein structure helps to recognize evolutionary relationships which are helpful in classification and tracing the evolution of protein prediction techniques. These comparative algorithms are divided into three branches. Intermolecular method, used for the comparison of matrix against another molecule by the construction of an inter-residue distance matrix within a molecule. DALI, a comparison web server, is the most widely used among all comparison algorithms. Class, Architecture, Topology, and Homologous (CATH andStructural classification of proteins (SCOP)) are the schemes for classification that are also widely used despite some variations have common grounds as well [14, 15]. The results can be compared from both these systems to obtain correct structure [16–18].

21.3.3 Protein Secondary and Tertiary Structure Prediction

For the secondary structure predictions, three generations of algorithms were developed. Ab initio based algo was the first gen that examined residue propensities which further had three categories: helices, strands, and coils [19, 20]. Since the propensities were obtained from small databases, it allowed the development of second gen. algorithms for growing databases. The evolution of third gen. algorithms brought revolutionary changes as it uses multiple sequence alignment information, shown in Fig. 21.1, which took the long-range intra-protein interactions for consideration. By combining neural networks and other complex algorithms, the efficacy of prediction was improved rationally [21]. To get highly accurate predictions, combined results from various top third gen. algorithms must be considered. Since the prediction of NMR structure or crystallography of globular proteins are complicated so secondary structure prediction is more common. Evolutionary information, hydrophobicity, and neural networks are used for the prediction of transmembrane segments [22, 23].

21.3.4 RNA Structure Prediction

The functions of RNA in the cell cannot be known until we develop an understanding of its structure. Therefore, to solve this problem a large number of prediction

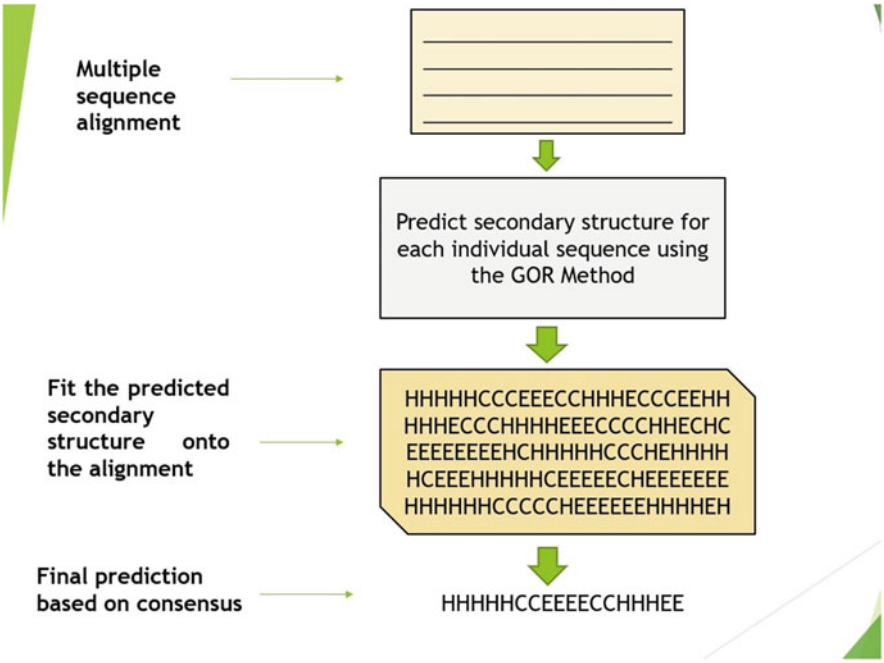


Fig. 21.1 Schematic representation of secondary structure prediction using multiple sequence alignment

algorithms have been developed and are still developing [24]. Presently, the prediction for RNA is mainly focused on the secondary structure for the reason being prediction difficulties in the tertiary structure [25]. Ab initio or comparative is the classification for secondary structure prediction based on energetic (single sequence query) and multiple sequence query, respectively, which makes the ab initio method more accurate as compared to the comparative one. This method has also been used for the gene prediction program. Indeed, the obligation for a unique set of the homologous sequence is an evident drawback. More research is still being conducted focusing to overcome the current drawbacks in RNA structure prediction [26].

21.4 Genomics and Proteomics

Methodical study of genes, their functions, and their interactions is called Genomics. Data assembly tools are used for the analysis of a large number of genes simultaneously which therefore characterize genomic studies. Genomics includes genome sequencing, mapping, and analysis till comparative analysis of the genome.

Structural genomics and functional genomics are two categories of genomics. Former includes initial analysis, construction, and mapping of genomic structure, gene identification, gene annotation, and comparative study of genome structure

whereas functional genomics refers to global gene expression analysis and function of the gene in a genome.

21.4.1 Genome Mapping, Assembly, and Comparison

Without using the sequence data, mapping the genome by comparative positions of genetic markers is a weak approach. The genome can only be described fully by a complete genome sequence. A full shotgun or hierarchical approach can be used for the whole-genome sequence. The former needs very extensive computing powers for the assembly, but the latter is also not efficient due to physical mapping requirement [27]. For genome assembly programs, ARACHNE and EULER are some of the best technology [28–30]. Genome annotation consists of finding the gene, assigning functions to the genes. Functional assignment relies upon homology researching and literature information and data mining. Even the human genome is still not completely known; it is supposed to be in the range of other eukaryotes. Genome economy is defined as when a large number of proteins are synthesized than the availability of genes for their coding. Alternative splicing and exon shuffling are the mechanisms behind this genome economy. For genome comparison, so many programs have been introduced. LAGAN and BLASTZ are some of the best in performance based on accuracy and speed [31, 32]. The comparison of genome order helps to find out the potent operons and assign them putative function. Different types of genome mapping are done to describe a genome like cytological map, genetic map, physical map, etc. as shown in Fig. 21.2. Cytologic maps are generated microscopically. Genetic maps are achieved through genetic crossing trials in which chromosome recombination is analyzed. Physical maps are obtained from overlapping clones which are detected by hybridizing the clone fragments with common probes.

21.4.2 Functional Genomics

Functional genomics is based on the gene functions at the genome level via high-throughput approaches. High throughput refers to the simultaneous analysis of all genes in the genome [33, 34]. The frame and core of the functional genomics consist of transcriptome analysis by ESTs (expressed Sequence Tags), SAGE (Serial Analysis of gene expression), and DNA microarrays [35–38]. It is essential to understand the gene interaction and the regulation at the whole-genome level. The high-throughput DNA and SAGE provide the quantitatively sound measure of gene expression. There are three techniques for the study of global gene expression, but DNA microarray is the most widely accepted method. It can fetch additional information which cannot be obtained by the other techniques [39–42]. Like other techniques, this too has limitations. Since it is a multi-step procedure, the chances of error at each step also increase. After the hypothesis is obtained by this method, the traditional and molecular biological methods are used to verify it.

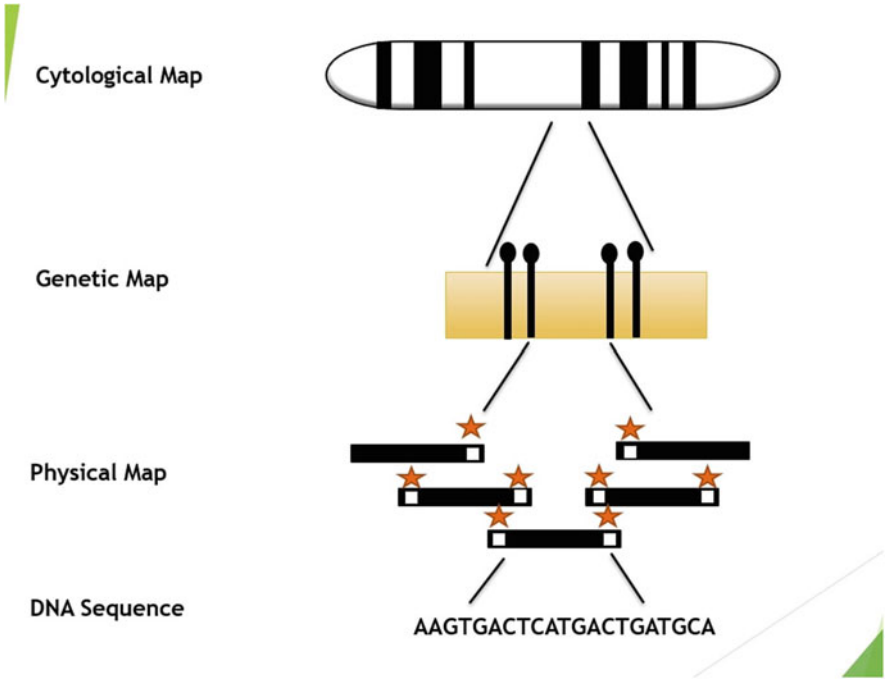


Fig. 21.2 Outline of various genome maps relative to the genomic DNA sequence. The maps correspond to various levels of resolution to describe a genome using genetic markers

21.4.3 Proteomics

Proteomics is the analysis of proteins, protein complexes, their localization, their interactions, and posttranslational modifications. Cellular functions can better be understood by the analysis of protein expression at the proteome level [43]. MS and 2D-PAGE that can be used for the identification of protein at a large scale are the conventional experimental approaches [44]. The process of identification requires the incorporation of bioinformatics tools to find out the databases for complimenting peptides. Protein microarrays and DIGE (Difference Gel Electrophoresis) are included in recent protein expression profiling methods [44–46]. Very often the sequence-based prediction out-turns in a large rate of non-true positives, which therefore results in very limited knowledge of the required structural features for modification. The use of a SVM classifier (Support Vector Machine) is certainly one step ahead to get more positive rates in prediction [47]. Proteomics has also application in the protein subcellular localization signals. There are many web tools like SignalP, PSORT, and TargetP that provide an accurate prediction of the signal peptides [48]. For the general determination of protein interactions, yeast two-hybrid experimental methods are used. The recent methods for prediction are based on the principles like sequence homology, phylogenetic information, gene

linkage pattern, and domain fusion. The prediction of protein interactions is of great importance in solving the functional interpretation of genes, encoded proteins, and genome annotation.

21.4.4 Pharmacogenomics

It is the study of how the genetic inheritance of an individual affects the body's response to drugs. It deals with the influence of the genetic variations on the drug responses by relating the drug efficacy or toxicity with gene expressions or polymorphism. It includes the study of varying targeted genes or functionally related genes for variability in the response of the drug. It can use genetic information to help the choice of drug and dosage regimen or dose on an individual basis. The changes in the efficacy are due to the genetic polymorphism in drug-metabolizing enzymes, receptors, drug targets, and transporters [49]. The main objective of pharmacogenomics is to maximize drug efficacy, minimize toxicity, aid in drug developments and patient prediction for faster response to intervention.

21.4.5 Ribosome Profiling and Its Applications

Ribosome profiling, which is also known as Ribo-Seq (Ribosome sequencing), is a method developed to monitor translation in vivo. It is currently based on deep sequencing of mRNA fragments which are ribosome protected [50, 51]. It facilitates the discovery of how by regulating gene expression, understanding of complicated biological processes which is also important to unfold the mechanisms of protein synthesis and provide an organized approach for experimental annotation of coding regions. It can also help to find out the positions of ribosomes that can be used to generate information about the translation mechanism or identify translated open reading frames (ORFs) [52]. It can further be used for the discovery of new translational products of a large variety. A simple mechanism of ribosome profiling is shown in Fig. 21.3.

21.5 Transcriptomics Analysis

Study of transcriptome, i.e., RNA Transcripts that are produced by the genome, under specific circumstances or in a specific cell. RNA in any of its forms is fundamentally defined as transcriptomics. It involves sets of all RNA molecules like mRNA, tRNA, rRNA, and the other non-coding RNAs produced in a cell or its population.

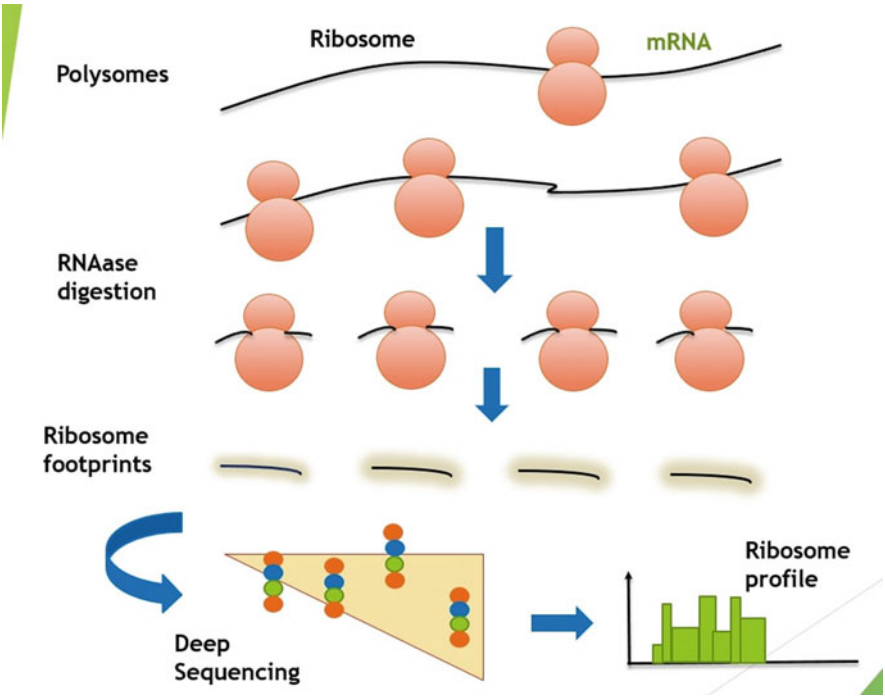


Fig. 21.3 Schematic representation of different steps in Ribosome Profiling

21.5.1 Aim and Scope

It can help catalog all species of transcripts including the non-coding RNA and small RNAs. It is used for determining the transcriptional structure of genes, posttranslation modifications, and splicing patterns. Also, it can measure the varying expression levels of each transcript under different conditions. It is also referred to as expression profiling in which we examine the expression levels of mRNA in a cell population. Transcriptome also indicates the genes that are expressed at any given time, but except for mRNA degradation phenomena like transcriptional attenuation [53].

21.5.2 Hybridization-Based Approaches

This method of technology has been introduced in many variants but only two of them remain to be most popular: two-color microarrays (cDNA or two-channel) and one-color microarray (oligonucleotide or one-channel) [54]. The former is based upon the competitive hybridization of two samples each one of which has been labeled with a fluorescent dye. After the hybridization is done, the array is then exposed to the laser light which may be red or green. The fluorescence emitted by the

array is proportional to the quality of RNA. The produced image is scanned yielding after some correction. Whereas one-color microassays are based on the RNA of the single sample which is labeled with the fluorescent dye. After exposure to scanner and laser light, the concentration of each location is measured yielding a value that represents an absolute measure of expression. Gene expression microassay has been proved to be very useful in providing an overall view of changes in the expression of gene between two or more biological conditions. However, as the understanding of expression of gene has progressed, it has also become apparent that more complex events than transcription and splicing occur within specific genes in a sample.

21.5.3 Sequence-Based Approaches

In RNA-seq transcriptomics sequencing individual cDNAs produced from the target RNA reinstates the hybridization of nucleotide probes. Evolving methods for these completely quantitative transcriptomic analyses have the scope in overcoming the drawbacks of microarray technology and there is an ongoing discussion about whether sequencing approaches may replace microarray in the meantime or even it may do so in the short term [55]. A process going parallelly on massively, next-generation sequencing generates hundreds of megabases to gigabases of nucleotide sequence output in a single instrument run, depending upon the platform.

21.5.4 Microarray Chips and Application

It is a technology in which thousands of hundreds of nucleic acids are attached to the surface and then is the relative concentration of nucleic acid sequence is measured in a mixture by hybridization and successive detection of hybridization events. By the use of this technique, the presence of even one genomic or cDNA sequence in 100,000 or maybe more sequences can be screened in a single hybridization. DNA microarrays are solid supports, generally made up of silicon or glass upon which the DNAs are attached in an organized grid; each of the spots presents called a probe is representative of a single gene. DNA chips, gene arrays, and biochips are synonyms for microarray chips. This process requires a DNA chip, target sample which is fluorescent-labeled, enzymes, fluorescent dyes, probes, and scanner. There are two types of microarrays: cDNA-based and oligonucleotide-based. The cDNA type is prepared by using cDNA and is called cDNA chips or cDNA microarray [56]. Using PCR, cDNA is amplified and then immobilized on a solid nylon filter support. The DNA chips are used in many areas like expression profiling of gene, discovery of novel drugs, diagnostic and genetic engineering, alternative splicing detection, functional genomics, DNA sequencing, proteomics, and toxicological research. One of the examples on how DNA microarray assay is performed for the detection of cancerous cells is shown in Fig. 21.4. Apart from its widely recognized uses, the main disadvantage of DNA chips is that they may be very expensive to create, it also

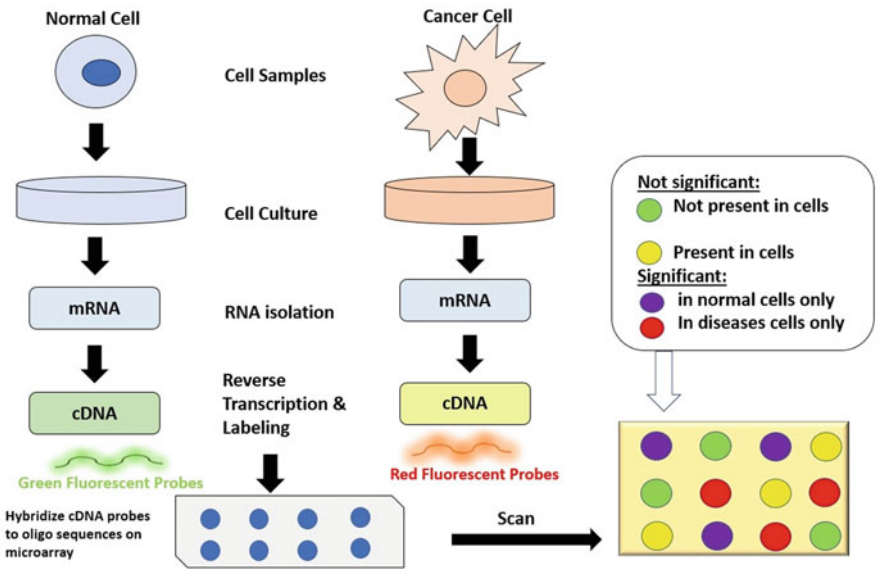


Fig. 21.4 Animated representation of a multistep procedure of a DNA Microarray assay experiment for the detection of cancerous cell and subsequent data analysis (color plates)

may require a long time for analysis which turns out to be very complex, DNA chips also have a very short shelf life which is another major difficulty.

21.5.5 Next-Generation Sequencing (NGS)

It is a process to figure out the imperative of DNA nucleotides, or their bases (A, T, G, C) in a particular genome that make up any organism's DNA. First gen sequencing is done in large gels or capillary tubing limits scale and it sequences many identical molecules. Second gen sequencing includes the sequence of millions of clonally amplified molecules per run [57]. It has stepwise sequencing chemistry. Next-generation sequencing is a high-throughput DNA sequencing technique. It uses micro- and nanotechnologies and massively parallel sequencing. It also sequences thousands of sequences at once. It produces a massive amount of data. Its workflow includes creating DNA fragments, then added platform-specific adapter sequence to every fragment. Advantage of NGS includes colony picking, no in vivo cloning, transformation, high degree of parallelism, reduced sample size, low reagent cost, and less time. Applications of NSG include mutational discovery, Transcriptome analysis, enabling metagenomics, discovering non-coding RNAs, gene regulation analysis, etc. NSG has transformed the way we carry out molecular biology and genomic studies. It has also permitted us to sequence and annotates the genome at a faster rate [58].

21.6 Bioinformatics in Treatment

The methods and tools using NGSs data are extensively used for the diagnosis and monitor of infectious diseases. Techniques like molecular identification, microbiome research, and antimicrobial resistance analysis are of great importance to find out the information on diseases. Therefore, the use of bioinformatics tools and methods for pathogen detection and typing is very important in treatment development.

21.6.1 Pathogen Identification and Strain Typing

The tools are widely used in the characterization, identification, and typing of all kinds of the pathogen. Genomic approaches are taken for the diagnosis and management of infective pathogens like bacteria, viruses, and fungus. Bioinformatics has application in the detection of virulence factors, strain typing, and resistome analysis [59]. NGS has also potency for pathogen identification from human specimens using the technique like whole-genome sequencing (WGS) [60]. Analysis of WGS and Ribosomal (rRNA) gene sequencing are very common in recent times. Tools like lasergene, CLCbio, and Geneious are most popular for sequence assembly, analysis, and microbiome studies. These tools are also used for the detection and removal of amplification-derived chimeric sequences. To facilitate the accurate identification of bacterial pathogens many comprehensive databases have also been compiled like "Greengenes" which has 1,049,126 aligned 16s rDNA records. For a better understanding of the quantitative understanding of microbial population, automatic analysis platforms like MG-RAST have been introduced as servers [61]. These servers have different options like quality control, upload, comparative analysis, and automated annotation. Even for the testing of novel bacterial pathogens, tools like PathogenFinder 1.1 are there which helps in the pathogenicity test. It has very high accuracy for prediction. Machine learned webservers like PaPrBaG is a very recent approach.

21.6.2 Antimicrobial Resistance

For understanding the antimicrobial resistance as well as resistance factors, the need for more accurate, fast detection is required. The presence of antibiotic resistance loci can be investigated by gene annotation services on genome contigs. Specialized search tools like Antibiotic Resistance Gene Search, ARBD (Antibiotic Resistance Genes Database), and Genome Feature Finder can be used to find the presence of antibiotic resistance loci in newly isolated bacterial pathogen [62]. Webservers like ResFinder 2.1 find the acquired antimicrobial resistance genes or chromosomal mutations in partially sequenced isolated bacteria. The approach of bioinformatics can be seen in both ways; clinical settings as well as to understand the molecular mechanism of AMR. And therefore, it can generate the patterns of phenotype-genotype data. The combination of machine learning algorithms and modern

molecular methods may give a better insight into the AMR at the molecular level for improved accurate and clinically relevant predictions. Therefore, it can lead to the development of more personalized medication and lesser side effects [63].

21.7 Molecular Phylogenetics

Molecular phylogenetics is a fundamental part of bioinformatics; phylogenetic trees are constructed to rationalize and visualize the convergence and divergence among similar biological sequences through sequence alignment. It is the subject of study of evolutionary relationships of the living organism in tree-like diagrams, which represents the pedigree. It can be studied by fossils that have morphological information, but it has certain limitations in the physical world. The other way is in the form of molecular data which are available in the form of protein sequences and DNA [64]. The continuous and slow accumulation of mutations leads to phenotypic changes in a long time.

21.7.1 Phylogenetics Basics

The basic mechanism of operation behind molecular phylogenetics is based on assumptions like the binary nature of the evolutionary tree and the independent evolution of sequence positions. Topology is defined as the branches of the phylogenetic tree; the number of these branches are proportional to the number of taxa; with a large number of taxa these branches also increase. Gene sequence-based trees do not assure the correlation with the species evolution in particular. Therefore, to extrapolate the result, caution must be taken care of. These phylogenetic trees can also be rooted or unrooted depending upon the genetic knowledge. The elementary step in the phylogenetic construction depends upon the selection of DNA sequence or protein sequence. Although both have their merits and disadvantages, the latter is generally selected in most cases. But, for the recent evolution studies, DNA is the marker of choice. The next step is the conduct of multiple sequence alignment, for the construction of phylogenetic tree accurate alignment must be obtained. Sometimes manual truncation of wrongly aligned regions is also required. Now, the selection of a proper substitution model is required which can provide the estimation for the tree with the consideration of multiple substitution events. Kimura and Jukes-Cantor are the most used nucleotide substitution method while PAM and JTT are commonly used amino acid substitution models [65, 66]. Rate heterogenicity can also be incorporated as an adjustment for the improved estimations.

21.7.2 Multiple Sequence Alignment

This technique is of great importance and has many bioinformatics applications. For the achievement of optimal alignment, so many of the algorithms have been worked

upon. Some of them are exhaustive nature-wise, some are heuristic [67, 68]. For most of the cases the latter one is used since it uses progressive, iterative, and block-based approach. Clustal is an important example of a progressive method that has an adjustable scoring matrix and gap penalty [69]. But this function has a drawback which relates to the early steps of error fixation in computation [70]. To overcome this bug T-coffee and DbClustal have been introduced which generates more sensitive alignment by combining both local and global alignments. The error fixation has been fixed using graphic profiles like in Poa. This is another improvement on the traditional progressive approach. Praline, a better profile than Clustal, is more accurate since it can restrict alignment based upon the information of protein structure. The block-based method identifies regional similarities whereas the iterative profile does repetitive refinement of suboptimal alignment. No alignment program is so advancing which can assure the correct alignment in case of large sequence and high divergence level. The alignment obtained from the automatic alignment profiles usually contains errors; therefore, the best-suited approach is performing alignment by combining multiple alignment programs. Rascal can be used to get increased accuracy after the result is obtained, or it can also be done manually [71].

21.7.3 Phylogenetic Tree Construction

To understand sequence evolution and relationships, molecular phylogenetics is an essential requirement. The accuracy for the phylogenetic analysis and tree-building is dependent on the assumed method. The next step is to know the mechanism of the working and its limitations [72, 73]. The better is the understanding of the assumed method, the better is the efficiency. Later, to obtain a correct picture of the phylogenetic tree, a suitable phylogenetic method is to be selected. The phylogenetic method can be subdivided into character-based and distance-based methods. NJ, UPGMA, and Fitch-Margoliash are part of the distance-based method [74]. The first 2 are cluster-based which are less time taking but not as accurate. The last one is accurate but not as fast. MP (Maximum Parsimony) and ML (Maximum Likelihood) are included in the character-based method. The parsimony principle is easily understood; it is not fast like distance methods. To overcome this drawback and speed up the calculation, heuristics and branch-and-bound search protocols are used. ML is the slowest but has a solid statistical base. Faster algorithms like NJML, Bayesian, GA, and quartet puzzling are used often to overcome the slow computation in ML. It must be noted that phylogenetic tree construction is a sophisticated process and not trivial. Factors like reliability, speed, and accuracy should be taken care of before selecting a particular phylogenetic method. It may also be possible that no method of phylogenetic reconstruction will give the correct tree. And to obtain a minimum error, the use of at least two methods is suggested for any phylogenetic analysis; in addition to this different weighting schemes, rate substitution models, and resampling strategies are recommended.

21.7.4 Bootstrapping and Jackknifing

For the testing of sampling errors in a phylogenetic tree, a statistical technique has been developed called bootstrapping. The mechanism behind this is the repeated sampling by perturbed datasets. Through this, the strength of the original tree can be estimated. The necessity of bootstrapping in a phylogenetic tree arises from the possibility of wrong alignment or incorrect distance measurements. To find out the real strength or reproducibility of a tree, trees are made with perturbed alignment which has some known fluctuations. A real robust phylogenetic tree must possess enough characters to pillar the relationship [75]. Therefore, this kind of analysis talks about the statistical confidence of a tree and its topology. The complimentary technique used along with bootstrapping is a re-sampling procedure called jackknifing. In this method, one-half of the sites in the given dataset are randomly deleted, therefore leaving behind datasets half of the original. Each of the datasets is put through phylogenetic tree construction using the same original method. The pros of jackknifing are it has a shorter time of computing due to shorter sequences and sites are not copied relative to the original dataset.

21.8 Personalized Medication and Cost Reduction

Personalized or Precision medicine (PM) or genomic medicine stands for the specialized medical care for a patient based upon its specific genetic characteristics. The development in high-throughput technologies like microarrays and next-generation sequencing (NGS) has brought revolutionary changes in PM and oncology [76]. The data produced in the enormous amount and the extent up to which we can process the data through the developed bioinformatics systems. It is customized care to each patient's genetic makeup. PM is dedicated to maximum benefit and minimum side effects compared to the standard treatment. It will also reduce the cost of treatment and reduce treatment time and patient suffering. But for the application of PM, each patient's genome has to be converted into digital data which can further be processed, stored, and retrieved when required. It can also open scope for the treatment of some incurable diseases.

21.8.1 Virtual-High-Throughput Screening (v-HTS)

Computational methods are rapidly developed with the increase in novel drug targets and are applied to pace up the drug discovery process. v-HTS is used to identify drugs from the library of drug collection. High-throughput screening (HTS) of small molecules permits the rapid interrogation of the effects of thousands to millions of small molecules in a variety of in vitro and cell-based assays. On the availability of structural data, v-HTS is done by either ligand-based screening methods or receptor-based screening. In the receptor-based method, potent candidates are searched using the 3D structure of target receptors which can then modulate the function of the

target receptor. The selected compound is then docked on the receptor binding sites with better electrostatic bonds than other compounds. In some cases when structural information of the target receptor is not available, the ligand-based method is applied. Substructure searching, 3D shape matching, pharmacophore matching, and similarity are used to recognize the similar structure to known inhibitors. Rigid protein docking may be irrelevant sometimes due to the conformational changes while ligand binding [77, 78].

21.8.2 Dosage Regimen to Reduce Treatment Time

Usage of data for pharmacokinetic and pharmacodynamic studies is used to reduce the treatment time by designing a more efficient dosage regimen system. Pharmacokinetic studies have recently proved the impact to reduce the human trial, cost, and time in the drug development. Pharmacokinetic modeling is used to design dosage regimens to reduce the treatment time as well as the cost. Opdivo (nivolumab), a Bristol-Myers Squibb's product, is the first and only FDA-approved PD-1 inhibitor to offer every 4-week dosing, a product used in different cancers. It is also approved for shorter 30-min infusion reducing previous infusion time in half. This is an example of how bioinformatics can bring a revolutionary change in drug designing.

21.8.3 Adverse Drug Reaction

Many existing ADR prediction studies have focused on examining the chemical properties of drug molecules. Although the mechanisms of UAW are complex and may not be well understood, machine learning techniques are hopeful solutions for understanding and analyzing these complex problems. In general, the basic steps of ADR prediction can be divided into two steps based on structural information. Firstly, each one of the drug molecule is characterized in an appropriate characteristic vector based on its chemical structure. Secondly, a machine learning algorithm is applied to the resulting feature space to predict ADRs [79]. So far, most of the available studies have focused on the second stage or the development of methods to improve predictive power. However, there is still relatively little research on the representation of drug molecules across several useful features and the interpretation of their effects on final ADR predictions. It should be noted that the search for specific drug molecule substructures associated with UAW can be exceptionally useful to find out the mechanism of action of the drug and can therefore be used in the early stages of drug design [80].

21.8.4 Drug Target Validation

Genomics and proteomics technologies have brought about an exemplar change in the drug discovery process, with bioinformatics playing a key position in using

genomic, transcriptomic, and proteomic data to gain insight into the molecular mechanisms underlying disease and to identify potential drug targets [81]. For the discovery of a drug, target identification is an essential step. To achieve this, many computational models are developed, and it also predicts target association on a broader scale. Web servers and biological databases are coordinated, and machine learning is used to make the process fast and accurate. Supervised and semi-supervised models have also been taken into consideration. The future majors of network-based drug discovery rely upon genome sequencing, personalized medication, and cancer hallmark network. The unique concept of drug discovery by using bioinformatics is the design of selective ligand for action on individual drug targets. The development in this area is creating the future of drug discovery [82, 83].

21.9 Conclusion

Bioinformatics focuses on identifying and evaluating some of the new or most important resources (i.e., information and software) that are critical in almost every aspect of drug discovery, drug testing, and drug development. It not only has a role in handling large data volumes, but also to predict, analyze, or facilitate clinical and preclinical findings. The clustering algorithms of bioinformatics can be employed in observing the functioning of genes or protein in healthy vs. diseased states and therefore may be utilized for the diagnosis and treatment of various diseases. The functions of the basic structural units in a living cell can be determined by the structural bioinformatics. Formats like molecular modeling database (MMDB) and macromolecular crystallographic information file (mmCIF) can retrieve the information from the complex proteins and the information can be stored in the protein database. Prediction of protein interactions is of great importance in solving the functional understanding of genes, encoded proteins, and genome annotation, while comparison of genome order helps to find out the potent operons and different types of genome mapping are done to describe a genome like cytological map, genetic map, physical map, etc. Ribosome profiling unfolds the mechanisms of protein synthesis and provides an organized approach for experimental annotation of coding regions. Molecular phylogenetics, next-generation sequencing, and strain typing are bringing revolution in the modern treatment algorithm, and moreover development of drug target validation and ADR identification via bioinformatics is playing vital role in the development of personalized medication and treatment cost reduction.

References

1. Altman R (2016) Current progress in bioinformatics. Brief Bioinform 17:1–1. https://doi.org/10.1093/bib/bbv105
2. Akalin PK (2006) Introduction to bioinformatics. Mol Nutr Food Res 50:610–619. https://doi.org/10.1002/mnfr.200500273

3. Foulkes AC, Watson DS, Griffiths CEM, Warren RB, Huber W, Barnes MR (2017) Research techniques made simple: bioinformatics for genome-scale biology. J Invest Dermatol 137:163–168. https://doi.org/10.1016/j.jid.2017.07.095
4. Teufel A (2015) Bioinformatics and database resources in hepatology. J Hepatol 62:712–719. https://doi.org/10.1016/j.jhep.2014.10.036
5. Menschaert G, Fenyö D (2017) Proteogenomics from a bioinformatics angle: a growing field. Mass Spectrom Rev 36:584–599. https://doi.org/10.1002/mas.21483
6. Huang X, Liu S, Wu L, Jiang M, Hou Y (2018) High throughput single cell RNA sequencing, bioinformatics analysis and applications. Adv Exp Med Biol 1068:33–43. https://doi.org/10.1007/978-981-13-0502-3_4
7. De R, Bush WS, Moore JH (2014) Bioinformatics challenges in genome-wide association studies (GWAS). Methods Mol Biol 1168:63–81. https://doi.org/10.1007/978-1-4939-0847-9_5
8. Fiser A (2004) Protein structure modeling in the proteomics era. Expert Rev Proteomics 1:97–110. https://doi.org/10.1586/14789450.1.1.97
9. Liddington RC (2004) Structural basis of protein-protein interactions. Methods Mol Biol 261:3–14. https://doi.org/10.1385/1-59259-762-9:003
10. Shi Y, Wu J (2007) Structural basis of protein-protein interaction studied by NMR. J Struct Funct Genomics 8:67–72. https://doi.org/10.1385/1-59259-762-9:003
11. Berman HM, Westbrook J, Feng Z (2000) The protein data bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/nar/28.1.235
12. Westbrook JD, Fitzgerald PM (2003) The PDB format, mmCIF, and other data formats. Methods Biochem Anal 44:161–179. https://doi.org/10.1002/0471721204.ch8
13. Ikeya T, Güntert P, Ito Y (2019) Protein structure determination in living cells. Int J Mol Sci 20:2442. https://doi.org/10.3390/ijms20102442
14. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38:545–549. https://doi.org/10.1093/nar/gkq366
15. Sillitoe I, Dawson N, Thornton J, Orengo C (2015) The history of the CATH structural classification of protein domains. Biochimie 119:209–217. https://doi.org/10.1016/j.biochi.2015.08.004
16. Dawson NL, Lewis TE, Das S (2017) CATH: an expanded resource to predict protein function through structure and sequence. Nucleic Acids Res 45:D289–D295. https://doi.org/10.1093/nar/gkw1098
17. Pearl FM, Lee D, Bray JE, Buchan DW, Shepherd AJ, Orengo CA (2002) The CATH extended protein-family database: providing structural annotations for genome sequences. Protein Sci 11:233–244. https://doi.org/10.1110/ps.16802
18. Knudsen M, Wiuf C (2010) The CATH database. Hum Genomics 4:207–212. https://doi.org/10.1186/1479-7364-4-3-207
19. Laasonen K (2013) Ab initio molecular dynamics. Methods Mol Biol 924:29–42. https://doi.org/10.1007/978-1-62703-017-5_2
20. Bonneau R, Baker D (2001) Ab initio protein structure prediction: progress and prospects. Annu Rev Biophys Biomol Struct 30:173–189. https://doi.org/10.1146/annurev.biophys.30.1.173
21. Kocincová L, Jarešová M, Byška J, Parulek J, Hauser H, Kozlíková B (2017) Comparative visualization of protein secondary structures. BMC Bioinformatics 18:1–12. https://doi.org/10.1186/s12859-016-1449-z
22. Jaroszewski L (2009) Protein structure prediction based on sequence similarity. Methods Mol Biol 569:129–156. https://doi.org/10.1007/978-1-59745-524-4_7
23. Baker D, Sali A (2001) Protein structure prediction and structural genomics. Science 294:93–96. https://doi.org/10.1126/science.1065659
24. Seetin MG, Mathews DH (2012) RNA structure prediction: an overview of methods. Methods Mol Biol 905:99–122. https://doi.org/10.1007/978-1-61779-949-5_8
25. Mathews DH, Turner DH, Watson RM (2016) RNA secondary structure prediction. Curr Protoc Nucleic Acid Chem 67:11.2.1–11.219. https://doi.org/10.1002/cpnc.19

26. Miklós I (2010) RNA structure prediction. Methods Mol Biol 673:19–36. https://doi.org/10.1007/978-1-60761-842-3_2.
27. Mychaleckyj JC (2007) Genome mapping statistics and bioinformatics. Methods Mol Biol 404:461–488. https://doi.org/10.1007/978-1-59745-530-5_22
28. Batzoglou S, Jaffe DB, Stanley K (2002) ARACHNE: a whole-genome shotgun assembler. Genome Res 12:177–189. https://doi.org/10.1101/gr.208902
29. Jaffe DB, Butler J, Gnerre S (2003) Whole-genome sequence assembly for mammalian genomes: Arachne 2. Genome Res 13:91–96. https://doi.org/10.1101/gr.828403
30. Stapleton G, Chapman P, Rodgers P (2019) The efficacy of Euler diagrams and linear diagrams for visualizing set cardinality using proportions and numbers. PLoS One 14:0211234. https://doi.org/10.1371/journal.pone.0211234
31. Brudno M, Do CB, Cooper GM (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res 13:721–731. https://doi.org/10.1101/gr.926603
32. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W (2004) Erratum: human-mouse alignments with BLASTZ. Genome Res 213:103–107
33. Guigo R, de Hoon M (2018) Recent advances in functional genome analysis. F1000Res 7. https://doi.org/10.12688/f1000research.15274.1
34. Del Giacco L, Cattaneo C (2012) Introduction to genomics. Methods Mol Biol 823:79–88. https://doi.org/10.1007/978-1-60327-216-2_6
35. Parkinson J, Blaxter M (2009) Expressed sequence tags: an overview. Methods Mol Biol 533:1–12. https://doi.org/10.1007/978-1-60327-136-3_1
36. Rudd S (2003) Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci 8:321–329. https://doi.org/10.1016/s1360-1385(03)00131-6
37. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484–487. https://doi.org/10.1126/science.270.5235.484
38. Horan MP (2009) Application of serial analysis of gene expression to the study of human genetic disease. Hum Genet 126:605–614. https://doi.org/10.1007/s00439-009-0719-5
39. Behzadi P, Ranjbar R (2019) DNA microarray technology and bioinformatic web services. Acta Microbiol Immunol Hung 66:19–30. https://doi.org/10.1556/030.65.2018.028
40. Pollack JR (2009) DNA microarray technology: introduction. Methods Mol Biol 556:1–6. https://doi.org/10.1007/978-3-319-17305-4_13
41. Kunz M, Ibrahim SM, Koczan D, Scheid S, Thiesen HJ, Gross G (2004) DNA microarray technology and its applications in dermatology. Exp Dermatol 13:593–606. https://doi.org/10.1111/j.0906-6705.2004.00243.x
42. Heller MJ (2002) DNA microarray technology: devices, systems, and applications. Annu Rev Biomed Eng 4:129–153. https://doi.org/10.1146/annurev.bioeng.4.020702.153438
43. Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH (2017) Proteomics: technologies and their applications. J Chromatogr Sci 55:182–196. https://doi.org/10.1093/chromsci/bmw167
44. Meleady P (1664) Two-dimensional gel electrophoresis and 2D-DIGE. Methods Mol Biol 2018:3–14. https://doi.org/10.1007/978-1-4939-7268-5_1
45. Atak A, Mukherjee S, Jain R (2016) Protein microarray applications: autoantibody detection and posttranslational modification. Proteomics 16:2557–2569. https://doi.org/10.1002/pmic.201600104
46. Blundon M, Ganesan V, Redler B, Van PT, Minden JS (1855) Two-dimensional difference gel electrophoresis. Methods Mol Biol 2019:229–247. https://doi.org/10.1007/978-1-4939-8793-1_20.
47. Huang S, Cai N (2018) Applications of support vector machine (svm) learning in cancer genomics. Cancer Genomics Proteomics 15:41–51. https://doi.org/10.21873/cgp.20063
48. Nielsen H (1611) Predicting secretory proteins with SignalP. Methods Mol Biol 2017:59–73. https://doi.org/10.1007/978-1-4939-7015-5_6
49. Wake DT, Ilbawi N, Dunnenberger HM, Hulick PJ (2019) Pharmacogenomics: prescribing precisely. Med Clin North Am 103:977–990. https://doi.org/10.1016/j.mcna.2019.07.002

50. Brar GA, Weissman JS (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. Nat Rev Mol Cell Biol 16:651–664. https://doi.org/10.1038/nrm4069
51. Ingolia NT (2016) Ribosome footprint profiling of translation throughout the genome. Cell 165:22–33. https://doi.org/10.1016/j.cell.2016.02.066
52. Silva J, Fernandes R, Romão L (2019) Translational regulation by upstream open reading frames and human diseases. Adv Exp Med Biol 1157:99–116. https://doi.org/10.1007/978-3-030-19966-1_5
53. Gollnick P, Babitzke P (2002) Transcription attenuation. Biochim Biophys Acta 1577:240–250. https://doi.org/10.1128/jb.183.20.5795-5802.2001
54. Samorodnitsky E, Jewell BM, Hagopian R (2015) Evaluation of hybridization capture versus amplicon-based methods for whole-exome sequencing. Hum Mutat 36:903–914. https://doi.org/10.1002/humu.22825
55. Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet 12:671–682. https://doi.org/10.1038/nrg3068
56. Tang P, Wang X, Shen X (2017) Use of DNA microarray chips for the rapid detection of Mycobacterium tuberculosis resistance to rifampicin and isoniazid. Exp Ther Med 13:2332–2338. https://doi.org/10.3892/etm.2017.4250
57. Le Gallo M, Lozy F, Bell DW (2017) Next-generation sequencing. Adv Exp Med Biol 943:119–148. https://doi.org/10.1007/978-3-319-43139-0_5
58. Yohe S, Thyagarajan B (2017) Review of clinical next-generation sequencing. Arch Pathol Lab Med 141:1544–1557. https://doi.org/10.5858/arpa.2016-0501-ra
59. Sintchenko V, Roper MP (2014) Pathogen genome bioinformatics. Methods Mol Biol 1168:173–193. https://doi.org/10.1007/978-1-4939-0847-9_10
60. Nakagawa H, Fujita M (2018) Whole genome sequencing analysis for cancer genomics and precision medicine. Cancer Sci 109:513–522. https://doi.org/10.1111/cas.13505
61. Fadal RG (1992) Experience with RAST-based immunotherapy. Otolaryngol Clin North Am 25:43–60. https://doi.org/10.1016/S0030-6665(20)31015-X
62. Liu B, Pop M (2009) ARDB-antibiotic resistance genes database. Nucleic Acids Res 37:443–447. https://doi.org/10.1093/nar/gkn656
63. McArthur AG, Waglechner N, Nizam F (2017) The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 57:3348–3357. https://doi.org/10.1128/aac.00419-13
64. Ajawatanawong P (2017) Molecular phylogenetics: concepts for a newcomer. Adv Biochem Eng Biotechnol 160:185–196. https://doi.org/10.1007/10_2016_49
65. Som A (2006) Theoretical foundation to estimate the relative efficiencies of the Jukes-Cantor +gamma model and the Jukes-Cantor model in obtaining the correct phylogenetic tree. Gene 385:103–110. https://doi.org/10.1016/j.gene.2006.03.027
66. Mirsky A, Kazandjian L, Anisimova M (2015) Antibody-specific model of amino acid substitution for immunological inferences from alignments of antibody sequences. Mol Biol Evol 32:806–819. https://doi.org/10.1093/molbev%2Fmsu340
67. Bawono P, Dijkstra M, Pirovano W, Feenstra A, Abeln S, Heringa J (2017) Multiple sequence alignment. Methods Mol Biol 1525:167–189. https://doi.org/10.1007/978-1-4939-6622-6_8
68. Chowdhury B, Garai G (2017) A review on multiple sequence alignment from the perspective of genetic algorithm. Genomics 109:419–431. https://doi.org/10.1016/j.ygeno.2017.06.007
69. Sievers F, Higgins DG (2018) Clustal Omega for making accurate alignments of many protein sequences. Protein Sci 27:135–145. https://doi.org/10.1002/pro.3290
70. Higgins DG, Thompson JD, Gibson TJ (1996) Using CLUSTAL for multiple sequence alignments. Methods Enzymol 266:383–402. https://doi.org/10.1016/s0076-6879(96)66024-8
71. Drinkwater B, Charleston MA (2016) RASCAL: a randomized approach for coevolutionary analysis. J Comput Biol 23:218–227. https://doi.org/10.1089/cmb.2015.0111
72. Francis A, Huber KT, Moulton V (2018) Tree-based unrooted phylogenetic networks. Bull Math Biol 81:936–937. https://doi.org/10.1007/s11538-018-0530-3
73. Hall BG (2013) Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol 30:1229–1235. https://doi.org/10.1093/molbev/mst012

74. Tsai MF, Yu SS (2016) Distance metric based oversampling method for bioinformatics and performance evaluation. J Med Syst 40:159. https://doi.org/10.1007/s10916-016-0516-3
75. Sun H, Huang R, Xia M, Shahane S, Southall N, Wang Y (2017) Prediction of hERG liability–using SVM classification, bootstrapping and jackknifing. Mol Inform 36:1600126. https://doi.org/10.1002/minf.201600126
76. König IR, Fuchs O, Hansen G, von Mutius E, Kopp MV (2017) What is precision medicine? Eur Respir J 50:1700391. https://doi.org/10.1183/13993003.00391-2017
77. Kontoyianni M (1647) Docking and virtual screening in drug discovery. Methods Mol Biol 2017:255–266. https://doi.org/10.1007/978-1-4939-7201-2_18
78. Badrinarayan P, Sastry GN (2011) Virtual high throughput screening in new lead identification. Comb Chem High Throughput Screen 14:840–860. https://doi.org/10.2174/138620711797537102
79. Alfirevic A, Pirmohamed M (2017) Genomics of adverse drug reactions. Trends Pharmacol Sci 38:100–109. https://doi.org/10.1016/j.tips.2016.11.003
80. Weinshilboum RM, Wang L (2017) Pharmacogenomics: precision medicine and drug response. Mayo Clin Proc 92:1711–1722. https://doi.org/10.1016/j.mayocp.2017.09.001
81. Chen X, Yan CC, Zhang X (2016) Drug-target interaction prediction: databases, web servers and computational models. Brief Bioinform 17:696–712. https://doi.org/10.1093/bib/bbv066
82. Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 4:682–690. https://doi.org/10.1038/nchembio.118
83. Sams-Dodd F (2005) Target-based drug discovery: is something wrong? Drug Discov Today 10:139–147. https://doi.org/10.1016/s1359-6446(04)03316-1

Statistical Modeling Techniques

22

Pooja Arora and Ambulge Sheetal

Contents

Abstract

Statistical modeling data is the process of applying statistical analysis to observed data and using mathematical equations to obfuscate information derived from the data. Some statistical models can act as baseline-predictive models that help to understand advanced modeling techniques. Machine learning methods like neural networks and optimal designs can eventually provide outcomes with accurate predictions. So, it is safe to say that there is a thin line between machine learning and statistical modeling. Statistical modeling is used in a number of domains like genomics, metabolomics, and proteomics and other omics data. Statistical modeling is used extensively in many domains like pharmacogenomics, which is also called stratified health care that describes about the treatment strategies of different patient behavior genetic variability and in precision medicine. Statistical

P. Arora (✉) · A. Sheetal
National Institute of Pharmaceutical Education and Research, S. A. S. Nagar, Mohali, Punjab, India
e-mail: poojaarora@niper.ac.in

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_22

studies also showed a promising screening of drugs in drug discovery and drug development. The data collected after screening by using MS, NMR is combined with statistical techniques such as univariate analysis, multivariate analysis, PCA, variability analysis, probabilistic modeling, and support vector machines which help in decision-making process. Advantage of these techniques can help in identification of biomarkers through predictive modeling that can increase the capacity of patient's survival and analyzing high-dimensional scaled profiling data.

Keywords

Statistical modeling · Predictive modeling · Pharmacogenomics · Biomarkers · Omics data

Chapter Objectives

- To understand the advances of biological and biomedical sciences
- To integrate and refine the data with statistical techniques in machine learning
- Understanding statistical methods to provide insights into precision medicine
- Availability of different statistical software and their applications
- To gain information on basic methods and procedures used for statistical calculations

22.1 Introduction

Advances in biological and medical sciences in conjunction with computer science enhance the analysis of big data in a wide variety of research fields. Researchers are focusing on computerizing every part of biomedical sciences in various formats for better results. Because open data is unstructured, statistical analysis is one of the best techniques to analyze the data. One should have knowledge of omics data prior to data scraping and traversing big data, which in terms is a challenging task. The use of machine learning (ML) is rapidly evolving to combine integrated or refined data with statistical approaches. These ML algorithms are vital due to their ability to mine (integrate) large-scale data and statistically analyze it.

Machine learning in omics sciences involves developing and applying computational algorithms that improve predictability, identification, and annotation of genetic and proteomic data. A machine learning prediction model is the most effective way to make sense of complex sets of omics data for genomics/proteomics analysis. Physiological characteristics and drug response can also be described with population models. It is calculated by either combining data from all the individuals, ignoring individual differences, or by calculating each individual's response data and then combining parameters to calculate mean parameters [1].

By using computer-aided techniques, it is possible to extract good representative sample information by studying certain statistical concepts. A parameter that sums up or describes a characteristic of a population, such as a mean or a standard

deviation, is needed in order to generalize the results. When a population's distribution is known and defined, such as the normal distribution, then a small number of parameters can be measured which can be used to construct a probability distribution. These parameters are called statistical parameters [2] for the purposes of extracting samples from the population. For various estimation experiments, an optimal design can be generated without bias and with the smallest variance differences. Based on statistical criteria, the optimality of a design depends on the model selected and validated. It is important to have a good understanding of both statistical theory and the use of experimental design [3] when dealing with this issue. Additionally, sensitivity allows us to determine how a target variable is affected by other variables and identifies all true positives. To predict the outcome of a decision, it is used. This analysis analyzes all the variables and examines how input variables affect target variables. By using all the variables and outcomes together, sensitivity analysis allows forecasting as well as the decision to be made [4].

Confidence region is an additional measure of multidimensional categorization of a confidence interval. Often represented as an ellipsoid around a point, an n-dimensional set of points is an estimated solution to a problem. When a set of points was repeated several times and a confidence region was established on each set of points, then accurate limits with minimum error between 95 and 99% are calculated. THIS is called confidence region that would include the point representing the "true" value for the set of variables being estimated. The fact that a confidence region has been calculated does not negate the possibility that the "true" values lie within that region, since we have not assumed a probability distribution of the "true" values and we may or may not have other information about where they are likely to lie [5].

Procedure and outcomes The collected raw data from various databases will undergo preprocessing steps which involve cleaning. In order to extract and comprehend the original data, R programming and packages like Bioconductor will be used in applying conditions and statements. Data analytics platforms like Hadoop or Spark are also useful for preprocessing big data.

Analyses, machine learning, and statistical modeling will be the next step in using data. Feature extraction, selection, and normalization can be performed on the compiled data prior to processing it. In order to extract features from genomic data, we will use various tools, including PyFeat, R packages (Genomic Tools, Seqinr), and in-house tools like Physicochemical n-Grams Tool (PnGT). Feature selection is a very critical step in ML/statistical modeling, which is followed by implementing various statistical methods like factor analysis, principal component analysis, and heat map development on constructed feature vectors. The data will be split into training and test sets (3:2, 7:3 or 8:2). Various supervised machine learning methods will be applied, such as support vector machines, random forests (RFs), artificial neural networks (ANNs), K-nearest neighbor (kNN) etc. [6]. The developed model is validated using *n*-fold cross-validation (n-CV), leave-one-out cross-validation (LOO-CV), MCC, and other traditional statistical methods like confusion matrix generation to calculate its accuracy, sensitivity, specificity, and precision.

Additionally, benchmark unprocessed genomic datasets [7, 8] will be used for external validation.

Through statistical modeling, the clinical similarities and anomalies between target and disease will be categorized and explained. It will improve the ability to detect diseases early and correctly. The precise classification and preliminary diagnosis of the particular disease of an individual will be vital to precision medicine [9]. They can be used to study various physiological processes and their reactions to various stimuli, such as infection, disease, or drugs.

22.2 Univariate Statistical Analysis

In the science of metabolomics, metabolites are measured in an organism in order to determine how the organism responds to various stimuli like infection, disease, and drugs. In these studies, it was shown that drugs can be safely and effectively screened for safety and effectiveness in the pharmaceutical industry and pharmaceutics. In these screenings, MS and NMR data are combined with statistical techniques such as univariate analysis, multivariate analysis, PCA, variability analysis, probability modeling, and support vector machines.

In small, focused studies (e.g., small clinical trials or animal studies), univariate statistical methods were used in metabolomics experiments based on LC/MS. In Bridgit Crews et al. metabolomics data based on mass spectrometry revealed statistical significance in changes between plasma and cerebral spinal fluid. In this study, we determined the reproducibility scheme for metabolic profiling of human biofluids to detect disease biomarkers. The univariate technique used the median coefficient of variation, which provided a means of comparing variations of plasma and CSF. In addition, boxplots were used to study the associations of metabolite intensity, which is done in R [10].

It is ultimately the goal to constrain the number of initially detected mass-to-charge ratio (mzRT) features [11] to a level that makes future MS/MS identification experiments feasible. Only mzRT features with a minimum FC and statistically significant changes will be retained.

Step 1: In this step, a quality control (QC) check is conducted to eliminate bioinformative features from mzRT. It is important to measure the QC samples. Perform both the continuous variable quantum calculation (CVQC) and continuous variable transmission (CVT) [12].

Step 2: Select the most suitable statistical test by selecting the most appropriate experimental design. You should check your data to see if it is paired. Then, check that the variances are equal and normal. In LC-MS untargeted metabolomics studies, low sample size datasets can affect the performance of normality tests.

Step 3: Attempt to apply statistical tests thoughtfully rather than mechanically by comparing the mean or median of your dataset and performing statistical inference. When applying a test, try to be aware of its limitations. The rest of the data analysis should be carried out using the same approach regardless of whether we used

parametric or nonparametric tests. If we choose to use a nonparametric test, we will plot our results as medians instead of means.

Step 4: Account for multiple testing. After false discovery rate correction (FDR) is applied, report the number of positive false findings. The frequency distribution of *p*-values can be plotted to determine whether there are significant differences in a dataset. Establish a threshold for accepting FDRs. A general consensus is to accept 5% of FDR level, but there is nothing special about this value and each researcher might justify their assumed FDR value, which should be fixed before data is collected.

Calculate the mean or median fold change (FC), depending on the test used for statistical inference. Fix the FC cutoff value [13].

$$\mathbf{FC_{mzrt}} = \frac{\overline{\overline{X}}}{\overline{\overline{Y}}},\ \ \overline{\mathbf{X}} > \overline{\mathbf{Y}}; \mathbf{FC_{mzrt}} = -\frac{\overline{Y}}{\overline{\overline{X}}}, \overline{\mathbf{X}} < \overline{\mathbf{Y}} \tag{22.1}$$

where X represents average raw intensities across case groups and Y represents average raw intensities across control groups.

Univariate analysis is performed using the above steps. It is necessary to take logarithms and convert them into lognormal distributions when working with non-parametric data. In order for biomarker studies to be generalized, there is a need for a large number of features. When interpreting fold changes between sample groups, filters should be applied. The minimum fold change cut off should be 2 to minimize the component of variation in treatment and healthy controls.

A number of multidimensional and multivariate statistical analyses, as well as pattern-recognition programs, have been developed to analyze the large amounts of data and figure out how the metabolic pathway information of the measurements can be understood. In this context, univariate statistics can be helpful, but can only examine a single variable at a time. Since univariate analysis is not always accurate, there are usually multiple variables involved. Many processes or any set of samples with a multitude of measurements require observing, measuring, or observing simultaneously more than one variable. As a result, multivariate analysis can offer a better depiction of highly correlated data and can indicate when a system or process is having problems. In metabolomics, multivariate analysis (MVA) may help determine which of these metabolites are important to identify. In GC/MS and HPLC/MS statistical data analysis, the following parameters are required: peak retention time, intensity, and mass/charge ratio (*m*/*z*). You may use bin integrations or metabolite concentrations for NMR chemical shifts.

22.3 Multivariate Analysis (MVA)

Multivariate statistics is simply the statistical analysis of more than one statistical variable simultaneously. For an introduction to commonly used statistical models (PCA, SIMCA, PLS-DA, KNN, OPLS, etc.), the reader is referred to three excellent reviews by Lindon et al., who compared PLS-DA MVA with support vector

machines (SVM) for the analysis of NMR data. Their results indicated that with less number of features, SVM gave a better predictive model when compared to that of PLS-DA. Multivariate statistical methods and pattern-recognition programs were developed to handle the acquired data and to search for the discriminating features between data acquired from two sample sets, healthy and diseased. Multivariate techniques deals with real applications with reasonably large number of measurements made on each object in one or more samples with multiple features simultaneously. Analyzing interdependence and analyzing dependence are the two general types of MVA techniques [14].

1. **Analysis of dependence**: Where one (or more) variables are dependent variables, to be explained or predicted by others, for example, Multiple regression, Discriminant analysis, Multivariate analysis of variance (MANOVA), Partial Least Square.

Multilinear regression (MLR) In multiple linear regressions, several independent variables are used to predict one direct variable using a least squares approach. Since the independent variables are orthogonal, there are only a few univariate regressions. Partially correlated independent variables are valued for their importance. Whenever an independent variable can be predicted based on the dependent one, there is a problem. Multicollinearity is the result of this.

Discriminant Analysis A discriminant analysis is a statistical technique for grouping individuals or objects into mutually exclusive and exhaustive categories based on independent variables. As much as possible, several discriminant functions must be determined, which are linear combinations of independent variables that allow individuals and groups to be distinguished or separated. It is determined by a multivariate test statistic known as Wilks' lambda. To derive conclusions about which variables are necessary to discriminate among groups, we can standardize the coefficients of the discriminant functions on the same basis as in multiple regression. It helps investigators identify groups or categories of subjects using several explanatory variables and discriminant analysis.

Each group is classified by determining a separate prediction equation based on the explanatory variables that indicates the probability of belonging to that group. To classify a future subject, a prediction is calculated for each group, and the individual is classified according to which group they most closely resembles.

Partial least square regression (PLSR) Rather than compute latent vectors explaining both independent and dependent variables, it computes latent vectors that address the multicolinearity problem. Multiple dependent variables are predicted using multiple regression. The model incorporates features from multiple linear regression and principal component analysis. Principal component analysis can be used to determine the scores of the units as well as the loadings of the variables, while multiple linear regression analysis can estimate the dependent variables. It is used to find the fundamental relationship between two matrices

(X and Y). In an iterative process, the matrices X and Y are decomposed into latent structures.

2. **Analysis of interdependence:** It is not considered a dependent variable. Analyze the relationships between variables, objects, or cases. For example, cluster analysis, factor analysis, and principal component analysis.

Cluster analysis Identification of groups of cases with similar characteristics. Also used to summarize data by defining segments of similar cases. The process of dissecting a cluster is known as cluster analysis.

Factor analysis It is a statistical method used to describe variability among observed variables in terms of a smaller number of unobserved variables called factors. Analyzing factors helps identify the variables that contribute to the correlations observed within a set of variables. In addition to data reduction, factor analysis can also be used to develop hypotheses about causal mechanisms or to screen variables for further analysis.

Principal Component Regression In principal component regression, the independent variables are first submitted to a principal component analysis and the scores of the units are then used as predictors in a standard multiple linear regression.

First step is to run principal component analysis so as to reduce dimensionality of data.

The second step is to run an ordinary least squares regression on selected components; factors most correlated with dependent variables are selected.

The third step is to compute the parameters of the model for the selected explanatory variables [15].

22.4 Principal Component Analysis

Principal component analysis is a technique to explain the variance-covariance structure through a few linear orthogonal combinations of the original genomic variables. This mathematical technique is used in gene expressions as a clustering tool. This technique is divided further into supervised PCA and unsupervised PCA. Supervised learning is a learning in which we teach or train the machine using data which is well labeled that means some data is already tagged with the correct answer. After that, the machine is provided with a new set of sample data for extracting principal components.

Supervised learning algorithm analyzes the training data (set of training examples) and produces a correct outcome from labeled data, whereas unsupervised learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data. Principal component analysis (PCA) is an unsupervised technique used to preprocess and reduce the

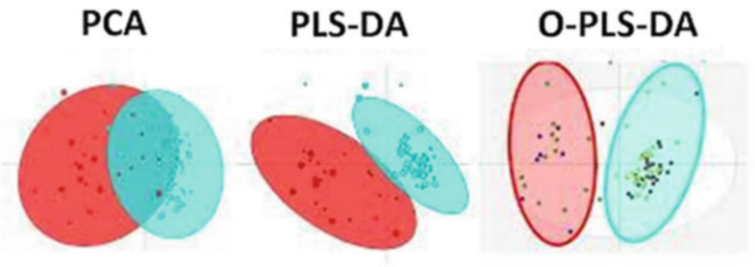


Fig. 22.1 Comparison of PCA, PLS-DA, and OPLSDA between two different classes

dimensionality of high-dimensional datasets while preserving the original structure and relationships inherent to the original dataset so that machine learning models can still learn from them and be used to make accurate predictions. Principal component analysis (PCA) falls under the category of unsupervised machine learning algorithms where the model learns without any target variable. PCA has been specifically used in the area of dimensionality reduction to avoid the curse of dimension. Supervised PCA also investigates the quantitative value of target variable and thus it is applicable on regression problems. PCA is predominantly used as a dimensionality reduction technique in domains like facial recognition, computer vision, and image compression. It is also used for finding patterns in data of high dimension in the field of finance, data mining, bioinformatics, psychology, etc.

Van et al. used 2-D total correlation spectroscopy NMR and statistical analysis to compare the global metabolic profiles of urines obtained from wild-type and ABCC6-knockout mice. Three statistical methods were used to analyze the NMR spectra; PCA, PLS-DA, and OPLSDA.

The PLS-DA and OPLS-DA gave almost identical results while PCA gave slightly different results. However, the three methods could successfully discriminate between the two groups. For profile differences between healthy and diseased states, the resulting *m/z* values, retention times, and intensity of every peak in the chromatogram of each sample set were submitted to PCA and OPLS-DA for statistical data analysis. OPLS-DA is a supervised procedure that constructs a linear combination of all peak intensities which maximizes the separation between healthy and diseased samples, while PCA is an unsupervised procedure where linear combinations of all peak intensities are constructed to produce orthogonal components that maximize the total variance in the samples independent of their group labels. The results of statistical analysis (Fig. 22.1) indicated that OPLS-DA gave better discrimination than PCA; however, both procedures could discriminate between the two groups at greater than 90% sensitivity and specificity. These techniques can lead to the discovery of disease markers and to the identification of different metabolic pathways which may prove useful in diagnostic and prognostic clinical settings. It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences since patterns in data can

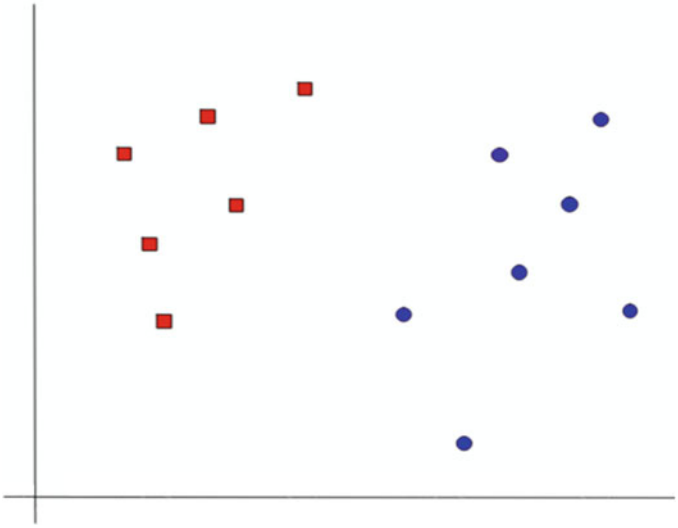


Fig. 22.2 Diagrammatic representation of PCA components showing the pattern of data distribution

be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data (Fig. 22.2). Steps involved in calculating this are:

Step 1: Acquisition of data—collect the data needed for analysis.

Step 2: Adjust the data—adjust the acquired data simply by subtracting the mean of the particular data from the acquired data.

Step 3: Find the covariance matrix—the covariance matrix describes all relationships between pairs of measurements in the considered data set.

$$\mathrm{Cov}(X, Y) = \frac{\sum_{i=1}^{n}(Xi - X)(Yi - Y)}{n - 1} \tag{22.2}$$

Step 4: Find the Eigenvalues and Eigenvectors (feature vector).

Step 5: Find the row feature vector—we can easily find row feature vector; it is just the transpose of Eigenvectors matrix.

Step 6: Find the new data set.

NEW (FINAL) DATA = ROW FEATURE VECTOR * ROW DATA ADJUST

Step 7: Calculate the principal component analysis [15, 16].

22.5 Support Vector Machines

Support vector machines are known to have excellent generalization abilities when compared to other statistical multivariate methods, such as PCA or PLS-DA. For generalization, we can use descriptive modeling and mechanistic modeling to classify the subcategories in the dataset.

If the purpose is just to provide a reasonable description of the data in some appropriate way without any attempt at understanding the underlying phenomenon, then descriptive modeling is used. The model here is selected based on its adequacy to represent the data structure. These models are very useful for discriminating between alternative hypotheses but are totally not helpful for capturing the fundamental characteristic of a mechanism. Whereas in mechanistic modeling, the purpose is to understand the mechanisms of action, which is critical for any data processing. Mechanistic models provide updated, rich, and reliable information about the problem, whereas descriptors are trained for translating scientific information in mathematical models [17]. SVMs are a relatively new machine learning supervised technique for classification of data [18].

Unlike PCA or PLS-DA, SVMs can be extended to nonlinear cases with the help of kernels, whereas PCA and PLSDA have an assumption of linearity. SVM or support vector machine is a linear model for classification and regression problems. It can solve linear and nonlinear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyperplane which separates the data into classes. At first approximation, what SVMs do is to find a separating line (or hyperplane) between data of two classes. SVM is an algorithm that takes the data as an input and outputs a line that separates those classes if possible. Suppose you have a dataset as shown below and you need to classify the red rectangles from the blue ellipses (let us say positives from the negatives). So your task is to find an ideal line that separates this dataset in two classes (say red and blue).

But, as you notice, there is not a unique line that does the job. In fact, we have infinite lines that can separate these two classes. So how does SVM find the ideal one? Let us take some probable candidates and figure it out ourselves (Fig. 22.3).

We have two candidates here, the green colored line and the yellow colored line. Which line according to you best separates the data?

If you selected the yellow line, then congrats, because that is the line we are looking for. It is visually quite intuitive in this case that the yellow line classifies better. But we need something concrete to fix our line. The green line in the image above is quite close to the red class. Though it classifies the current datasets, it is not a generalized line, and in machine learning, our goal is to get a more generalized separator.

According to the SVM algorithm, we find the points closest to the line from both the classes. These points are called support vectors. Now, we compute the distance between the line and the support vectors. This distance is called the margin. Our goal is to maximize the margin. The hyperplane for which the margin is maximum is the optimal hyperplane (Fig. 22.4).

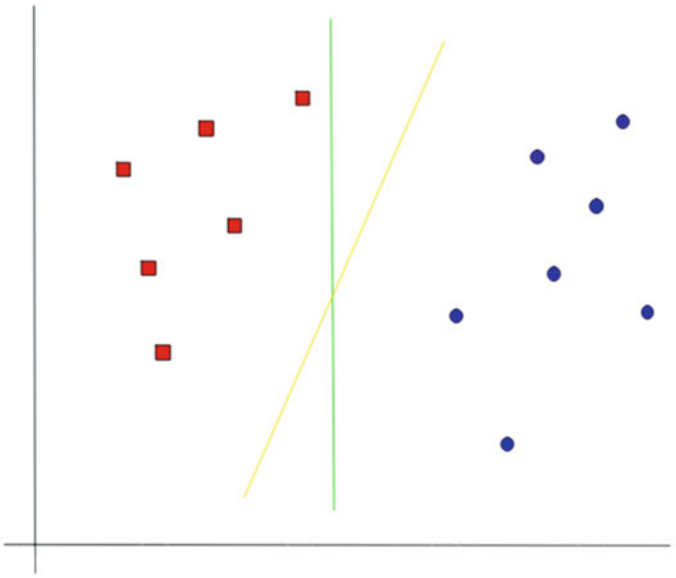


Fig. 22.3 SVM model showing the classification of two classes, that is, red rectangle points with blue circular points

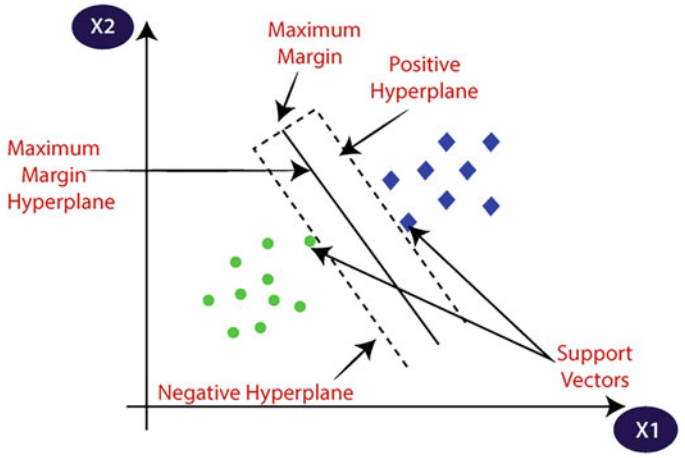


Fig. 22.4 Diagrammatic representation of SVM algorithm with maximized margin showing the plane for classification

Thus, SVM tries to make a decision boundary in such a way that the separation between the two classes (that street) is as wide as possible [19].

22.5.1 Applications

The use of support vector machine algorithms and their examples are used in many technologies which incorporate the use of segregation and distinction. Compared to the other ML methods, SVM is very powerful at recognizing subtle patterns in complex datasets. SVM can be used to recognize handwriting, recognize fraudulent credit cards, identify a speaker, as well as detect faces. Cancer is a genetic disease where the genomic feature patterns or feature function patterns may represent the cancer subtypes, the outcome prognosis, drug benefit prediction, tumor genesis drivers, or a tumor-specific biological process. Therefore, the Aartificial Iintelligence of SVM can help us in recognizing these patterns in a variety of applications.

22.6 Probabilistic Modeling

A probabilistic method or model is based on the theory of probability or the fact that randomness plays a role in predicting future events. The opposite is deterministic, which is the opposite of random. It tells us something can be predicted exactly, without the added complication of randomness.

Probabilistic models incorporate random variables and probability distributions into the model of an event or phenomenon. While a deterministic model gives a single possible outcome for an event, a probabilistic model gives a probability distribution as a solution. These models take into account the fact that we can rarely know everything about a situation. There is nearly always an element of randomness to take into account. For example, life insurance is based on the fact we know with certainty that we *will* die, but we do not know when. These models can be part deterministic and part random or wholly random. Random variables from the normal distribution, binomial distribution, and Bernoulli distribution form the foundation for this type of modeling (Fig. 22.5).

A normal distribution curve, sometimes called a bell curve, is one of the building blocks of a probabilistic model.

The probabilistic method is a way to prove the existence of a structure with certain properties in combinations. Condition of normality under the model create a probability space by choosing elements at random and further proves that any random element from the space will attain a positive probability. The method is widely used in a variety of disciplines, including: statistical physics, quantum mechanics, and theoretical computer science [20].

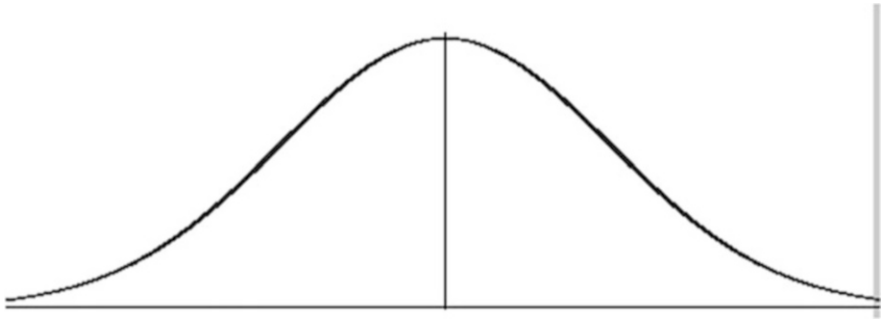


Fig. 22.5 A probabilistic model showing a bell-shaped curve between *x* and *y* axis

22.7 Partial Least Squares (PLS)

Similar to PCA, partial least squares is a data reduction technique used to identify latent structures of both predictors and responses by maximizing the covariance between them. It constructs a linear combination, for example, genomic variables inserting survival outcomes as weights in the study. Partial least squares (PLS) regression is a technique that reduces the data to a smaller set of uncorrelated components and performs least squares regression on these components, instead of the original data. Advanced tool called partial least squares-discriminant analysis (PLS-DA) is a versatile algorithm that can be used for predictive and descriptive modeling as well as for discriminative variable selection. PLS-DA is gaining popularity in metabolomics, and in other integrative omics analyses, both chemo metrics and omics data sets are characterized by large volume, large number of features, noise, and missing data. These data sets also often have a lot fewer samples than features. PLS-DA can be thought of as a "supervised" version of principal component analysis (PCA) in the sense that it achieves dimensionality reduction but with full awareness of the class labels. Besides its use as for dimensionality reduction, it can be adapted to be used for feature selection as well as for classification. It is a regression method with a special binary dummy "*y*-variable," and it is commonly used for classification purposes and biomarker selection in metabolomics studies. Four diagnostic statistics of PLS-DA, namely the number of misclassifications (NMC), the area under the receiver operating characteristic (AUROC), Q2 and discriminant Q2 (DQ2) are several statistical approaches currently in use with PLS-DA method discussed. All four diagnostic statistics are used in the optimization and the performance assessment of PLS-DA models and to validate outcomes of PLS-DA analyses, for example, double cross-validation procedures or permutation testing. PLS-DA models with NMC or AUROC as diagnostic statistics are more accurate in finding biomarkers responsible for two classes' discrimination [21].

PLS is used to find the fundamental relations between two matrices (X and Y), that is, a latent variable approach to modeling the covariance structures in these two spaces.

Suppose $\mathbf{X}$ is a mean-centered $n \times m$ matrix and $\mathbf{Y}$ is a mean-centered $n \times p$ matrix. The sample covariance matrix of $\mathbf{X}$ and $\mathbf{Y}$ is given by

$$\text{cov}(X, Y) = \frac{1}{n-1} X^{\mathrm{T}} Y. \tag{22.3}$$

The variance of X is given by

$$\text{var}(X) = \text{cov}(X, X). \tag{22.4}$$

(The reason the sample covariance matrix has $n - 1$ in the denominator rather than n is to correct for the fact that we are using sample mean instead of true population mean to do the centering.) $S = \text{var}\ (X)$ is symmetric. The diagonal entry $S_{j,j}$ is called the variance of xj. The total variance of the data in X is given by the trace of S: $\text{tr}(S) = \Sigma_j S_{j,j}$. The value $S_{i,j}$, $i \neq j$, is called the covariance of x_i and x_j. The correlation between X and Y is defined by

$$\text{corr}(X, Y) = \sqrt{\text{var}\ (X)}\text{cov}(X, Y)\sqrt{\text{var}\ (Y)} \tag{22.5}$$

22.8 Some Statistical Softwares

Softwares	Description	URL
R	It is a powerful package used for complete statistical operations	https://www.r-project.org/
SPSS	It is descriptive statistics, parametric and nonparametric analyses, as well as graphical depictions of results	https://www.ibm.com/analytics/spss-statistics-software
Stata	It is used for statistics and data sciences	https://www.stata.com/
Statistica	Statistica provides data analysis, data management, statistics, data mining, machine learning, text analytics, and data visualization	https://www.tibco.com/
Minitab	The Minitab software offers a range of both basic and fairly advanced statistical tools for data analysis	https://www.minitab.com/en-us/
SAS	To create scripts for more advanced analyses	https://www.sas.com/en_us/software/stat.html

22.9 Conclusions and Future Prospects

Statistical studies have shown a promising screening of drugs in the pharmaceutical industry and drug development. Machine learning methods like neural networks and optimal designs with statistical methods can eventually provide better outcomes with accurate predictions. So, it is safe to say that there is a thin line between machine learning and statistical modeling. Statistical modeling is used in a number of domains like genomics, metabolomics, and proteomics and other omics data. Machine learning in omics sciences is concerned with the development and application of computational algorithms that improve the predictability, identification, and annotation of genetic and proteomic-based data. Analysis, ML/statistical modeling, and interpretation will be the next step using data. The compiled data can be preprocessed by performing feature extraction and selection followed by normalization of features. Feature selection is a very critical step in ML/statistical modeling which will be preceded by implementing various statistical methods like factor analysis, principal component analysis, heat map development on constructed feature vectors followed by model validation. External validation will also be performed using benchmark untouched genomic dataset which would be compiled from the different databases.

The developed statistical models will help us in classification and explanation of targets and disease. This will enhance the knowledge level for correct and early identification of the diseased condition. This precise classification and preliminary diagnosis of the particular disease of an individual will be a key factor for precision medicine. The above techniques are most widely used statistical methods in machine learning for prediction of outcomes. The most common softwares used in statistical modeling are explained in the above different statistical software table.

References

1. Stanski DR, Maitre PO (1990) Population pharmacokinetics and pharmacodynamics of thiopental: the effect of age revisited. Anesthesiology 72:412–4222. https://link.springer.com/article/10.1208/aapsj070239
2. Belsley DA (2004) Weak data. Encyclopedia of statistical sciences. https://doi.org/10.1002/0471667196.ess0352.pub2
3. Guttorp P, Lindgren G (2009) Karl Pearson and the Scandinavian school of statistics. Int Stat Rev 77:64. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-5823.2009.00069.x
4. Saltelli A (2002) Sensitivity analysis for importance assessment. Risk Anal 22(3):1–122. https://onlinelibrary.wiley.com/doi/abs/10.1111/0272-43322.00040
5. Draper NR (1966) H. Smith, Applied regression analysis. Wiley, New York, 407, p 90
6. Frank LE, Friedman JH (1993) A statistical view of some chemometrics regression tools. Technometrics 35(2):109–135. https://www.tandfonline.com/doi/abs/10.1080/00401706.1993.10485033
7. Crews B, Wikoff WR, Patti GJ, Woo HK, Kalisiak E, Heideker J, Siuzdak G (2009) Variability analysis of human plasma and cerebral spinal fluid reveals statistical significance of changes in mass spectrometry-based metabolomics data. Anal Chem 81(20):8538–8544. https://pubs.acs.org/doi/abs/10.1021/ac9014947

8. Vinaixa M, Samino S, Saez I, Duran J, Guinovart JJ, Yanes O (2012) A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data. Metabolites 2(4): 775–795. https://www.mdpi.com/2218-1989/2/4/775
9. Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD (2009) Analytical and statistical approaches to metabolomics research. J Sep Sci 32(13):2183–2199. https://www.sciencedirect.com/science/article/pii/B9780128123645000080
10. Patti GJ, Yanes O, Siuzdak G (2012) Innovation: metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol 13:263–269. https://www.nature.com/articles/nrm3314
11. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78:779–787. https://pubs.acs.org/doi/abs/10.1021/ac051437y
12. Kristian Hovde L (2011) Multivariate methods in metabolomics—from pre-processing to dimension reduction and statistical analysis. TrAC Trends Anal Chem 30:827–841. https://www.sciencedirect.com/science/article/abs/pii/S0165993611000914
13. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125:279–284. https://www.sciencedirect.com/science/article/abs/pii/S0166432801002972
14. Multivariate Analysis; Methods and Applications—William R Dillon and Matthew Goldstein; Wiley series 2nd edn. https://rss.onlinelibrary.wiley.com/doi/abs/10.2307/2347476
15. Sandeep K, Siddharth KS, Prashant M (1971) Multivariate analysis: an overview. J Multivar Anal 316–346. https://www.journals.elsevier.com/journal-of-multivariate-analysis
16. Varun G, Manish M, Piyush CO, Pankaj K (2015) Principal component analysis & factor analysis as an enhanced tool of pattern recognition. Int J Elec Electr Eng Telecoms 23:1–15. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792409/
17. Wang B (2006) Computer applications in pharmaceutical research and development, vol 2. Wiley. https://www.wiley.com/en-us/Computer+Applications+in+Pharmaceutical+Research+and+Development-p-9780470037225
18. Mahadevan S, Shah SL, Marrie TJ, Slupsky CM (2008) Analysis of metabolomic data using support vector machines. Anal Chem 80(19):7562–7570. https://pubs.acs.org/doi/abs/10.1021/ac800954c
19. Joachim T (2002) Learning to classify text using support vector machines. Springer Science & Business Media, p 2122. https://www.springer.com/gp/book/9780792376798
20. Grigoriu M, Cornell CA, Veneziano D (2007) Probabilistic modelling as decision making. J Eng Mech Div 105(4):585–596. https://www.osti.gov/servlets/purl/1137255
21. Rannar S, Lindgren F, Geladi P, Wold S (1994) A PLS Kernel algorithm for data sets with many variables and fewer objects. Part 1: theory and algorithm. J Chemometrics 8(2):111–125. https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/cem.1180080204

Molecular Modeling of Nanoparticles

23

Bilge Bicak, Bahar Gok, Serda Kecel-Gunduz,
and Yasemin Budama-Kilinc

Contents

B. Bicak
Institute of Graduate Studies in Sciences, Istanbul University, Istanbul, Turkey
e-mail: bbicak@istanbul.edu.tr

B. Gok
Graduate School of Natural and Applied Science, Yildiz Technical University, Istanbul, Turkey

S. Kecel-Gunduz
Physics Department, Faculty of Science, Istanbul University, Istanbul, Turkey
e-mail: skecel@istanbul.edu.tr

Y. Budama-Kilinc (✉)
Department of Bioengineering, Faculty of Chemical and Metallurgical Engineering, Yildiz Technical University, Istanbul, Turkey

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_23

Abstract

In recent years, nanotechnology has opened new horizons in the field of biomedicine by designing and using nanoparticles with different physicochemical properties for clinical diagnosis and therapeutics. To play a role in various biofunctions, nanoparticles need to be transported across the membrane and into the target cell or region. However, it is not clear enough, how the nanoparticles affect the cell, and what kind of interactions are there between cell and nanoparticles. Therefore, it is useful to understand how nanoparticles interact with lipid membranes in order to obtain safe applications in nanobiomedicine. Computer modeling and simulation of nanoparticles quantitatively describes the correlation between particle microstructure and properties. With computational modeling, it is possible to manage each parameter individually and define the mechanisms responsible for the experimental result, so it is a powerful tool compared to experimental constraints. For different conditions, which are not always possible to examine in a laboratory environment, interactions are possible with simulated computerized calculations. Computational approaches, such as molecular dynamics (MD) simulations, as a natural complement to experimental techniques, are among the approaches used in the modeling of nanoparticles by providing various factors such as accessible time scales, the full atomistic description of the system, the dynamic behavior of the system, and the inclusion of environmental influences. In this chapter, the approaches developed for modeling nanoparticles will be explained in detail.

Keywords

Molecular modeling · Nanoparticle · MD · Docking

Chapter Objectives

- To describe computer modeling and simulation of nanoparticles
- To introduce different models used for computer modeling
- To compare the models used for computer modeling
- To show the different software in molecular modeling
- To explain how there are applications for nanoparticles
- To present the use of docking analysis for nanoparticles

23.1 Introduction

As a result of the rapid developments in computer technology, computerized methods have facilitated the examination of both regular structures and irregular structures. In cases where experiments and theory are not sufficient, computer simulation methods have started to be used to obtain numerical solutions of problems whose analytical solution cannot be fully realized. It is necessary to choose the most accurate method to meet the requirements of the system and to create the closest simulation to the system [1, 2]. Many simulation methods, from quantum

mechanical calculation methods to quasi-experimental methods, have been developed to model atomic behaviors most similar to the system structure [2–5]. Methods such as molecular dynamics (MD), molecular mechanics (MM), Monte Carlo (MC), mesh optimization, and energy minimization (EM) can be cited as examples [6–8]. The common feature of these methods is to give information about the microscopic or macroscopic structure of the system, taking into account the classical interactions [9]. Especially MD simulation method is one of the most used methods because it pioneers experimental studies in examining the mechanical, thermodynamic, and structural properties of matter. This method, which aims to obtain the trajectory of the system in the phase space by examining the time-dependent changes of the particles in the system, is applied by integrating Newtonian equations of motion.

MD simulations as a function of time are one of the most preferred methods in the investigation of both mechanical and thermodynamic properties (such as temperature, pressure) of substances. Looking at the development of this method historically, it was first used in 1950 by Alder and Wainwright to study the interactions in the hard sphere approach. Then, in 1964, Rahman made the first simulation using a realistic potential for liquid argon. Since that day, studies with MD simulation methods have been intensely continued. MD simulation method is preferred in many fields of study to model various molecules in every branch of science and to analyze the properties of the modeled molecules. Some of these working areas are given below [10, 11].

Basic research: Balancing, testing molecular chaos, kinetic theory, diffusion, transport properties, size dependence, testing models and potential functions

Phase transitions: First- and second-order phase transitions, simultaneous phases, layout parameters, critical events

Collective behaviors: Distortion of space-time-correlation function, coupling of rotation and translation, vibration, spectroscopic measurements, orientation order, dielectric properties

Complex liquids: Structure and dynamics of glasses, molecular liquids, pure water and aqueous solutions, liquid crystals, ionic liquids, liquid interfaces, films and very thin layers.

Polymers: Chains, rings and branched molecules, membranes, balance configurations and migration processes

Solids: Crystal defect formation and movement, fracture, grain boundaries, structural transformations, radiation damages, elastic and plastic mechanical properties, friction, shock waves, molecular crystals, clastics, epitaxial growth

Biomolecules: Protein structure and dynamics

Fluid dynamics: Laminar flow, boundary layers, rheology of non-Newtonian fluids, unstable flow

23.2 Nanoparticles

Nanotechnology is a very popular, novel multidisciplinary science that has entered our lives in many fields with the millennium. Thanks to nanotechnology, the matter can be regulated and controlled at the molecular level. In this way, the matter can

behave extraordinarily in the nanoscale. These opportunities offered by nanotechnology have made this branch of science the center of attention for many scientists. Nanoparticles, the most important product of this new science, have found wide application opportunities in electronics, chemistry, textile, energy production, and many engineering fields, especially in biomedicine. Thanks to nanotechnology, it has been possible for researchers from many different research fields to develop new treatment and diagnostic methods for diseases. In addition, nanotechnology has enabled the production of more biocompatible and durable materials. Especially, with the widespread use of nanoscale particles in the diagnosis and treatment of diseases, the term "nano-medicine" has entered our lives.

Nanoparticles have greater chemical reactivity compared to their conventional forms, due to their greater surface area to volume ratio. Nanoparticles can deliver drugs to the desired site within the optimum dose range, therefore therapeutic efficacy of drugs and patient compliance improves. Moreover, their side effects are poor. Their sizes range from 1 to 100 nm. Although it can be divided into different classes, the most current classification of nanoparticles can be divided into three categories as organic nanoparticles, inorganic nanoparticles, and carbon-based nanoparticles. Each of these nanoparticles has unique strengths and limitations.

Organic nanoparticles are widely known as dendrimers, micelles, and liposomes or polymeric nanoparticles. These nanoparticles are biodegradable and nontoxic [12]. Dendrimers are nanosized (1–100 nm) macromolecules, which possess complex spherical structures that consist of three main units (internal space, repeating units, and end groups) [13]. They are effective controlled release agents with cavities for slow release of drug molecules [14]. Dendrimers can be used for the delivery of both hydrophilic and hydrophobic molecules due to their versatile design [15, 16]. They can be produced with different surface charges (positive, negative, or neutral). Cationic dendrimers are known to cause high cytotoxicity compared to anionic or neutral dendrimers due to their binding with negatively charged cell membrane [16, 17]. Micelles are spherical amphiphilic structures that consist of a hydrophilic head and hydrophobic tail. They are obtained by self-assembly process. Micelles have hydrophobic tails on the inside with the hydrophilic heads on the outside in an aqueous medium [18]. The hydrophobic domain of micelles in the core protects the poorly soluble drug molecules. Polymeric micelles are biocompatible and have high stability [19]. Liposomes are described as vesicles with the ability to encapsulate both the hydrophilic and the lipophilic drugs [20]. Although they have several advantages such as ability to protect drugs from degradation, target the site of action, and reduce the noxiousness and other side effects, some limitations related to low stability and low drug-loading capacity of liposomes appear as disadvantages [20–22]. Polymeric nanoparticles have great attention over recent years due to their modifiable size, shape, and surface properties [23]. They are mostly nanospheres or nanocapsules shaped. Polymeric nanoparticles protect drugs against chemical degradation [24]. In addition, they have come into prominence in nano-medicine due to their biocompatibility and their ability to encapsulate the high amount of drug molecules [25]. Drugs that pose a threat to healthy cells, whose solubility and absorption are low and whose chemical structure is disrupted at physiological pH and body temperature, can be easily administered to the body by using polymeric

nanoparticles [26, 27]. Polymeric nanoparticles enable drugs to be released in a controlled manner after they enter the living system. In the presence of conditions such as magnetic induction, light irradiation, ultrasonic agitation, temperature difference, enzyme activation, redox potential, or pH difference, depending on the structure of the polymer, the drug molecules can be released out of the nanoparticles to the target region [27]. Since the drug candidate designed in the form of polymeric nanoparticles will not disperse in the body thanks to the nanoparticle system, it is possible to reach an effective drug dosage in the target tissue or organ by using it at low doses [28].

Inorganic nanoparticles can be used in numerous biological, biomedical, and pharmaceutical applications [29]. The most important advantage of inorganic nanoparticles is that they can be used in both diagnosis and treatment. These nanoparticles include metals, metal oxides, and metal salts. Metallic nanoparticles are gold, silver, iron, zinc, and silica [30]. Metallic nanoparticles can be synthesized in different morphologies such as spherical, triangle, rod, and star [31]. The disadvantage of these nanoparticles is that they can show some toxic effects by causing accumulation in vital organs such as liver, lungs, and kidney [32]. There are many studies that aimed to reduce or eliminate these effects by coating these nanoparticles with various biocompatible materials [33].

Carbon-based nanoparticles are found in different morphologies. Fullerenes, carbon nanotubes, carbon nanofibers, carbon black, graphene, and carbon onions are included under the carbon-based nanoparticles category [34]. The carbon nanotubes are used in biomedical applications due to their easy translocation across the cell membrane and low toxicity [35–37].

Nanoparticles make a great contribution to the pharmaceutical technology. For example, nanoparticles allow the combination of two drug molecules to be used together for therapy [38]. Besides, labeled nanoparticles can be followed in the living system. The labeled nanoparticles which interact with the cells and tissues can be imaged. Moreover, by functionalization of the surfaces of the nanoparticles, the drug molecules can be directed to the desired target region. In this way, the properties and activities of the drug molecules are preserved until they reach the target area [39]. Thanks to the targeting strategy, not only it is possible to reduce the amount of the drug to be used and eliminate side effects but also to send the drug to the right target [40]. Thus, the effectiveness of the drug increases in treating without other tissues and cells damaging them [41–43]. A specific chemical interaction occurs between antigen-specific antibodies and antigens [44]. In this regard, the surface of the nanoparticles can be functionalized with antibodies specific to the surface antigens of the target cells. The nanoparticles functionalized with specific antibodies bypass other cells and come into contact with the target cell, the bonding occurs with the cell surface, and the nanoparticles are taken into the cell. Thus, the drug is delivered to the targeted region [44]. Also, it is possible to obtain more functional agents with the combined use of nanoparticles. For example, dendrimer molecules labeled with magnetic metal oxide can be internalized into stem cells in order to follow the behavior of stem cells in certain tissues and organs. The stem cells containing these functionalized nanoparticles can be introduced into the living

system, and their interactions with the target tissue or organ and their behavior can be monitored [45].

The outcomes about the long-term effects of all these nanoparticles in the living system are limited [46]. These interactions are determined by both in vitro toxicity and in vivo toxicity tests [47]. As the most important and possible side effects of nanoparticles, it has been defined as their potential to go beyond their targeted distribution areas, to be small enough to escape phagocytosis, to cause inflammatory and toxic effects by modifying their protein structures [47, 48]. However, there are limitations in studying intermolecular interactions of nanoparticles in biological systems at the nanoscale using existing experimental techniques. The use of calculation methods that minimize these limitations has gained importance. The spatial and temporal resolutions that computational techniques currently allow make it possible to study the specific interactions and dynamics of nanoparticles in biological environments.

The use of various calculation techniques makes it possible to determine both the possible cytotoxic effects and the beneficial effects of nanoparticles by better understanding the interactions and dynamics of biological systems and nanoparticles. We will explain the crucial calculation techniques in detail in the next section.

23.3 Molecular Modeling of Nanoparticles

Significant advances in the development of high-speed computers and modeling have made it possible to model nanoparticles used in an important area such as drug delivery. In relation to the calculation method used, studies on optical properties and electronic structure, nanoparticles, and biomolecule-nanoparticle interaction can be carried out. Modeling with various computational methods is carried out to obtain information about the electronic structure and time-dependent behavior of molecular systems like these and the like [49]. Molecular modeling consists of a molecular model and a computational technique. This computational technique is used to characterize the behavior of the molecules. The first step in molecular modeling is the construction of suitable molecular model. According to all knowledge about this area, molecular modeling is divided into two parts. One part is the full atomistic model, and this part gives highest level of detail because all atoms are considered. Another part is the coarse-grained model that abridge the atomic detail by enclosing groups of atoms into beads that lump the fundamental features (such as charge, polarity) of the atoms that they embed. Although there are less details about the system, a coarse-grained model can provide significant insights [50]. Models that include electrons in calculations (like DFT) are not explained in this chapter, because their application in molecular modeling of nanoparticle is blocked by their computational inefficiency [50].

According to Fig. 23.1, there is inversely proportional relationship between resolution and size scale. High-resolution methods give the information about systems having more details. But these methods have the lowest performance limit and system size. When size scale increases, the performance limit increases because

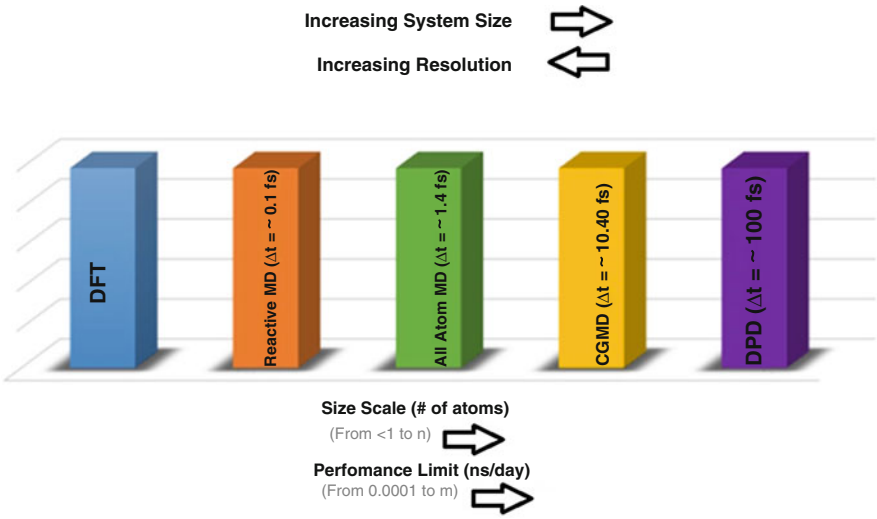


Fig. 23.1 Comparison of performance limit and system size for different molecular-modeling methods

the systems are explained in less detail in relation to the size scale. If it is explained in more detail, increase in the performance limit is due to the increased time steps (Δt) that are possible when describing a system with less detail [49].

In general terms, molecular models contain simplifications that allow the simulation of complex systems. Otherwise, it is not possible to perform the simulation. Another part of the molecular modeling consists of computational methods that help the definition of dynamics, energetics, and conformational sampling of the system. Both full atomistic models and coarse-grained MD have some advantages and disadvantages [49]. In this section, it is aimed to present general information about models preferred in molecular modeling studies to the reader.

23.3.1 MD (Full Atomistic Models)

The MD method enables the simulation of the kinetic and thermodynamic properties of molecular systems with the help of Newton's equations of motion. Molecular dynamic simulations play an important role in predicting and understanding the structure and interactions of molecular systems at the atomic level. They are important tools to make possible predictive molecular design [51].

In modeling the complex systems with large dimensions, methods such as MD are often used to obtain information about binding behavior. In molecular dynamics, we come across traditional MD and all-atom MD uses, that is, nonreactive MD. Traditional MD, by using classical methods, namely Newton's equations of motion, treats atoms as bonded (with spring-like bonds) spheres and enables modeling studies to be carried out. All-atom MD simulations are used to determine the behavior of molecular systems with too many atoms [49, 50].

These computer simulations act as bridge between theoretical and experimental models and between microscopic length and time scales and the macroscopic world of the laboratory [52].

MD simulations are based on the solution of classical equations of motion. These simulations calculate the motion of the atoms in a molecular assembly using Newton's equations to determine the net force and acceleration experienced by each atom (see Eqs. 23.1 and 23.2).

$$\text{Energy function} : U(\rightarrow r_1, \rightarrow r_2, \ldots \rightarrow r_N) = U\left(\vec{R}\right) \tag{23.1}$$

$$\text{Used to determine the force on each atom} : m_i \frac{d^2 \rightarrow r_i}{dt^2} = \rightarrow F_i$$

$$= -\vec{\nabla} U\left(\vec{R}\right) \tag{23.2}$$

In simulations using molecular dynamics, atoms thought to be connected by a spring-like bond are defined by potential energy functions called force fields (FF). Molecular coordinates and velocities connected with a function of simulation time can be appraised via the solving Newton's equation of motion. It is understood from all these narratives that molecular dynamic simulations will be successful by using appropriate force field. The most commonly used force fields such as AMBER [53, 54], CHARMM [55], OPLS [56], and GROMOS [57] are carefully parameterized and corrected for macromolecules such as proteins, carbohydrates, and lipids. However, it is much more difficult to develop force fields to describe inorganic-biology interactions. The reason for these difficulties is that the parameters are variable according to a function of nanoparticle inorganic core material, surface and ligand chemistry, and solvent medium. Current force-field parameters have been obtained from some theoretical and experimental data [49].

The potential energy whose gradient plays a key role in molecular dynamic studies is often expressed as a sum of intramolecular and intermolecular energies. In a classical, MM approach, the intermolecular potential energies consist of electrostatic energy and repulsion and attraction energies, modeled by a Lennard-Jones potential. The intramolecular energies consist of bonding interactions, involving stretching energy, bending energy, and torsion energy (see Eq. 23.3).

$$U_{\text{intramolecular}} = \frac{1}{2} \sum_{\text{bonds}} k_{ij}^{r} \left(r_{ij} - r_{\text{eq}}\right)^2 + \frac{1}{2} \sum_{\substack{\text{bend} \\ \text{angles}}} k_{ijk}^{\theta} \left(\theta_{ijk} - \theta_{\text{eq}}\right)^2$$

$$+ \frac{1}{2} \sum_{\substack{\text{torsion} \\ \text{angles}}} \sum_{m} k_{ijkl}^{\varphi,m} \left(1 + Cos\left(m\varphi_{ijkl} - \gamma_m\right)\right) \tag{23.3}$$

The potential energy function;

$$U\left(\vec{R}\right) = \sum_{\text{bonds}} k_i^{\text{bond}}(r_i - r_0)^2 + \sum_{\text{angles}} k_i^{\text{angle}}(\theta_i - \theta_0)^2 + \sum_{\text{dihedrals}} k_i^{\text{dihedral}}[1 + \text{Cos}\,(n_i\varphi_i + \delta_i)] + \sum_i \sum_{j \neq i} 4\in_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\in r_{ij}} \tag{23.4}$$

In Eq. 23.4, the terms express bond, angle, dihedral, and nonbonded, respectively. Since electrons are not taken into account in molecular studies, it is not possible to conduct studies on chemical reactions and excited states. It is only applicable for systems managed by electrostatic and van der Waals interactions. In addition, simulations can be performed by adding solvent molecules, ions and other molecules to the system in MD studies, taking into account environmental conditions. As a result of the studies, conformation information, attachment poses, and energetic information can be obtained [50].

23.3.1.1 Basic Steps in MD Simulation Studies

There are basically four steps in MD simulations. These are system preparation, minimization, equilibration, and production.

System preparation focuses on preparing the starting state of the desired system for input to an appropriate simulation package, including building a starting structure, solvating (if necessary), applying a force field, etc.

The aim of minimization is to get rid of "clashes," that is, relax the structure toward a minimum. The forces on any atom in the simulation are not large enough to carry atoms to an unreasonable distance in a single time step. In this step standard minimization algorithms such as Steepest Descent [51] are used.

MD simulations can be carried out to generate different statistical ensembles as specified by different conditions [58, 59].

If the total number of atoms N and the volume of the unit cell are kept constant, MD simulations are said to be made in the microcanonical (NVE) ensemble. In large NVE systems, fluctuations in temperature are small and can be considered approximately constant. There are also situations in which volume and temperature should be kept constant. Therefore, for these classes, MD must reproduce an isothermal ensemble such as the canonical NVT ensemble where the number of particles, volume, and temperature are fixed, or the isothermal isobaric NPT ensemble where the number of particles, pressure, and temperature are fixed [58].

After the preparation, minimization, and equilibration steps of the system, the production step is passed to simulate the desired conditions and to collect data for subsequent analysis.

23.3.2 Enhanced Sampling Method

Simulations using MD methods have time scales up to 1000 ns and lengths up to 20 nm. In simulations, long time intervals are required for many events (such as molecular bonding) to occur, and MD may be unable to resolve such situations. In simulations such as binding of proteins, there are usually local minimums separated by high energy barriers, and because of the minimum, simulations can limit complete sampling as the simulations can get stuck to the local energy minimum. There are different approaches to solve this type of problems [50]. The enhanced sampling method has been developed for large conformational changes of biological systems, which is a very time-consuming and complex process for cases where MD simulations are insufficient [60]. Also, the enhanced sampling methods allow better statistics to be collected [61].

23.3.2.1 Replica-Exchange MD (REMD)

The enhanced sampling technique, developed to assist in cases where MD simulations are insufficient, is also classified into classes. In the replica exchange MD (REMD) method, independent system states dependent on a particular variable such as temperature and defined by the positions of the atoms are exchanged in relation to temperature and energy differences between simulations based on the MC acceptance scoring algorithm. REMD identifies these probabilities of change in system states and enables these possibilities to be determined quickly from the properties of the system. These situations help to investigate the conformational area in more detail by overcoming the energy barriers on the potential energy surface. Independent simulations will experience at least one variable change and therefore continuity will not be possible. But it will fit a suitable Boltzmann distribution [60]. REMD method has been widely used in recent years in developing conformational sampling of MD simulations. The REMD method is common in MD packages that are widely used around the world, such as AMBER [53], GROMACS [62], and NAMD [63].

23.3.2.2 Metadynamics

The metadynamic method, which was put forward by Alessandro Laio and Michele Parrinello [64] and also known by different names such as MTD, METAD, or MEtaD, is a method used in MD studies. Metadynamics prevents many samplings of local energy minimization, that is, after a state is sampled, it prevents the sampling of the same states again. This inhibition is accomplished by adding a positive Gauss potential to the real energy landscape. Metadynamics accept Gaussian-like bias to overcome high energy barriers in free energy surface (FES). The metadynamic technique can be expressed as filling free energy wells with computational sand. In this method, where the local minimum is avoided, it adds the Gaussian hill to the potential energy of the existing region of state space. The local minimum is filled by adding Gaussian hills, allowing the system to achieve different configurations. As the simulation extends, all potential energy fills with the accumulated Gaussian hills, and eventually the total potential energy flattens [60, 65, 66]. Applications of

metadynamics in MD programs such as NAMD and GROMACS have provided opportunities for use in many different fields from solid state to biology.

23.3.3 Coarse-Grained Models

Coarse-grained modeling, based on the work of Michael Levitt and Ariel Warshel [67–69], aims to learn the behavior of the system using coarse-grained representations of complex systems. This model, which is frequently used in biomolecules [70, 71], enables modeling by providing various coarse-grained models for modeling molecules such as proteins, nucleic acids [72], and lipid membranes [73]. In these models, the molecules are represented by "pseudo-atoms" approximating groups of atoms, not by individual atoms, as in other models. CGMD (coarse-grained MD) allows for larger system sizes, time scales, and time steps, by decreasing the details of the molecular representation [49]. MARTINI, commonly used in the coarse-grained model, is the force field in which four heavy atoms are grouped in one bead [74]. It was originally developed for lipids [75], then modified for proteins [76, 77] and DNA [78]. While all atomic MD have time steps 1–4 fs, force fields such as MARTINI used in CGMD allow use of time steps of 10–40 fs. Force field parameters were obtained from experimental data [49]. Coarse-grained models have been used for practical applications in MD studies [71].

23.3.3.1 Dissipative Particle Dynamics

Dissipative particle dynamics (DPD), designed by Hoogerbrugge and Koelman [79, 80], is a technique for simulating the dynamic and rheological properties of fluids. The technique was modified by P. Español [81] to ensure proper thermal equilibrium condition. DPD, which has longer time and length scales than traditional MD simulations, has been used in studies on the formation of nanoparticles. However, acceptance of multiple atoms and even molecular clusters as single beads, as in CGMD, is an indication that atomic details are weak in this technique [49]. The DPD method has been shown in literature studies to simulate liposome formation [82] and to model heterogeneous multiphase fluids containing polymer micelles [83].

23.3.4 Comparison of Models for Nanomaterial-Biology Interactions

Full atomistic models have the advantages like the highest level of detail as these models comprise distinctly all atoms in the system. These models help to brighten suitable description of the nano-bio interface with those fundamental interactions such as van der Waals, electrostatic, hydrogen bonding, pi-pi stacking, and pi-cation interactions. Also, these models help to take into account environmental effects by incorporating solvent molecules, ions, and other molecules into the system. In studies with proteins in nano-bio interface interactions, it is possible to emphasize amino acids that direct structural changes of proteins at a single amino acid level

with the help of MD simulation and atomic scale resolutions. Simulation results can be compared with experimental results and their accuracy can be evaluated. We know that MD simulations have the limited range for time and length. Also, the enhanced sampling methods are inhibitive for large and complex systems.

Coarse-grained models have much longer time and length scales than full-atom models, and for these reasons, it is a preferable model, although it is difficult. As with the other model, this model has some flaws in itself. This model roughly explains interactions such as electrostatic interactions, hydrogen bonds, and solvent effects, and in this case is an indication that atomic details are ignored. In the CG model, if the properties of the system are correctly reproduced and the interaction potentials are parameterized, simulations made in this model scale are important tools that can be used to complete MD simulations. In addition, for more accurate analysis at the atomic level, the CG model can help obtain some input structures for protein-protein interactions that are difficult to obtain in MD studies. Enhanced sampling methods can support simulations on the CG scale when the time scale is not accessible [50]. The comparison of the models is summarized in Table 23.1.

23.3.5 Solvent Model

It is important to determine the appropriate solvent models in modeling studies. Parameterization should be done meticulously for accurate representation of solvents and solvent-solver interactions, and the choice of force field for the solvent becomes an imperative [85]. Even for the most commonly used solvent, water, there are many force fields. SPC [86], SPC/E [87], TIP3P [88], TIP4P [89], and TIP5P [90] are different models used for water. Each model represents different properties of water with different degrees of sensitivity.

Explicit solvent can give the prediction about solvent effects (e.g., at solute−solvent interactions). However, as the number of particles to be simulated will increase, models to be used are important in such systems. The nucleic acid simulations (the system consists of charged molecules) indicate that the force field parameters of the salts must be selected meticulously. Otherwise, it may cause simulation operation to be incorrect [91].

Table 23.1 The comparison of full atomistic and coarse-grained models

Full atomistic model	Coarse-grained model
Quantitative	Semi-quantitative
Microscopic level	Mesoscopic level [84]
Large range interactions only	Short range interactions only
Highest level of atomic detail	–
–	Longer time and length scales

23.4 The Software Used in Molecular Modeling

With the development of more advanced technologies and simulation algorithms, molecular simulations have also been significantly improved and diversified in size, length, and system complexity. With the advent of many molecular modeling software packages, including GROMACS, Amber, CHARMM, GROMOS, and NAMD, it has become available for further research. There are various empirical force fields for MD simulations. The most common are CHARMM (Chemistry at HARvard Macromolecular Mechanics) [92], AMBER (Assisted Model Building with Energy Refinement) [93], GROMACS (GROningenMAchine for Chemical Simulations) [94], and NAMD (NAnoscale MD) [63]. While CHARMM, AMBER, and GROMACS have force fields in their names, NAMD can be used with any force field. Within these force fields, the CHARMM and AMBER force fields are developed for proteins and nucleic acids and are very similar to each other. AMBER software package was written in Fortran 90, and C programming languages perform biomolecule simulations using the AMBER force field. CHARMM program also enables the realization and analysis of a wide variety of molecular simulations. Initially, GROMACS improved for lipid simulations and less successful for proteins, but later versions were developed. GROMACS is also an MD program package designed for simulations of proteins, lipids, and nucleic acids. This program works faster than many simulation programs. GROMACS is run through the command line interface, and its files are used as input and output. It provides a comprehensive view for calculation, progression and estimated end time (ETA) feedback, trajectory viewer, and trajectory analysis. In addition, the availability of different force fields makes GROMACS a preferable program [95].

The GROMACS program was first created at the University of Groningen in the Netherlands in the early 1990s, using the source code system and its expansion is "GROningenMAchine for Chemical Simulation." Although it was primarily designed for biochemical molecules such as proteins, lipids, and nucleic acids that contain complex linked interactions, it was later used to compute unbound interactions as well. In addition to these, it is used for nonbiological systems such as polymers. The GROMACS program can be used with GROMOS, OPLS, AMBER, and ENCAD force fields, and it consists of many subprograms, but does not have any visual interface [95].

NAMD is a parallel program where calculations are made using large parallel machines. In the most common methods, the shear distance is used to calculate binary interactions between atoms. Nonbonding interactions between atoms outside this shear radius are not calculated or can be calculated very rarely. NAMD uses molecular structure files, XPLOR coordinates, and CHARMM force fields. It can also use periodic boundary conditions on any combination of the three coordinate axes, along with nonperiodic simulations. It performs all electrostatic simulations and cutting simulations. The VMD program [96] can be used to observe the

interactions in the ongoing simulations in the NAMD program [97]. Apart from these programs, LAMMPS and GROMOS programs can also be used for MD calculations.

23.5 Applications of Molecular Modeling for Nanoparticles

In 2015, a study of citrate-coated gold nanoparticles (AuNPs) in interaction with insulin and fibrinogen was carried out with DL_POLY_2.20 software by Tavanti et al. using coarse-grained MD simulations. At the end of study, a detailed description of the process of corona formation was achieved [98]. Same group (with same computational model and software) carried out the study of computational simulations of the ubiquitin corona around gold nanoparticles. In this study the effects of nanoparticle size and environment were taken into account. It is reported that the ubiquitin binding styles and changes depend on the nanoparticle environment and size [99]. A study in 2016 was carried out by Yu and Zhou with GROMACS software program using MARTINI force field. In this study, lysozyme adsorption on different-sized silica nanoparticles (SNPs) was simulated. It was reported that orientation distribution and conformation change of adsorbed lysozyme relate to the increase of nanoparticle size [100]. Looking at the studies using the dissipative particle dynamics, the interactions between the nanoparticle-protein corona complex and cell membranes of different types were investigated by Ding and Ma in 2014. This study revealed that the biological importance of protein corona can be understood thoroughly, and there are some suggestions for the design of nanoparticles in drug delivery [101]. The interactions of single-wall carbon nanotubes with human serum proteins were researched with MD simulations using NAMD2 software program. In this study, both experimental and theoretical studies were achieved. According to the results of study, important information was presented about the design of safe carbon nanotube nanomaterials [102]. The studies are detailed in Table 23.2. Additionally, information of other software for nanostructures modeling with some applications is provided (see Table 23.3).

23.5.1 Docking Analysis of Nanoparticles

Docking analysis allows to determine the binding conformations during the binding of the known ligand to the active site of the target protein and the binding energy that occurs during this placement. Today, there are many programs for docking studies, but since this program uses different algorithms and scoring functions to determine the docking position, they differ from each other. The predicted binding conformation and orientation of the ligand to the active binding site of the targeted receptor is called the binding pose. The algorithm of the docking program is performed by calculating and placing the ligand in the flexible position into the targeted region; these algorithms are performed in three basic categories.

Table 23.2 Some applications of nanoparticles in molecular modeling

Device material	Protein	Method	Outcomes	Reference
Carbon nanotubes	Bovine fibrinogen Immunoglobulin Transferrin Bovine serum albumin	MD	Protein affinity with the surface Structural changes	[102, 103]
Generic hydrophobic nanoparticle	a1-Antitrypsin Human serum albumin Transferrin Immunoglobulin G Fibrinogen a2-Macroglobulin	CG	Binding energies Structural changes	[104]
Gold nanoparticles	Insulin Fibrinogen	CG	Competitive binding Structural changes	[98, 105]
Silica nanoparticles	Lysozyme	CG (MARTINI)	Curvature effects on lysozyme adsorption	[100]
Generic hydrophobic/ hydrophilic nanoparticle	Bovine serum albumin	CG (DPD)	Binding energy as a function of size and surface characteristics	[101]

Table 23.3 Information on other software for nanostructures modeling

Software	Initial release	Written in	Developer (s)	Website	Applications
CP2K	2000	Fortran	CP2K developers group	cp2k.org	[106, 107]
LAMPPS	1995	C++	Sandia National Laboratories Temple University	lammps.sandia.gov	[108, 109]
Materials Studio	2000		Accelrys, BIOVIA	3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-materials-studio/	[110, 111]
Scigress	2016	C++, C, Java, Fortran	Fujitsu Limited	scigress.com	[112, 113]

1. Systematic methods
2. Random methods
3. Simulation methods

Molecular docking analysis is often used to reveal the interaction between metallic NPs and protein targets, for example, between cadmium oxide NPs and cancerous proteins [114] and between palladium NP and viral protein templates (VP6) [115]. In both studies, the docking of nanoparticles was carried out using AutoDock (The Scripps Research Institute, La Jolla, CA, USA) [116], which is a frequently used program in docking studies. The aims of nanoparticle studies are to examine the interaction with proteins in the target region and to determine the binding sites of nanoparticles. Carreño-Fuentes et al. investigated the metal nanoparticles and interactions of metal-binding sites in the specified protein. Gowri et al. examined the interaction of nanoparticles with various proteins involved in cancer activity. Apart from this, quantum mechanics (QM) calculations also provide detailed information about the intrinsic reactivity of various complex structures. Therefore, molecular docking calculations and QM calculations are very important in drug design and development, in the prediction of unknown molecular structures, in determining drug interactions, and in revealing enzyme reactions [117].

It has been observed that CuO NPs cause hepatotoxicity and nephrotoxicity and also cause neurotoxicity and genotoxicity [118–122]. TiO_2 NPs were also revealed to cause similar toxicity, but to a lesser extent compared to CuO NPs. For this reason, molecular docking analysis method was used to understand the interaction mechanism of serum albumin, which is one of the important transport proteins in human blood and plays an important role for drug delivery systems by nanoparticles with high toxic effects such as CuO NP and TiO_2NPs.

23.6 Conclusion

Recently one of the top-priority tasks of nano-medicine is the creation of drug with "ideal" properties. Though nanoparticles present many advantages for nano-medicine applications, still there are some health hazard concerns due to their unrevealed long-term effects and toxicity on tissues. In this context, before in vitro and in vivo studies, at the first stage using theoretical methods with in silico studies may reduce many experimental steps. Consequently, in silico methods will be used more frequently and effectively by researchers in the near future.

23.7 Credible Online Resources for Further Reading

URL	Category of source	About
https://www.gromacs.org/	Institution	Molecular dynamics simulations
https://ambermd.org/	Institution	Molecular dynamics simulations
https://www.charmm.org/	Institution	Molecular simulation
https://www.ks.uiuc.edu/Research/namd/	Institution	Molecular dynamics simulations
http://gaussian.com/	Gaussian, Inc.	MM, quantum chemistry method calculations
www.hyper.com	Hypercube, Inc.	Molecular modeling
https://www.schrodinger.com/	Schrödinger, Inc.	Molecular modeling and drug design
http://www.arguslab.com/arguslab.com/ArgusLab.html	Institution	Molecular modeling, graphics, and drug design
http://autodock.scripps.edu/	Institution	Molecular modeling

References

1. Aydın G (2009) Nikel-Rodyum alaşımının bazı termodinamik ve termoelastik özelliklerinin moleküler dinamik simülasyon yöntemi ile incelenmesi Yüksek Lisans Tezi, Ankara
2. Uğurluoğlu N (2006) Altın ve aliminyumun bazı termoelastik özelliklerinin moleküler dinamik simülasyon tekniği ile incelenmesi: Yüksek lisans tezi, Gazi Üniversitesi
3. Heermann DW (1990) Computer-simulation methods. Computer simulation methods in theoretical physics. Springer, pp 8–12
4. Karagöz H (2010) Ni-Au Alaşımlarına Paladyum Katkısının Termodinamik Ve Mekanik Özelliklere Etkisinin Moleküler Dinamik İncelenmesi Yüksek Lisans Tezi, Sakarya Üniversitesi
5. Kart HH, Tomak M, Uludoğan M, Çağın T (2005) Thermodynamical and mechanical properties of Pd–Ag alloys. Comput Mater Sci 32(1):107–117. https://doi.org/10.1016/j.commatsci.2004.07.003
6. Cong H-R, Bian X-F, Zhang J-X, Li H (2002) Structure properties of Cu Ni alloys at the rapid cooling rate using embedded-atom method. Mater Sci Eng A Struct Mater 326(2):343–347. https://doi.org/10.1016/S0921-5093(01)01527-1
7. Fruio E, Çeviri Çolakoğlu K, Kara C (1997) Moleküler Dinamik El-Kitabı. Springer College in Computational Physics, ICTP, Trieste, p 65
8. Grujicic M, Dang P (1995) Computer simulation of martensitic transformation in Fe-Ni face-centered cubic alloys. Mater Sci Eng A Struct Mater 201(1–2):194–204. https://doi.org/10.1016/0921-5093(94)09735-6
9. Catlow C (1990) An introduction to computer modelling of condensed matter. In: Computer modelling of fluids polymers and solids. Springer, pp 1–23
10. Haile JM (1992) Molecular dynamics simulation: elementary methods. Wiley
11. Yıldırım H (2012) Bakır metali nanoparçacıkların moleküler dinamik simülasyonu: Pamukkale Üniversitesi Fen Bilimleri Enstitüsü

12. Parveen S, Misra R, Sahoo SK (2012) Nanoparticles: a boon to drug delivery, therapeutics, diagnostics and imaging. Nanomedicine 8(2):147–166. https://doi.org/10.1016/j.nano.2011.05.016
13. Sadjadi S (2016) Dendrimers as nanoreactors. In: Organic nanoreactors. Elsevier, pp 159–201. https://doi.org/10.1016/B978-0-12-801713-5.00006-9
14. Singh TG, Sharma N (2016) Nanobiomaterials in cosmetics: current status and future prospects. In: Nanobiomaterials in galenic formulations and cosmetics. Elsevier, pp 149–174. https://doi.org/10.1016/B978-0-323-42868-2.00007-3
15. Garg T, Singh O, Arora S, Murthy R (2011) Dendrimer—a novel scaffold for drug delivery. Int J Pharm Sci Rev Res 7(2):211–220
16. Janaszewska A, Lazniewska J, Trzepiński P, Marcinkowska M, Klajnert-Maculewicz B (2019) Cytotoxicity of dendrimers. Biomolecules 9(8):330. https://doi.org/10.3390/biom9080330
17. Ciolkowski M, Rozanek M, Bryszewska M, Klajnert B (2013) The influence of PAMAM dendrimers surface groups on their interaction with porcine pepsin. Biochim Biophys Acta 1834(10):1982–1987. https://doi.org/10.1016/j.bbapap.2013.06.020
18. Hanafy NA, El-Kemary M, Leporatti S (2018) Micelles structure development as a strategy to improve smart cancer therapy. Cancers 10(7):238. https://doi.org/10.3390/cancers10070238
19. Lu Y, Zhang E, Yang J, Cao Z (2018) Strategies to improve micelle stability for drug delivery. Nano Res 11(10):4985–4998. https://doi.org/10.1007/s12274-018-2152-3
20. Daraee H, Etemadi A, Kouhi M, Alimirzalu S, Akbarzadeh A (2016) Application of liposomes in medicine and drug delivery. Artif Cells Nanomed Biotechnol 44(1):381–391. https://doi.org/10.3109/21691401.2014.953633
21. Liu X, Huang G (2013) Formation strategies, mechanism of intracellular delivery and potential clinical applications of pH-sensitive liposomes. Asian J Pharm Sci 8(6):319–323. https://doi.org/10.1016/j.ajps.2013.11.002
22. Sercombe L, Veerati T, Moheimani F, Wu SY, Sood AK, Hua S (2015) Advances and challenges of liposome assisted drug delivery. Front Pharmacol 6:286. https://doi.org/10.3389/fphar.2015.00286
23. Zielińska A, Carreiró F, Oliveira AM, Neves A, Pires B, Venkatesh DN et al (2020) Polymeric nanoparticles: production, characterization, toxicology and ecotoxicology. Molecules 25(16):3731. https://doi.org/10.3390/molecules25163731
24. Campos PM, Bentley MVLB, Torchilin VP (2016) Nanopreparations for skin cancer therapy. In: Nanobiomaterials in cancer therapy. Elsevier, pp 1–23. https://doi.org/10.1016/B978-0-323-42863-7.00001-3
25. Jawahar N, Meyyanathan S (2012) Polymeric nanoparticles for drug delivery and targeting: a comprehensive review. Int J Health Allied Sci 1(4):217. https://doi.org/10.4103/2278-344X.107832
26. Chenthamara D, Subramaniam S, Ramakrishnan SG, Krishnaswamy S, Essa MM, Lin F-H et al (2019) Therapeutic efficacy of nanoparticles and routes of administration. Biomater Res 23(1):1–29. https://doi.org/10.1186/s40824-019-0166-x
27. Karimi M, Ghasemi A, Zangabad PS, Rahighi R, Basri SMM, Mirshekari H et al (2016) Smart micro/nanoparticles in stimulus-responsive drug/gene delivery systems. Chem Soc Rev 45(5):1457–1501. https://doi.org/10.1039/C5CS00798D
28. Senapati S, Mahanta AK, Kumar S, Maiti P (2018) Controlled drug delivery vehicles for cancer treatment and their performance. Signal Transduct Target Ther 3(1):1–19. https://doi.org/10.1038/s41392-017-0004-3
29. Mody VV, Siwale R, Singh A, Mody HR (2010) Introduction to metallic nanoparticles. J Pharm Bioallied Sci 2(4):282. https://doi.org/10.4103/0975-7406.72127
30. Khandel P, Yadaw RK, Soni DK, Kanwar L, Shahi SK (2018) Biogenesis of metal nanoparticles and their pharmacological applications: present status and application prospects. J Nanostruct Chem 8(3):217–254. https://doi.org/10.1007/s40097-018-0267-4

31. Khodashenas B, Ghorbani HR (2019) Synthesis of silver nanoparticles with different shapes. Arab J Chem 12(8):1823–1838. https://doi.org/10.1016/j.arabjc.2014.12.014
32. Ajdary M, Moosavi MA, Rahmati M, Falahati M, Mahboubi M, Mandegary A et al (2018) Health concerns of various nanoparticles: a review of their in vitro and in vivo toxicity. Nanomaterials 8(9):634. https://doi.org/10.3390/nano8090634
33. Długosz O, Szostak K, Staroń A, Pulit-Prociak J, Banach M (2020) Methods for reducing the toxicity of metal and metal oxide NPs as biomedicine. Materials 13(2):279. https://doi.org/10.3390/ma13020279
34. Bhaviripudi S, Mile E, Steiner SA, Zare AT, Dresselhaus MS, Belcher AM et al (2007) CVD synthesis of single-walled carbon nanotubes from gold nanoparticle catalysts. J Am Chem Soc 129(6):1516–1517. https://doi.org/10.1021/ja0673332
35. Chowdhury SM, Lalwani G, Zhang K, Yang JY, Neville K, Sitharaman B (2013) Cell specific cytotoxicity and uptake of graphene nanoribbons. Biomaterials 34(1):283–293. https://doi.org/10.1016/j.biomaterials.2012.09.057
36. Lalwani G, Cai X, Nie L, Wang LV, Sitharaman B (2013) Graphene-based contrast agents for photoacoustic and thermoacoustic tomography. Photoacoustics 1(3–4):62–67. https://doi.org/10.1016/j.pacs.2013.10.001
37. Lalwani G, Patel SC, Sitharaman B (2016) Two-and three-dimensional all-carbon nanomaterial assemblies for tissue engineering and regenerative medicine. Ann Biomed Eng 44(6):2020–2035. https://doi.org/10.1007/s10439-016-1623-5
38. Patra JK, Das G, Fraceto LF, Campos EVR, del Pilar Rodriguez-Torres M, Acosta-Torres LS et al (2018) Nano based drug delivery systems: recent developments and future prospects. J Nanobiotechnol 16(1):1–33. https://doi.org/10.1186/s12951-018-0392-8
39. Friedman AD, Claypool SE, Liu R (2013) The smart targeting of nanoparticles. Curr Pharm Des 19(35):6315–6329
40. Roma-Rodrigues C, Mendes R, Baptista PV, Fernandes AR (2019) Targeting tumor microenvironment for cancer therapy. Int J Mol Sci 20(4):840. https://doi.org/10.3390/ijms20040840
41. Farokhzad OC, Langer R (2006) Nanomedicine: developing smarter therapeutic and diagnostic modalities. Adv Drug Deliv Rev 58(14):1456–1459. https://doi.org/10.1016/j.addr.2006.09.011
42. Ferrari M (2005) Cancer nanotechnology: opportunities and challenges. Nat Rev Cancer 5(3):161–171. https://doi.org/10.1038/nrc1566
43. ud Din F, Aman W, Ullah I, Qureshi OS, Mustapha O, Shafique S et al (2017) Effective use of nanocarriers as drug delivery systems for the treatment of selected tumors. Int J Nanomed 12:7291. https://doi.org/10.2147/IJN.S146315
44. Janeway Jr CA, Travers P, Walport M, Shlomchik MJ (2001) The interaction of the antibody molecule with specific antigen. Immunobiology: the immune system in health and disease, 5th edn. Garland Science
45. Kraitchman DL, Bulte JW (2008) Imaging of stem cells using MRI. Basic Res Cardiol 103(2):105–113. https://doi.org/10.1007/s00395-008-0704-5
46. Chen N, Wang H, Huang Q, Li J, Yan J, He D et al (2014) Long-term effects of nanoparticles on nutrition and metabolism. Small 10(18):3603–3611. https://doi.org/10.1002/smll.201303635
47. Kumar V, Sharma N, Maitra S (2017) In vitro and in vivo toxicity assessment of nanoparticles. Int Nano Lett 7(4):243–256. https://doi.org/10.1007/s40089-017-0221-3
48. Buzea C, Pacheco II, Robbie K (2007) Nanomaterials and nanoparticles: sources and toxicity. Biointerphases 2(4):MR17–MR71. https://doi.org/10.1116/1.2815690
49. Nash JA, Kwansa AL, Peerless JS, Kim HS, Yingling YG (2017) Advances in molecular modeling of nanoparticle–nucleic acid interfaces. Bioconjug Chem 28(1):3–10. https://doi.org/10.1021/acs.bioconjchem.6b00534
50. Casalini T, Limongelli V, Schmutz M, Som C, Jordan O, Wick P et al (2019) Molecular modeling for nanomaterials–biology interactions: opportunities, challenges and perspectives. Front Bioeng Biotechnol 7:268. https://doi.org/10.3389/fbioe.2019.00268

51. Braun E, Gilmer J, Mayes HB, Mobley DL, Monroe JI, Prasad S et al (2019) Best practices for foundations in molecular simulations [Article v1. 0]. Living J Comput Mol Sci 1(1):5957. https://doi.org/10.33011/livecoms.1.1.5957
52. Allen MP (2004) Introduction to molecular dynamics simulation. Computational soft matter: from synthetic polymers to proteins. 23(1):1–23
53. Case D, Ben-Shalom I, Brozell S, Cerutti D, Cheatham T III, Cruzeiro V et al (2018) AMBER 2018. University of California, San Francisco
54. Ponder JW, Case DA (2003) Force fields for protein simulations. Adv Protein Chem 66:27–85. https://doi.org/10.1016/S0065-3233(03)66002-X. Elsevier
55. MacKerell AD Jr, Banavali N, Foloppe N (2000) Development and current status of the CHARMM force field for nucleic acids. Biopolymers 56(4):257–265. https://doi.org/10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W
56. Jorgensen WL, Tirado-Rives J (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 110(6):1657–1666
57. van Gunsteren WF, Berendsen HJ (1987) Groningen molecular simulation (GROMOS) library manual. Biomos, Groningen 24(682704):13
58. Halitoğlu S (2008) Poliimid Gaz Ayırma Membranlarının Moleküler Simülasyonu: Fen Bilimleri Enstitüsü
59. Kotelyanskii M, Theodorou DN (2004) Simulation methods for polymers. CRC Press
60. Bernardi RC, Melo MC, Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochim Biophys Acta 1850(5):872–877. https://doi.org/10.1016/j.bbagen.2014.10.019
61. Lie HC, Quer J (2017) Some connections between importance sampling and enhanced sampling methods in molecular dynamics. J Chem Phys 147(19):194107. https://doi.org/10.1063/1.4989495
62. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718. https://doi.org/10.1002/jcc.20291
63. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E et al (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802
64. Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci 99(20):12562–12566. https://doi.org/10.1073/pnas.202427399
65. Musiani F, Giorgetti A (2017) Protein aggregation and molecular crowding: perspectives from multiscale simulations. Int Rev Cell Mol Biol 329:49–77. https://doi.org/10.1016/bs.ircmb.2016.08.009. Elsevier
66. Zheng L, Alhossary AA, Kwoh C-K, Mu Y (2019) Molecular dynamics and simulation. Academic Press, pp 550-566.
67. Levitt M (2014) Birth and future of multiscale modeling for macromolecular systems (Nobel Lecture). Angew Chem Int Ed 53(38):10006–10018. https://doi.org/10.1002/anie.201403691
68. Levitt M, Warshel A (1975) Computer simulation of protein folding. Nature 253(5494):694–698. https://doi.org/10.1038/253694a0
69. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol 103(2):227–249. https://doi.org/10.1016/0022-2836(76)90311-9
70. Ingólfsson HI, Lopez CA, Uusitalo JJ, de Jong DH, Gopal SM, Periole X et al (2014) The power of coarse graining in biomolecular simulations. Wiley Interdiscip Rev Comput Mol Sci 4(3):225–248. https://doi.org/10.1002/wcms.1169
71. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A (2016) Coarse-grained protein models and their applications. Chem Rev 116(14):7898–7936
72. Potoyan DA, Savelyev A, Papoian GA (2013) Recent successes in coarse-grained modeling of DNA. Wiley Interdiscip Rev Comput Mol Sci 3(1):69–83. https://doi.org/10.1002/wcms.1114

73. Baron R, Trzesniak D, de Vries AH, Elsener A, Marrink SJ, van Gunsteren WF (2007) Comparison of thermodynamic properties of coarse-grained and atomic-level simulation models. ChemPhysChem 8(3):452–461. https://doi.org/10.1002/cphc.200600658
74. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B 111(27):7812–7824. https://doi.org/10.1021/jp071097f
75. Marrink SJ, De Vries AH, Mark AE (2004) Coarse grained model for semiquantitative lipid simulations. J Phys Chem B 108(2):750–760. https://doi.org/10.1021/jp036508g
76. de Jong DH, Singh G, Bennett WD, Arnarez C, Wassenaar TA, Schäfer LV et al (2013) Improved parameters for the martini coarse-grained protein force field. J Chem Theory Comput 9(1):687–697. https://doi.org/10.1021/ct300646g
77. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S-J (2008) The MARTINI coarse-grained force field: extension to proteins. J Chem Theory Comput 4(5):819–834. https://doi.org/10.1021/ct700324x
78. Uusitalo JJ, Ingólfsson HI, Akhshi P, Tieleman DP, Marrink SJ (2015) Martini coarse-grained force field: extension to DNA. J Chem Theory Comput 11(8):3932–3945. https://doi.org/10.1021/acs.jctc.5b00286
79. Hoogerbrugge P, Koelman J (1992) Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. Europhys Lett 19(3):155
80. Koelman J, Hoogerbrugge P (1993) Dynamic simulations of hard-sphere suspensions under steady shear. Europhys Lett 21(3):363
81. Espanol P, Warren P (1995) Statistical mechanics of dissipative particle dynamics. Europhys Lett 30(4):191
82. Sevink G, Fraaije J (2014) Efficient solvent-free dissipative particle dynamics for lipid bilayers. Soft Matter 10(28):5129–5146. https://doi.org/10.1039/C4SM00297K
83. Tang Y-H, Li Z, Li X, Deng M, Karniadakis GE (2016) Non-equilibrium dynamics of vesicles and micelles by self-assembly of block copolymers with double thermoresponsivity. Macromolecules 49(7):2895–2903. https://doi.org/10.1021/acs.macromol.6b00365
84. Harmandaris V, Kalligiannaki E, Katsoulakis M, Plecháč P (eds) (2017) From atomistic to systematic coarse-grained models for molecular systems. Eccomas Proc UNCECOMP 394–405. https://doi.org/10.7712/120217.5378.17211
85. Levy RM, Gallicchio E (1998) Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. Annu Rev Phys Chem 49(1):531–567
86. Berendsen HJ, Postma JP, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. Intermolecular forces. Springer, pp 331–342
87. Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. J Phys Chem 91(24):6269–6271
88. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935
89. Jorgensen WL, Madura JD (1985) Temperature and size dependence for Monte Carlo simulations of TIP4P water. Mol Phys 56(6):1381–1392
90. Mahoney MW, Jorgensen WL (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J Chem Phys 112(20):8910–8922. https://doi.org/10.1063/1.481505
91. Joung IS, Cheatham TE III (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J Phys Chem B 112(30):9020–9041. https://doi.org/10.1021/jp8001614
92. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616
93. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE III, DeBolt S et al (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode

analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun 91(1–3):1–41
94. Lindahl E, Hess B, Van Der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 7(8):306–317. https://doi.org/10.1007/s008940100045
95. Kökcü Y (2018) Antitümör ve Antioksidant Özellikli Tripeptid Yüklü Polimerik Nanopartiküllerin Geliştirilmesi: Yüksek lisans tezi, İstanbul Üniversitesi
96. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(1):33–38
97. Mercan Tez Ö (2014) Dengede olmayan biyolojik sistemler için crooks teoremi yardımıyla serbest enerji hesaplamasının uygulanması: Doktora tezi, Süleyman Demirel Üniversitesi
98. Tavanti F, Pedone A, Menziani MC (2015) Competitive binding of proteins to gold nanoparticles disclosed by molecular dynamics simulations. J Phys Chem C 119(38):22172–22180. https://doi.org/10.1021/acs.jpcc.5b05796
99. Tavanti F, Pedone A, Menziani MC (2015) A closer look into the ubiquitin corona on gold nanoparticles by computational studies. New J Chem 39(4):2474–2482. https://doi.org/10.1039/C4NJ01752H
100. Yu G, Zhou J (2016) Understanding the curvature effect of silica nanoparticles on lysozyme adsorption orientation and conformation: a mesoscopic coarse-grained simulation study. Phys Chem Chem Phys 18(34):23500–23507. https://doi.org/10.1039/C6CP01478J
101. Ding HM, Ma YQ (2014) Computer simulation of the role of protein corona in cellular delivery of nanoparticles. Biomaterials 35(30):8703–8710. https://doi.org/10.1016/j.biomaterials.2014.06.033
102. Ge C, Du J, Zhao L, Wang L, Liu Y, Li D et al (2011) Binding of blood proteins to carbon nanotubes reduces cytotoxicity. Proc Natl Acad Sci 108(41):16968–16973. https://doi.org/10.1073/pnas.1105270108
103. Gu Z, Yang Z, Chong Y, Ge C, Weber JK, Bell DR, Zhou R (2015) Surface curvature relation to protein adsorption for carbon-based nanomaterials. Sci Rep 5:10886. https://doi.org/10.1038/srep10886
104. Lopez H, Lobaskin V (2015) Coarse-grained model of adsorption of blood plasma proteins onto nanoparticles. J Chem Phys 143(24):243138. https://doi.org/10.1063/1.4936908
105. Quan X, Peng C, Zhao D, Li L, Fan J, Zhou J (2017) Molecular understanding of the penetration of functionalized gold nanoparticles into asymmetric membranes. Langmuir 33(1):361–371. https://doi.org/10.1021/acs.langmuir.6b02937
106. Laasonen K, Panizon E, Bochicchio D, Ferrando R (2013) Competition between icosahedral motifs in AgCu, AgNi, and AgCo nanoalloys: a combined atomistic–DFT study. J Phys Chem C 117(49):26405–26413. https://doi.org/10.1021/jp410379u
107. Mozaffari S, Li W, Dixit M, Seifert S, Lee B, Kovarik L et al (2019) The role of nanoparticle size and ligand coverage in size focusing of colloidal metal nanoparticles. Nanoscale Adv 1(10):4052–4066. https://doi.org/10.1039/C9NA00348G
108. Demir B, Chan K-Y, Yang D, Mouritz A, Lin H, Jia B et al (2019) Epoxy-gold nanoparticle nanocomposites with enhanced thermo-mechanical properties: an integrated modelling and experimental study. Compos Sci Technol 174:106–116. https://doi.org/10.1016/j.compscitech.2019.02.020
109. Kamiński M, Jurkiewicz K, Burian A, Bródka A (2020) The structure of gold nanoparticles: molecular dynamics modeling and its verification by X-ray diffraction. J Appl Crystallogr 53(1). https://doi.org/10.11071/S1600576719014511
110. Al Hasan N (ed) (2018) Prediction of mechanical properties of EPON 862 (DGEBF) cross-linked with curing agent (TETA) and SiO2 nanoparticle based on materials studio. In: Proceedings of the 1st international conference on materials engineering and science (IConMEAS), Istanbul, Turkey

111. Mofradnia SR, Tavakoli Z, Yazdian F, Rashedi H, Rasekh B (2018) Fe/starch nanoparticle-*Pseudomonas aeruginosa*: bio-physiochemical and MD studies. Int J Biol Macromol 117:51–61. https://doi.org/10.1016/j.ijbiomac.2018.04.191
112. Bisht R, Jaiswal JK, Oliver VF, Eurtivong C, Reynisson J, Rupenthal ID (2017) Preparation and evaluation of PLGA nanoparticle-loaded biodegradable light-responsive injectable implants as a promising platform for intravitreal drug delivery. J Drug Deliv Sci Technol 40:142–156. https://doi.org/10.1016/j.jddst.2017.06.006
113. Shityakov S, Förster C (2013) Multidrug resistance protein P-gp interaction with nanoparticles (fullerenes and carbon nanotube) to assess their drug delivery potential: a theoretical molecular docking study. Int J Comput Biol Drug Des 6(4):343–357. https://doi.org/10.1504/IJCBDD.2013.056801
114. Gowri S, Gopinath K, Arumugam A (2018) Experimental and computational assessment of mycosynthesized CdO nanoparticles towards biomedical applications. J Photochem Photobiol B 180:166–174. https://doi.org/10.1016/j.jphotobiol.2018.02.009
115. Carreño-Fuentes L, Bahena D, Palomares LA, Ramírez OT, José-Yacamán M, Plascencia-Villa G (2016) Molecular docking and aberration-corrected STEM of palladium nanoparticles on viral templates. Metals 6(9):200. https://doi.org/10.3390/met6090200
116. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455–461. https://doi.org/10.1002/jcc.21334
117. Wasukan N, Kuno M, Maniratanachote R (2019) Molecular docking as a promising predictive model for silver nanoparticle-mediated inhibition of cytochrome P450 enzymes. J Chem Inf Model 59(12):5126–5134. https://doi.org/10.1021/acs.jcim.9b00572
118. Bulcke F, Thiel K, Dringen R (2014) Uptake and toxicity of copper oxide nanoparticles in cultured primary brain astrocytes. Nanotoxicology 8(7):775–785. https://doi.org/10.3109/17435390.2013.829591
119. Chibber S, Ahmad I (2016) Molecular docking, a tool to determine interaction of CuO and TiO2 nanoparticles with human serum albumin. Biochem Biophys Rep 6:63–67. https://doi.org/10.1016/j.bbrep.2016.03.004
120. Doudi M, Setorki M (2014) Acute effect of nano-copper on liver tissue and function in rat. J Nanosci Nanotechnol 1(5). https://doi.org/10.7508/nmj.2015.05.007
121. Pal A (2014) Copper toxicity induced hepatocerebral and neurodegenerative diseases: an urgent need for prognostic biomarkers. Neurotoxicology 40:97–101. https://doi.org/10.1016/j.neuro.2013.12.001
122. Pal A, Siotto M, Prasad R, Squitti R (2015) Towards a unified vision of copper involvement in Alzheimer's disease: a review connecting basic, experimental, and clinical research. J Alzheimers Dis 44(2):343–354. https://doi.org/10.3233/JAD-141194

Pharmaceutics Informatics: Bio/Chemoinformatics in Drug Delivery

24

Rania M. Hathout and Abdelkader A. Metwally

Contents

R. M. Hathout (✉)
Department of Pharmaceutics and Industrial Pharmacy, Faculty of Pharmacy, Ain Shams University, Cairo, Egypt
e-mail: rania.hathout@pharma.asu.edu.eg

A. A. Metwally
Department of Pharmaceutics and Industrial Pharmacy, Faculty of Pharmacy, Ain Shams University, Cairo, Egypt

Department of Pharmaceutics, Faculty of Pharmacy, Health Sciences Center, Kuwait University, Kuwait City, Kuwait
e-mail: Abdelkader_ali@pharma.asu.edu.eg

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_24

Abstract

Whilst bioinformatics deals with the biological receptors and proteins (biological macromolecules) and their manipulation, the chemoinformatics is usually concerned with the computational handling of small molecules and chemical compounds. The integration of the two approaches has been performed in several studies in order to study the interaction of drugs (therapeutically active compounds) with their macromolecular carriers or matrices on one hand and the biological targeted receptors and proteins on the other hand. Accordingly, the term "pharmaceutics informatics" can be introduced to compass the application of both the bioinformatics and chemoinformatics tools in drug delivery. This new approach would save the researchers and formulators the wet experiments for the preparation and characterization of drug-loaded delivery systems that consume a lot of resources, time and effort. The currently used techniques include data mining, computing physico-chemical descriptors, machine learning methods, molecular dynamics and docking experiments.

Keywords

Pharmaceutics · Informatics · Drug delivery · Descriptors · Machine learning · Computational

Chapter Objectives

- Understanding the concept and importance of pharmaceutics Informatics
- Integrating bioinformatics and chemoinformatics in drug delivery
- Recognizing the important tools to implement the pharmaceutics informatics
- Getting acquainted with the different softwares and online resources that are implemented in the pharmaceutics informatics approach
- Learning the different aspects of applying the pharmaceutics informatics in drug delivery

24.1 Introduction

The word "informatics" includes a lot of tools such as: softwares, data mining, artificial intelligence, data sciences, machine learning natural language processing, optimization and cloud computing [1, 2]. The use of any of these aforementioned tools solely or in combinations in the field of drug delivery and formulation introduces a new term which could be named "pharmaceutics informatics". This novel approach can revolutionize the drug formulation and delivery field leading to dramatic reduction in the resources costs and the time of development of new drug formulations.

Pharmaceutics informatics encompasses, in particular, two overlapping and pharmaceutically related areas: bioinformatics and chemoinformatics. This novel approach mainly focuses on: (1) data mining to obtain and/or calculate important

drugs data such as the isomeric SMILES and the constitutional, electronic and physical descriptors of the drug molecules, (2) softwares and webpages to perform molecular dynamics to simulate the different drug carriers whether lipidic, polymeric or proteinaceous, (3) molecular docking of drugs on their potential carriers and finally (4) machine learning methods such as the artificial neural networks, partial least squares and principal component analysis to correlate the drug descriptors with their loading and/or entrapment efficiencies and optimize the obtained results [3]. Pharmaceutics informatics is more general and encompasses other sub-related areas such as the nanoinformatics [4] and the biopharmaceutical informatics [5].

This chapter provides the readers with a deep insight on this new computational approach and demonstrates several research-based studies that are associated with drug delivery issues and applications.

24.2 Bio/Chemoinformatics Tools in Detecting the Drug Loading in Lipidic and Polymeric Nanoparticulate Matrices

The solid lipid and PLGA nanoparticles are very important hydrophobic drug delivery systems that are used through several routes of administration [6]. The concept of utilizing bio/chemoinformatics tools in drug formulations, dosage form design and drug delivery, and specifically for the determination of drug loading in nano-carrier matrices (Fig. 24.1), was introduced in 2015 [7] Where the study of literature gathered drugs loading on two different hydrophobic nanoparticulate matrices: namely the tripalmitin and the PLGA (poly lactic acid-*co*-glycolic acid) was performed. These tools included data mining in medical and chemical sources such as PubMed and PubChem, calculation of the constitutional, electronic, topological and physico-chemical descriptors of the loaded drugs such as the molecular weight, total polar surface area (TPSA), fragment complexity and log P using a bioinformatics software such as Bioclipse®, molecular dynamics modelling of the matrices using an open source tool (GROMACS®) and finally docking of the investigated molecules on the constructed lipophillic nano-carrier matrices using two open-sourced softwares: ArgusLab® and Autodock vina®.

Figure 24.1 demonstrates the scheme of the complete study that was introduced in [7].

The work in this study demonstrated the successful mathematical modelling of the masses of the loaded drugs per 100 mg of the matrix whether in the tripalmitin matrix or within the PLGA counterpart. Projecting the investigated drugs to their main descriptors was performed and the obtained calculated descriptors were successfully correlated with the generated binding energies (ΔG) resulting from docking the constructed minimized energy 3D chemical structures of the drug molecule on the molecular dynamics simulated carriers using a machine learning method, namely, the artificial neural network (ANN). Moreover, the prediction of the loading of an experimental drug, curcumin, was successfully accomplished scoring a percentage bias of only 12%.

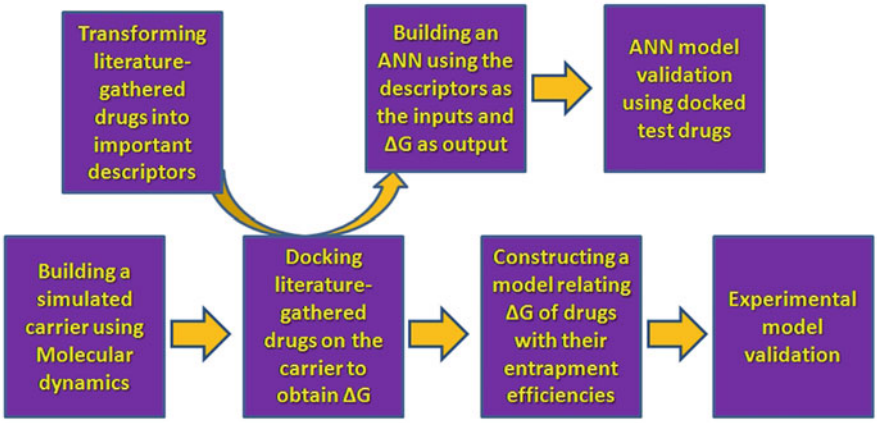


Fig. 24.1 Study scheme of Metwally and Hathout, 2015 [7]. Reprinted with permission from Metwally and Hathout, Mol Pharm 2015 Aug 3;12(8):2800–10, copyright 2020, ACS through RightsLink®

In 2016, Hathout and Metwally utilized a different commercial software for the docking experiments: MOE® of the same drug molecules and on the same matrices. Again the masses of the loaded drugs per 100 mg of the carrier were successfully mathematically modelled [8]. The authors also utilized another type of machine learning methods (Gaussian processes) that led to better prediction of the validating experimented drugs scoring this time a lower percentage bias (7%).

Similarly in 2020, Ritu et al. [9] conducted a research study to estimate the optimum accommodating nanoparticulate system to load filgrastim as a drug molecule in order to prepare a targeted delivery to the bone marrow aiming for neutropenia treatment.

The prepared targeted delivery of filgrastim also comprised tripalmitin solid lipid nanoparticulate system. The chemical structure of tripalmitin and the PLGA polymer was generated using Marvin Sketch package, then its Mol2 and pdb files were obtained using Schrodinger application Maestro® software. Likewise to the study performed by Metwally and Hathout [7], the simulation of the lipid and the polymer, separately but each with the drug; filgrastim, was achieved using GROMACS®. The parameters of the constructed carriers were obtained using SwissParam available online (http://www.swissparam.ch/). The energy minimization of the modelled systems was performed using the steepest descent method.

The modelling of the lipid and the polymer matrices together with utilizing the docking experiments was proven to predict successfully the mass loading of different drugs. The superiority of tripalmitin solid lipid nanoparticles in accommodating filgrastim to the PLGA nanoparticles was demonstrated (Fig. 24.2).

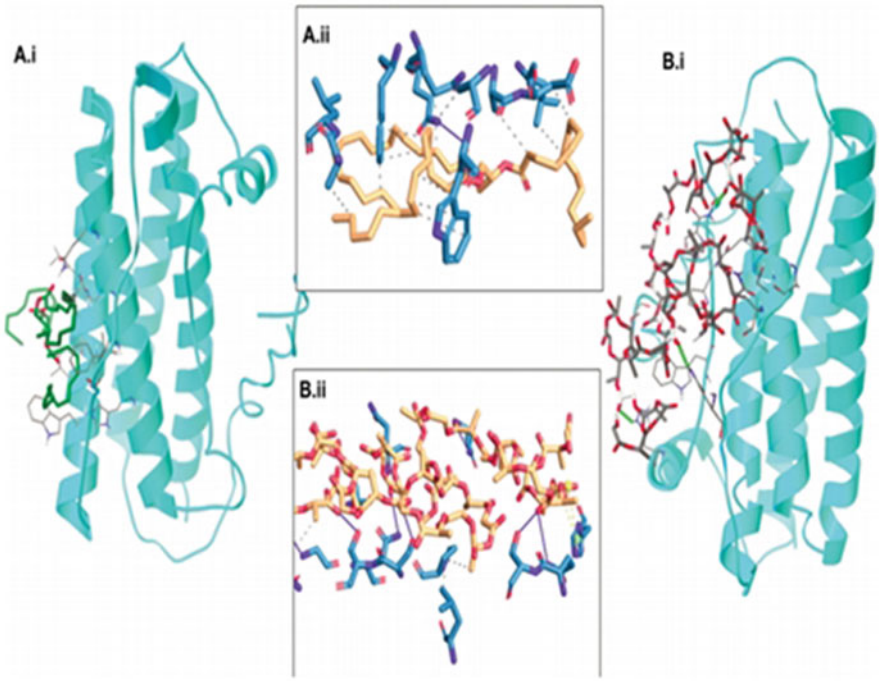


Fig. 24.2 Intermolecular interactions of filgrastim with the investigated carriers. (A.i) Docked conformation of filgrastim with the constructed tripalmitin matrix. (A.ii) The filgrastim–tripalmitin interactions. (B.i) The docked conformation of filgrastim and the PLGA polymer highlighting three H-bond formation and the interacting residues (dotted lines). (B.ii) The filgrastim–PLGA interactions. Reprinted with permission from Taylor and Francis® through RightsLink® licence number: 4936781264691

24.3 Bio/Chemoinformatics Tools in Selecting the Optimum Oil for Drugs Solubilization in Microemulsion Systems

Usually, selecting the best oily domain to accommodate a certain hydrophobic drug molecule is the rate-limiting step of any new microemulsion research study. This is usually attributed to the exhausting and time- and resources-consuming drug analysis studies especially in the viscous used oils. Accordingly, the bio/chemoinformatics approach was proven a help in this issue. In order to select the optimum oil to solubilize the hormonal drug testosterone, the investigated oils, that is, oleic acid, ethyl oleate, isopropyl myristate and mineral oil, were modelled using molecular dynamics utilizing GROMACS®. Water was also simulated as a control. The parameters of the oils were obtained using Cgenff [10] available online (https://cgenff.paramchem.org/).

Figure 24.3 demonstrates the docking energies of testosterone on the investigated oils and on water on the right *y*-axis. On the left *y*-axis, the drug solubility in the oils

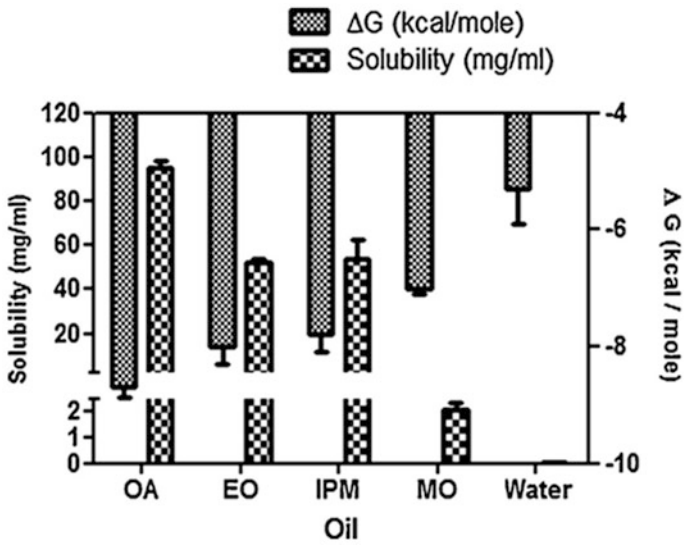


Fig. 24.3 Successful correlation of the experimental solubility results of testosterone in different oils and in water with the obtained corresponding binding energies. Reprinted with permission from Elsevier® through RightsLink® licence number: 4934200622977

in mg/mL was demonstrated. It is obvious from the data that the experimental ranking of the oils ability to solubilize the drug correlated well with the rank of the same oils with respect to the negative values of the binding energy (ΔG) of testosterone docked on their matrices. A more negative binding energy in Kcal/mole indicates better affinity to the drug and more accommodation and hence better solubilization.

24.4 Bio/Chemoinformatics in Selecting the Optimum Protein Carrier for Polyphenolic Drugs

Polyphenolic drugs are a mine of treasures in the treatment and therapy of many diseases [11]. Specifically, their highly potent discovered anti-cancer properties are being extensively investigated [12, 13]. However, the optimum exploitation of these valuable compounds was always hindered by their unfavourable physico-chemical properties, such as low solubility and stability, that in turn lead to low bioavailability and poor biological performance [14, 15]. This fact challenged the drug delivery scientists and formulators to search for the best advanced carrier systems that can manage these hindrances and target these molecules to the cancerous site of action with therapeutically sufficient amounts associated with the least possible side effects.

Example of the highly investigated polyphenolic drugs are curcumin and resveratrol [16, 17].

One of the safest, biocompatible and feasible drug delivery systems are the protein carriers such as gelatin, albumin, collagen, zein and casein [18, 19].

In a proof-of-concept study that was conducted in 2016 [20], gelatin and albumin proteinaceous matrices were compared in their ability to accommodate curcumin and resveratrol using the bio/chemoinformatics methods. In this work, the mass loading of the cargos on both carriers was literature-gathered except for the curcumin/gelatin combination that was experimentally prepared and characterized. Combined bioinformatics and chemoinformatics tools were harmonized in order to select the optimum carrier for the chemo-preventive agents as follows:

- The bovine serum albumin crystal structure was obtained from RSCB pdb (www.rscb.org) where it was coded 4F5S.
- The gelatin (as a de-natured protein) matrix was simulated using molecular dynamics using GROMACS®. The utilized parameters of gelatin were obtained using CgenFF available online (https://cgenff.paramchem.org/) [10]. The simulated matrix comprised 48 molecules. Each molecule was composed of an 18 amino-acid peptide, representing gelatin, and containing the building block amino-acid sequence of the protein: ALA, GLY, PRO, ARG, GLY, GLU, PRH, GLY, PRO, ALA, GLY, PRO, ASP, GLY, GLU, PRH, GLY and PRO. The system was subjected to energy minimization using the steepest descent method. The system was then subjected to a molecular dynamics simulation run, with a time step of 2 fs, full periodic boundary conditions and a cut-off distance for Van der Waal's and electrostatic interactions of 1.2 nm. PME was chosen to treat long-range electrostatic interactions. The LINCS algorithm was used to constrain all bonds. The system was equilibrated at 373 K using a v-rescale thermostat, and at a pressure of 1 bar using a Berendsen barostat for 3 ns.
- Obtaining the isomeric SMILES (Simplified Molecular Input Line System) of the investigated polyphenols from Pubchem® database and using them in the chemical structures builder tool of the commercial software package (MOE®) where the 3D structures were obtained and then subjected to energy minimization using MMFF94x forcefield of the same software.
- Docking of the investigated drugs on the virtually simulated carriers using MOE® using the "triangular matcher" as the placement method and utilizing the London score as the docking algorithm.

Figure 24.4 summarizes the pharmaceutics informatics (integrated bio/chemoinformatics) tools that were used in the study.

The results in this study presented a strong correlation between the masses of loaded drugs that resulted experimentally (Table 24.1) and their corresponding obtained binding energies that resulted from the molecular dynamics and docking experiments (Table 24.2).

The work in this study provided an evidence ultimate goal that informatics tools can provide a means to replace the highly tedious, time-consuming and resources-wasting wet-lab experiments.

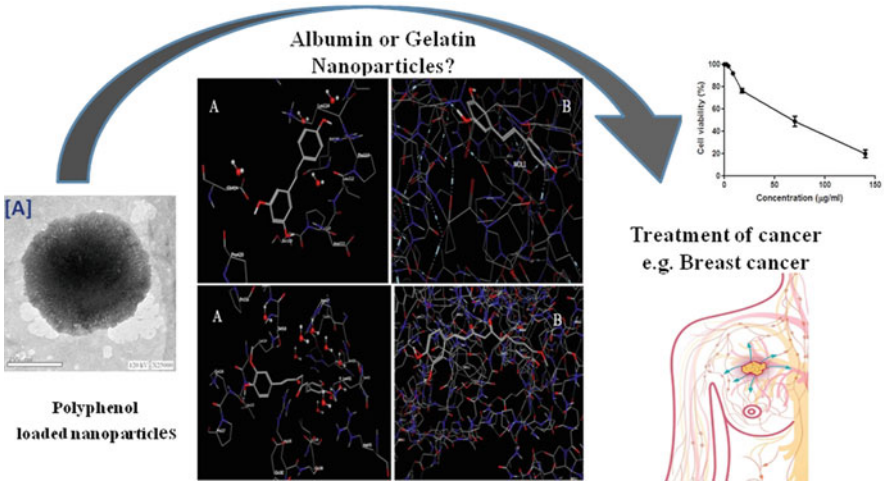


Fig. 24.4 Summary of comparing albumin to gelatin nanoparticulate matrices for loading polyphenols using pharmaceutics informatics (combined bio/chemoinformatics) tools. Reprinted with permission from Elsevier® through RightsLink® licence number: 4934210398372

Table 24.1 Mass of the loaded investigated polyphenols per 100 mg of the protein nano-carrier. Taken with permission from Elsevier® through RightsLink® licence number: Taken with permission from Elsevier® through RightsLink® licence number: 4934210398372

	Mass loaded per 100 mg protein nano-carrier (mg)	
Drug	Albumin	Gelatin
Resveratrol	39.4	1.96
Curcumin	10	3.5

Table 24.2 The obtained binding energies after docking the investigated polyphenols on the virtual protein nano-carriers. Taken with permission from Elsevier® through RightsLink® licence number: 4934210398372

	Binding energies (Kcal/mole)			
	Albumin		Gelatin	
Drug	London ΔG score	ASE score	London ΔG score	ASE score
Resveratrol	-9.5 ± 0.5	-14.2 ± 0.1	-8.3 ± 0.2	-9.21 ± 0.1
Curcumin	-13.2 ± 2.1	-21.28 ± 0.2	-8.11 ± 0.1	-19.32 ± 0.01

24.5 Bio/Chemoinformatics in Comparing the Biopharmaceutical Behaviour of Curcuminoids in Alzheimer's Disease

The newly addressed pharmaceutics informatics approach integrating the bio- and the chemoinformatics was also proven to help in virtually simulating a complete pharmaceutics study at three sequential levels: the formulation level, the biopharmaceutical level and finally the therapeutic level. The investigated disease was the Alzheimer's and the drugs to be compared were of the curcuminoids class, namely, curcumin and bisdemethoxycurcumin, and the investigated route was the nose-to-brain delivery [21].

At the formulation level, the loading of the two drugs on a polymeric carrier was evaluated by docking on a simulated PLGA matrix composed of 32 molecules of polylactic acid and 32 molecules of polyglycolic acid; each of them is composed of 35 monomers and is end capped with a methyl group.

At the biopharmaceutical level, the interaction of the two molecules with mucin that is present in the nose and with the P-gp efflux pumps that are responsible for the expulsion of drugs out of the brain cells were investigated.

At the therapeutic level, the interaction of the two curcuminoids with the amyloid plaques and which are responsible for blocking the neuronal transmission in the brain were also studied.

Figure 24.5 demonstrates the scheme of the bio/chemo steps that were conducted in this study.

The results demonstrated the superiority of curcumin to bisdemethoxycurcumin at three levels as shown in Table 24.3.

The differences in the results between the two investigated drugs were attributed to their differences in their main constitutional, electronic and physico-chemical descriptors that are displayed in Table 24.4. A number of 5 more hypothesized molecules were introduced and their similar descriptors were calculated.

Since the number of descriptors (factors or variables) was 6 (i.e. more than two dimensions), then an informatics tool (multi-variate) statistics method, viz. principal component analysis (PCA) [22, 23], was utilized to analyse the results and cluster the investigated molecules according to their descriptors into groups (Fig. 24.6).

Interestingly, the docking behaviour of the clustered hypothesized molecules with curcumin was very close to their parent compound on all of the investigated macromolecules.

It was concluded from this work that the use of several pharmaceutics (bio/chemo) informatics tools have confirmed the superiority of curcumin PLGA nanoparticles to bisdemethoxycurcumin for nose-to-brain formulation targeting Alzheimer disease. Moreover, a newly proposed synthetic analogue diethoxybisdemethoxycurcumin was proposed with a potential for excellent delivery and augmented biological performance. The current study could shed the light on the usage of the new approach in the search of drug derivatives or biosimilars aiming for better alternatives considering all aspects of drug delivery research: formulation, biopharmaceutical or therapeutic. The study would also offer a new methodology in

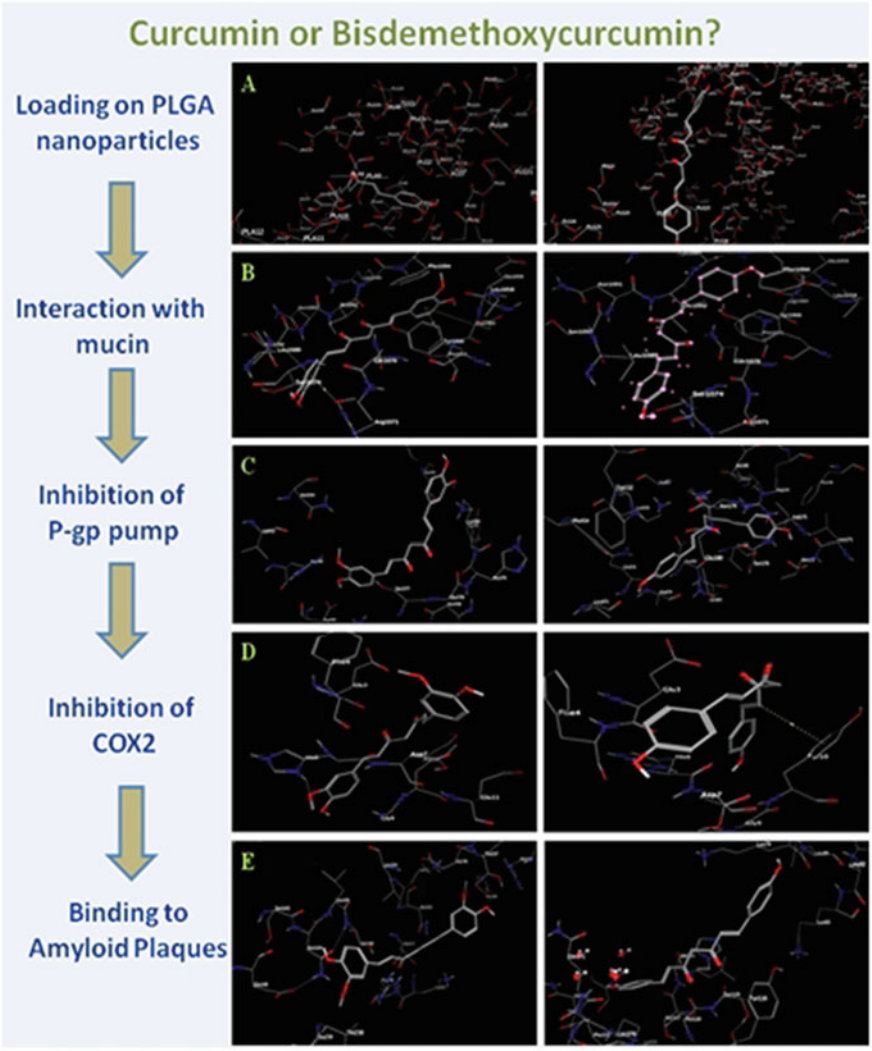


Fig. 24.5 The study scheme in (Hathout et al., 2017) demonstrating the interaction of curcumin (left panel) and bisdemethoxycurcumin (right panel) with the different investigated macromolecules. Modified with permission from Taylor and Francis® through RightsLink® licence number: 4938430787711

medicines design which can save formulators, phyto-chemists (especially those working with extracts producing large number of similar chemical compounds) and pharmacists' huge time, efforts and resources spent on wet-lab trials. Moreover, it can serve as a validating or a confirmatory tool for the adopted biological experiments.

Table 24.3 The obtained binding energies after docking the investigated curcuminoids on the virtual proposed nano-carrier and the nose-to-brain-related macromolecules

Macromolecule (carrier/peptide/protein)/PDB ID code (if available)	Binding energy (kcal/mole)	
	Curcumin	Bisdemethoxycurcumin
PLGA polymer (nanoparticles matrices)	-13.05 ± 0.1	-11.56 ± 0.2
Mucin PDB ID: 2ACM	-12.03 ± 0.2	-12.76 ± 0.3
P-gp efflux pump PDB ID: 3G6I	-8.92 ± 0.05	-8.71 ± 0.06
Amyloid peptide PDB ID: 1IYT	-8.20 ± 0.1	-8.23 ± 0.05
Cyclooxygenase 2 PDB ID: 5IKQ	-16.41 ± 0.1	-11.38 ± 0.03

Table 24.4 The calculated descriptors of the natural and proposed curcumoinids

Molecule	Main descriptors					
	Total hydrophobic surface area	Number of H-bond acceptors	Number of H-bond donors	Molecular flexibility	Log P (O/W)	Molecular weight
Curcumin	665.79	6.00	2.00	5.30	3.72	368.38
BDMC	554.72	4.00	2.00	3.86	3.74	308.33
Diamino-BDMC	585.82	4.00	4.00	3.91	2.46	338.36
Dicarboxy-BDMC	633.97	8.00	6.00	4.98	3.16	396.35
Diethoxy-BDMC	703.59	6.00	2.00	6.37	4.40	396.44
Dihydroxy-BDMC	571.88	6.00	4.00	4.31	3.19	340.33
Dimethyl-BDMC	609.84	4.00	2.00	4.31	4.40	336.39

24.6 Bio/Chemoinformatics Tools in Selecting the Optimum Natural Bio-macromolecular Carrier of Doxorubicin

The calculation of encapsulation or entrapment of therapeutic molecules in nanoparticulate matrices are of the most important characterization steps during nanoparticles formulation experiments. The loaded amount of any drug in the nanoparticles determines the amount of drug that would reach the site of action. However, wet experiments that are usually conducted to select the optimum carrier amongst a lot of available ones could be highly time, effort and resources consuming. Accordingly, the possibility of using molecular dynamics and docking experiments for the selection a suitable natural bio-macromolecule for loading a chemotherapeutic agent, namely doxorubicin, into nanoparticles was investigated.

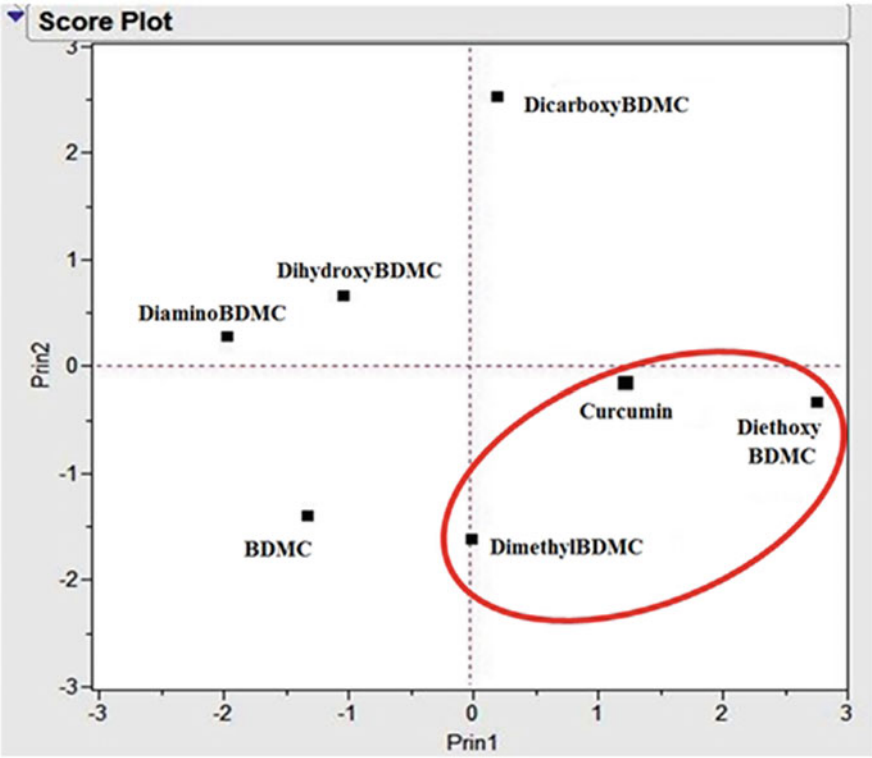


Fig. 24.6 Applying principal component analysis (PCA) to select the closest hypothesized molecules to curcumin. Reprinted with permission from Taylor and Francis® through RightsLink® licence number: 4938430787711

The results of docking the chemotherapeutic agent doxorubicin on three natural bio-macromolecules, namely albumin, hyaluronic acid and surfactin, were compared to the loading data obtained from literature [24].

Figure 24.7 represents a schematic diagram of the three investigated natural bio-macromolecules that accommodated doxorubicin in the literature together with the corresponding constructed docking poses of the drug in the matrices as obtained from MOE® commercial software package.

The results revealed that doxorubicin-loading rank that was gathered from literature correlated perfectly well with the resultant docking binding energies (Fig. 24.8). In conclusion, docking experiment provided an excellent tool for selecting an optimum natural biocompatible macromolecular carrier for loading the doxorubicin. The used informatics tools can be projected to other natural macromolecular carriers in the future to select the best drug-carrier pair [24].

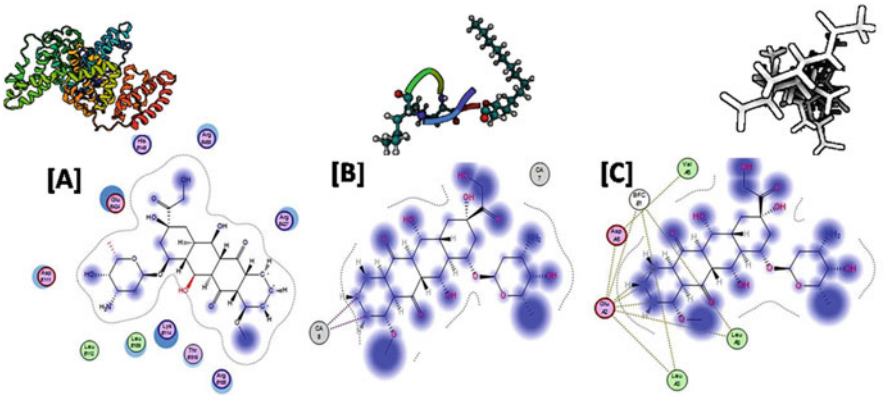


Fig. 24.7 Schematic diagrams of the investigated bio-macromolecules are displayed at the upper panel and the docking optimum of doxorubicin on the investigated natural bio-macromolecular carriers: (**a**) albumin, (**b**) hyaluronic acid and (**c**) surfactin. Reprinted with permission from Springer Nature® through RightsLink® licence number: 4938481172784

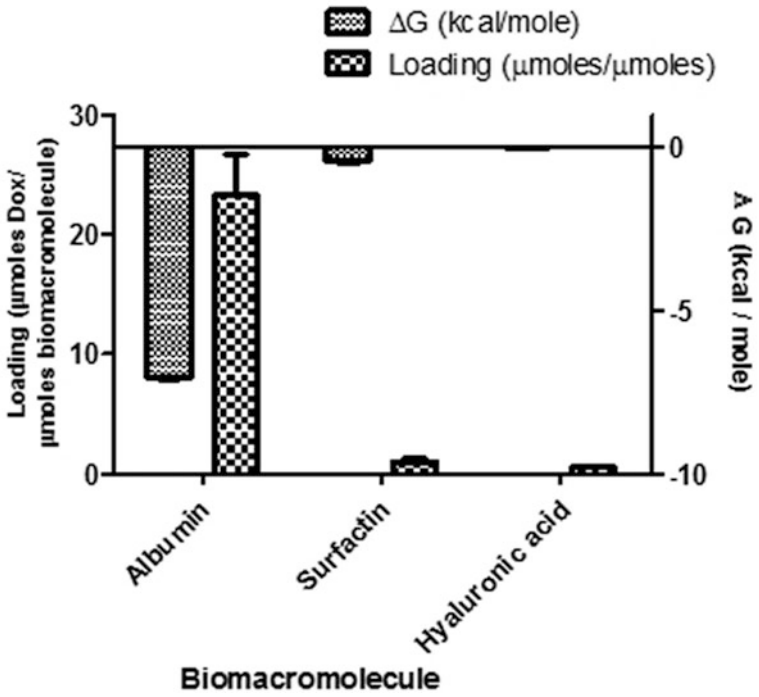


Fig. 24.8 Synchronized results of loaded masses of doxorubicin with the corresponding binding energies (ΔG in Kcal/mole) on the investigated natural bio-macromolecules in [24]. Reprinted with permission from Springer Nature® through RightsLink® licence number: 4938481172784

24.7 Bio/Chemoinformatics for Re-purposing and Re-formulation of an Old Molecule in Order to Combat COVID-19

In 2019, the world was stricken by the COVID-19 pandemic caused by the SARS-COV2 virus. During this crisis, pharmacists and biologists endeavours have not stopped, exerting enormous efforts and facing great challenges in the battle to fight this alarming disease [25, 26]. Racing time to save the lives of as much people as possible, re-purposing old molecules to combat the disease was highly recommended. In this context, chloroquine and its derivative hydroxychloroquine were at the front lines [27–29]. However, a big debate started about the efficacy of these molecules and some inconsistent results were obtained [30].

Investigating the two molecules, it was found that they suffer from very high volume of distribution in the biological bodies (reaching a level of 100 kg/L) indicating the drugs distribution even to the fatty tissues [31]. This high volume of distribution leads to decreasing the therapeutic efficacy at the disease sites of action on one hand and to the associated side effects on the other hand [32].

Accordingly, in a work conducted by Hathout et al. [33], the intranasal and the pulmonary routes were proposed as an alternative route to the oral delivery. Gelatin micro- and nanoparticulate matrices were suggested as carriers for the studied drugs in order to concentrate the drugs at the site of action, sustain the drugs release and exploit the mucoadhesive properties of gelatin [34] in order to increase the contact time with the nasal mucosa [33]. Molecular docking on the gelatin-simulated matrix demonstrated high binding potential leading to high loading and sustained release profile.

Moreover, the drug's efficacy was studied through different reported mechanisms. Moreover, good binding to mucin as well as various important disease-related receptors comprising angiotensin-converting enzyme 2 (ACE-2), heparan sulphate proteoglycan and phosphatidylinositol binding clathrin assembly protein (PICALM), which are expressed in the lung and intranasal tissues and represent crucial sites of attachment of the viral particles to the surface of respiratory cells, were confirmed.

The idea of the work is summarized in Fig. 24.9.

The gelatin matrix was simulated using GROMACS® open source software. The crystal structure of the delivery matrices and receptor targets were obtained from the RSCB protein data bank (http://www.rcsb.org), where mucin, ACE-2 and PICALM were coded 2ACM, 6m17 and 3zyk, respectively. The IUPAC structure of HSPG were obtained from PubChem®, constructed using the ChemDraw® Ultra package version 10 composed of two molecules (resembling the HSPG present in cell membranes), energy-minimized using MMFF94x forcefield found in MOE® version 2014.0901 (Chemical Computing Group Inc., Montreal, Canada) and finally used for docking.

The results of the molecular docking experiments are displayed in Table 24.5.

According to this study, the efficacy of chloroquine and its related derivative hydroxychloroquine in the prophylaxis and treatment of COVID-19 is confirmed.

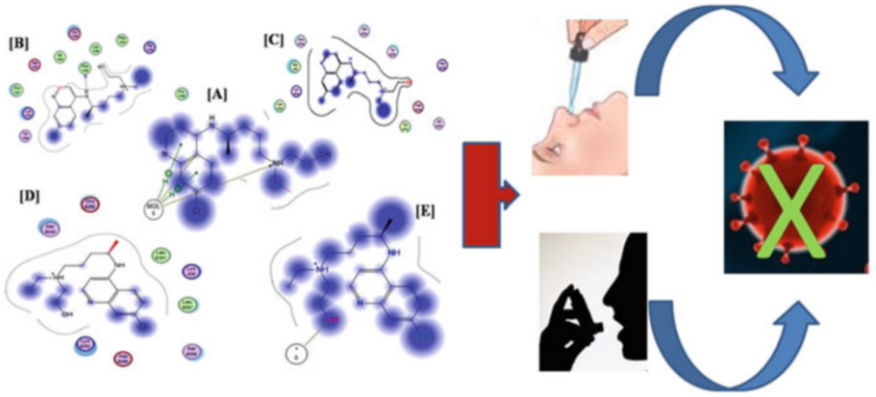


Fig. 24.9 Interaction of hydroxychloroquine with (**a**) gelatin, (**b**) mucin, (**c**) PICALM, (**d**) ACE-2 and (**e**) heparan sulphate. Loading on gelatin matrices for the intranasal or the pulmonary routes is postulated to increase effectiveness. Reprinted with permission from Elsevier ® through RightsLink® licence number: 4934201081620

Table 24.5 Docking results of chloroquine and hydroxychloroquine on the investigated macromolecules in [33]. Taken with permission from Elsevier ® through RightsLink® licence number: 4934201081620

Macromolecule (carrier/protein/proteoglycan)—PDB code	Binding energy (kcal/mole)	
	Chloroquine	Hydroxychloroquine
Gelatin matrix (gelatin nanospheres)	−8.72 ± 0.1	−10.09 ± 0.01
Mucin—2ACM	−9.22 ± 0.1	−10.10 ± 0.1
ACE-2—6m17[a]	−8.71 ± 0.2	−8.75 ± 0.2
PICALM—3zyk[b]	−8.29 ± 0.1	−10.49 ± 0.2
Heparan sulphate proteoglycan	−8.83 ± 0.1	−11.65 ± 0.1

[a]*ACE-2* angiotensin-converting enzyme-2
[b]*PICALM* phosphatidylinositol binding clathrin assembly protein

Drug targeting using gelatin particulate systems through the intranasal and pulmonary routes was recommended to augment and optimize the effect.

The software that could be used in the context of pharmaceutics informatics can be classified into: (1) software that is used to simulate carrier systems by MDS, such as GROMACS and NAMD which are available as free open source packages, (2) software packages used to calculate molecular descriptors (Chemoinformatics software), e.g. MOE and Bioclipse (which includes a CDK plug-in), (3) software that are used to build machine learning models such as JMP and the R statistical language and (4) software that is used in molecular docking studies such as ArgusLab, Autodock Vina and MOE. A list of the important examples of such packages is presented in Table 24.6. A list of examples of credible and useful online sources are provided at the end of this chapter.

Table 24.6 List of important softwares that are used in pharmaceutics informatics

Software name	Latest version	Open-source/ proprietary	Vendor/provider	Application(s)	References
MOE	2019.0102	Proprietary	Chemical Computing Group Inc., Montreal, Canada	Descriptors calculations/energy minimization of drugs and carriers/ molecular docking	[8, 35]
Autodock Vina	1.1.2. (May 11, 2011)	Open-source/ Apache License	The Scripps Research Institute, La Jolla, California, USA	Molecular docking	[7, 36]
GROMACS	2021-rc1	Open-source/LGPL free-software license	University of Groningen Royal Institute of Technology, Groningen, The Netherlands and Uppsala University, Uppsala, Sweden	Molecular dynamics simulations	[8, 20, 35, 36]
NAMD	2.14	Propitiatory freeware for non-commercial use	Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, IL, USA	Molecular dynamics simulations	[38]
R	4.04	Open-source/free	R Foundation for statistical computing	R is a software environment for statistical computing and graphics	[39]
JMP	13.0 (2016)	Proprietary	SAS Campus Drive, Cary, North Carolina, USA	Machine learning methods such as principal component analysis and artificial neural networks	[7, 21, 22]
Bioclipse	2.6.2. (2016)	Open-source/Free	SOURCEFORGE, San Diego, California, USA	Descriptors calculations	[7, 40]
ChemDraw	20.0. (November 3, 2020)	Proprietary	PerkinElmer, Waltham, Massachusetts, USA	Drawing chemical structures in 3D/energy minimization	[7, 8, 21, 36]
ArgusLab	4.0.1. (2018)	Freeware	Planaria Software LLC, Seattle, Washington, USA	Molecular docking	[7, 41]

24.8 Conclusion

The integration of bioinformatics and chemoinformatics in drug delivery resulted in a novel approach, pharmaceutics informatics. This new concept was introduced and a detailed overview of its applications in the drug delivery field was provided. The implementation of this new platform could lead to huge saving in resources, time and efforts exerted in the pharmaceutical industry field.

24.9 Credible Online Resources for Further Reading

URL	Category of source	Importance
https://www.rcsb.org/	Worldwide Protein Data Bank Foundation	Source of all protein structures and receptors
https://pubchem.ncbi.nlm.nih.gov/	National Center for Biotechnology Information	Source of chemical names, structures and canonical and isomeric SMILES of all chemical compounds
http://www.swissparam.ch/	Swiss Institute of Bioinformatics	Provides topology and parameters for molecular dynamics
https://cgenff.paramchem.org/	Personal web page maintained by Alex Mackerell and Kenno Vanommeslaegthe	Provides topology and parameters for molecular dynamics
https://www.molinspiration.com/cgi-bin/properties	Molinspiration company focuses on the development and application of cheminformatics techniques	Calculation of molecular descriptors and prediction of bioactivity
https://chemminetools.ucr.edu/	Online service for analysing and clustering of small molecules	Many tools including the calculation of molecular descriptors

References

1. Huang S, Zhou J, Wang Z, Ling Q, Shen Y (2017) Biomedical informatics with optimization and machine learning. EURASIP J Bioinforma Syst Biol 2017:4. https://doi.org/10.1186/s13637-017-0058-0
2. Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B et al (2017) Deep learning for health informatics. IEEE J Biomed Health Inform 21:4–21. https://doi.org/10.1109/JBHI.2016.2636665
3. Duarte Y, Maírquez-Miranda V, Miossec MJ, Gonzaílez-Nilo F (2019) Integration of target discovery, drug discovery and drug delivery: a review on computational strategies. WIREs Nanomed Nanobiotechnol 11:e1554. https://doi.org/10.1002/wnan.1554
4. Sason H, Shamay Y (2020) Nanoinformatics in drug delivery. Isr J Chem 60:1108–1117. https://doi.org/10.1002/ijch.201900042
5. Kumar S, Plotnikov NV, Rouse JC, Singh SK (2018) Biopharmaceutical informatics: supporting biologic drug development via molecular modelling and informatics. J Pharm Pharmacol 70:595–608. https://doi.org/10.1111/jphp.12700

6. Mishra V, Bansal KK, Verma A, Yadav N, Thakur S, Sudhakar K et al (2018) Solid lipid nanoparticles: emerging colloidal nano drug delivery systems. Pharmaceutics 10:191. https://doi.org/10.3390/pharmaceutics10040191
7. Metwally AA, Hathout RM (2015) Computer-assisted drug formulation design: novel approach in drug delivery. Mol Pharm 12:2800–2810. https://doi.org/10.1021/mp500740d
8. Hathout RM, Metwally AA (2016) Towards better modelling of drug-loading in solid lipid nanoparticles: molecular dynamics, docking experiments and Gaussian Processes machine learning. Eur J Pharm Biopharm 108:262–268. https://doi.org/10.1016/j.ejpb.2016.07.019
9. Karwasra R, Fatihi S, Raza K, Singh S, Khanna K, Sharma S et al (2020) Filgrastim loading in PLGA and SLN nanoparticulate system: a bioinformatics approach. Drug Dev Ind Pharm 46:1354–1361. https://doi.org/10.1080/03639045.2020.1788071
10. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J et al (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 31:671–690. https://doi.org/10.1002/jcc.21367
11. Abdel-Hafez SM, Hathout RM, Sammour OA (2018) Curcumin-loaded ultradeformable nanovesicles as a potential delivery system for breast cancer therapy. Colloids Surf B Biointerfaces 167:63–72. https://doi.org/10.1016/j.colsurfb.2018.03.051
12. Abdelhamid HN, El-Bery HM, Metwally AA, Elshazly M, Hathout RM (2019) Synthesis of CdS-modified chitosan quantum dots for the drug delivery of Sesamol. Carbohydr Polym 214:90–99. https://doi.org/10.1016/j.carbpol.2019.03.024
13. Ossama M, Hathout RM, Attia DA, Mortada ND (2019) Enhanced allicin cytotoxicity on HEPG-2 cells using glycyrrhetinic acid surface-decorated gelatin nanoparticles. ACS Omega 4:11293–11300. https://doi.org/10.1021/acsomega.9b01580
14. Yehia R, Hathout RM, Attia DA, Elmazar MM, Mortada ND (2017) Anti-tumor efficacy of an integrated methyl dihydrojasmonate transdermal microemulsion system targeting breast cancer cells: in vitro and in vivo studies. Colloids Surf B: Biointerfaces 155:512–521. https://doi.org/10.1016/j.colsurfb.2017.04.031
15. Mehanny M, Hathout RM, Geneidi AS, Mansour S (2016) Exploring the use of nanocarrier systems to deliver the magical molecule; curcumin and its derivatives. J Control Release 225:1–30. https://doi.org/10.1016/j.jconrel.2016.01.018
16. Abdel-Hafez SM, Hathout RM, Sammour OA (2020) Attempts to enhance the anti-cancer activity of curcumin as a magical oncological agent using transdermal delivery. Adv Tradit Med. https://doi.org/10.1007/s13596-020-00439-5
17. Salehi B, Mishra AP, Nigam M, Sener B, Kilic M, Sharifi-Rad M et al (2018) Resveratrol: a double-edged sword in health benefits. Biomedicine 6:91. https://doi.org/10.3390/biomedicines6030091
18. Hathout RM, Metwally AA (2000) Gelatin Nanoparticles. Methods Mol Biol 2019:71–78. https://doi.org/10.1007/978-1-4939-9516-5_6
19. Hathout RM, Omran MK (2016) Gelatin-based particulate systems in ocular drug delivery. Pharm Dev Technol 21:379–386. https://doi.org/10.3109/10837450.2014.999786
20. Metwally AA, El-Ahmady SH, Hathout RM (2016) Selecting optimum protein nano-carriers for natural polyphenols using chemoinformatics tools. Phytomedicine 23:1764–1770. https://doi.org/10.1016/j.phymed.2016.10.020
21. Hathout RM, El-Ahmady SH, Metwally AA (2018) Curcumin or bisdemethoxycurcumin for nose-to-brain treatment of Alzheimer disease? A bio/chemo-informatics case study. Nat Prod Res 32:2873–2881. https://doi.org/10.1080/14786419.2017.1385017
22. Hathout RM, Metwally AA, Woodman TJ, Hardy JG (2020) Prediction of drug loading in the gelatin matrix using computational methods. ACS Omega 5:1549–1556. https://doi.org/10.1021/acsomega.9b03487
23. Hathout RM (2014) Using principal component analysis in studying the transdermal delivery of a lipophilic drug from soft nano-colloidal carriers to develop a quantitative composition effect

permeability relationship. Pharm Dev Technol 19:598–604. https://doi.org/10.3109/10837450.2013.813544
24. Gad HA, Hathout RM (2020) Can the docking experiments select the optimum natural bio-macromolecule for doxorubicin delivery? J Clust Sci. https://doi.org/10.1007/s10876-020-01910-8
25. Hathout RM, Kassem DH (2020) Positively charged electroceutical spun chitosan nanofibers can protect health care providers from COVID-19 infection: an opinion. Front Bioeng Biotechnol 8:885. https://doi.org/10.3389/fbioe.2020.00885
26. Hathout RM, Kassem DH (2021) Uniting electroceutical and cosmeceutical interventions in combating coronavirus using e-poly-l-lysine. Sci Pharm 89:2. https://doi.org/10.3389/fbioe.2020.00885
27. Gao J, Tian Z, Yang X (2020) Breakthrough: chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. Biosci Trends 14:72–73. https://doi.org/10.5582/bst.2020.01047
28. Gautret P, Lagier JC, Parola P, Hoang VT, Meddeb L, Mailhe M et al (2020) Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int J Antimicrob Agents 105949. https://doi.org/10.1016/j.ijantimicag.2020.105949
29. Goyal P, Choi JJ, Pinheiro LC, Schenck EJ, Chen R, Jabri A et al (2020) Clinical characteristics of Covid-19 in New York City. N Engl J Med 382:2372–2374. https://doi.org/10.1056/NEJMc2010419
30. Geleris J, Sun Y, Platt J, Zucker J, Baldwin M, Hripcsak G et al (2020) Observational study of hydroxychloroquine in hospitalized patients with Covid-19. N Engl J Med 382:2411–2418. https://doi.org/10.1056/NEJMoa2012410
31. Haladyj E, Sikora M, Felis-Giemza A, Olesinska M (2018) Antimalarials—are they effective and safe in rheumatic diseases? Reumatologia 56:164–173. https://doi.org/10.5114/reum.2018.76904
32. Haeusler IL, Chan XHS, Guerin PJ, White NJ (2018) The arrhythmogenic cardiotoxicity of the quinoline and structurally related antimalarial drugs: a systematic review. BMC Med 16:200. https://doi.org/10.1186/s12916-018-1188-2
33. Hathout RM, Abdelhamid SG, Metwally AA (2020) Chloroquine and hydroxychloroquine for combating COVID-19: investigating efficacy and hypothesizing new formulations using bio/chemoinformatics tools. Inform Med Unlocked 21:100446. https://doi.org/10.1016/j.imu.2020.100446
34. Shokry M, Hathout RM, Mansour S (2018) Exploring gelatin nanoparticles as novel nanocarriers for Timolol Maleate: augmented in-vivo efficacy and safe histological profile. Int J Pharm 545:229–239. https://doi.org/10.1016/j.ijpharm.2018.04.059
35. Hathout RM, Abdelhamid SG, El-Housseiny GS, Metwally AA (2020) Comparing cefotaxime and ceftriaxone in combating meningitis through nose-to-brain delivery using bio/chemoinformatics tools. Sci Rep 10:21250. https://doi.org/10.1038/s41598-020-78327-w
36. Metwally AA, Hathout RM (2015) Replacing microemulsion formulations experimental solubility studies with in-silico methods comprising molecular dynamics and docking experiments. Chem Eng Res Des 104:453–456. https://doi.org/10.1016/j.cherd.2015.09.003
37. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455–461. https://doi.org/10.1002/jcc.21334
38. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JËV, Bernardi RC et al (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. J Chem Phys 153:044130. https://doi.org/10.1063/5.0014475
39. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical. Computing, Vienna, Austria, 1-1-2021

40. Spjuth O, Helmus T, Willighagen EL, Kuhn S, Eklund M, Wagener J et al (2007) Bioclipse: an open source workbench for chemo- and bioinformatics. BMC Bioinform 8:59. https://doi.org/10.1186/1471-2105-8-59
41. Bitencourt-Ferreira G, de Azevedo WF (2019) Molecular docking simulations with ArgusLab. In: de Azevedo WF Jr (ed) Docking screens for drug discovery. Springer, New York, pp 203–220. https://doi.org/10.1007/978-1-4939-9752-7_13
42. Shah S, Firlak M, Berrow S, Halcovitch N, Baldock S, Yousafzai B, Hathout R, Hardy J (2018) Electrochemically enhanced drug delivery using polypyrrole films. Materials 11(7):1123. https://doi.org/10.3390/ma11071123
43. Abd-algaleel SA, Metwally AA, Abdel-Bar HM, Kassem DH, Hathout RM (2021) Mol Pharm 18(10):3763–3776. https://doi.org/10.1021/acs.molpharmaceut.1c00276
44. Abd-algaleel SA, Metwally AA, Abdel-Bar HM, Kassem DH, Hathout RM (2021) Mol Pharm 18(10):3763–3776. https://doi.org/10.1021/acs.molpharmaceut.1c00276

Computer-Aided Development and Testing of Human Extra-Thoracic Airway Models for Inhalation Drug Delivery

25

Xiuhua April Si and Jinxiang Xi

Contents

X. A. Si
Department of Aerospace, Industrial, and Mechanical Engineering, California Baptist University, Riverside, CA, USA
e-mail: asi@calbaptist.edu

J. Xi (✉)
Department of Biomedical Engineering, University of Massachusetts, Lowell, Lowell, MA, USA
e-mail: Jinxiang_Xi@uml.edu

V. A. Saharan (ed.), *Computer Aided Pharmaceutics and Drug Delivery*,
https://doi.org/10.1007/978-981-16-5180-9_25

Abstract

Knowing the deposition rate and distribution of administered pharmaceuticals in the human respiratory tract is crucial to establish dose-response relationships and optimize therapeutic outcomes. Recent advances in imaging technology have made it practical to generate respiratory tract models with great details and anatomical accuracy. Pharmaceutics and inhalation drug delivery can benefit from such image-based computer-generated airway models by making available detailed information on airflow, particle motion, and wall interactions. In this chapter, we will introduce image-based computational methods to better understand respiratory aerosol dynamics. Specifically, the mouth-throat, mouth-lung, and nasal airway will be discussed. In the mouth-throat (MT) models, we will illustrate the importance of realistic airway morphologies on the accuracy of predictive inhalation dosimetry. In the mouth-lung models, experimental studies using 3D-printed hollow casts will be presented to cross-validate the predictive dosimetry models. Numerical techniques of hyper-morphing, probability analysis (including input sensitivity analysis and output uncertainty quantification), and statistical shape modeling will be introduced along the way. In the nasal model, experimental studies will be presented with different sprays, nebulizers, and delivery techniques (normal and bidirectional). New techniques that use ferromagnetic particles and pulsating flows for targeted delivery to the olfactory region and maxillary sinus will also be discussed.

Keywords

Inhalation dosimetry · CFD · Inhalation drug delivery · Respiratory dynamics

Chapter Objectives

- Demonstrate how the respiratory airway models (mouth-throat, nose, and lung) were developed.
- Explain the physics underlying pharmaceutical delivery and the numerical methods to simulate respiratory aerosol dynamics.
- Introduce the computational software and solvers used for pharmaceutical research and development, such as ANSYS Fluent and COMSOL.
- Demonstrate the computer-aided testing of the morphological effects of the upper airway on particle deposition in the mouth-throat region.
- Optimize pulmonary drug delivery using both computational modeling and in vitro experiments.
- Develop and test the nose-to-brain (N2B) drug delivery using both modeling and experimental testing.

25.1 Introduction

25.1.1 Significance of Airway Models in Pharmaceutical Development

The delivery efficiency of inhalation delivery of pharmaceutical aerosols is generally low (5–30%) and can be complicated by many factors. The human pharynx is an approximately 90° bend changing from a horizontal oral cavity to a vertical trachea. Large particles cannot maneuver through this 90° bend well very well due to their large inertia, and a large portion of them will impact onto the back wall of the pharynx and are wasted [1]. Likewise, 90% of nasal sprays will be lost in the anterior nose due to the constriction of the nasal valve as well as the large inertia of nasal spray droplets [2]. Besides the respiratory anatomical effects, inhalers and patient compliance can also affect the delivery efficiencies. In this chapter, we will focus on the effects of respiratory anatomy on inhalation drug delivery, and specifically on the effects of the flow-limiting regions such as the nasal valve, pharyngeal curvature, and glottis. In doing so, respiratory models of the nasal airway, mouth-throat model, and mouth-lung model were developed based on medical images (CT or MRI) and in vitro casts, which was explained in detail in Sect. 25.1.2.

An important factor in inhalation drug delivery is knowing where administered pharmaceuticals deposit and whether sufficient doses reach the target tissue. Both the total and locally delivered doses can be crucial to the therapeutic outcomes of the inhalation drug deliveries. Thus, it is critical to characterize aerosol deposition distributions in various respiratory regions. However, the airways are inaccessible to standard quantification and visualization instruments except for the radiological imaging approaches. However, radiological imaging also surfer from setbacks such as radioactive risk, availability, cost, and operation complexity [2, 3]. Computational predictions nowadays can approximate the in vivo conditions to high accuracies while providing tremendous details of the airflow and particles which are usually laborious to acquire in experiments. In recent years, computational simulations have gained more popularity in the design/optimization of inhalation devices and assessment of environmental health risks [4–7]. Nonetheless, drug delivery to the human respiratory tract is a complex phenomenon, which can be further compounded by numerous causes, such as patient health, respiration, device, and drug property. Large numbers of assumptions are needed to make numerical modeling feasible. As a result, many factors have to be excluded or neglected, which might have remarkably modified the transport and deposition of particles if retained otherwise. Also note that previous in vitro deposition experiments regularly used casts that were not transparent, nor could be opened apart, rendering it impractical to visualize the deposition patterns on the inner airway surfaces. It is thus highly desirable to develop techniques that can visualize and quantify the inhalation dosimetry inside the respiratory airways both accurately and simply.

The current in vitro testing of orally inhaled drug products (OIDPs) often uses a USP IP (United States Pharmacopeia induction port) followed by a cascade impactor

or a multistage liquid impinger [8]. There are attempts to quantify lung dosimetry directly from in vitro measurements. But estimations of regional dosimetry of pharmaceutical agents with the USP IP shape limited success mainly due to their underestimation of the doses in the oropharyngeal region and the subsequent overestimation in the lungs. There exists an urgent need for a new generic mouth-throat (MT) geometry in place of the USPIP to characterize the OIDP dosimetry. The USP IP has been implemented as a standard for comparison of different OIDPs and filters out the majority of large, high-speed aerosols. However, dosimetry tests with the USP IP are not satisfactory, which generally underestimates the complementary in vivo deposition data [9–11]. For the DPIs (i.e., dry powder inhalers) and pMDIs (i.e., pressurized metered-dose inhalers), mouth-throat deposition composes a large portion of the deposited dose in the human respiratory tract [12, 13]. The simple structure of the USP IP (a 90° bend with uniform circular cross-sections) fails to capture the intricate morphology of the mouth-throat airway, whose cross-sections are highly irregular in shape with variable areas. As a result, generic upper airway models with improved physical realism are needed that can predict inhalation dosimetry in the MT and lung similar to in vivo measurements. The selection of the best MT model will simulate the mean value of aerosol deposition in adult volunteers. The results will provide improved in vitro testing results for the pharmaceutical industry. A realistic human MT model will also be useful in the laboratory to study inhaled toxic particles for exposure assessment in ambient and occupational environments.

To successfully develop a generic upper airway model that is sufficiently representative of the population, we need to know the oropharyngeal factors that most affect the behavior and fates of orally inhaled drug products (OUDPs) and rank these factors according to their relative importance. However, it is inconsistent in the literature regarding such factors and their importance. Key morphological factors proposed for OIDPs include the airway volume (or the mean equivalent diameter) [14], minimum cross-sectional area [1, 15], total airway surface area [1, 16], average cross-sectional area [15, 17], airway axial length [16, 17], and oral airway curvature [1, 15, 18]. Concerning submicron aerosols, two more factors were suggested, that is, the average hydraulic diameter and the average airway perimeter of the airway [19]. Clearly, it will be ideal to know the relative importance of each factor in determining the deposition of particles of varying sizes in hope that the resultant generic airway model can capture the most important dosimetry features in most drug delivery scenarios. However, the success of inhalation drug delivery depends on the interactions between the respiratory physiology, breathing activities, medical devices, and pharmaceutical agents. Even though extensive works have been done to understand these interactions, a clear picture is still missing.

25.1.2 Previous Works

Inhalation dosimetry of orally inhaled agents can be affected by many factors, such as aerosol properties, breathing conditions, devices, and the respiratory physiology,

as demonstrated in many previous works [20–32]. Efforts in developing anatomically accurate human MT models have been undertaken based either on dental impression or medical imaging [9, 14, 17, 33–36]. Cheng et al. [33] evaluated the deposition rate of micrometer aerosols in a hollow oral cast that was reconstructed from a dental impression in a human volunteer, and the detailed dimension of these models, such as the perimeter and cross-sectional area along the main flow direction, was reported and referred to as LRRI model since then [15]. This model has been widely employed in studying the delivery efficiencies of the DPI [37] and pMDI [38]. Stapleton et al. [34] put forward an alternative but simplified mouth-throat model by retaining several physiological features like the epiglottis, variable oral size, and MT airway curvature. With this model, airflow and deposition for 3–5 μm aerosols were measured by Heenan et al. [18], revealing a strong association between the local flow characteristics and regional deposition distributions. Grgic et al. [17] demonstrated significant intersubject variability in both average and subregional depositions of OIDPs in seven MT models reconstructed from MRI scans. Efforts to quantify the intersubject variability also included Golshahi et al. [16], who tested CT-based MT and nose replicas of 6–14-year-old children. CT-based lung models were also proposed that retained lung bifurcations from G2-G9 [22, 23, 25, 39–46].

Computational fluid-particle dynamic (CFPD) simulations have been extensively used to investigate inhalation dosimetry in human upper airway models, which have been evolving from idealized, simple structures to anatomically accurate, highly sophisticated ones. Kleinstreuer and Zhang [47] developed an idealized MT model featured by a 180° bend and circular cross-sections consistent with the hydraulic diameters of the LRRI model [15]. A much higher deposition was predicted in this idealized model than the USP IP, and it is suggested that the USP IP might be inadequate to simulate in vivo dosimetry of OIDPs [48]. Xi and Longest [1] studied the effects of geometrical complexity on particle deposition using four MT models with decreasing physical realism (i.e., realistic, elliptic, circular, and constant-diameter) and demonstrated that the realistic model matched the measurements by Cheng et al. [15] better than the three idealized models. Furthermore, large discrepancies in the local or cellular-level deposition were predicted among MT models with varying levels of complexities, which could be one order of magnitude higher in the realistic MT model than in the idealized geometries. The triangular glottal aperture and the dorsal-sloped trachea were found to significantly enhance the deposition in the realistic model. Geometrical details are needed to assess the formation of deposition hot spots that are essential in accurately estimating the tissue response at the cellular level. Considering that the elliptic model retained key geometrical factors and can reasonably capture the deposition variations, it has been applied to investigate the deposition of DPI aerosols in the upper airway and was later labeled as the VCU (Virginia Commonwealth University) model [49, 50].

Validated methods exist for the measurement of inhalation dosimetry in respiratory airway hollow casts [51, 52]. By contrast, methods to visualize and quantify regional or local deposition fractions have rarely been reported. Dye-based approach with methylene blue can visualize the regional drug bioequivalence in terms of the

intensity of the blue stain [53]. However, this method cannot quantify the dosage and the deposited dye can drip and diffuse. Gamma scintigraphy using technetium-99m (^{99m} Tc) agents have been used to envisage drug distributions in the human noses both in vivo [54–56] and in vitro [57–59]. One setback of this method is that the gamma rays attenuate quickly in the human body and up to 50% energy can be scattered by the body tissue [60]. Such energy scattering can also distort the resultant images and give rise to biased dosimetry [61]. Besides, scintigraphy images are 2D and cannot distinguish deposition in different depths. Kundoor and Dalby [62, 63] demonstrated the utility of a simple, effective method to visualize droplet deposition patterns using Sar-Gel, a water-sensitive paste (Sartomer Arkema Group, Exton, PA) that turns into red upon water contact. Sar-Gel is highly sensitive to liquid water and can detect deposited water as low as 0.5 μL in volume, which is about the size of the smallest droplets in nasal sprays. Moreover, color spreading occurs only when the volume of a single droplet exceeds 25 μL. In light of other advantages like short reaction time, safe, and easy cleanup, Sar-Gel is well suited for the visualization of particle deposition.

Variability in the pulmonary drug delivery efficiency can result from different factors such as patient airway structure, respiration condition, device usages, and drug properties [64, 65]. Certain sources can be hard to control and will considerably alter drug delivery efficiency and therapeutic outcome consistency. Unfortunately, deterministic models cannot take into account the inherent uncertainties of the inputs directly [66–68]. Considering the increasing demand for reliable quantification of inhalation dosimetry, alternative models are needed that can determine the output uncertainty but are still numerically efficient. To meet these challenges, we combined two distinct methods, that is, the deterministic models [69–71] and the probabilistic analyses including the sensitivity analysis and uncertainty quantification [72, 73]. Valuable information can be yielded regarding the reliability (or confidence level) of the estimated dosimetry [74].

Most previous studies have excluded the input probability distributions in estimating the intersubject dosimetry variability; instead, variability was considered by comparing different input parameter values or between different airway geometries, and thus was still deterministic. One exception was Guo et al. [75], who used Monte Carlo simulations to estimate the propagation of input uncertainties in nasal sprays. They concluded that the regression approach and resulting standard deviations overestimated the output uncertainty, and Monte Carlo simulations could predict statistically significant confidence levels. The setback of the Monte Carlo method required prohibitive computational resources to solve the fluid-particle dynamics in a large number of test cases for statistical analysis. There exist more efficient sampling methods, such as the response surface method (RSM) and the algorithm of most probable point (MPP) [74]. Probabilistic analysis for the dosimetry of pulmonary drug delivery has been scarce.

The objective of this chapter is to introduce the fundamental theory and principles behind the inhalation drug delivery, as well as the latest advances in the computer-aided development of human upper airway models and their applications in aerosol inhalation dosimetry. Both numerical simulations and in vitro experiments will be

presented. It is hoped that this chapter will provide a comprehensive introduction to respiratory aerosol dynamics for the general audience and at the same time provide a review of state-of-the-art techniques for active practitioners in this field. Specific objectives include:

1. Demonstrate how the respiratory airway models (mouth-throat, nose, and lung) were developed.
2. Explain the physics underlying pharmaceutical delivery and the numerical methods to simulate respiratory aerosol dynamics.
3. Introduce the computational software and solvers used for pharmaceutical research and development, such as ANSYS Fluent and COMSOL.
4. Demonstrate the computer-aided testing of the morphological effects of the upper airway on particle deposition in the mouth-throat region.
5. Optimize pulmonary drug delivery using both computational modeling and in vitro experiments.
6. Develop and test the nose-to-brain (N2B) drug delivery using both modeling and experimental testing.

25.2 Methods

25.2.1 Computer-Aided Development of Airway Models

25.2.1.1 Mouth-Throat Model

We will use three examples to demonstrate the methods to reconstruct surface geometries of the human airway from either preexisting casts or on-shelf medical images (i.e., CT or MRI scans). Figure 25.1 illustrates the procedures of developing a mouth-throat model. The oral airway was developed from a dental impression with a half-way mouth opening. Detailed dimensions of the slices along the airway (second panel in Fig. 25.1a) such as cross-sectional shapes, perimeters, areas, and hydraulic diameters were provided in Cheng et al. [33]. The shape and dimension of the contours followed those measured in Cheng et al. [33]. These contours were then connected with smooth lines, and faces were built to cover the entire mouth cavity, giving rise to the realistic model as shown in the third panel of Fig. 25.1a. The throat model was developed from the CT images of a healthy adult that was acquired at a resolution of 1 mm. The software package MIMICS (Materialise, Leuven, Belgium) was applied to convert CT scans to a solid body model. The combined mouth-throat (MT) model was then imported as an IGES file into the software ANSYS ICEM CFD (ANSYS, Inc., Canonsburg, PA) for meshing. Geometric surface smoothing was conducted to remove artifacts or unnecessary anatomical details to avoid excessive grid elements. Major characteristics retained in this realistic MT model include a half opening mouth, a realistic MT airway curvature, a triangular glottal aperture, and a sloped upper trachea (Fig. 25.1a). Landmark structures like the mouth opening, larynx, and trachea were scaled to match the reported data in

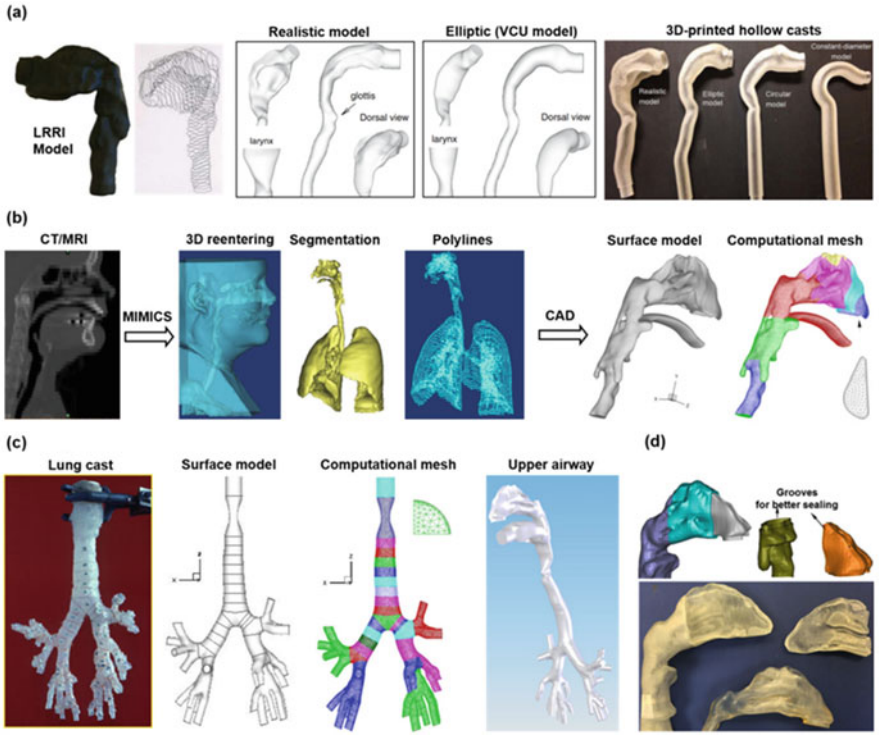


Fig. 25.1 Airway model reconstruction: (**a**) mouth-throat (MT) model that was based on the LRRI (Lovelace Respiratory Research Institute) model, (**b**) nasal model reconstructed from CT scans, and (**c**) lung model reconstructed from the lung cast provided by Dr. Cohen at New York University. The three parts have been combined to form a complete extra-thoracic airway model. Computational meshes were generated using the meshing software ANSYS ICEM CFD for numerical analysis and hollow casts were fabricated using 3D printing techniques for in vitro tests. Step grooves were developed at each interface for easy assemble and good sealing, as shown in (**d**). The elliptic MT model has evolved into the current VCU model

previous deposition studies, leading to a mouth opening diameter of 2.2 cm, a glottal area of 0.87 cm^2, and a tracheal area of 2.0 cm^2 (Fig. 25.2a).

In the first step of simplification, the cross-sections of the realistic model were replaced with ellipses of the same hydraulic diameters (Fig. 25.1). The resulting elliptic model is much simpler by removing the anatomical details of the cheek-teeth lumen and replacing the triangular glottis using an oval glottal aperture. This elliptic model has evolved into the current VCU model, as widely accepted in pharmaceutical industries [49, 50]. Further simplifications were made to generate the circular and constant-diameter models, as illustrated in the right panel of Fig. 25.1a. In the circular model, all cross-sections are circular with equivalent hydraulic diameters to those in the realistic and elliptic models. The constant-diameter model consists of a 180° bend (curvature radius: 3.2 cm) of a constant-diameter tube (area: 4.43 cm^2).

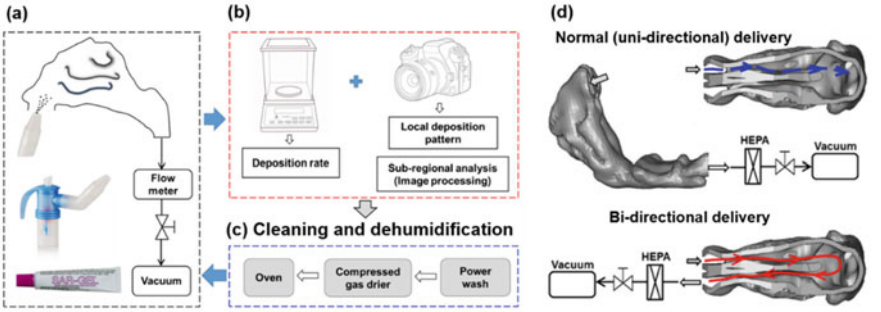


Fig. 25.2 Intranasal deposition test diagram: (**a**) drug delivery testing platform with a vacuum pump to simulate steady-state nasal inhalation, (**b**) dosimetry and distribution analyses, (**c**) cast cleaning and dehumidification, and (**d**) two delivery strategies: normal (or unidirectional) and bidirectional

The airway volumes of the four models (i.e., realistic, elliptic, circular, and constant-diameter) are 65.1, 56.8, 52.5, and 35.4 cm^3, respectively.

25.2.1.2 Nasal Model

The second example shows the reconstruction of a nasal model from MRI head scans of a 53-year-old male with no known respiratory diseases (Fig. 25.1b). This image data were originally reported in Guilmette et al. [76] and had since then been used in several in vitro tests [77–80] and numerical simulations [69, 81–85]. MIMICS (Materialise, Leuven, Belgium) was used to convert the image data set into a solid airway model, which was further converted into a group of polylines defining the solid model. From these polylines, a surface geometry was reconstructed using Gambit (ANSYS, Inc., Canonsburg, PA) (Fig. 25.1b). The boy-fitted mesh was generated for the nasal model, with five layers of prismatic cells in the near-wall region (right panel, Fig. 25.1b).

25.2.1.3 Lung Model

The third example shows the reconstruction of the lung. The original lung cast had been developed postmortem from a 34-year-old male [86] and was scanned using a Discovery LSCT scanner (GE Medical Systems) with 512 × 512 pixel resolution, 0.7-mm slice spacing, and 0.65-mm overlap. Acquired scans were segmented using MIMICS (Materialise, Leuven, Belgium) into a solid geometry and its defining polylines, from which a surface geometry was reconstructed using Gambit (ANSYS, Inc., Canonsburg, PA). The resulting lung geometry extends from the trachea to the bifurcation generations four to six (G4-G6). In total, there are 23 outlets and 44 bronchi retained.

The three regions of the respiratory tract (mouth-throat, nose, and lung) were also connected into one model (right panel in Fig. 25.1c). Computational meshes can be generated using ANSYS ICEM CFD (Canonsburg, PA) for numerical analysis. Similarly, hollow casts can be fabricated from these models for in vitro deposition

tests. Figure 25.1d shows the nasal airway casts built with an in-house 3D printer (Stratasys Objet30 Pro, Northville, MI). The casts have a uniform wall thickness (4 mm). The 3D the printing resolution is 16 μm (0.0006 in) and the printing material is polypropylene (Veroclear, Northville, MI).

25.2.2 Experimental Setup

Figure 25.2 illustrates the experimental setup for nasal deposition tests with three steps: drug administration, deposition analysis, and cast cleaning and dehumidification. Steady inhalation was simulated using a vacuum (Robinair 3 CFM, Warren, MI) connecting to the nasopharynx. A flow meter (Omega, FL-510, Stamford, CT) was utilized to monitor the volumetric flow rate (Fig. 25.2a). Before testing, the inner walls of the nasal airway cast were coated with a thin layer of Sar-Gel and then put together and fastened with a clamp. A high-precision electronic scale (Sartorious, 0.01 mg precision, Elk Grove, IL) was employed to measure the weight of the nasal cast (W_0), as illustrated in Fig. 25.2b. Aerosols were released into the nostrils for a specified time (e.g., 20 s) at 30° from the vertical orientation [63]. In the second step (immediately after the aerosol release), the new weight (W_1) of the cast was measured. The added weight ($\Delta W = W_1 - W_0$) denoted the deposited aerosols, and the deposition fraction was computed as ΔW over the output from the spray/nebulizer. The cast was then disassembled and photos were taken of the Sar-Gel color on the inner walls (Fig. 25.2b). In the third step, a power washer (Karcher, 1600 psi, West Allis, WI) was used to wash away the Sar-Gel on the cast surface (Fig. 25.2c). To remove the moisture, the cast was positioned in an oven (ThermolyneFurnatrol*18200*, Dubuque, IA) at 55 °C for 60 min. The cast was then left in the lab for another 60 min to allow the relative humidity and temperature of the cast to become fully equivalent to the environment. The third step was essential in the deposition measurement to avoid the complication from hygroscopic effect (dry surface) or water evaporation (wet surface), both of which could lead to fluctuations of the electronic scale reading. In nasal drug delivery, two strategies were presented, with one being the normal, unidirectional approach, where drugs were released into one nostril and inhaled into the lung, and the other being bidirectional, where drugs were released into one nostril and exited via the other nostril, as shown in Fig. 25.2d.

25.2.3 Computational Fluid-Particle Governing Equations

Respiration airflows were assumed to be Newtonian flows, incompressible, and isothermal. The low-Reynolds k-ω turbulence model was adopted to resolve the flow field [52, 83, 87]. The trajectories of particles were tracked using the discrete-phrase Lagrangian approach [85]. The general governing equation of particle motion is:

$$\frac{d(m_p v_i)}{dt} = \frac{f}{\tau_p C_c} m_p(u_i - v_i) + m_p g_i(1-\alpha) + F_{i,\text{Lift}} + F_{i,\text{Brownian}} + F_{i,\text{Magnet}} + F_{i,\text{Acous}} \quad (25.1)$$

where u_i and v_i are the velocity of the airflow and particle, respectively, τ_p (i.e., $\rho_p\ d_p{}^2/18\mu$) is the particle reaction time to airflow, C_c is the Cunningham slip correction factor, and f is the drag calculated following Morsi and Alexander [88]. The Saffman lift force was also included for micrometer particles and the Brownian motion force was included for nanoparticles [89]. Also, the magnetophoretic force can be considered for magnetic particles and acoustic force for pulsating aerosols.

25.2.3.1 Magnetophoretic Force

The magnetic flux density B depends on the permeability of free space μ_0, the magnetic intensity H, and the magnetization M [90]:

$$B = \mu_0(H + M) \quad (25.2)$$

The effective magnetic dipole moment caused by H is calculated as:

$$m_{\text{eff}} = \frac{4}{3}\pi d_p^3 KH \quad (25.3)$$

Here, K is the Clausius-Mossotti factor$K = (\mu_p - \mu_f)/(\mu_p + 2\mu_f)$. The magnetophoretic force exerted on a spherical particle in a nonuniform magnetic field is [91]:

$$F_{i,\text{Magnetophoretic}} = (\mu_0 m_{\text{eff}} \cdot \nabla)H \quad (25.4)$$

The symbols μ_p and μ_f are the magnetic permeability of particles and fluid, respectively. From Eqs. 25.3 and 25.4, the magnetophoretic force is proportional to the particle volume, and the product of the magnetic field and its gradient. The direction of the force is along the magnetic field gradient.

25.2.3.2 Acoustophoretic Force

For the pressure acoustic computation, boundary conditions are the sound hard boundary for the wall, the plane wave radiation for the outlet, and the pressure for the inlet. The governing equations to calculate the acoustophoretic force is shown below [92]:

$$f_{i,\text{acoust}} = -\nabla\left[V_\text{p}\left(f_1\frac{P^2}{2\rho c^2} - f_2\frac{3}{4}\rho v^2\right)\right];\ \ f_1 = 1 - \frac{k_0}{k_\text{p}};\ \ f_2 = \frac{2\left(\rho_\text{p} - \rho\right)}{2\rho_\text{p} + \rho} \tag{25.5}$$

Here ρ is the air density, ρ_p is the particle density, V_p is particle volume, v is particle velocity, P is the instantaneous pressure, k_0 and k_1 are the bulk moduli, and f_1 and f_2 are the scattering coefficients for the monopole and dipole, respectively. Multiphysics software COMSOL (Burlington, MA) was implemented to simulate the pulsating pressure distributions and associated particle motions.

25.2.4 Numerical Methods

Both ANSYS Fluent (Canonsburg, PA) and COMSOL (Burlington, MA) were used to resolve the airflow and track the particle motions. In ANSYS Fluent, user-defined functions (UDFs) were developed to consider the near-wall velocity interpolation (NWI) and anisotropic turbulent effect. Near-wall interpolation is a linear function of the particle velocity from zero at the wall to the velocity at the cell center. Second-order spatial discretization or higher were used for all transport terms. The convergence of airflow solution was achieved when the mass residual decreased by five orders of magnitude and the residual variation profiles for both mass and momentum became flat. Convergence sensitivity analyses were also conducted to establish grid-independent and particle-count-independent results following the method of Xi et al. [93].

The computational mesh of respiratory models was created with ANSYS ICEM CFD (Canonsburg, PA) with tetrahedral cells in the main flow region and body-fitted multi-layer prismatic cells in the near-wall region. It has been demonstrated that the near-wall prismatic cells are critical in establishing grid-independent results. To achieve a similar level of numerical accuracy, the total number of cells needs to be five times higher in the all-tetrahedral mesh than a hybrid tetrahedral mesh with near-wall prismatic cells [89].

25.2.4.1 ANSYS Fluent and ICEM CFD

ANSYS Fluent (Canonsburg, PA) is a popular engineering software to simulate fluid flow, heat transfers, fluid-structure interactions, aeroacoustics, and electromagnetic flows. The latest version is ANSYS Fluent 2020R2. Due to its high accuracy and robustness, ANSYS Fluent has been widely used in both academia and industries. In recent years, ANSYS Fluent has been providing a student version, which is freely available to all students in the world. This version can be installed on any Microsoft Windows 64-bit machine. This product can be downloaded via the link: https://www.ansys.com/academic/free-student-products. Examples of using ANSYS Fluent in pharmaceutical design and delivery can be found in [94–96].

ANSYS ICEM CFD (Canonsburg, PA) is the mesh generation software bundled with ANSYS Fluent. It can efficiently mesh large, complex models with either structured or unstructured elements. ANSYS ICEM CFD can read geometries generated in most computer-aided design software, such as Gambit, Solidworks, CATIA, blender, etc. The geometry format include step, IGES, sat, xml, STL, VRML, and other third-party mesh format. The generated mesh can be exported for either computational fluid dynamics or finite element analysis. More information about ANSYS ICEM CFD can be found in: https://www.ansys.com/services/training-center/fluids/introduction-to-ansys-icem-cfd.

25.2.4.2 COMSOL Multiphysics

COMSOL Multiphysics (Burlington, MA) is a computation software that gains popularity in recent years due to its easy-to-use features and its capacity in simulating multiple physics. As of November 2020, the latest version is COMSOL 5.6. COMSOL was started as "FEMLAB" around 20 years ago and took root in the finite element method to solve engineering problems. The software was later changed to the current name. COMSOL does not provide free, student version. The trail period is generally 15 days. The software comprises the base module to conduct general engineering simulations, as well as a variety of specialty modules to tackle specific physics, such as electrical, mechanical, particles, acoustics, and chemical applications. The system requirements for installation include at least 4 GB of RAM and 2–13 GB of disk space, depending on your installation options of specialty modules.

The physics included in the software covers almost all aspects of engineering. Besides electric problems (or AC/DC: alternating current/direct current), it also includes magnetic, acoustic, solid mechanics, flow, heat, and chemical applications. COMSOL's interface allows the user to include different physics easily. Another salient feature is that there are a large number of tutorials that are readily available to new users to start, either in the library embedded with the software or through the gallery of the COMSOL website: https://www.comsol.com/models. Either new users with no prior experience with COMSOL or experienced users can learn from these tutorials. The Fluid-flow module and particle-tracing module are the two modules needed to study the transport and deposition of pharmaceutical particles. If there are electrostatic charges in the particles, AC/DC module should be used. For those who are interested in product designs of inhalers, the optimization module is also needed. For readers who are interested in learning more about this software, please refer to https://www.comsol.com/comsol-multiphysics. Interested readers are also referred to a video tutorial that explains step by step how to COMSOL and ANSYS Fluent to simulate olfactory drug delivery with active control of pharmaceutical aerosols [97]. More examples can be found in [98, 99].

25.2.5 Statistical Analysis

Statistical analysis software Minitab (State College, PA) was utilized to evaluate the variance in dosimetry. To assess the sample variability and determine the major influencing factors, one-way analysis of variance (ANOVA) and Tukey's method with stacked data were applied. To compare the relative importance between different factors, the mean effect analysis of ANOVA was performed, for example, on the MT model, particle size, geometrical factor, and inhalation rate [100]. To evaluate the interactive effects of dominating factors, the interaction analysis of ANOVA was undertaken. When the p-value was <0.05, a statistically significant difference was reached.

25.3 Applications

25.3.1 Model Validation

The computational models of the nasal, mouth-throat, and mouth-lung geometries had been extensively validated, as demonstrated in Fig. 25.3a–c, respectively. All models were meshed with five layers of prismatic elements near the wall, as illustrated in the left panel in Fig. 25.3a. Near-wall grid convergence was established at the near-wall height of 0.05 mm for both 150 nm and 1 m particles (middle panel,

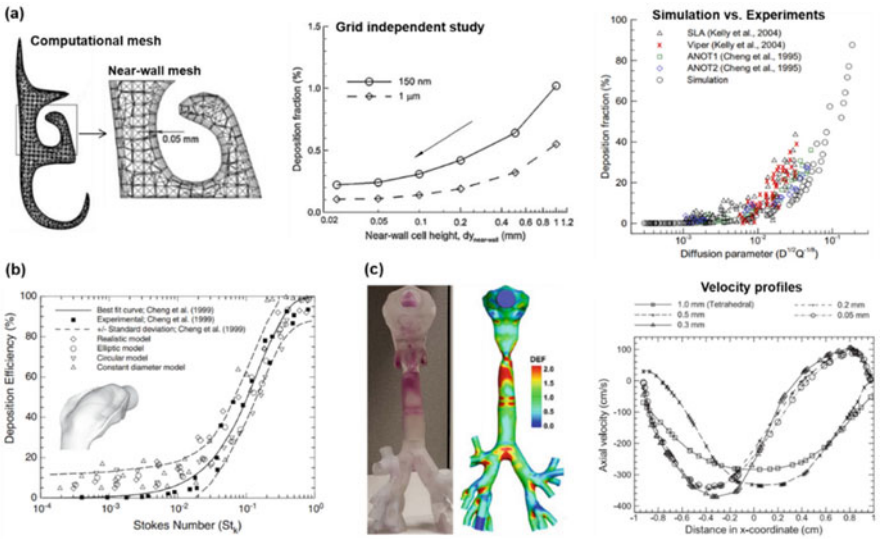


Fig. 25.3 Validation for the nasal, mouth-throat (MT), and mouth-lung models: (**a**) computational mesh, grid-independent study, and simulation versus experiment, (**b**) MT model validation for micrometer particles as a function of Stokes number, and (**c**) mouth-lung model validation in terms of deposition visualization, and velocity profile variation versus mesh density

Fig. 25.3a) for the nasal cavity. Further refining the near-wall cells below 0.05 mm was demonstrated to have a small effect on the deposition rate. The third panel in Fig. 25.3a shows the comparison of the CFD-predicted deposition fractions and in vitro test results versus a diffusion parameter proposed by Cheng et al. [15].

Figure 25.3b shows the numerically predicted deposition rates versus the particle Stokes number for the four mouth-throat models as shown in Fig. 25.1a in comparison to measured data and Cheng's empirical correlation in the LRRI (Lovelace Respiratory Research Institute) MT model [15]. Nearly all predicted data from the four MT models fall within the experimental uncertainty bonds. In particular, the most idealized constant-diameter model (triangles) overestimated the in vitro data.

Deposition visualization was also used for model validation purposes, as illustrated by Fig. 25.3c. It was observed that the numerical predictions matched the experimentally determined deposition distribution to a high degree, suggesting that the computer models indeed captured the particle behaviors and fates. In particular, we observed several positions where the predictions and measurements are highly similar. For instance, the computer model successfully captured the crescent deposition hot spot immediately downstream of the glottis. Moreover, there were two streaks in the front middle trachea in both the predicted and experimental deposition patterns.

Mesh sensitivity analysis of the airflow velocity field is presented in the right panel of Fig. 25.3c, which compares the velocity profiles between different mesh densities in the tracheal mid-plane 3 cm below the glottis. Large discrepancies in the velocity profiles exist between coarse and fine near-wall heights, indicating that the airflow is highly sensitive to the near-wall grid size. Moreover, the velocity profiles in the two coarsest meshes (near-wall height: 1.0 and 0.5 mm) were relatively symmetric. By contrast, the velocity profiles became more asymmetric with increasingly finer meshes. It appeared that the coarse meshes failed to capture the skewed or reversed flows because of exceeding numerical diffusion. From Fig. 25.2, the convergence for the velocity field was achieved at the near-wall height of 0.05 mm. The good agreement between model simulations and experiments in both the deposition rate and deposition distributions, as well as the grid-independent airflow field, imparts confidence into the subsequent simulation results.

25.3.2 Mouth-Throat Model Development and Testing

Figure 25.4a shows the dimensions of different MT models in terms of perimeters and hydraulic diameters versus the axial distance from the oral opening. As expected, the largest variability was observed in the realistic model in all geometrical parameters. Dimension discrepancies between the realistic and elliptic models are due to the removal of the cheek-teeth cavities in the realistic MT model. The second panel shows the airflow fields (i.e., velocity vectors, mid-plane contours, and streamlines) in the realistic and constant-diameter MT models at 30 L/min. In the realistic model, the peak velocity downstream of the glottis (i.e., laryngeal jet) shifts to the front wall of the trachea due to fluid inertia (Fig. 25.4b, right panel). This jet

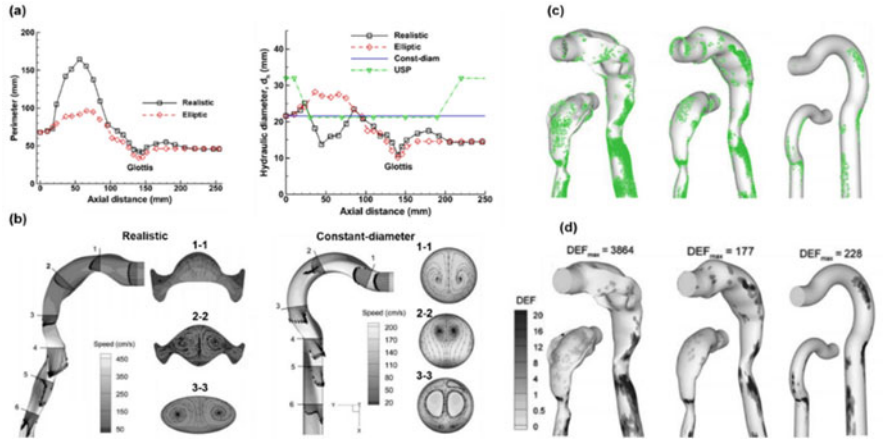


Fig. 25.4 Geometrical complexity effects of the mouth-throat (MT) models: (**a**) perimeter and hydraulic diameter of the MT models, (**b**) airflow patterns in the realistic and constant-diameter models, (**c**) surface deposition of 6 μm particles, and (**d**) local deposition in terms of the deposition enhancement factor (DEF)

effect induces reversal flows near the back wall of the trachea. Considering the sloped angle of the trachea in the realistic and elliptic models, increased particle impaction is expected from the laryngeal jet. In the constant-diameter model (Fig. 25.4b), a pair of counter-rotating helical vortices are observed in most cross-sections, which are absent in the realistic model. Due to a different curvature than that in the realistic model, the maximum velocity in the constant-diameter model is observed near the upper wall instead of the lower concave surface as in the realistic model (Fig. 25.4b).

A comparison of deposition distributions is shown in Fig. 25.4c, where both the deposition rate and distribution differ significantly among MT models with varying physical realism. For the realistic and elliptic models, elevated deposition is observed in the back of the pharynx due to the abrupt MT curvature and associated particle inertia impaction. By contrast, reduced deposition is observed throughout the pharynx in the constant-diameter model (right panel, Fig. 25.4c). Large variations in deposition distributions are found at other locations too, including the lower larynx and trachea. Specifically, high levels of particle accumulation occur in the realistic model below the glottis (right panel, Fig. 25.4c); however, the intensity of this particle accumulation decreases as the geometrical complexity decreases from the elliptic to the constant-diameter models.

To further investigate the impact of the geometrical factors on inhalation dosimetry [100], the shape of existing MT models was modified using Hypermorph, as illustrated in Fig. 25.5a. One advantage of Hypermorph is that geometry can be modified locally in a quantitative manner (Fig. 25.5b), and thus is well suited for a systematic parameter study. Figure 25.5c displays the modified realistic and elliptic MT models with three factors of interest: oral cavity volume, glottal aperture, and

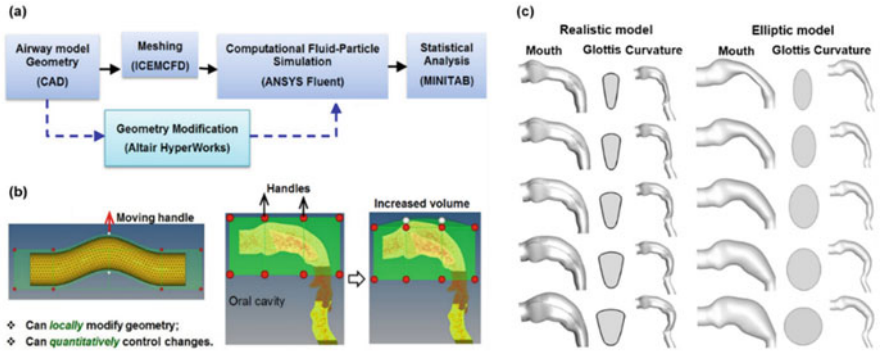


Fig. 25.5 MT model modification: (**a**) flowchart showing the airway models being systemically modified for subsequent parametric studies, (**b**) using Hypermorph to modify a model geometry locally and quantitatively, and (**c**) modified MT geometries for the realistic and elliptic models with varying oral cavity volumes, glottis areas, and MT curvatures. The effect of total airway volume was evaluated by uniformly scaling the MT models with four different scale factors (not shown due to unchanged shape)

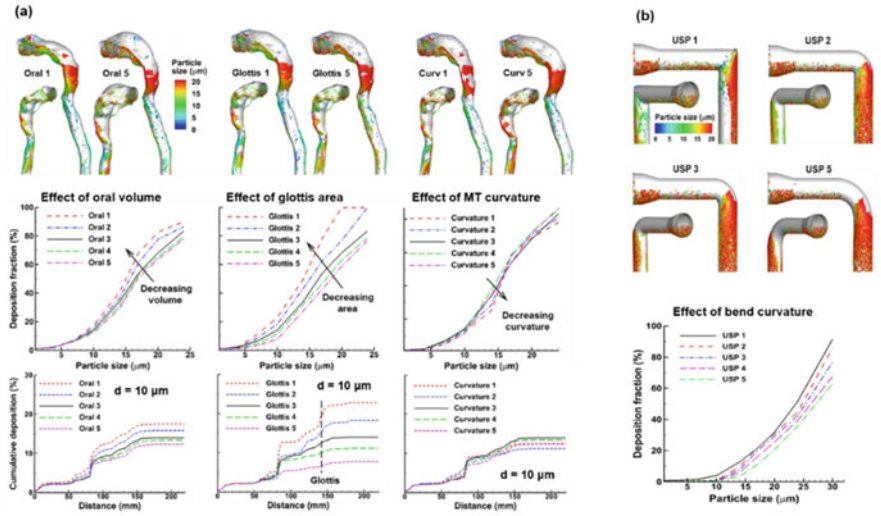


Fig. 25.6 Deposition variations in modified geometries: (**a**) USP and (**b**) realistic models. In USP, the effect of bend curvature on surface deposition and deposition fraction is shown at 30 L/min

airway curvature. The effect of total airway volume was also evaluated by uniformly scaling the MT models with four different scale factors, but figures were not shown due to unchanged shape.

Figure 25.6a shows the differences in deposition among the modified realistic geometries at 30 L/min in terms of (1) oral cavity volume, (2) glottal aperture area, and (3) airway curvature. These included the deposition pattern (upper panel), the

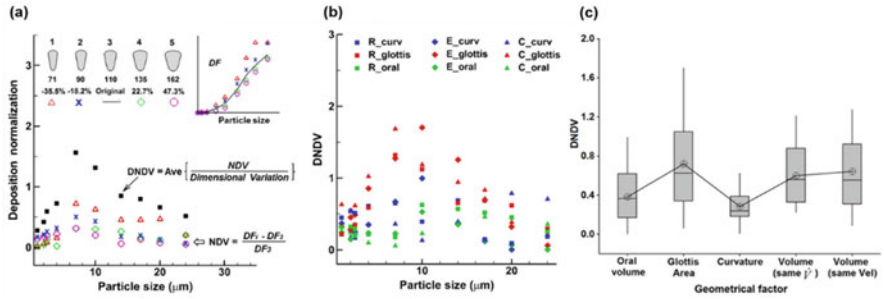


Fig. 25.7 Sensitivity analyses of the geometrical factors: (**a**) dimensionally normalized deposition variation (DNDVs); (**b**) DNDV versus particle size in three different MT models in realistic (R), elliptic (E), and constant-diameter (C) model geometries; and (**c**) box plots of DNVCs for different geometrical factors

deposition rate (middle panel), and the cumulative deposition fractions of 10-μm particles along the mean flow direction (lower panel). When varying the above three factors, deposition changed much differently in both magnitude and trend. For instance, reducing the glottal area and oral volume both increased the deposition fractions; by contrast, reducing the MT curvature (in both length and angle) decreased particle deposition (Fig. 25.6a, right panel). In this example, the most influencing factor was the glottal area, followed by the oral volume, while the impact from the MT curvature was the least among the three factors. Variations in the glottal area led to significant changes in particle deposition both before and after the glottal aperture (Fig. 25.6a).

Figure 25.6b displays the deposition patterns in the modified USP ducts by replacing the sharp 90° bend with smoothed ones of varying curvatures. The reduced deposition was predicted in the curved bend in comparison to that in the sharp bend, as expected. The USP deposition further decreased with increasing curvature radius, presumably resulting from smoother particle trajectories and diminished inertial impactions (Fig. 25.6b). In smoothed USP ducts, fewer particles were deposited on the back of the bend and more on the ventral and lateral sides.

To assess the impacts from the geometrical factors, the deposition variations, instead of the absolute deposition rate, were considered, which were first normalized by the control case deposition rate, and then by the magnitude of the dimensional variation, that is, dimensionally normalized deposition variation (DNDVs). Figure 25.7a shows the deposition normalization with respect to the glottal area in the realistic model. There are five variants of glottal apertures (Case 1–5) and Case 3 (original dimension) was used as the control case. Figure 25.7b shows the resultant DNDV in terms of the particle size in the realistic (R), elliptic (E), and constant-diameter (C) model geometries. The DNDV denotes the relative change in deposition per unit structural variation. To determine the major influencing factors, the DNDVs for all variables considered (geometry variation, particle size: 0.5–24 μm, and breathing condition: 15–60 L/min) were analyzed by means of ANOVA. Figure 25.7c shows the box plot of the five influencing factors. The glottal aperture

area and the airway volume were the most important factors in determining MT dosimetry of micrometer particles, while the oral volume and the MT curvature exerted an insignificant impact. Note that the "*volume*" in Fig. 25.7c represented the entire MT airway, as opposed to the oral volume. Two breathing conditions were considered when considering the MT volume effect: the same inlet velocity and the same inhalation flow rate. From Fig. 25.7c, the MT volume effect was observed to be significant (p-value <0.01) in both scenarios.

25.3.3 Pulmonary Drug Delivery

25.3.3.1 Probability Analysis

The dosimetry uncertainty was evaluated due to variability in three inputs: particle size, particle density, and inhalation flow rate (Fig. 25.8a). The input variability of each input variable has a normal distribution and a standard deviation σ that is 25% of the mean μ. Figure 25.8b shows the dosimetry variability at a 95% confidence level based on the prescribed input uncertainties.

A comparison of the input sensitivity of the three factors is illustrated in Fig. 25.8c in terms of mean and total effects. The index of the main effect represents a variable's individual impact on the response, while that of the total effect includes both the main effect index and the interactions with other influencing parameters [101]. As a result, the disparity between these two indexes is from the interactions among the inputs. For both indexes, the variability in particle size is the most sensitive factor in determining the pulmonary dosimetry, while the variability in the airflow speed is the least among the three factors. Interestingly, even though the airflow speed has a very small main effect, it still affects the pulmonary dosimetry by its strong interactions with other parameters, as shown by the relatively larger total effect (Fig. 25.8c). Only a parameter that has a near-zero total-effect index should be considered to be trivial.

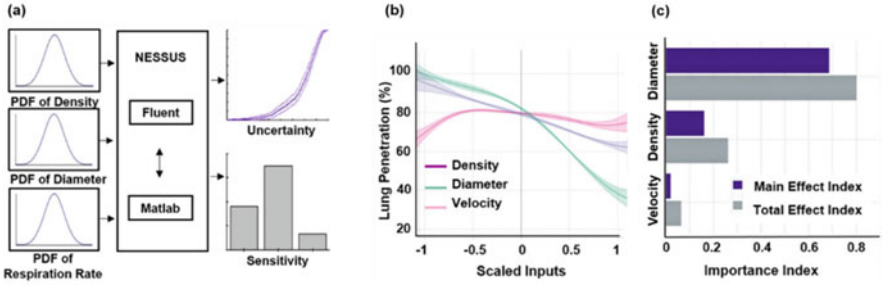


Fig. 25.8 Probability analysis of deposition in the mouth-lung model: (**a**) the flowchart of probability analysis that combines NESSUS, ANSYS Fluent, and Matlab, with normal distributions of particle density, particle diameter, and inlet airflow velocity; (**b**) dosimetry uncertainty at 95% confidence level; and (**c**) input sensitivity of the three factors

25.3.3.2 Deposition Visualization: CFD Versus Experiments

Both numerical simulations and in vitro experiments were conducted in the mouth-lung model (Fig. 25.9a). Complex flow fields arose from the complex airway structures, as shown in Fig. 25.9b. Salient features included an abrupt (90°) change in the main flow direction, a laryngeal jet, and flow instability downstream of the glottis [102]. Significant inertial deposition in the velopharynx dorsal wall was expected from the abrupt change in the flow direction (Slice 1, Fig. 25.9b).

The laryngeal jet induced a flow recirculation zone in the upper trachea, which could increase the resident time of particles in this region and thereby increase particle deposition (Slice 3, Fig. 25.9b). Considering the coherent structures, stream-wise vortex filaments formed above the glottal aperture from secondary swirling flows and continued to grow because of the energy inputs from the laryngeal jet, inducing vortex rings within the upper trachea. These coherent structures had a direct impact upon the generation and decay of turbulence, as well as the pressure drop across the glottal aperture. Considering the aerosol transport (right panel in Fig. 25.9b), the majority of particles were found to follow the high-speed main flow and fewer particles were found in the low-speed or recirculation regions. Aerosols reached the lung carina ridge about $T = 0.16$ s after being inhaled.

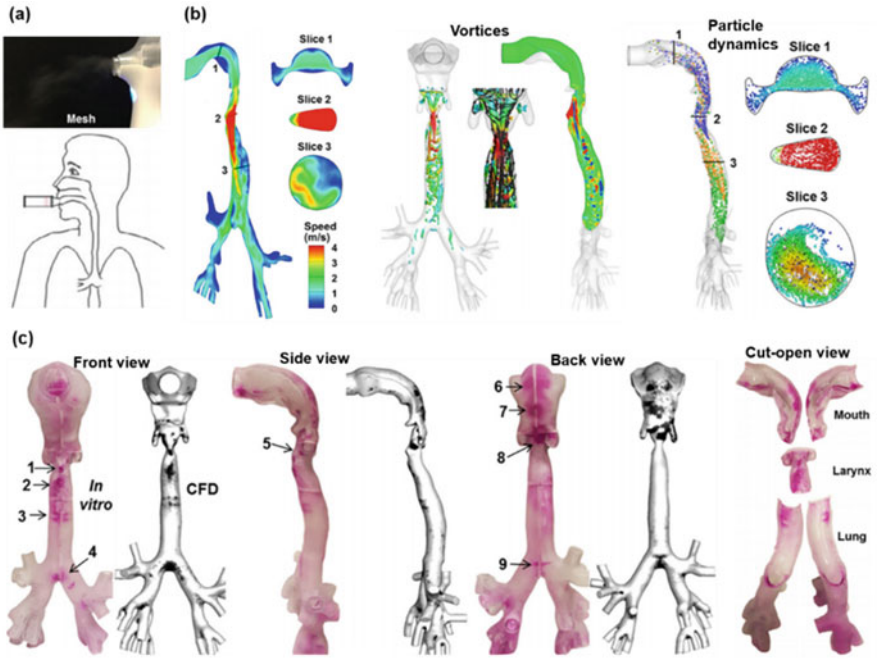


Fig. 25.9 Comparison of in vitro measurements and computational predictions in a mouth-lung cast under normal breathing condition: (**a**) in vitro experimental diagram; (**b**) computationally predicted airflow, vortex coherent structures, and particle dynamics; (**c**) experimental measurements versus CFD predictions of the particle deposition distribution in the front, side, back, and cut-open views

Due to the laryngeal jet, particles were shifted to the ventral wall of the trachea (Slice 3, right panel, Fig. 25.9b). As a result, enhanced particle depositions were expected in the ventral wall of the trachea, which was corroborated by the Sar-Gel visualizations shown in Fig. 25.9c.

Figure 25.9c compares side by side between the in vitro experiments and computational predictions of deposition distributions of nebulized aerosols at 30 L/min, which closely resembled each other and validated the computer models hereof in capturing the particle behaviors and fates in human airways. In particular, Fig. 25.9c listed nine locations, where CFD and Sar-Gel visualization compared favorably [102]. These included the deposition hot spots at the glottis tip (point 1), downstream of the glottis (point 2), middle trachea (two strips, point 3), front of the lung carina ridge (point 4), lateral larynx (point 5), the roof of the oral cavity (points 6, 7), dorsal pharynx (point 8), and back of the carina ridge (point 9, Fig. 25.9c). A cut-open view of the aerosol deposition in the mouth-lung model is illustrated in Fig. 25.9c. Since the aerosol depositions on the inner walls are directly exposed and are not biased by the transparency of the casts, Fig. 25.9c provides a more accurate representation of the delivered pharmaceuticals inside the airways. In addition, as the Sar-Gel color depth depends directly on the mass of the applied water, delivered doses can be quantified, as demonstrated in Xi et al. [2].

25.3.3.3 Deposition Visualization: Normal and Diseased Breathing Conditions

Various lung compliance and airflow resistor (PneuFlo Parabolic) were used varied in the breath simulator to simulate diseased breathing conditions (Fig. 25.10a). For instance, to simulate asthmatic respiration, the default resistors for normal breathing Rp 5 (i.e., 5 cm H_2O/L/s) were replaced by Rp 25 (i.e., 25 cm H_2O/L/s) [103]. The increased airflow resistance led to a longer time for the lung to fully exhale before starting a new breath (i.e., obstructive lung disease). To mimic a restrictive respiratory disease like fibrosis, the lung compliance was changed from the control setting of 0.1 L/cm H_2O (health: 0.05–0.1 L/cm H_2O) to 0.03 L/cm H_2O to simulate the higher lung stiffness [104–106].

Sar-Gel visualizations of the nebulized aerosol deposition are compared in Fig. 25.10b between the normal, asthmatic, and fibrosis cases for 30 respiration cycles [103]. The highest deposition in the upper airway was noted in the normal case, while the asthma case led to the least deposition in the upper airway. In comparison to the normal case, the slow inhalation in fibrosis gave rise to lower deposition in the mouth-throat region. It is noted that deposition patterns on the left and right walls were not symmetric for all of the three cases herein.

Figure 25.10c shows measured masses under the normal and pathological breathing conditions for 30 respiration cycles with each case repeating five times. The upper airway dose under asthmatic breathing conditions was significantly lower than (~50% of) the control case. By comparison, the fibrosis case led to similar doses as the control case. In light of the dose variability, no outlier existed in these three cases, suggesting good repeatability of the measured dosages. Figure 25.10d shows the subregional deposition in different regions of the airway cast, that is, the right upper,

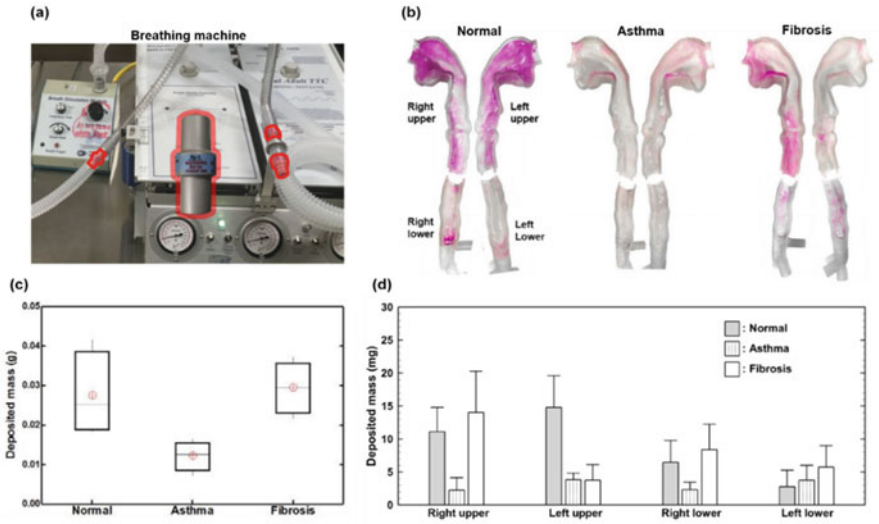


Fig. 25.10 Characterization of the deposition of aerosols from a mesh nebulizer in a mouth-lung cast under normal and pathological conditions: (**a**) PneuFlo parabolic flow resistor and their locations in the breathing machine; (**b**) surface deposition under three breathing conditions: normal, asthma, and fibrosis; (**c**) measured masses under the normal and pathological breathing conditions for 30 respiration cycles; and (**d**) deposited masses in the four subregions of the mouth-lung cast

left upper, right lower, and left lower. Apparently, the dosimetry of nebulized aerosols was highly sensitive to the breathing conditions. The asymmetric ventilations to the left and right lungs were presumably responsible for the left-right asymmetry in deposited doses. Again, the dosimetry variability was not significant with no obvious outliers, indicating satisfactory repeatability of the in vitro measurements.

25.3.3.4 Statistical Shape Modeling for Lung Morphing

Statistical shape modeling (SSM) has been demonstrated to be a useful method to study lung structural remodeling and resultant dosimetry variability. Even though SSM was originally developed as an algrothm in computer graphics and image processing [107, 108], it has been increasingly used in biomechanics, such as evaluation of fracture risks or implant performance [109, 110], face recognition [111], forensics [112], anthropology [113], and evolutional biology [114, 115]. SSM has been demonstrated to be an effective tool to evaluate a large number of subjects based on a limited number of original samples.

In theory, a training set was used to find the principal components (or eigenvectors/features) that were further used to generate new samples by varying the coefficients (or weights) of major eigenvectors. Figure 25.11a shows four examples of the SSM-generated lung models, which have a similar architecture as the training set but has a remodeled left lower lung lobe. The first two models

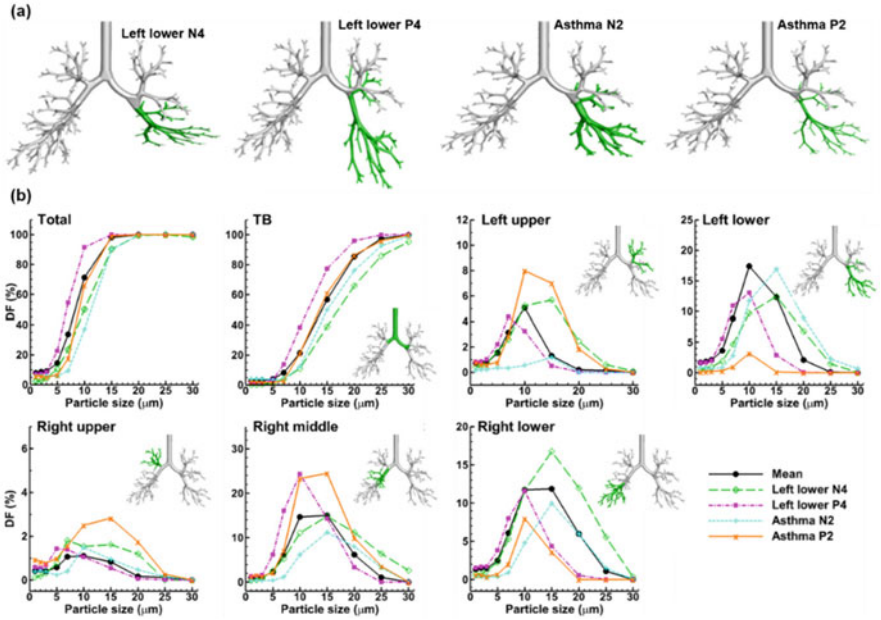


Fig. 25.11 Statistical shape modeling in morphing lung diseases and aerosol inhalation dosimetry: (**a**) four lung geometries generated from statistical shape modeling, with first two being rigid or over compliant and the last two being dilated and constricted, and (**b**) total and subregional deposition fractions (DFs) of the four remodeled geometries in comparison to the normal geometry. The numbers N2 and P2 represent the weights of the principal components (i.e., features) that reconstruct the two asthmatic lung models

represent the lungs with different compliances, while the last two models represent the over-inflation or constriction of the left lower lobe, respectively.

Figure 25.11b compares the doses between the normal and remodeled lung geometries. The highest total DF was predicted in the P4 model (pink solid squares in Fig. 25.11b), while the lowest total DF was predicted in the N2 model. Larger differences due to lung structural remodeling were observed in subregional DFs, as shown in the second to seventh panels in Fig. 25.11b. Moreover, large DF variations occurred not only in the regions with structural remodeling (i.e., the two left lobes), but also in the right upper and middle lobes, suggesting a systematic influence from local airway remodeling.

25.3.4 Nasal Drug Delivery

25.3.4.1 Deposition Visualization for Nasal Sprays and Nebulizers

Figure 25.12a visualizes the deposition of nasal sprays using Sar-Gel [2]. The plume angle of the sprays was $19° \pm 0.6°$, $35° \pm 0.8°$, $33° \pm 0.8°$, $20° \pm 0.5°$ for Miaoling, Astelin, Apotex, and Nasonex, respectively. It was observed that most spray droplets

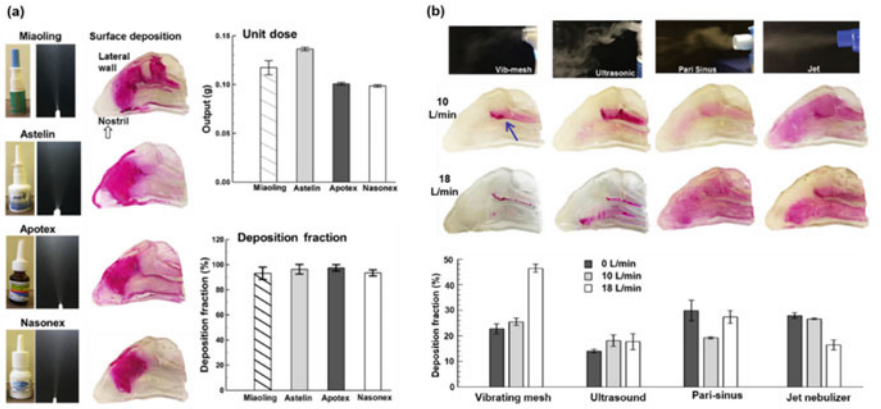


Fig. 25.12 In vitro tests of nasal sprays and nebulizers: (**a**) plumes, deposition patterns, and doses in the nose cast using different nasal spray products; (**b**) soft mists, deposition patterns, and deposition fractions of different types of nebulizers (i.e., vibrating mesh, ultrasound, Pari Sinus, and jet nebulizer)

deposited in the anterior nose, especially in the nasal valve region. The unit output (per stroke) of the nasal sprays is 0.12 ± 0.15 g. Close to 100% of the nasal sprays that were administered into the nostril(s) deposited inside the nose (right lower panel, Fig. 25.12a).

Figure 25.12b shows the Sar-Gel visualization of nasal deposition using four nebulizers of different mechanisms in the aerosol generation. The nozzle angle to the nostril was 60° from the horizontal direction. Low-speed soft mists were noted in all nebulizers except the jet nebulizer (upper right, Fig. 25.12b). Downward droplet motions occurred after discharging from the ultrasonic nebulizer due to the slow velocity and large size of the droplets. The deposition patterns were remarkably different among the four nebulizers (middle panel, Fig. 25.12b), from focused (mesh nebulizer) to widespread (jet-type) patterns. At a low inhalation rate (10 L/min), the strip of deposition on the edge of the middle turbinate (blue arrow) is consistent with the main inspiratory flow. At 18 L/min, more droplets were entrained by the main flow into the median passage, reducing the deposition on the turbinate edge. Considering the ultrasonic nebulizer, similar patterns were obtained at 10 and 18 L/min. Considering the Pari Sinus, the deposition distribution varied from less diffusive at 10 L/min to more widespread at 18 L/min. In the jet nebulizer, the highly dispersed deposition was found at both flow rates. Also, there was less deposition in the upper nose at 18 L/min. The core flow mainly occurred in the lower and median passages and entrained aerosols that otherwise went to the upper nose.

The lower panel of Fig. 25.12b shows the deposition fractions (DFs) at three respiration flow rates, with each test case being repeated five times. In comparison to the nasal spray products, the deposition of nebulized aerosols is much lower. The maximum DF is 46% for the mesh nebulizer at 18 L/min and the minimum DF is 15% for the ultrasonic nebulizer at 0 L/min (breath-holding). Interesting trends are

noted in the DF variation with flow rate for different nebulizers: the DF increases with flow rate for the mesh and ultrasonic nebulizer, while it decreases with flow rate for Pari Sinus and Philips jet nebulizers.

25.3.4.2 Normal Versus Bidirectional Nasal Drug Delivery

Different patterns of particle deposition are expected in the two passages under the bidirectional breathing pattern [116, 117]. To visualize the deposition pattern in the second passage, the direction of the bidirectional delivery was reversed, namely from "left in right out" to "right in left out" (Fig. 25.13a), so that particle deposition in the left (exiting) passage could be revealed. As expected, much fewer aerosols were deposited in the second passage. Figure 25.13b shows the deposition pattern after 1 minute's administration. This diminished deposition is reasonable considering that the majority of the administered droplets have been filtered out in the first (entrance) passage. Overall, the bidirectional technique yielded higher depositions in both the nasal cavity and the olfactory region compared to the normal technique. It is interesting to see from Fig. 25.13b that PARI Sinus performs better with the normal delivery method, while the mesh nebulizer is better with the bidirectional technique. This trend is valid in both the nasal cavity and the olfactory region.

To further understand the bidirectional effects on particle behaviors, snapshots of particle motion at various instants after administration were computed (Fig. 25.13c). The velocity vectors were also plotted on particles at selected instants. After 30 ms, particles approached the nasopharynx and changed directions sharply in the bi-directional mode to enter the second passage, leading to greater resistance than the normal mode. This increased resistance would affect the airflow and particle dynamics in both nasal passages, and the second passages in particular due to the ambient pressure at its exit. The diffusive surface deposition predicted by numerical

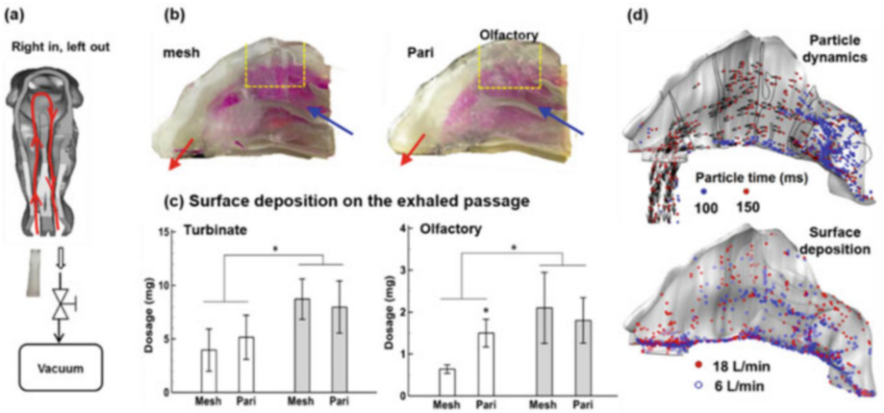


Fig. 25.13 Deposition pattern in the second (exiting) nasal passage using the bidirectional delivery protocol for the mesh (Voyager Pro) and PARI Sinus nebulizers: (**a**) delivery diagram (right in, left out), (**b**) deposition pattern, (**c**) deposition fractions in the turbinate and olfactory region, and (**d**) simulation results of the particle dynamics and deposition distribution in the exhaled passage

simulations (lower panel, Fig. 25.13c) is consistent with that obtained via in vitro tests (Fig. 25.13b).

25.3.4.3 Nasal Passage Dilation Effects

The human nose is a compliant structure that can change its shape/size either actively or passively to adjust airflow ventilation. Nasal expansion can alter the dosimetry of inhaled aerosols within the nasal cavity [118]. To quantify the dosimetry variation from nasal dilations, both in vitro tests and numerical simulations were undertaken in three nasal models (N0, N1, N2) with progressive dilation in the nasal passages (Fig. 25.14a, b). Specifically, gradual expansion was made to the nasal valve region, as evidenced by the front nose width of 2.49, 2.90, and 3.40 com in N0, N1, and N2, respectively (Fig. 25.14a). Relative to the control case N0, the expansion rate of the nasal valve was 30% and 50% in N1 and N2 (Fig. 25.14b).

First, valve dilation on flow partition in the nose was investigated. Both nasal valves were split into the upper and lower zones and their ventilation rates were quantified (Fig. 25.14c). In contrast to a similar upper-lower area ratio (~39%) for the three models, the flow ventilation to the upper zone was much lower, that is, 17–22%. Moreover, valve dilation enhanced the flow partition to the upper zone, which was 17% in N0 and 20% in N2 for the normal (unilateral) delivery and was 18.5% in N0 and 21% in N2 for the bidirectional delivery. The flow partition to the upper valve increased significantly in the left nose (the exhalation nasal passage) for the bidirectional delivery, with a 30.5% flow partition in N0 and 36% in N2. In

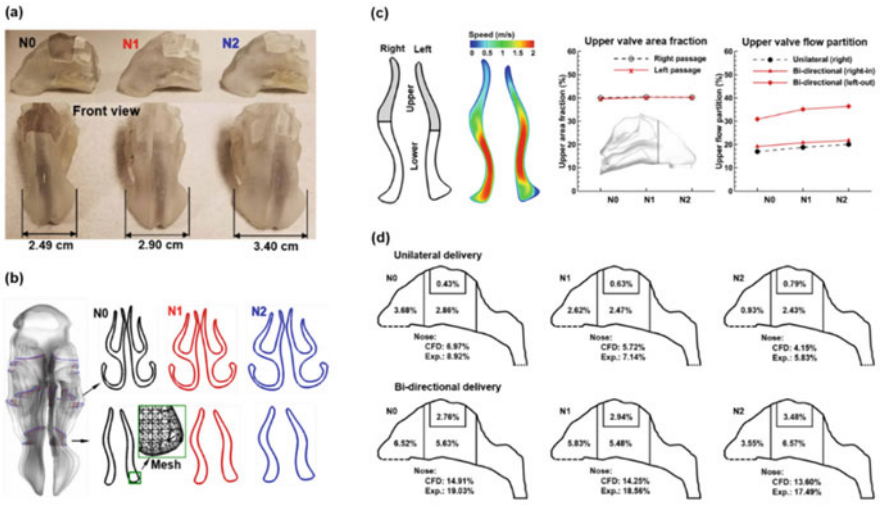


Fig. 25.14 Nasal passage dilation effects: (**a**) 3D-printed casts of three nasal models (N0, N1, N2) with increasing passage dilations, (**b**) cross-sections of the three nose dilatation models at the nasal valve and turbinate, (**c**) CFD-predicted flow partition, and (**d**) comparison of CFD and in vitro measured deposition fractions among the three models under unilateral and bidirectional deliveries

comparison to the right nose (inspiratory nasal passage), the high-speed flow zone in the left valve shifted upward, as demonstrated by the speed contour in Fig. 25.14c.

Numerically predicted DFs are shown in Fig. 25.14d for the three models in comparison to in vitro measurements. Good agreement was attained between the predicted and measured DFs both in magnitude and trend. The predicted DFs slightly, but consistently, underestimated the experimental data. For both delivery methods, the vestibular dosage decreased with the valve dilation. Specifically, a large decrease in the vestibular dosage occurred from N1 to N2 with the normal delivery method. On the other hand, the olfactory dosage increased with the valve dilation for all cases herein, which was 0.43%, 0.63%, and 0.79% in N0, N1, and N2, respectively. Further increased olfactory DFs were achieved using the bidirectional method, with the olfactory DF being 2.76%, 2.94%, and 3.48% in N0, N1, and N2, respectively. The optimal olfactory DF with bidirectional delivery (i.e., 3.48% in N2) was around four times the unilateral olfactory DF in N2 (0.79%) and eight times that in N0 (0.43%).

25.3.4.4 Targeted Olfactory Delivery Using Magnetic Guidance

To evaluate whether it is practical to use magnetic field to guide ferromagnetic particles to the olfactory region [99, 119], a two-plate channel with a 10-mm height and 150-mm length was tested with permanent magnets above the channel (Fig. 25.15a). The particle density was 1500 kg/m^3 as a mixture of drug and iron nanoparticles, the relative magnetic permeability of the particles was 700, and the particle size was 15 μm. Without a magnetic field, particles followed parabolic paths due to the gravity (Fig. 25.15a). By imposing an appropriate magnetic field, particle motions could be controlled, for instance, to move horizontally versus settling vertically (layout 1). Particles can also be guided to a specific site by imposing a strong attraction force locally (layout 2). The magnet field was nonuniform; it was

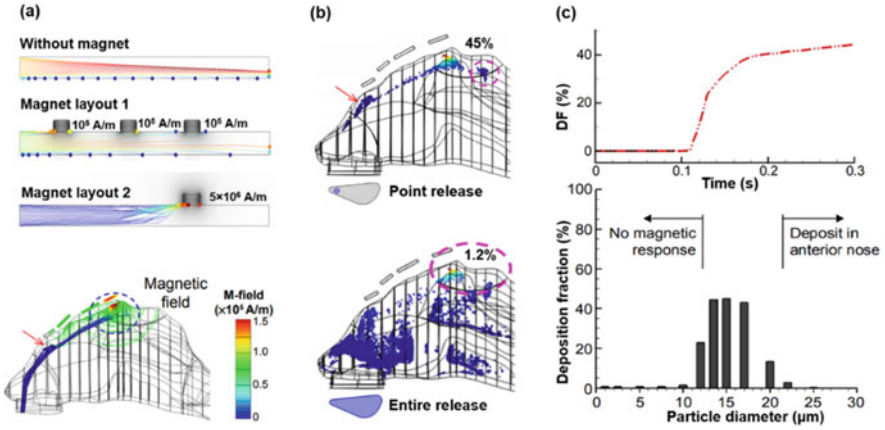


Fig. 25.15 Delivering magnetic particles to the olfactory region: (**a**) feasibility study in a two-plate channel, (**b**) point release of particles versus releasing particles from the entire nostril, and (**c**) olfactory dose as a function of the particle size

stronger near the magnet bars and decayed quickly away from the magnet bars. It was demonstrated that by releasing particles from a specific point (point release) and imposing a proper magnetic strength, the olfactory dose could be as high as 45% (Fig. 25.15b, c). By comparison, only 1.2% of particles could reach the olfactory region if being released from the entire nostril, even with an optimal magnetic strength (Fig. 25.15b).

The delivery efficiency of magnetic particles to the olfactory region is highly sensitive to the particle size. The most effective olfactory delivery occurred for particles of 10–20 μm. Decreased olfactory doses occurred for particles $\leq$10 μm or $\geq$20 μm. This is because that the particles $\leq$10 μm have low response to magnetic force, while particles $\geq$20 μm will be filtered out by the nasal valve from inertia impaction. Note that the magnetic force experienced by a particle varies with the third power of the particle size (d_p^3), in comparison to the first power of the magnetic intensity. The optimal olfactory dosimetry was shown to be 13–17 μm [99, 119].

25.3.4.5 Pulsating Aerosol Delivery to Maxillary Sinuses

The delivery efficiency of the drugs to the paranasal sinuses is unacceptably low at this stage. Pulsating flows have been shown to improve drug delivery into these secluded cavities due to flow-sinus resonance, but a manageable control on the delivery efficiency is still impractical. The maxillary sinuses are located peripherally at both sides of the nose and connected to the middle meatus through an ostium. To gain a better understanding of the pulsating aerosol delivery to the sinus, the variation of resonance frequency and sinus deposition with the ostium size, sinus shape, and pulsation frequency were systemically investigated [120]. The sinus model was built based on MRI head scans of a 53-year-old healthy male and connected to an existing nasal cavity model (Fig. 25.16a, b). To evaluate the geometrical effect of the sinus, the shape and size of the original sinus were modified using HyperMorph 10.0 (Troy, MI). Figure 25.16a shows the 3D-printed hollow casts of the sinuses with a 4-mm wall thickness (Fig. 25.16a).

To find the resonance frequencies of the nose-sinus geometry, a sampling point was specified in the ostium that connected the nasal cavity and the sinus. The variation of pressure (dashed line) and local acceleration (solid line) versus the input frequencies are shown in Fig. 25.16c for the sampling point. Two resonant frequencies are identified: 545 and 2260 Hz. Figure 25.16d shows the COMSOL-predicted sinus deposition in comparison to in vitro tests. As expected, the optimal sinus DF occurs at 545 Hz (resonant frequency) and decreases as the pulsating frequency drifts away from 545 Hz. The sinus DF changes insignificantly when varying the sinus dimensions as in this study.

Figure 25.16e shows the acoustic pressures in the nose-sinus ($D = 3$ mm, $L = 3$ mm, and $\mathrm{Vol}/\mathrm{Vol}_0 = 1$) at 545 Hz. A large pressure gradient was noted across the ostium. Figure 25.16f shows the distribution of the particle deposition within the sinus after 1 minute's pulsating drug delivery. Even though at a small amount, some particles indeed enter the sinus driven by the oscillating pressure waves (Fig. 25.16f), which have the largest amplitudes at the resonant frequency. Particles remaining inside the nasal passage are also dispersed by the acoustic waves

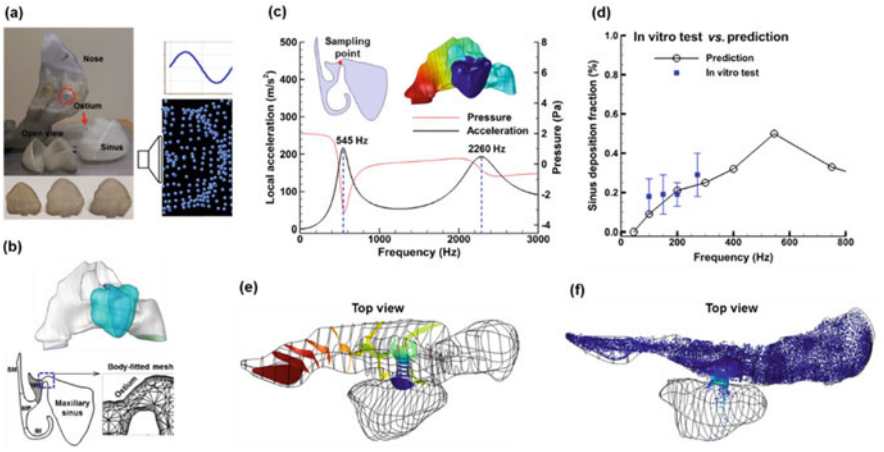


Fig. 25.16 Pulsating aerosol delivery to maxillary sinuses: (**a**) hollow casts of nose-sinus models, (**b**) computational model with near-wall prismatic mesh, (**c**) resonance occurs at 545 Hz, (**d**) comparison between in vitro measurements and numerical predictions, (**e**) acoustic pressure, and (**f**) sinus deposition distribution

(Fig. 25.16f). In light of the extremely low sinus dose at the current stage, further studies are needed in harnessing the pulsating aerosols. With a proper sound frequency and amplitude, it is promising to control the drug delivery to the paranasal sinuses or the olfactory region and achieve clinically relevant doses.

25.3.5 Computational Fluid-Particle Dynamics for COVID-19: Effect of Mask-Wearing

The year 2020 has witnessed several surges of COVID-19 pandemic in the USA and around the world. This pandemic has been caused by the SARS-CoV-2 virus, or coronavirus, and has been demonstrated to transmit from person to person through respiratory droplets and small particles, which are produced when an infected person speaks, breathes, coughs, or sneezes. These aerosols can be breathed into the respiratory tract and lead to infection. Community spread of COVID-19 has been reported frequently and is the major reason for COVID-transmission. Considering that the generation and transport of respiratory droplets share the same mechanisms of pharmaceutical aerosols, extensive studies have been undertaken since the pandemic to understand better the transmission routes and curb the transmission momentum. The example presented below demonstrated how computer-aided modeling and simulations added to our understanding of the effects of wearing a facemask on protection efficiencies.

In addition to social distancing and frequent handwashing, facemask-wearing has been an effective preventive measure to reduce viral community transmission. Wearing a mask both lower the inhalation of ambient virus-laden droplets and

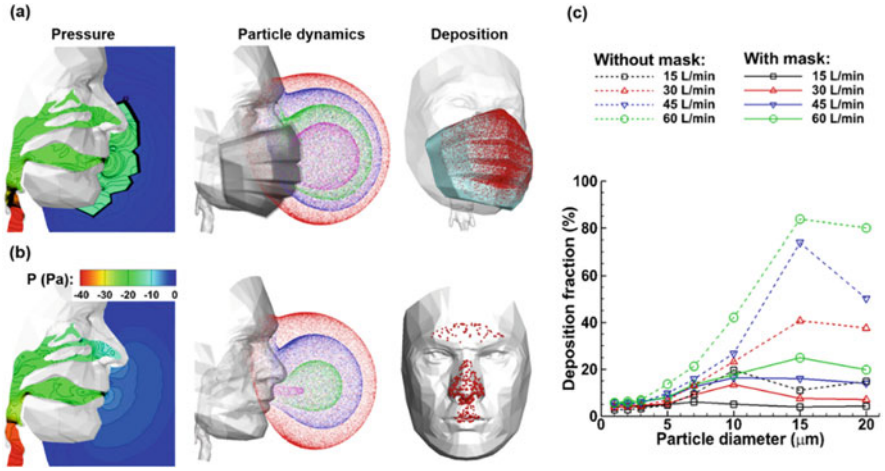


Fig. 25.17 Mask-wearing effects on respiration dynamics: (**a**) with a surgical mask, (**b**) without a mask, and (**c**) comparison of deposition of airborne aerosols with versus without a mask for different particle sizes and inhalation flow rates

minimize the exhalation of virus-laden droplets from a COVID patient [121, 122]. However, when talking about the facemask protective effectiveness, we often refer to the filtration efficiency of the mask material (like N95 being 95% effective), not the effectiveness it protects people from airborne bacteria/viruses. Till now, we do not have quantitative data on mask protection efficiencies when we wear them, which are more relevant to evaluate the risk and curb the COVID transmission [123, 124]. In this instance, we aimed to quantify the actual protection efficiencies to the upper airway when wearing a three-layer surgical mask [125].

To this aim, we developed a physiologically accurate model of a subject wearing a pleated surgical mask (Fig. 25.17a, b). Using numerical methods, we tracked aerosols through the facemask and studies their behavior and fates after crossing the facemask (Fig. 25.17a). Among these aerosol droplets, some will deposit on the face and the others will enter the airway. We further follow these droplets into the respiratory airway and find out where they will deposit, for instance, in the nose, pharynx, or into the lung. The protection efficiency of different regions of the airway can then be calculated (Fig. 25.17c).

The results may be surprising to many. Wearing an old or low-filtration mask (say $<35\%$) can inhale more aerosol particles that are smaller than 2.5 μm ($PM_{2.5}$) [125]. Figure 25.17c compares the deposition fraction of airborne aerosol between wearing a 65%-filtration mask and with no mask. While good protection by mask was observed for particles >10 μm, equivalent deposition of 0–2.5 μm was predicted with and without a mask. It appears intuitive to premise that with a mask, regardless old or new, will always be better off than without a facemask. However, this premise is valid only for aerosols ≥ 5 μm, not for $PM_{2.5}$. The modified (slower) airflow when wearing a mask helps the particles to be inhaled into the nose. The filtration efficiency of a three-layer surgical mask can be any value between 25% (used) to

95% (new). While wearing a 95% mask will protect well, wearing a 25%-filtration mask can be worse than without [125]. Moreover, for environments with high concentrations of virus-laden $PM_{2.5}$ aerosols, such as COVID critical care units or COVID-specialized hospitals, face-covering with only a three-layer surgical mask would not provide enough protection.

25.4 Conclusion

In this chapter, we presented the fundamental theories in respiratory aerosol dynamics and demonstrated their usages through a series of efforts in our lab aiming to improve the physical realism of respiratory airway models. Both numerical and experimental tests were performed to this aim by exploring inhalation dosimetry in different regions of the airway (i.e., the mouth-throat, nose, and lung) and using different delivery techniques (i.e., normal vs. bidirectional nasal delivery, aerodynamic vs. magnetophoretic vs. acoustophoretic guidance, etc.). Specifically, computer-aided image-based design of the airway models opened a new door to the pharmaceutical industries to test the delivery efficiency and variability that was not even feasible less than 20 years ago. With anatomically accurate airway models and complementary numerical simulations, a better understanding has been attained in existing inhalation devices and delivery protocols, and improvements and optimization have been achieved in some cases. Furthermore, new devices or protocols can be designed for personalized diagnosis and targeted drug delivery.

25.5 Credible Online Resources for Further Reading

URL	Category of source	What to read or refer
https://www.ansys.com/products/fluids/ansys-fluent	Private company	Software to simulate airflow and particle deposition
https://www.ansys.com/services/training-center/fluids/introduction-to-ansys-icem-cfd	Private company	Tutorial for developing airway models, set up numerical simulations, and visualize results
https://www.altair.com/hypermesh	Private company	Software to change model morphology
https://www.comsol.com	Private company	Software to simulate Multiphysics phenomenon
https://www.comsol.com/models	Private company	Learn more about how to set up COMSOL simulations
https://www.slicer.org/	Open source	Software to segment medical images
https://www.slicer.org/wiki/Documentation/4.10/Training	Open source	Learn more about how to Slicer to process medical images
https://www.materialise.com	Private company	Software for image segmentation

References

1. Xi J, Longest PW (2007) Transport and deposition of micro-aerosols in realistic and simplified models of the oral airway. Ann Biomed Eng 35:560–581. https://doi.org/10.1007/s10439-006-9245-y
2. Xi J, Yuan JE, Zhang Y, Nevorski D, Wang Z, Zhou Y (2016) Visualization and quantification of nasal and olfactory deposition in a sectional adult nasal airway cast. Pharm Res 33:1527–1541. https://doi.org/10.1007/s11095-016-1896-2
3. Darquenne C, Fleming JS, Katz I, Martin AR, Schroeter J, Usmani OS et al (2016) Bridging the gap between science and clinical efficacy: physiology, imaging, and modeling of aerosols in the lung. J Aerosol Med 29:107–126. https://doi.org/10.1089/jamp.2015.1270
4. Xi J, Talaat M, Si XA, Han P, Dong H, Zheng S (2020) Alveolar size effects on nanoparticle deposition in rhythmically expanding-contracting terminal alveolar models. Comput Biol Med 121:103791. https://doi.org/10.1016/j.compbiomed.2020.103791
5. Talaat K, Xi J, Baldez P, Hecht A (2019) Radiation dosimetry of inhaled radioactive aerosols: CFPD and MCNP transport simulations of radionuclides in the lung. Sci Rep 9:1–21. https://doi.org/10.1038/s41598-019-54040-1
6. Zhang Z, Martonen TB (1997) Deposition of ultrafine aerosols in human tracheobronchial airways. Inhal Toxicol 9:99–110. https://doi.org/10.1080/089583797198295
7. Inthavong K, Tian ZF, Li HF, Tu JY, Yang W, Xue CL et al (2006) A numerical study of spray particle deposition in a human nasal cavity. Aerosol Sci Technol 40:1034–U1033. https://doi.org/10.1080/02786820600924978
8. Pharmacopeia US (2005) Physical tests and determinations: aerosols, nasal sprays, metered-dose inhalers, and dry powder inhalers. In: United States Pharmacopeia First Supplement, 28-NF, General Chapter (601), vol 28. United States Pharmacopeial Convention, Rockville, pp 3298–3316
9. Zhou Y, Sun J, Cheng Y-S (2011) Comparison of deposition in the USP and physical mouth-throat models with solid and liquid particles. J Aerosol Med Pulm Drug Deliv 24:277–284. https://doi.org/10.1089/jamp.2011.0882
10. Clark A, Newman S, Dasovich N (1998) Mouth and oropharyngeal deposition of pharmaceutical aerosols. J Aerosol Med 11:S116–S120. https://doi.org/10.1089/jam.1998.11.Suppl_1.S-116
11. Newman SP (1998) How well do in vitro particle size measurements predict drug delivery in vivo? J Aerosol Med Deposit Clear Eff Lung 11:S97–S104
12. Ali M, Mazumder MK, Martonen TB (2009) Measurements of electrodynamic effects on the deposition of MDI and DPI aerosols in a replica cast of human oral-pharyngeal-laryngeal airways. J Aerosol Med 22:35–44. https://doi.org/10.1089/jamp.2007.0637
13. Longest PW, Tian G, Walenga RL, Hindle M (2012) Comparing MDI and DPI aerosol deposition using In vitro experiments and a new stochastic individual path (SIP) model of the conducting airways. Pharm Res 29:1670–1688. https://doi.org/10.1007/s11095-012-0691-y
14. Burnell PKP, Asking L, Borgstrom L, Nichols SC, Olsson B, Prime D et al (2007) Studies of the human oropharyngeal airspaces using magnetic resonance imaging IV—the oropharyngeal retention effect for four inhalation delivery systems. J Aerosol Med Deposit Clear Eff Lung 20:269–281. https://doi.org/10.1089/jam.2007.0566
15. Cheng YS, Zhou Y, Chen BT (1999) Particle deposition in a cast of human oral airways. Aerosol Sci Technol 31:286–300. https://doi.org/10.1080/027868299304165
16. Golshahi L, Noga ML, Finlay WH (2012) Deposition of inhaled micrometer-sized particles in oropharyngeal airway replicas of children at constant flow rates. J Aerosol Sci 49:21–31. https://doi.org/10.1016/j.jaerosci.2012.03.001
17. Grgic B, Finlay WH, Burnell PKP, Heenan AF (2004) In vitro intersubject and intrasubject deposition measurements in realistic mouth-throat geometries. J Aerosol Sci 35:1025–1040. https://doi.org/10.1016/j.jaerosci.2004.03.003

18. Heenan AF, Finlay WH, Grgic B, Pollard A, Burnell PKP (2004) An investigation of the relationship between the flow field and regional deposition in realistic extra-thoracic airways. J Aerosol Sci 35:1013–1023. https://doi.org/10.1016/j.jaerosci.2004.03.004
19. Cheng KH, Cheng YS, Yeh HC, Swift DL (1997) Measurements of airway dimensions and calculation of mass transfer characteristics of the human oral passage. J Biomech Eng 119: 476–482. https://doi.org/10.1115/1.2798296
20. Borgstrom L, Olsson B, Thorsson L (2006) Degree of throat deposition can explain the variability in lung deposition of inhaled drugs. J Aerosol Med 19:473–483. https://doi.org/10.1089/jam.2006.19.473
21. De Backer JW, Vos WG, Burnell P, Verhulst SL, Salmon P, De Clerck N et al (2009) Study of the variability in upper and lower airway morphology in Sprague-Dawley rats using modern micro-CT scan-based segmentation techniques. Anat Rec 292:720–727. https://doi.org/10.1002/ar.20877
22. Garcia GJM, Schroeter JD, Segal RA, Stanek J, Foureman GL, Kimbell JS (2009) Dosimetry of nasal uptake of water-soluble and reactive gases: a first study of interhuman variability. Inhal Toxicol 21:607–618. https://doi.org/10.1080/08958370802320186
23. Garcia GJM, Tewksbury EW, Wong BA, Kimbell JS (2009) Interindividual variability in nasal filtration as a function of nasal cavity geometry. J Aerosol Med Pulm Drug Deliv 22: 139–155. https://doi.org/10.1089/jamp.2008.0713
24. Hofmann W, Asgharian B, Winkler-Heil R (2002) Modeling intersubject variability of particle deposition in human lungs. J Aerosol Sci 33:219–235. https://doi.org/10.1016/S0021-8502(01)00167-7
25. Schroeter JD, Garcia GJM, Kimbell JS (2010) A computational fluid dynamics approach to assess interhuman variability in hydrogen sulfide nasal dosimetry. Inhal Toxicol 22:277–286. https://doi.org/10.3109/08958370903278077
26. Thekedar B, Oeh U, Szymczak W, Hoeschen C, Paretzke HG (2011) Influences of mixed expiratory sampling parameters on exhaled volatile organic compound concentrations. J Breath Res 5:016001. https://doi.org/10.1088/1752-7155/5/1/016001
27. Vijverberg SJH, Koenderman L, Koster ES, van der Ent CK, Raaijmakers JAM, Maitland-van der Zee AH (2011) Biomarkers of therapy responsiveness in asthma: pitfalls and promises. Clin Exp Allergy 41:615–625. https://doi.org/10.1111/j.1365-2222.2011.03694.x
28. Longest PW, Xi J, Hindle M (2013) Effective delivery of nanoparticles and micrometer-sized pharmaceutical aerosols to the lung through enhanced condensational growth. U.S. patent number: 5522385
29. Xi J, Kim J, Si XA, Zhou Y (2013) Diagnosing obstructive respiratory diseases using exhaled aerosol fingerprints: a feasibility study. J Aerosol Sci 64:24–36. https://doi.org/10.1016/j.jaerosci.2013.06.003
30. Xi J, Yang T (2019) Variability in oropharyngeal airflow and aerosol deposition due to changing tongue positions. J Drug Deliv Sci Technol 49:674–682. https://doi.org/10.1016/j.jddst.2019.01.006
31. Xi J, Si XA, Kim JW (2015) Characterizing respiratory airflow and aerosol condensational growth in children and adults using an imaging-CFD approach. In: Heat transfer and fluid flow in biological processes. Academic, pp 125–155
32. Xi J, Kim J, Si XA (2016) Effects of nostril orientation on airflow dynamics, heat exchange, and particle depositions in human noses. Eur J Mech B/Fluids 55:215–228
33. Cheng KH, Cheng YS, Yeh HC, Swift DL (1997) An experimental method for measuring aerosol deposition efficiency in the human oral airway. Am Ind Hyg Assoc J 58:207–213. https://doi.org/10.1080/15428119791012856
34. Stapleton KW, Guentsch E, Hoskinson MK, Finlay WH (2000) On the suitability of k-epsilon turbulence modeling for aerosol deposition in the mouth and throat: a comparison with experiment. J Aerosol Sci 31:739–749. https://doi.org/10.1016/S0021-8502(99)00547-9
35. DeHaan WH, Finlay WH (2004) Predicting extrathoracic deposition from dry powder inhalers. J Aerosol Sci 35:309–331. https://doi.org/10.1016/j.jaerosci.2003.09.002

36. Xi J, Longest PW (2008) Effects of oral airway geometry characteristics on the diffusional deposition of inhaled nanoparticles. J Biomech Eng 130:011008
37. Cheng YS, Yazzie D, Gao J, Muggli D, Etter J, Rosenthal GJ (2003) Particle characteristics and lung deposition patterns in a human airway replica of a dry powder formulation of polylactic acid produced using supercritical fluid technology. J Aerosol Med Deposit Clear Eff Lung 16:65–73. https://doi.org/10.1089/089426803764928374
38. Cheng YS, Fu CS, Yazzie D, Zhou Y (2001) Respiratory deposition patterns of salbutamol pMDI with CFC and HFA-134a formulations in a human airway replica. J Aerosol Med 14: 255–266. https://doi.org/10.1089/08942680152484180
39. Xi J, Berlinski A, Zhou Y, Greenberg B, Ou X (2012) Breathing resistance and ultrafine particle deposition in nasal–laryngeal airways of a newborn, an infant, a child, and an adult. Ann Biomed Eng 40:2579–2595. https://doi.org/10.1007/s10439-012-0603-7
40. Segal R, Kepler GM, Kalisak DL, Richardson RB, Kimbell JS (2003) Effects of differences in nasal anatomy on airflow distribution: a comparison of three individuals. Toxicol Sci 72:37–38. https://doi.org/10.1007/s10439-008-9556-2
41. Segal RA, Kepler GM, Kimbell JS (2008) Effects of differences in nasal anatomy on airflow distribution: a comparison of four individuals at rest. Ann Biomed Eng 36:1870–1882. https://doi.org/10.1007/s10439-008-9556-2
42. Miyawaki S, Hoffman EA, Wenzel SE, Lin C-L (2019) Aerosol deposition predictions in computed tomography-derived skeletons from severe asthmatics: a feasibility study. Clin Biomech 66:81–87. https://doi.org/10.1016/j.clinbiomech.2017.10.020
43. Choi J, Hoffman EA, Lin C-L, Milhem MM, Tessier J, Newell JD Jr (2017) Quantitative computed tomography determined regional lung mechanics in normal nonsmokers, normal smokers and metastatic sarcoma subjects. PLoS One 12:e0179812–e0179812. https://doi.org/10.1371/journal.pone.0179812
44. Miyawaki S, Hoffman EA, Lin C-L (2017) Numerical simulations of aerosol delivery to the human lung with an idealized laryngeal model, image-based airway model, and automatic meshing algorithm. Comput Fluids 148:1–9. https://doi.org/10.1016/j.compfluid.2017.02.008
45. Wu D, Tawhai MH, Hoffman EA, Lin C-L (2014) A numerical study of heat and water vapor transfer in MDCT-based human airway models. Ann Biomed Eng 42:2117–2131. https://doi.org/10.1007/s10439-014-1074-9
46. Lambert AR, O'Shaughnessy PT, Tawhai MH, Hoffman EA, Lin C-L (2011) Regional deposition of particles in an image-based airway model: large-Eddy simulation and left-right lung ventilation asymmetry. Aerosol Sci Technol 45:11–25. https://doi.org/10.1080/02786826.2010.517578
47. Kleinstreuer C, Zhang Z (2003) Laminar-to-turbulent fluid-particle flows in a human airway model. Int J Multiph Flow 29:271–289. https://doi.org/10.1016/S0301-9322(02)00131-3
48. Zhang Y, Finlay WH, Matida EA (2004) Particle deposition measurements and numerical simulations in a highly idealized mouth-throat. J Aerosol Sci 35:789–803. https://doi.org/10.1016/j.jaerosci.2003.12.006
49. Delvadia RR, Longest PW, Hindle M, Byron PR (2013) In vitro tests for aerosol deposition. III: effect of inhaler insertion angle on aerosol deposition. J. Aerosol Med 26: 145–156. https://doi.org/10.1089/jamp.2012.0989
50. Delvadia R, Hindle M, Longest PW, Byron PR (2013) In vitro tests for aerosol deposition II: IVIVCs for different dry powder inhalers in normal adults. J Aerosol Med 26:138–144. https://doi.org/10.1089/jamp.2012.0975
51. Zhou Y, Guo M, Xi J, Irshad H, Cheng Y-S (2014) Nasal deposition in infants and children. J Aerosol Med 27:110–116. https://doi.org/10.1089/jamp.2013.1039
52. Zhou Y, Xi J, Simpson J, Irshad H, Cheng Y-S (2013) Aerosol deposition in a nasopharyngolaryngeal replica of a 5-year-old child. Aerosol Sci Technol 47:275–282. https://doi.org/10.1080/02786826.2012.749341

53. Valentine R, Athanasiadis T, Thwin M, Singhal D, Weitzel EK, Wormald PJ (2008) A prospective controlled trial of pulsed nasal nebulizer in maximally dissected cadavers. Am J Rhinol Allergy 22:390–394. https://doi.org/10.2500/ajr.2008.22.3191
54. Moller W, Schuschnig U, Celik G, Munzing W, Bartenstein P, Haussinger K et al (2013) Topical drug delivery in chronic rhinosinusitis patients before and after sinus surgery using pulsating aerosols. PLoS One 8:e74991. https://doi.org/10.1371/journal.pone.0074991
55. Djupesland PG, Skretting A, Winderen M, Holand T (2006) Breath actuated device improves delivery to target sites beyond the nasal valve. Laryngoscope 116:466–472. https://doi.org/10.1097/01.mlg.0000199741.08517.99
56. Washington N, McGlashan JA, Jackson SJ, Bush D, Pitt KG, Rawlins DA et al (2000) The effect of nasal patency on the clearance of radiolabeled saline in healthy volunteers. Pharm Res 17:733–736. https://doi.org/10.1023/a:1007590501540
57. Suman JD, Laube BL, Lin TC, Brouet G, Dalby R (2002) Validity of in vitro tests on aqueous spray pumps as surrogates for nasal deposition. Pharm Res 19:1–6. https://doi.org/10.1023/a:1013643912335
58. Guo Y, Laube B, Dalby R (2005) The effect of formulation variables and breathing patterns on the site of nasal deposition in an anatomically correct model. Pharm Res 22:1871–1878. https://doi.org/10.1007/s11095-005-7391-9
59. Laube BL, Sharpless G, Shermer C, Nasir O, Sullivan V, Powell K (2010) Deposition of albuterol aerosol generated by pneumatic nebulizer in the Sophia Anatomical Infant Nose-Throat (SAINT) model. Pharm Res 27:1722–1725. https://doi.org/10.1007/s11095-010-0171-1
60. Fleming JS, Conway JH, Bolt L, Holgate ST (2003) A comparison of planar scintigraphy and SPECT measurement of total lung deposition of inhaled aerosol. J Aerosol Med 16:9–19. https://doi.org/10.1089/089426803764928310
61. Bondesson E, Bengtsson T, Borgstrom L, Nilsson LE, Norrgren K, Trofast E et al (2003) Planar gamma scintigraphy—points to consider when quantifying pulmonary dry powder aerosol deposition. Int J Pharm 258:227–240. https://doi.org/10.1016/s0378-5173(02)00579-3
62. Kundoor V, Dalby RN (2010) Assessment of nasal spray deposition pattern in a silicone human nose model using a color-based method. Pharm Res 27:30–36. https://doi.org/10.1007/s11095-009-0002-4
63. Kundoor V, Dalby RN (2011) Effect of formulation- and administration-related variables on deposition pattern of nasal spray pumps evaluated using a nasal cast. Pharm Res 28:1895–1904. https://doi.org/10.1007/s11095-011-0417-6
64. Courrier HM, Butz N, Vandamme TF (2002) Pulmonary drug delivery systems: recent developments and prospects. Crit Rev Ther Drug Carrier Syst 19:425–498. https://doi.org/10.1615/CritRevTherDrugCarrierSyst.v19.i45.40
65. Labiris NR, Dolovich MB (2003) Pulmonary drug delivery. Part I: physiological factors affecting therapeutic effectiveness of aerosolized medications. Br J Clin Pharmacol 56:588–599. https://doi.org/10.1046/j.1365-2125.2003.01892.x
66. Chan HK, Chew NY (2003) Novel alternative methods for the delivery of drugs for the treatment of asthma. Adv Drug Deliv Rev 55:793–805. https://doi.org/10.1016/s0169-409x(03)00078-4
67. Agertoft L, Pedersen S, Nikander K (1999) Drug delivery from the Turbuhaler and Nebuhaler pressurized metered dose inhaler to various age groups of children with asthma. J Aerosol Med 12:161–169. https://doi.org/10.1089/jam.1999.12.161
68. Gibson PG, Wlodarczyk JH, Borgas T (1995) Drug delivery in asthma: a comparison of spacers with a jet nebuliser. Aust N Z J Med 25:324–325. https://doi.org/10.1111/j.1445-5994.1995.tb01897.x
69. Si XA, Xi J, Kim J, Zhou Y, Zhong H (2013) Modeling of release position and ventilation effects on olfactory aerosol drug delivery. Respir Physiol Neurobiol 186:22–32. https://doi.org/10.1016/j.resp.2012.12.005

70. Kim J, Xi J, Si X, Berlinski A, Su WC (2014) Hood nebulization: effects of head direction and breathing mode on particle inhalability and deposition in a 7-month-old infant model. J Aerosol Med Pulm Drug Deliv 27:209–218. https://doi.org/10.1089/jamp.2013.1051
71. Xi J, Kim J, Si XA, Corley RA, Kabilan S, Wang S (2015) CFD modeling and image analysis of exhaled aerosols due to a growing bronchial tumor: towards non-invasive diagnosis and treatment of respiratory obstructive diseases. Theranostics 5:443–455. https://doi.org/10.7150/thno.11107
72. Easley SK, Pal S, Tomaszewski PR, Petrella AJ, Rullkoetter PJ, Laz PJ (2007) Finite element-based probabilistic analysis tool for orthopaedic applications. Comput Methods Programs Biomed 85:32–40. https://doi.org/10.1016/j.cmpb.2006.09.013
73. Laz PJ, Browne M (2010) A review of probabilistic analysis in orthopaedic biomechanics. Proc Inst Mech Eng H 224:927–943
74. Lu J, Xi J, Langenderfer JE (2017) Sensitivity analysis and uncertainty quantification in pulmonary drug delivery of orally inhaled pharmaceuticals. J Pharm Sci 106:3303–3315. https://doi.org/10.1016/j.xphs.2017.06.011
75. Guo C, Doub WH, Kauffman JF (2010) Propagation of uncertainty in nasal spray in vitro performance models using Monte Carlo simulation: part II. Error propagation during product performance modeling. J Pharm Sci 99:3572–3578. https://doi.org/10.1002/jps.22094
76. Guilmette RA, Wicks JD, Wolff RK (1989) Morphometry of human nasal airways in vivo using Magnetic Resonance Imaging. J Aerosol Med 2:365–377. https://doi.org/10.1089/jam.1989.2.365
77. Guilmette RA, Cheng YS, Yeh HC, Swift DL (1994) Deposition of 0.005–12 micrometer monodisperse particls in a computer-milled, MRI-based nasal airway replica. Inhal Toxicol 6 (Suppl 1):395–399
78. Cheng KH, Cheng YS, Yeh HC, Swift DL (1995) Deposition of ultrafine aerosols in the head airways during natural breathing and during simulated breath-holding using replicate human upper airway casts. Aerosol Sci Technol 23:465–474. https://doi.org/10.1080/02786829508965329
79. Kelly JT, Asgharian B, Kimbell JS, Wong B (2004) Particle depositon in human nasal airway replicas manufactured by different methds. Part II: ultrafine particles. Aerosol Sci Technol 38:1072–1079. https://doi.org/10.1080/027868290883432
80. Kelly JT, Asgharian B, Kimbell JS, Wong B (2004) Particle depositon in human nasal airway replicas manufactured by different methds. Part I: inertial regime particles. Aerosol Sci Technol 38:1063–1071. https://doi.org/10.1080/027868290883360
81. Schroeter JD, Kimbell JS, Asgharian B (2006) Analysis of particle deposition in the turbinate and olfactory regions using a human nasal computational fluid dynamics model. J Aerosol Med 19:301–313. https://doi.org/10.1089/jam.2006.19.301
82. Shi H, Kleinstreuer C, Zhang Z (2006) Laminar airflow and nanoparticle or vapor deposition in a human nasal cavity model. J Biomech Eng 128:697–706. https://doi.org/10.1115/1.2244574
83. Xi J, Longest PW (2008) Numerical predictions of submicrometer aerosol deposition in the nasal cavity using a novel drift flux approach. Int J Heat Mass Transf 51:5562–5577. https://doi.org/10.1016/j.ijheatmasstransfer.2008.04.037
84. Si X, Xi J, Kim J (2013) Effect of laryngopharyngeal anatomy on expiratory airflow and submicrometer particle deposition in human extrathoracic airways. Open J Fluid Dyn 3:286–301. https://doi.org/10.4236/ojfd.2013.34036
85. Kim J, Xi J, Si XA (2013) Dynamic growth and deposition of hygroscopic aerosols in the nasal airway of a 5-year-old child. Int J Numer Method Biomed Eng 29:17–39. https://doi.org/10.1002/cnm.2490
86. Cohen BS, Asgharian B (1990) Deposition of ultrafine particles in the upper airways: an empirical analysis. J Aerosol Sci 21:789–797. https://doi.org/10.1016/0021-8502(90)90044-X

87. Xi J, Si X, Kim J, Su G, Dong H (2014) Modeling the pharyngeal anatomical effects on breathing resistance and aerodynamically generated sound. Med Biol Eng Comput 52:567–577
88. Morsi SA, Alexander AJ (1972) An investigation of particle trajectories in two-phase flow systems. J Fluid Mech 55:193–208. https://doi.org/10.1016/j.mvr.2010.05.002
89. Xi J, Longest PW (2009) Characterization of submicrometer aerosol deposition in extrathoracic airways during nasal exhalation. Aerosol Sci Technol 43:808–827. https://doi.org/10.1007/s10439-008-9552-6
90. Ribeiro P, Si X (2012) Section 2: electric and magnetic circuit. In: Beaty HW, Fink DG (eds) Standard handbook for electrical engineering, 16th edn. McGraw Hill, pp 2.1–2.58
91. Gao Y, Jian YC, Zhang LF, Huang JP (2007) Magnetophoresis of nonmagnetic particles in ferrofluids. J Phys Chem C 111:10785–10791. https://doi.org/10.1021/jp0705673
92. Settnes M, Bruus H (2012) Forces acting on a small particle in an acoustical field in a viscous fluid. Phys Rev E 85:016327
93. Xi J, Si X, Kim JW, Berlinski A (2011) Simulation of airflow and aerosol deposition in the nasal cavity of a 5-year-old child. J Aerosol Sci 42:156–173
94. Xi J, Si X, Longest WJP (2014) Electrostatic charge effects on pharmaceutical aerosol deposition in human Nasal–Laryngeal airways. Pharmaceutics 6:26–35
95. Rygg A, Longest W (2016) Absorption and clearance of pharmaceutical aerosols in the human nose: development of a CFD model. J Aerosol Med 29:416–431. https://doi.org/10.1089/jamp.2015.1252
96. Graham LJ, Taillon R, Mullin J, Wigle T (2010) Pharmaceutical process/equipment design methodology case study: cyclone design to optimize spray-dried-particle collection efficiency. Comput Chem Eng 34:1041–1048. https://doi.org/10.1016/j.compchemeng.2010.04.004
97. Si XA, Xi J (2016) Modeling and simulations of olfactory drug delivery with passive and active controls of nasally inhaled pharmaceutical aerosols. J Vis Exp e53902. https://doi.org/10.3791/53902
98. Xi J, Si X, Gaide R (2014) Electrophoretic particle guidance significantly enhances olfactory drug delivery: a feasibility study. PLoS One 9:e86593. https://doi.org/10.1371/journal.pone.0086593
99. Xi J, Zhang Z, Si XA, Yang J, Deng W (2016) Optimization of magnetophoretic-guided drug delivery to the olfactory region in a human nose model. Biomech Model Mechanobiol 15:877–891. https://doi.org/10.1007/s10237-015-0730-9
100. Xi J, Yuan JE, Yang M, Si X, Zhou Y, Cheng Y-S (2016) Parametric study on mouth–throat geometrical factors on deposition of orally inhaled aerosols. J Aerosol Sci 99:94–106. https://doi.org/10.1016/j.jaerosci.2016.01.014
101. Preacher KJ, Curran PJ, Bauer DJ (2006) Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. J Educ Behav Stat 31:437–448. https://doi.org/10.3102/10769986031004437
102. Xi J, Yang T, Talaat K, Wen T, Zhang Y, Klozik S et al (2018) Visualization of local deposition of nebulized aerosols in a human upper respiratory tract model. J Vis 21:225–237. https://doi.org/10.1007/s12650-017-0456-0
103. Wells B, Guest N, Martinez J, Hakizimana E, Si X, Su W et al (2018) Visualization and quantification of nebulized aerosol deposition in mouth-lung casts under healthy and abnormal breathing conditions. Lungs Breath 2:1–6. https://doi.org/10.15761/LBJ.1000127
104. Bernasconi M, Ploysongsang Y, Gottfried SB, Milic-Emili J, Rossi A (1988) Respiratory compliance and resistance in mechanically ventilated patients with acute respiratory failure. Intensive Care Med 14:547–553. https://doi.org/10.1007/bf00263528
105. Storstein O, Field AS Jr, Massumi R, Gray FD Jr (1959) Airway resistance and lung compliance; the effects of histamine and hyperventilation in normal subjects and patients with chronic lung disease. Yale J Biol Med 31:387–396
106. Papandrinopoulou D, Tzouda V, Tsoukalas G (2012) Lung compliance and chronic obstructive pulmonary disease. Pulm Med 2012:542769. https://doi.org/10.1155/2012/542769

107. Mardia KV, Kent JT, Walder AN (1991) Statistical shape models in image analysis. Comput Sci Stat 550–557
108. Slice DE (2007) Geometric morphometrics. Annu Rev Anthropol 36:261–281. https://doi.org/10.1146/annurev.anthro.34.081804.120613
109. Fitzpatrick C, FitzPatrick D, Lee J, Auger D (2007) Statistical design of unicompartmental tibial implants and comparison with current devices. Knee 14:138–144. https://doi.org/10.1016/j.knee.2006.11.005
110. Castro-Mateos I, Pozo JM, Cootes TF, Wilkinson JM, Eastell R, Frangi AF (2014) Statistical shape and appearance models in osteoporosis. Curr Osteoporos Rep 12:163–173. https://doi.org/10.1007/s11914-014-0206-3
111. Maghari AYA, Liao IY, Belaton B (2012) Quantitative analysis on PCA-based statistical 3D face shape modeling. In: Computational modelling of objects represented in images: fundamentals, methods and applications Iii. pp 13–18
112. Hauser R, Smolinski J, Gos T (2005) The estimation of stature on the basis of measurements of the femur. Forensic Sci Int 147:185–190. https://doi.org/10.1016/j.forsciint.2004.09.070
113. Krishan K, Chatterjee PM, Kanchan T, Kaur S, Baryah N, Singh RK (2016) A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. Forensic Sci Int 261:165.e161–165.e168. https://doi.org/10.1016/j.forsciint.2016.02.007
114. Sella G, Hirsh AE (2005) The application of statistical physics to evolutionary biology. Proc Natl Acad Sci U S A 102:9541–9546. https://doi.org/10.1073/pnas.0501865102
115. Macdonald-Wallis C (2014) Statistical analysis of human growth and development. Int J Epidemiol 43:635–636. https://doi.org/10.1093/ije/dyu068
116. Xi J, Yuan J, Alshaiba M, Cheng D, Firlit Z, Johnson A, Nolan A, Su W (2016) Design and testing of electric-guided delivery of charged particles to the olfactory region: experimental and numerical studies. Curr Drug Deliv 13:265–274. https://doi.org/10.2174/1567201812666150909093050
117. Xi J, Wang Z, Nevorski D, White T, Zhou Y (2017) Nasal and olfactory deposition with normal and bidirectional intranasal delivery techniques: in vitro tests and numerical simulations. J Aerosol Med 30:118–131. https://doi.org/10.1089/jamp.2016.1295
118. Xi J, Wang Z, Si XA, Zhou Y (2018) Nasal dilation effects on olfactory deposition in unilateral and bi-directional deliveries: in vitro tests and numerical modeling. Eur J Pharm Sci 118:113–123. https://doi.org/10.1016/j.ejps.2018.03.027
119. Xi J, Zhang Z, Si X (2015) Improving intranasal delivery of neurological nanomedicine to the olfactory region using magnetophoretic guidance of microsphere carriers. Int J Nanomed 10:1211–1222. https://doi.org/10.2147/IJN.S77520
120. Xi J, Si XA, Peters S, Nevorski D, Wen T, Lehman M (2017) Understanding the mechanisms underlying pulsating aerosol delivery to the maxillary sinus: in vitro tests and computational simulations. Int J Pharm 520:254–266. https://doi.org/10.1016/j.ijpharm.2017.02.017
121. Gandhi M, Havlir D et al (2020) Open Forum Infect Dis 7:ofaa131. https://doi.org/10.1093/ofid/ofaa131
122. Eikenberry SE, Mancuso M, Iboi E, Phan T, Eikenberry K, Kuang Y et al (2020) To mask or not to mask: modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infect Dis Model 5:293–308. https://doi.org/10.1016/j.idm.2020.04.001
123. Asadi S, Cappa CD, Barreda S, Wexler AS, Bouvier NM, Ristenpart WD (2020) Efficacy of masks and face coverings in controlling outward aerosol particle emission from expiratory activities. Sci Rep 10:15665. https://doi.org/10.1038/s41598-020-72798-7
124. Liu Z, Yu D, Ge Y, Wang L, Zhang J, Li H et al (2019) Understanding the factors involved in determining the bioburdens of surgical masks. Ann Transl Med 7:754–754. https://doi.org/10.21037/atm.2019.11.91
125. Xi J, Si X, Nagarajan R (2020) Effects of mask-wearing on the inhalability and deposition of airborne SARS-CoV-2 aerosols in human upper airway. Phys Fluids 32:123312. https://doi.org/10.1063/5.0034580