Interdisciplinary Biotechnological Advances

Mithun Rudrapal
Johra Khan *Editors*

CADD and Informatics in Drug Discovery

Springer

Interdisciplinary Biotechnological Advances

This series is an authoritative resource that explains recent advances and emerging directions in biotechnology, reflecting the forefront of research clearly and reliably, without excessive hype. Each volume is written by authors with excellent reputations and acknowledged expertise in the topic under discussion. The volumes span the entire field from an interdisciplinary perspective, covering everything from biotechnology principles and methods to applications in areas including genetic engineering, transgenic plants and animals, environmental problems, genomics, proteomics, diagnosis of disease, gene therapy, and biomedicine. The significance of these applications for the achievement of UN Sustainable Development Goals is highlighted. The series will be highly relevant for Master's and PhD students in Biotechnology, Nanochemistry, Biochemical Engineering, and Microbiology, medical students, academic and industrial researchers, agricultural scientists, farmers, clinicians, industry personnel, and entrepreneurs.

Mithun Rudrapal • Johra Khan
Editors

CADD and Informatics in Drug Discovery

Springer

Editors
Mithun Rudrapal
Department of Pharmaceutical Sciences
Vignan's Foundation for Science,
Technology & Research
Guntur, Andhra Pradesh, India

Johra Khan
College of Applied Medical Sciences
Majmaah University
Majmaah, Saudi Arabia

ISSN 2730-7069 ISSN 2730-7077 (electronic)
Interdisciplinary Biotechnological Advances
ISBN 978-981-99-1315-2 ISBN 978-981-99-1316-9 (eBook)
https://doi.org/10.1007/978-981-99-1316-9

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

Preface

This book titled *CADD and Informatics in Drug Discovery* has been proposed to impart updated knowledge on recent advances in computational and bioinformatics tools/techniques and their practical applications in modern drug design and discovery programme. This book encompasses fundamental principles, advanced methodologies, and applications of various CADD approaches including several cutting-edge areas. This book also presents recent developments covering ongoing trends in the field of computer-aided drug discovery. Having contributions by a global team of experts (academicians, scientists, and researchers), the book is expected to be an ideal resource for drug discovery scientists, medicinal chemists, pharmacologists, toxicologists, phytochemists, biochemists, biologists, R&D personnel, researchers, students, teachers, and those working in the field of drug discovery. This book is also expected to fill the knowledge gaps that exist in the current CADD approaches and methodologies/protocols being widely used in both academic and research practices. Further, a special focus on current status of various computational drug design approaches (SBDD, LBDD, *De novo* drug design, Pharmacophore-based search), bioinformatics tools and databases, computational screening and modeling of phytochemicals/natural products, artificial intelligence and machine learning, and network pharmacology and system biology would certainly guide researchers, students, or readers to conduct their research in the emerging area(s) of interest. By design, this book is expected to be highly beneficial to different stakeholders working in the pharmaceutical and biotechnology industries (R&D), the academic as well as research sectors. This book would be the best choice for majority of readers among various communities such as scientists, researchers, R&D personnel, students, and teachers. This book would, therefore, meet the basic needs of teaching, learning, and research practices particularly in the subject/area of computer-aided drug design and discovery. Looking into the focal theme of the proposed book, the following three unique selling points could be attributed to address the reader's demands up to the level of their satisfaction:

1. Highlights advanced computational approaches, tools, and techniques for drug design and discovery.
2. Details fundamental knowledge, methodologies, and practical applications of CADD approaches.
3. Depicts several cutting-edge areas of CADD approaches and bioinformatics tools/techniques.

Guntur, India — Mithun Rudrapal
Majmaah, Saudi Arabia — Johra Khan

Contents

About the Editors

Mithun Rudrapal, PhD, FIC, CChem, is Associate Professor at the Department of Pharmaceutical Sciences, School of Biotechnology and Pharmaceutical Sciences, Vignan's Foundation for Science, Technology and Research (Deemed to be University) Guntur, India. Dr. Rudrapal has been actively engaged in teaching and research the field of Pharmaceutical and Allied Sciences for more than 12 years. He has over a hundred publications in peer-reviewed international journals to his credit and has filed a number Indian and International patents. In addition, Dr. Rudrapal is the author of dozen published or forthcoming books. Dr. Rudrapal works in the areas of Medicinal Chemistry, CADD, Drug Repurposing, Phytochemistry, Herbal Drugs, and Nanophytotherapeutics.

Johra Khan is Assistant Professor currently working at the Department of Medical Laboratory Sciences, College of Applied Medical Sciences, Majmaah University, Majmaah, Saudi Arabia. She has published a number of publications in various national and international peer-reviewed journals and presented scientific papers at national and international conferences. Dr. Johra Khan works in the area of Molecular Biology.

Chapter 1
Role of Bioinformatics in Drug Design and Discovery

Pinkal H. Patel, Adarsh Jha, and G. S. Chakraborthy

1.1 Introduction

The first phase within the procedure of finding new drugs is the diagnosis of illness along the nicely defined signs that lower life condition. An ideal medicine is typically defined as compound (that might be a straightforward molecule/complex polypeptide) or chemical mixture that reduces discomfort despite having a significant negative impact on the patient. Other desirable drug characteristics include a low probability of drug resistance, which would dramatically reduce the medicine's commercial value (David et al. 2009; Drews and Ryser 1997), and low negative environmental effects (Davies and Davies 2010), such as no reactivation of bacterial species after human usage (Boxall et al. 2012). As a result, a desired drug is the one that not only works well but also causes furthermore few brief consequences upon the individual, community, as well as surroundings negative impacts.

The main object of this review will be how bioinformatics might speed up the search for such desirable medications. The fields of population gene fingerprint, evolutionary biology, genetics, metagenomics, and omics are all included in the interdisciplinary science of bioinformatics. Information from increased output molecules (Fig. 1.1) is used by bioinformaticians in drug discovery to compare patients, animal disease models, cancer cell lines, and those who carry symptoms, as well as normal subjects. Main goals parallel between: (1) relate signs of illness to sequence variation, transcriptional regulation, and additional external conditions that affect genomic regulation; (2) specify targeted therapies that both remove or recover cellular activity, such as cancer cells; (3) anticipate or further develop drug

P. H. Patel (✉) · A. Jha · G. S. Chakraborthy
Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Parul Institute of Pharmacy and Research, Parul University, Vadodara, Gujarat, India

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_1

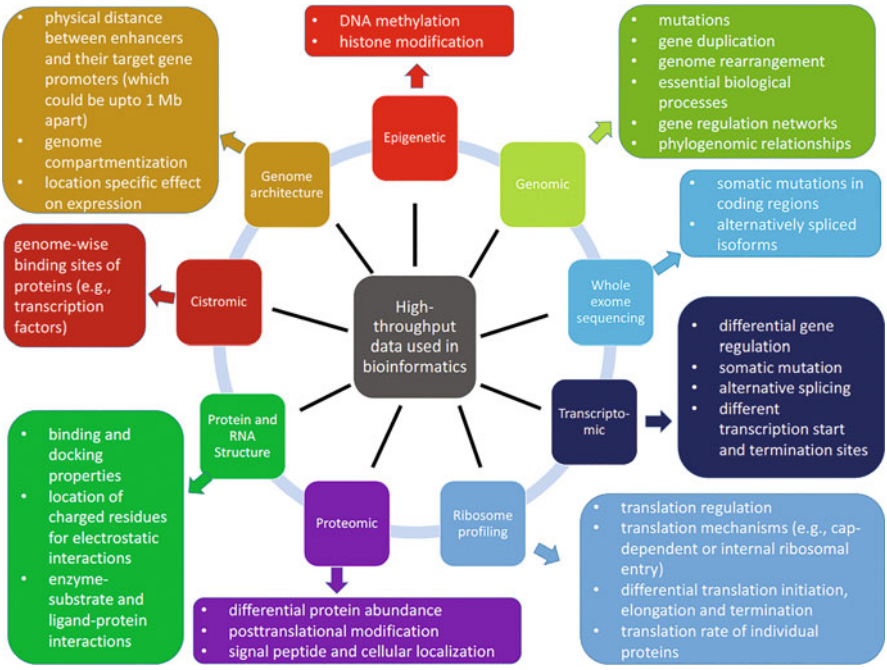


Fig. 1.1 Important high-throughput data types and their important details pertinent to the drug discovery. The exclusion of the metabolic data is because they pertain to cheminformatics. From "Bioinformatics and Drug Discovery" by Xuhua Xia, 2017, Copyright 2017 by Xuhua Xia

compounds that interact to the target and produce the intended beneficial benefit; and minimize adverse results; and (4) evaluate the efficacy of the candidate drugs.

1.2 Genome Sequencing and Genomic Exons Information in Drug Discovery

The discovery of sequence homology between a platelet-derived growth factor (PDGF) and the cancerous gene of the simian sarcoma virus, v-sis, by straightforward string matching is among of the earliest bioinformatics applications to drug target finding (Doolittle et al. 1983; Waterfield et al. 1983). This discovery prompted two new ways of thinking in addition to the use of PDGF as a target for cancer drugs (Pietras et al. 2003; Bergsten et al. 2001; Ehnman et al. 2013). First, the viral transforming factor might simply convert a growth factor's temporary expression to constitutive expression, indicating that growth factors could be potential targets for the creation of anti-cancer drugs. Second, anything that alters the way genes are expressed may be a factor in cancer. The advancement of mechanism-based anti-cancer medication discovery throughout the ensuing years was aided by this new

conceptualization of cancer biology (Gibbs 2000; Shoemaker 2006; Moffat et al. 2014).

1.2.1 Genetic Diseases

Patients with inherited disorders whose genomes and complete exomes sequenced have revealed several somatic mutations that are linked to genetic diseases (Ow et al. 2014; Song et al. 2014; Zhang et al. 2016) and may serve as therapeutic targets one day.

Many genomic differences identifying illness causing variations between matched patients' and healthy controls are the main challenge in bioinformatics research on somatic mutations (Brucher and Jamall 2016). High-genetic heterogeneity can be seen in some disorders, including cancer (Garraway and Lander 2013), even within the cells that make up a single tumour (Ling et al. 2015). Many of these somatic mutations can result from cellular dysfunction rather than being the cause of it (Brucher and Jamall 2016).

There are three different types of somatic mutations that need to be distinguished: those that cause the disease and may be drug targets; individuals who are directly connected to the illness genotype are consequently related to the disease; who are consequently linked to the illness, and those who are not linked to the illness yet unavoidably appear in the patient group and not in the control group. Although not a pharmacological target, the second category of mutations can be utilised to diagnose disease. There are two ways to exclude the third category. The first is by expanding the sample size. If several forms of sarcoma share a similar somatic variation, its relation to that sarcoma is greater than the one which only affects one form of sarcoma (Pereira et al. 2016).

Without knowledge of the illness process distinguishing among the initial and latter types of genetic differences between patients and controls is much difficult. A damage variation may arise in the coding sequence (CDS), a regulatory motif (such as the response elements for ligand-activated nuclear receptors), or an enhancer that may be up to 1 million bases away from the CDS. Three methodologies are frequently used by bioinformaticians to determine whether a mutation has a significant effect on gene activity: (1) If either the variation substitutes a very different amino acid for a generally conserved position, such as replacing the a positively charged arginine paired by a non-polar, uncharged glycine (Baird et al. 2015), if the mutation affects the change impacts and information in order; (2) the variation is a highly conserved non-coding sequence, which is typically determined by comparing the genomes of humans and non-human species; or for cellular machinery, such as a start and end point for expression, a junction point, or a control sequence (e.g., ribosome, spliceosome, degradosome). The available libraries of known compliance sequences that have been painstakingly assembled and evaluated facilitate the last method (Daily et al. 2011; Huang et al. 2006; Xie et al. 2009). Genomes are frequently searched for regulatory patterns using bioinformatics methods. Support

vector machines (Xia 2012; Rouchka 1997), whereby the differences between two groups of sequences can be extracted (like, sequence-present and sequence-absent), may be utilised to identify the distinctions between two sets of genomes (like motif-present and motif-absent), and the information gathered from this process can be utilised to find or analyse structures in sequences (Hua and Sun 2001; Zien et al. 2000). Nuclear receptor responses can be control sequence components, whose discovery frequently leads to the improvement of therapeutic targets (Kotokorpi et al. 2010). Software like DAMBE (Xia 2013), which with only a few mouse clicks, it is possible to retrieve codons, ribosomal RNAs, transfer RNAs, introns, exons, 5′ and 3′ crossover sites, upwards or downward regions of genomic features, etc., makes these investigations easier. DAMBE has tools for estimating the minimal folding energy, Gibbs sampler, and PWM, but it also calculates the protein isoelectric point and measures protein translation efficiency.

A distantly related genome or a different biological pathway that can negate the effects of variation can be found using bioinformatics a distantly related genome or a different biological pathway which can negate the effects of variation. Mammals frequently have redundant or partially redundant functions. For instance, the mouse paralogous genes USP4 and USP15 have redundant functions (Vlasschaert et al. 2015). The deposition of relatively lengthy fatty acids that causes Human adrenoleukodystrophy (ALD) is brought on by the selective elimination of the 10-exon ABCD1 genome (Krasemann et al. 1996). This implies that in addition to restricting VLCFA intake through diet, alternative metabolic pathways for VLCFA should also be activated by controlling the expression of another lipolytic genome (ABCD2) and inhibiting the activity of an enzyme that produces VLCFA (Morita et al. 2011). Sickle-cell anaemia, which is brought when the human beta-globin gene changes by just one peptide (Pauling et al. 1949), is another instance of switching on alternative biochemical activities or partially redundant genomes (Steinberg and Rodgers 2001; Kutlar 2007). Because HbF inhibits haemoglobin (HbF) polymerisation and clumping, it makes the foetal HbF gene an attractive therapeutic target. Adults with sickle-cell anaemia and thalassemia might experience less discomfort if there was a medication that could restore the silent HbF (Kioussis et al. 1983; Taramelli et al. 1986). It is interesting to note that some individuals with thalassemia carry the normal copy of the β-globin gene, but due to mutations far from it, the gene does not express itself. Later on, epigenetic modification and genomic architecture will be discussed, along with such long-range gene regulation.

1.2.2 Human Diseases Caused by Pathogens

For the development of target-based drugs against pathogens, well-annotated genomes are crucial. The three key steps in the overall bioinformatics methodology are as follows. The first step is to determine which pathogen genes are crucial for therapeutic targets. Identification of these crucial genes can be facilitated by a genome, especially one that has been thoroughly annotated. Because pathogenic

organisms use the salvage process in place of a new method of producing nucleosides, genes involved in nucleotide synthesis, for instance, are well recognised yet frequently absent in pathogenic species. The genes for ATP, GTP, and TTP de novo synthesis have been lost in *Trypanosoma brucei*, but the pathogen still has a small ability to synthesize CTP from scratch. It is most probably as CTP cannot be dependably acquired via recovery and normally has a lower concentration in cells than the other oligonucleotides. This suggests a potential therapeutic target: the CTP production route. In fact, preventing CTP synthesis stops the pathogen's development and replication (Hofer et al. 2001). When infections and their evolutionary relatives are compared genomically, essential genes are frequently found to be highly conserved. Occasionally, it also can be inferred from information collected from biological systems with specifically and purposefully scooped genomes, such as *Escherichia coli*, *Bacillus subtilis*, or *Saccharomyces cerevisiae*. Any bacterium most likely needs the same genome that is needed for the two bacterial strains.

Checking if such crucial genes have host homologues is the second step in the development of medications to combat pathogens. If so, inhibiting these crucial pathogen genes could have a negative impact on how the host homologue functions. As a result, we must compare the sequences and structures to find parasite and host-associated proteins have any distinctive features that can help thus in the creation of medications tailored to certain infections.

Third, it is crucial that the medicine specifically targets neither of its evolutionary descendants who are not harmful but the infection itself in order to lessen the chance that the infection will develop antibiotic immunity. This is why the preferred source of drug targets has evolved into unlike its non-pathogenic cousins; infection islands are different in harmful microbes (Hacker et al. 1997, Hacker and Kaper 2000).

A glutamate transport mechanism, which lacks in humans and birds, but presents in the disease *Clostridium perfringens*, was discovered using bioinformatic research (Bhatia et al. 2014). Humans as well as tamed creatures and birds will be protected by drugs created to combat such a conveyance mechanism. The phosphoinositide-3 kinase (PI3K) signalling route is crucial for the human parasite *Giardia intestinalis* and may be used as a therapeutic target. It is crucial to determine what distinguishes the PI3K sequences (Gipi3k1 and Gipi3k2) in *G. intestinalis* from animals as the PI3K route is also crucial in many microbes. Comparisons of the parasite's sequences with other organisms showed a special insertion that can be used as a pathogen-specific therapeutic target. *Pseudomonas aeruginosa* is also targeted using the same strategy (Fernandez-Pinar et al. 2015). Similar to this, when creating anti-HIV-1 drugs may aim the genomes associated in reverse transcription and the proteolytic breakdown of the viral' translated polypeptide because those functions are not just essential for viral viability and spreading but are also not closely related to human functions homologs, meaning that blocking them should have few negative effects on people.

The use of existing medications against different pathogens can also benefit from genomic analysis. A number of bacterial pathogens' cell surfaces include galactofuranose (Galf), whose synthesis needs the enzyme UDP-galactopyranose

mutase (UGM) (Gruber et al. 2009; Kincaid et al. 2015). Galf is not present in humans, making UGM a desirable pharmacological target (Pedersen and Turco 2003). Later, nematodes and other eukaryotic unicellular pathogens (Beverley et al. 2005) were shown to contain UGM encoded by the gene GLF (Wesener et al. 2013). Drugs created to combat bacterial infections are repurposed to combat eukaryotic unicellular pathogens. Drug development is more affordable when drugs are repurposed. Genomic study reveals that while eukaryotic UGMs are comparable to one another and to prokaryotic UGMs, they are also very dissimilar from one another, suggesting difficulty in translating drugs from bacterial pathogens to eukaryotic pathogens. However, there is a very significant probability that a medicine developed to treat one eukaryotic UGM could also be used to treat other eukaryotic pathogens.

Additionally, genomics has helped us better understand how drugs work. *Plasmodium falciparum* growth might be inhibited by the venom protein PcFK1 of the spider *Psalmopoeus cambridgei*; however, the exact mechanism was unknown. Sequence similarities between PcFK1 and the protein substrate of the *P. falciparum* enzyme PfSUB1 were discovered through sequence analysis, supporting the idea that PcFK1 is an antagonist of PfSUB1. This theory is supported by further docking predictions and in vitro tests, which identify PfSUB1 as a potential therapeutic target (Bastianelli et al. 2011).

Understanding functional redundancy in essential cellular processes is essential for creating pathogen-fighting medications that are successful. Ethambutol is a medication that targets the *Mycobacterium tuberculosis* arabino-furanosyltransferases Mt-EmbA and Mt-EmbB, which are involved in the formation of the cell wall compound mycolyl-arabinogalactan-peptidoglycan. Another arabino-furanosyltransferase, Mt-AftA, was discovered by bioinformatic analysis. This enzyme is not inhibited by ethambutol, making it a potential therapeutic target. In addition to being more effective against the virus, medications that target all three arabinofuranosyltransferases also lessen the possibility that the pathogen may acquire drug resistance (Alderwick et al. 2006). Because the genomic shift between glucose and lactose as well as the lac-operon was discovered, bacterial species have been known to activate varied biochemical routes to meet the demand for vitality and development (Jacob and Monod 1961); however, without knowledge about how cells act with respect to alternate routes which may be triggered in regard to the medicine, a compound cannot be successful towards a bacterium or a sarcoma.

Together with integrated genomic evolutionary relationships and constructed chronological outlook (Xia 2007, Higgs and Attwood 2013), bioinformatics frequently contributes to settling disputes over molecular mechanisms. The causative analysis of CpG methylation causing CpG depletion by the following CT alteration caused by random deamination is one example of this. *Mycoplasma genitalium* and *Mycoplasma pneumoniae* both lack DNA CpG methyltransferase (Cardon et al. 1994; Goto et al. 2000), but *M. genitalium* exhibits a much stronger CpG deficiency than *M. pneumoniae*, raising the possibility so the two groups' different levels of CpG deficit has no bearing on CpG methylation. Without taking evolutionary theory

into account, such a conclusion from genetic investigations is frequently incorrect. Multiple CpG methyltransferases are found in *M. pulmonis* as well as other relatives that diverge before *M. genitalium* and *M. pneumoniae*, so a thorough phylogenetic analysis using the programme DAMBE demonstrated thus the two groups' forebears ought to have numerous CpG methyltransferases (Xia 2013). Both *M. pneumoniae* and *M. genitalium* started to gain CpG following their mutual descendants' inactivation of the CpG methyltransferases. Compared to *M. genitalium*, *M. pneumoniae* developed very swiftly (with a longer branch) and recovered CpG quicker. Such findings validated a normal relationship between CpG-specific DNA methylation and CpG deficiency and emphasise the importance of viewing biological processes from an evolutionary perspective. Lately, a brand-new bilateral genome alignment-based evolutionary approach was created to make it easier to conduct comparative genomic research on extremely divergent populations. Because many of these are due to the difficulty in obtaining accurate multiple sequence alignment with highly divergent sequences while studying widely dispersed bacterial or viral species (Xia 2016a, 2016b).

1.3 Epigenetics, Genome Architecture, and Cistromes in Drug Discovery

Identical deleterious variants in monozygotic twins, such as the aforementioned ALD mutation, frequently exhibit significant phenotypic variation (Korenke et al. 1996; Petronis 2004; Petronis 2006; Petronis et al. 2003). These findings show the connection between epigenetic changes and human diseases (Zoghbi and Beaudet 2016; Jiang et al. 2004). DNA methylation and histone modification are two related processes that make up epigenetic modification. Maintaining the DNA methylation pattern is human DNA methyl transferase 1 (DNMT1), for whom CatD region recognises hemi-methylated CpG motifs in mammals in order to keep the DNA methylation sequence between parent and daughter cells. Histone deacetylase is instead drawn towards the methylated CpG in animal species, where it removes the acetyl group and restores the positive charge of the lysine residues (or histone N-terminal) in histone so that the negatively charged DNA backbone can tightly wrap around the positively charged histone to silence the genotype. In many aspects, a silent gene is analogous due to some kind of variant with failure (Wade and Wolffe 2001). Blockers of histone deacetylase are being used as targeted therapies in an effort to reactivate the apoptotic process since it appears that some malignancies are induced via DNA methylation and histone deacetylation, which permanently silence apoptosis-related genes (Insinga et al. 2005a, 2005b). Given that deacetylase inhibitors frequently have a significant impact on the regulation of many other genes (Bolden et al. 2006), the fundamental issue with this strategy is specificity, which may help to clarify how some medications frequently avoid enrolling in clinical testing (Voelter-Mahlknecht 2016). Currently being developed are techniques for

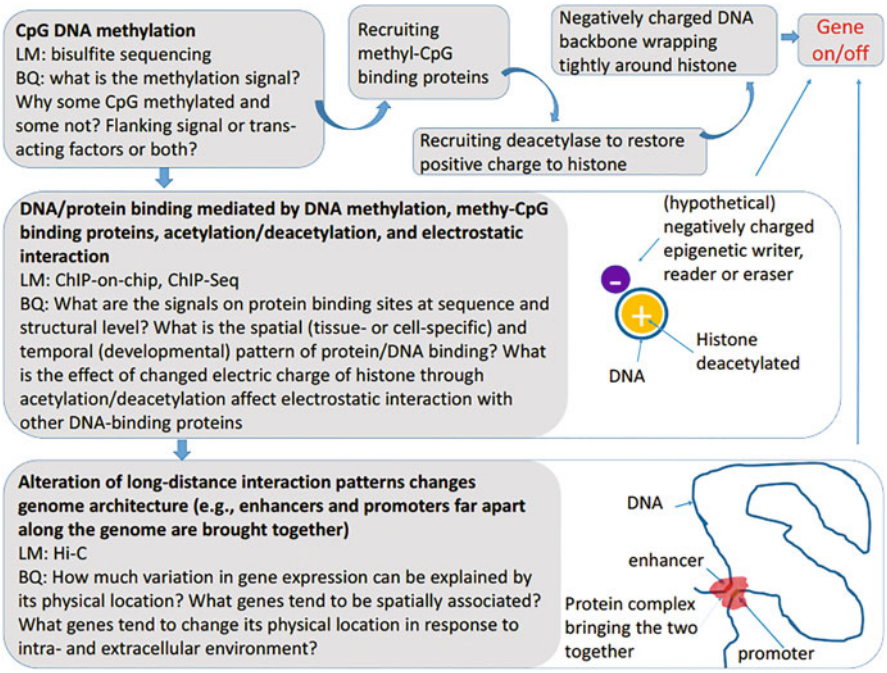


Fig. 1.2 An outline of how epigenetic changes to proteins that bind to DNA can affect gene expression. These changes can affect short-distance interactions like enhancer–promoter connections, as well as long-distance interactions like protein–protein and protein–DNA interactions. BQ: example bioinformatic queries; LM: laboratory technique. From "Bioinformatics and Drug Discovery" by Xuhua Xia, 2017, Copyright 2017 by Xuhua Xia

precision epigenome editing that involve DNA-binding components and particular detection and modification of sequences (Kungulovski and Jeltsch 2016).

A very thorough conceptualisation of protein expression and genomic alteration has now taken the place of the widely held notion that irreversible genetic suppression is the primary function of DNA methylation and histone deacetylation (Fig. 1.2). This conceptual shift necessitates the combined study of a number of high-throughput data sources, including data on DNA/protein binding (cistrome) (Grigg and Clark 1994; Grigg 1996) from ChIP-on-chip and ChIP-Seq, methylation pattern from bisulfite sequencing (Robertson et al. 2007), and genomic architecture from Hi-C or its derivatives (Lieberman-Aiden et al. 2009). DNA methylation changes DNA/protein binding, which in turn changes genome architecture by enabling the joining of two DNA segments that are spatially separated along linear DNA. The ability to analyse the spatial relationship inducers and promoters can sometimes be up to 1 million units distant from each other, which is made possible by genome architecture data. Since 1930, studies of translocation have shown that the location of a gene's promoter or enhancer on the genome affects how that gene is expressed (Muller and Altenburg 1930). However, it was not until much later that

empirical data emerged to support this theory. The genetic regulatory promoter cluster theory was developed as a result of this. In other words, a gene whose promoter loops stated near the centre. As the hub contains one or more enhancers, the function of all genomes will be silenced by the loss of this type of hub whose expression is dependent based on its vicinity toward the base physically.

What the methylation signal on DNA is and if it can be modulated to change epigenetic alterations are the two main questions from a bioinformatics perspective. As I've previously indicated, the beta-globin genome is correctly copied in certain thalassemia cases, although a few thalassemia cases get the proper version of the beta-globin genome, the genome is not activated because of alterations that happened far from it (Kutlar 2007; Kioussis et al. 1983). One may come up with two theories. First, the patient has a mutation or deletion in the promoter which regulates production of the β-globin gene (Kioussis et al. 1983; Forrester et al. 1990). Second, aberrant epigenetic changes and protein/DNA binding cause the enhancer that is normally situated close to the promoter of the β-globin gene to be moved to a different location. It would be clearer if these hypotheses were tested, which has only been made possible by the accessibility of high-output data on methylation frequencies, cistromes (a compendium of all protein/DNA receptor), and genomic structure, if we knew how to readjust the beta-globin promotor and the regulator to create the genome expression (Deng et al. 2012, Deng et al. 2014a, 2014b, Hou et al. 2008). Similar to the previous example, if DNA methyl activation quiets the beta-globin genome, understanding we can reawaken the genome by learning where to adjust the signal to change the methylation sequence that has been silenced. According to the same line of thinking, the understanding of location-selective methyl removal process particularly useful if methyl activation silences the prenatal globin genome and reactivating these prenatal globin genomes can lessen this problem brought upon with alterations in adult globin genomes (Kungulovski and Jeltsch 2016).

Given that the mammalian genomes contain both methylated and unmethylated CpG, one basic bioinformatic approach would to find out if surrounding nucleotides add to methylation signs, analyse the surrounding locations of those two same components of CpG dinucleotides. Although done on a small scale, however, such examinations of the surrounding areas among methylated and unmethylated CpG have demonstrated a significant splicing sign in the adjacent regions of the 5′ and 3′ splice sites (Ma and Xia 2011; Vlasschaert et al. 2016) have not produced definitive conclusions (Bibikova et al. 2011; Eckhardt et al. 2006; Shoemaker et al. 2010). Although the idea of an imprinting centre (IC) has been around for a while (Ohta et al. 1999), its structural or sequence-level physical underpinnings continue to be a mystery.

Monozygotic twins with the same genetic abnormality frequently express the related disease differently (Korenke et al. 1996; Petronis 2004; Petronis 2006; Petronis et al. 2003); therefore, it makes sense to look for environmental factors like nutrition that may have an impact on epigenetic alteration (Chen et al. 2016; Sharma et al. 2016). A methionine deficit is likely to influence DNA methylation since S-adenosyl L-methionine (SAM) is required for methylation as the methyl

donor (Ingrosso et al. 2003; Ingrosso and Perna 2009). This has been proven to be the case. Similar to this, it would be expected that any significant methionine disturbance, such as the loss of the essential enzyme methylthioadenosine phosphorylase (MTAP), would likewise have an impact on DNA methylation, gene regulation, and cancer. Indeed, cancer cells frequently have MTAP loss (Bigaud and Corrales 2016). Therefore, every gene that affects methionine metabolism may be a potential target for drugs, and bioinformatics can efficiently discover such genes using databases like KEGG (Kanehisa 2013; Kanehisa et al. 2016; Tanabe and Kanehisa 2012).

A perfect drug (or containing nano-device) must be able to recognise and rectify the incorrect DNA methylation sequence in a targeted manner (Kungulovski and Jeltsch 2016). We must first determine the proper methyl activation sequence or, better yet, locate a set of compounds which carry out this exact sequence in order to create such a medicine or nanomachine. The evidence from experiments has grown in favour of RNA's function in adjustments to gene regulation (Jin et al. 2004). The epigenomic coding may not be on DNA since the zygote's DNA gets demethylated to recover pluripotency (Clark 2015). The epigenetic codes, particularly those that define new DNA methyl activation (Bao and Bedford 2016), are not likely to be present in proteins because protamine replaces most fundamental histones in male germ cells and since proteins don't seem to be very good at generating data. A group of extremely robust and physically preserved RNA sequences, though, may include such codes and is available from the egg and sperm phase onward. Long non-coding RNAs (lncRNAs) may regulate the chromatin state and take part in epigenetic modification. DNA that contains several sequential receptors were discovered during the lncRNAs such HOTAIR (Chu et al. 2011; Chu et al. 2012; Rinn et al. 2007), and Kcnq1ot1 (Pandey et al. 2008) engaging to such sites enhances the activation of Polycomb Repressive Complex 2 (PRC2) for regulating histone H3 lysine-27 trimethylation. Analysis of lncRNAs bound to DNA and protein utilising the ChiRP-seq method (Chu et al. 2011; Chu et al. 2012). Small RNAs too can regulate genetic modifications (Chen et al. 2016; Sharma et al. 2016; Rodgers et al. 2015; Gapp et al. 2014). A variety of short RNA species are found in mature sperm, and these small RNAs do have an impact on the phenotypic of children. These short RNAs on children also seem to be involved in epigenetic remodelling (Rodgers et al. 2015, Gapp et al. 2014). The ENCODE pilot project demonstrates "Most of the nucleotides in the gene could be found in main transcribed, demonstrating that the genome is pervasively expressed," according to one study (Birney et al. 2007). These, for data scientists searching to recognise epigenome-modifying RNAs as prospective drug targets, non-coding transcribed may be a veritable great resource.

Epigenetic alteration has a long history. In order to defend themselves in opposition to endogenic type II restriction endonucleases, many bacterial species methylate the DNA in their own cells (Murphy et al. 2013). Human viral diseases like HIV-1 can cause significant alteration in host epigenetic pattern (Abdel-Hameed et al. 2016). Some bacteriophages possess methyltransferases which can modify its own genomes to withstand human-restrictive digestion. It is now understood that many infections have the ability to alter host epigenetic patterns in order to survive

and reproduce in the host (Bierne et al. 2012; Arbibe and Sansonetti 2007), as well as some host defence systems against pathogens (Bierne et al. 2012). Uncertainty surrounds the ultimate course of such host cells that have undergone epigenetically driven pathogen modification. Do they stop the invasion of the virus, return the epigenetic pattern to normal, and resume normal function? Or do they start a specific apoptotic pathway and die? A model organism or cell line that can have its epigenetic pattern disturbed by external causes and subsequently returned to normal is what epigeneticists need.

1.4 Transcriptomics and Drug Discovery

Using transcriptomic data, it is now possible to distinguish between patient and matched control groups' transcription start and termination sites, alternatively spliced isoforms, and differentially regulated genes (Berger et al. 2010; Arvaniti et al. 2016; Bell et al. 2016; Furukawa et al. 2016; Haustead et al. 2016; Mlera et al. 2016). Drug discovery is primarily aided by transcriptomic data analysis in two ways: drug target identification and phenotypic screening to identify and hone drug candidates.

1.4.1 Phenotype Screening

Phenotypic screening has been the subject of controversy, although recently proposed definitions all share the following five characteristics (Moffat et al. 2014; Eder et al. 2014): (1) several different chemicals (chemical entity) should be selected in a systematic manner for the screening; (2) pheno-genomic alterations brought on every substance must be monitored; (3) a standard for requirement should be developed; (4) molecules that have favourable bioactivities (phenotypes) must be kept as active compounds for additional evaluation and validation; and (5) the method of action should not be the screening's main point of interest. The revelation of artemisinin, one of the most successful treatments for the *Plasmodium falciparum* pathogen that causes malaria, is one of the success stories of phenotypic screening's effectiveness in identifying active components in conventional medicine (Miller and Su 2011).

When developing medications for illnesses with multiple causes, such as genetic disorders, morphological separation is more effective, whereas the target-based approach is more effective when developing medications for illnesses with comparatively simple causes, such as mono-genetic conditions (Swinney 2013; Swinney and Anthony 2011). The genetic makeup of cancer is varied (Garraway and Lander 2013), and the genetic variety within a single tumour is incredibly significant (Ling et al. 2015). Phenotypic screening created especially for cancer has been utilised extensively in the discovery of cancer drugs for such complicated disorders

(Shoemaker 2006). The discovery of an effective chemical through screening frequently illuminated the molecular mechanism of action (Swinney et al. 2016).

Because FDA-approved drugs have already undergone the challenging regulatory process, phenotypic screening for which medication reuse is economical. This approach has led to the development of potential antagonists for enteroviruses (Ulferts et al. 2016), anti-aging therapies (Snell et al. 2016), anti-neoplastic drugs (Ozsvari et al. 2016), and allosteric Bcr-Abl inhibitors in the fight against chronic myeloid blood cancer (Singh et al. 2016).

In what ways can bioinformatics support phenotypic screening? The explanation is found in the fact that phenotype is frequently defined, most recent phenotypic separation studies (Wishart 2016a, 2016b, Xia et al. 2009), especially in separation for anti-neoplastic medicines, use either directly or indirectly a genomic (transcripts or protein) database (Shoemaker 2006) or metagenomic characteristics. From this vantage point, there are two more methods for treating cancer cells. The first step is to get cancer cells' gene expression back to normal cell levels. When the first option fails, the second is to eliminate cancer cells by causing apoptosis (Shoemaker 2006; Moffat et al. 2014). These two methods suggest the following two criteria for phenotypic drug discovery: (1) raising similarity in genetic regulation between normal and cancerous cells and (2) raising similarity in genetic expression among both tumour and dead cells.

By using genome expression data as characteristics to create a logical and rational indicator of drug desire (Idd) in morphological testing investigations, bioinformatics can support gene expression and drug development. For assessing pharmacological effects and safety at varied drug doses, therapeutic indices (Swinney 2009; Muller and Milton 2012) based on various pharmacokinetic theories might be supplemented by a similar Idd (Gabrielsson and Green 2009, Holford and Sheiner 1981a, 1981b). The poor success rate of medications revealed using phenotypic screen may have been caused by the absence of an explicit Idd (Eder et al. 2014). Due to this, I'll take an uncommon step in this review article and start the process of creating a drug-desirability index that includes both symptom relief and negative effects.

Assign the genomic assessments of Gp, a "control," and a "patient" who could be an animal illness model or a tumour cell line who is in a healthy state as Gn, and a patient who has taken a potential therapy as Gd. Today, it is simple to calculate a number of pairwise distances (Xia and Xie 2001) between Gn and Gp, Gd and Gp, and Gn and Gd (designated Dnp, Ddp, and Dnd, respectively, Fig. 1.3). The terms "Dnp" and "Dnp − Dnd" refer to the "Dnp − Dnd" correspond to the contender medication's ability to lessen the intensity of the disorders, while "drug efficacy" (Emax) in pharmacodynamics models correspond to the contender medication's ability to reduce the severity (Holford and Sheiner 1981a, 1981b). The difference between (Dnd + Ddp) and Dnp or (Dnd + Ddp − Dnp) could be used to measure the side effect. This suggests that Drug B in Fig. 1.3b has a more severe side effect than Drug A in Fig. 1.3a. We may create an inventory of medication desire (Idd) that uses these concepts as:

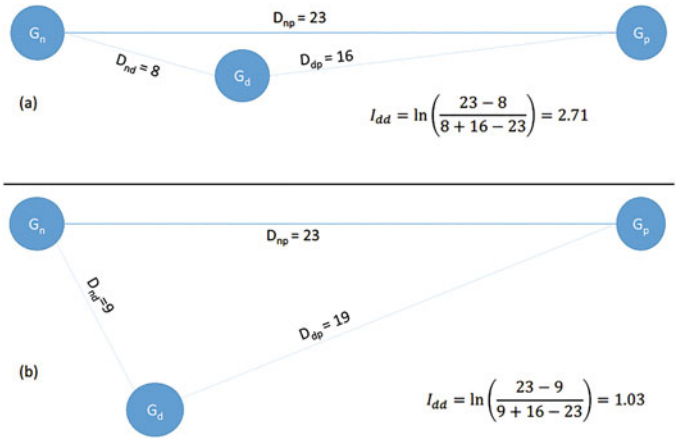


Fig. 1.3 Illustration of numbers of phenotypic screening using Idd in Eq. (1.1) for two sets of transcriptome data (**a**) and (**b**). Genetic expression in healthy cells, sick-cytes before medication, and sick-cytes after medication are referred to as Gn, Gp, and Gd, respectively. From "Bioinformatics and Drug Discovery" by Xuhua Xia, 2017, Copyright 2017 by Xuhua Xia

$$\text{Idd} = \ln\left(\frac{\text{Dnp} - \text{Dnd}}{\text{Dnd} + \text{Ddp} - \text{Dnp}}\right) \tag{1.1}$$

The use of Eq. (1.1) is seen in Fig. 1.3, where Drug A in Fig. 1.3a with Idd = 2.71 is preferred over Drug B in Fig. 1.3b with Idd = 1.03 (Fig. 1.3b), although this is not anticipated to occur in practise, one issue. The problem of Eq. (1.1) is whenever Gd = Gn or Gd = Gp, the fraction would be 0. To the equation, one might nonetheless add a small pseudonumber (c) as follows:

$$\text{Idd} = \ln\left(\frac{c + \text{Dnpp} - \text{Dnd}}{c + \text{Dnd} + \text{Ddp} - \text{Dnp}}\right) \tag{1.2}$$

The only prerequisite for c is that it must be small in relation to (Dnp − Dnd) in order for it to have a minimal impact on Idd. C could be set to 0.01*.

Idd is available to assess any test results where a set of data is available to depict the individual prior medication use, the individual during a classic control period, as well as the individual after medication use, such as blood ferritin and transferrin levels, calcium and iron levels, etc. Idd's use is not exclusively restricted to molecular techniques or genomic regulation profiles. It can be used to assess the desirability of various medications as well as those applied at various concentrations or given via various delivery methods (such as oral, subcutaneous injection, etc.). By simply substituting Gn with the gene expression of apoptotic cells, it is possible to obtain Idd for the second parameter, which is how far a drug may make cancerous cells result in cell death.

Accurate gene expression characterisation is necessary for the two criteria to be applied effectively. The growth of greater-output tools, like including next genomics and microarrays inside this history currently (Xia and Xie 2001; Gentleman et al. 2005), has been accompanied by the development of bioinformatic techniques and tools (Deng et al. 2014a, 2014b, Dobin et al. 2013, Langmead et al. 2010, Langmead and Salzberg 2012, Langmead et al. 2009, Roberts et al. 2013, Roberts et al. 2011, Trapnell et al. 2009, Trapnell et al. 2012). Sadly, the fundamental issue with allocating sequence reads to paralogous genes that has long plagued the analysis of microarray data has not been resolved. Instead, nearly all software provides users with two easy but ineffective options: either ignore sequence reads matching multiple genes or distribute such sequence reads equally among paralogous genes. Because multi-cellular eukaryotes have many duplicated genes, it follows that a large number of genes' expression cannot be accurately defined due to an improper allocation of sequence reads to paralogous genes.

1.4.2 Drug Target Identification

To identify client and regulate differences in genomic expression pattern, as well as transcriptional variants, transcriptomic information from RNA-Seq can be employed. Diseases like Alzheimer's disease (AD), which is linked to aberrant splicing from the parent peptide for amyloid, are frequently brought on by changes in the spatial and temporal distributions of various splicing isoforms (APP). Amyloid is produced by the proteolytic processing of APP, and it helps to build the extracellular neuritic plaques that are typically thought to be the cause of AD. The exon 7 (E7) of the multi-exon APP gene encodes for the Kunitz protein enzyme blocker. The animal central nervous system expresses three of at least eight variants created by variable splicing of such APP pre-mRNA that are distributed spatially and temporally (one deficit E7 and the rest two containing E7). While an E7-deficit isoform (APP695) is frequently exhibited in neurons, the E7-having isoforms (APP770 and APP751) are typically expressed in astrocytes as well as microglial cells. Contrary to the secreted E7-lacking isoform, the E7-containing APPs form persistent, non-covalent, inhibitory complexes with trypsin. Both in humans and mice, an increase in E7-containing isoforms is linked to AD symptoms. E7-skipping in APP695 is probably caused by the down regulation of U2AF expression during cellular development of brain tissues. However, according to the latest report, the RBFox1 protein has a binding domain (U)-GCAUG that is found both inside and ahead of E7, which is directly related to E7-skipping. RBFox1 is a splicing factor that is specific to neurons and muscles, and it causes exon skipping in a number of genomes, containing APP. Since both U2AF and RBFox1 are particularly expressed in brain tissues, a medication option, which either stimulates RBFox1 or inhibits U2AF, could lessen the chance of acquiring AD. These transcriptome investigations have considerably advanced our knowledge of the pathogenesis of numerous human diseases connected to alternative splicing, including AD.

Diseases frequently have abnormal gene expression or regulatory alterations. The fundamental challenge is in interpreting reason and consequence since the transcription of a genome that causes an illness may start at point t1, while many other genes could express differently at time t2, which could be years after time t1. As a result, there are virtually always many false positives when contrasting the gene expression patterns of diseased as well as healthy individuals. Sadly, despite the fact that it is easy to cut a wood chunk from the pine on a frequent basis, it is significantly more difficult to eliminate a client's liver on a short and regular basis.

Most of the human gene is translated according to an examination of transcriptomic data. Many RNA structures that are both little or large could be discovered by mining transcriptomic information medications or therapeutic targets because RNA interference has been shown to regulate a variety of biological functions. Which unannotated human transcripts are crucial for human biology among the vast number?

A region is considered to be integral when it is predicted to be conserved amongst subspecies, such as apes or primates, from an evolutionary perspective. Using one of the many bioinformatics techniques available, lakhs of unique transcribed contain highly expressed RNAs. Any RNA species that is crucial for a certain function might be a potential aim.

1.5 Proteomic Data and Drug Discovery

The workhorses of living cells, proteins, are the aberrant overproduction of which is frequently linked to disease. Transcriptomic data are frequently not a useful indicator of protein abundance since transcribed genes may be differently translated (or not translated) (Ingolia et al. 2014; Xia 2003), and various proteins degrade at different rates (Gilbert et al. 2007). Because of this, proteome analysis and comparison between patient and control groups is frequently more useful in finding pharmacological targets than genomic or transcriptome information. Nearly all model organisms have yielded proteomic data, which has been deposited in open databases like PaxDB (Wang et al. 2012). These data have made it much easier to create (Xia 2003) and use indices that predict translation efficiency (Prabhakaran et al. 2015, Chithambaram et al. 2014a, 2014b).

Without a cohort to track over time, proteomic data suffer from the same issue with regard to causal interpretation as genomic and transcriptomic data, as I've already indicated. It is particularly challenging to determine which protein is actually causing the disease given the variable expression that has been seen in several proteins. At different cell cycle stages, certain proteins fluctuate in abundance. Unless the time and geographical variability of cellular molecules are not accounted for, comparing proteome (or transcriptome) across patient/normal pairings may still result in misclassification with minimal benefit to drug research. Cells can be sampled over a range of time periods in animal models. Mono-transcriptase molecular analysis can rebuild a genome expression during the cell cycle (Heath et al.

2016; Saadatpour et al. 2015; Wu and Tzanakakis 2013) and proteome characterisation over time (ordering cell gene sequences from phase 3, 1, and 2 to phases 1, 2, and 3 might lead to significantly more illuminating findings).

1.6 Ribosome Profiling and Drug Discovery

Protein abundance statistics are limited by (1) the inability to identify peptides with lower concentrations, brief molecules, or transitory peptides and (2) the difficulty in isolating, separating, and purifying membrane proteins, which are frequently crucial elements in signal transduction. Once upon a time, transcriptomic data gave rise to the assumption that transcriptomic data might be used to estimate protein genomic data (Heath et al. 2016, Saadatpour et al. 2015, Wu and Tzanakakis 2013), yet varying mRNA conversion yields and protein degradation efficiencies skew the correlation among the ratio of peptide to mRNA availability. Although, it is anticipated that ribosome profiling data combined with transcriptome data will produce accurate forecasts of the rate of protein synthesis. The mRNA availability as well as translating rate data is available via transcriptomic and ribosome profiling data, respectively. If genes A and B have translation efficiencies of RA and RB, respectively, based on ribosome profiling data and Depending on transcriptome evidence as well as the mRNA availability levels of NA and NB, correspondingly, their corresponding related peptide formation efficiencies are NA*RA and NB*RB. Protein breakdown rate can be determined by comparing (Smircich et al. 2015) discrepancies amongst the peptide quantity estimated in this way as well as the actual peptide quantity. It is recommended that transcriptome and proteomic data be collected during a similar study, typically out of a single cell (Heath et al. 2016, Saadatpour et al. 2015, Wu and Tzanakakis 2013).

Traditionally derived from microarray(Arava et al. 2003; MacKay et al. 2004), ribosome profiling data are now mostly derived through profound decoding of mRNA's 30 nucleotide ribosome-protected fragments (RPF) (Ingolia et al. 2009; Ingolia et al. 2009; Ingolia et al. 2011). However, there is a good agreement between the two methods and the results from the yeast (Xia et al. 2011). The position of the ribosome on mRNA can be determined by mapping the sequenced RPFs to protein-coding genes. Translational efficiency may be proxied by ribosomal density. However, across an mRNA, ribosomes could glide and cluster densely with poor codon use. Elongation efficiency must therefore be taken into account, for example, by ribosome density being regressed against the transforming growth score (Xia et al. 2009). Control motifs like the poly(A) tract can be identified using information from ribosome sequencing that affect translation efficiency, for instance (Xia et al. 2011). Small poly(A) at 5' UTR may make it easier to engage gene transcription elements and speed up translation, but lengthy poly(A) can link to poly(A)-binding proteins and restrict translation. These regulatory motifs can serve as easily manipulable pharmacological targets that are quickly recognisable.

By combining two factors, there are four main models of translation initiation. The first is whether the internal ribosome entrance sites or the 5′ end of the mRNA (Kozak 1980; Jackson et al. 2010) are where the translation machinery begins scanning for the start codon (Gilbert et al. 2007; Doudna et al. 2007; Sonenberg and Meerovitch 1990; Yu et al. 2011; Elroy-Stein et al. 2007). The short ribosomal subunit's ability to look for the initiation codon is the second, if the scan can also be done by a fully developed ribosome. Even while when inner ribosome uptake occurrence is now largely accepted, only new ribosome profiling studies have provided compelling empirical evidence for fully formed ribosomes along mRNAs' 5' UTRs (Ingolia et al. 2014), indicating that these ribosomes may also search for the start codon.

Many viral internal ribosome entry sites (IRESs) have robust secondary structure, in contrast to eukaryotic IRESs for whom IRES action decreases as second structural stability increases (Xia et al. 2009). The inter-cistronic region of the cricket paralysis virus (CrPV) contains the IRES that can interact with the ribosome directly without the need for translation initiation components (Jan and Sarnow 2002; Jan et al. 2001; Pestova et al. 2004; Schuler et al. 2006). In order to avoid requiring initiation components necessary for cap-dependent translation (Pestova et al. 1998; Boehringer et al. 2005), the IRES of hepatitis C virus (HCV) can mimic the gene transcription unit (Komar and Hatzoglou 2005).

An essential biological mechanism for reacting to the extracellular environment is translation regulation. A dozen or so genome of the yeast *S. cerevisiae* is ordinarily transcribed but not translated; nevertheless, when the surface nutrients are depleted, these genes are translated, and the resulting proteins allow the ability of yeast cells to scavenge nutrition from the growth media and proliferate (Gilbert et al. 2007). Ribosome profiling information can show if these translation-regulated messages are being translated, which will help us comprehend how animals employ translational control to acclimate to the surroundings.

The best method for finding several novel peptide-coding genomes that would make good therapeutic targets is ribosome profiling. The discovery that unlabelled protein-coding genes can exist in the extensively researched phage λ gene emphasises how many protein-coding genes may still be unstudied (Liu et al. 2015). Usually, genes that are not categorised as translating segments contain ribosome in both human and mouse, which results in the creation of polypeptides (Yoon et al. 2014). Given that the majority of the human genome is really transcribed (Birney et al. 2007), bioinformatic examination of the ribosome profiling data may lead to the identification of numerous additional protein-coding genes (Popa et al. 2016).

Assigning RPFs to paralogous genes when an RPF matches many genes equally well is a critical issue in the analysis of ribosome profiling data. A polypeptide segment may bind to various peptides with perfect results when protein identification is commonly done using peptide mass fingerprinting, which is a difficulty with transcriptomic and proteomic data. Most programs provide two subpar choices: Genome runs with several paralogous genome matches ought to not be used, and

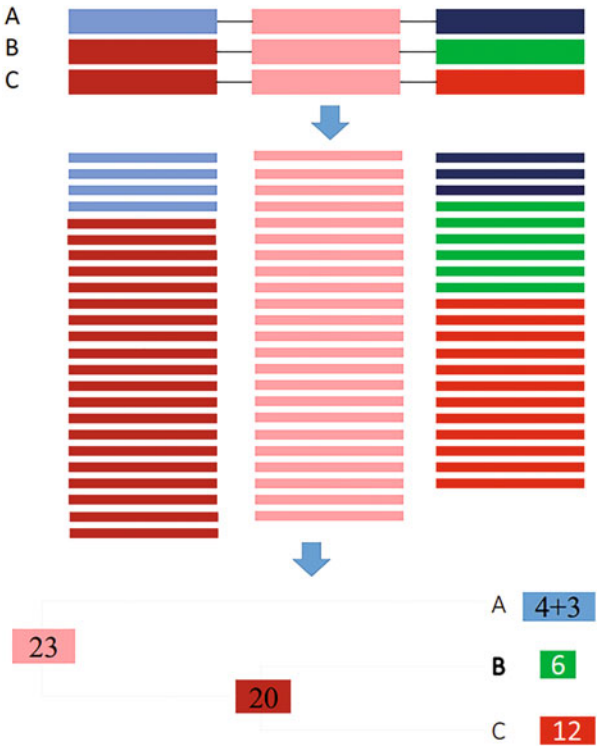


Fig. 1.4 Sharing of data within a genome family containing the paralogous genes A, B, and C, each of which has three idealised segments: a centre segment that is conserved and identical, a first segment that is strongly homologous to both B and C, and a third segment that has diverged. The colour of the reads matches that of the gene segment they match. From "Bioinformatics and Drug Discovery" by Xuhua Xia, 2017, Copyright 2017 by Xuhua Xia

these runs should be distributed evenly across the matching paralogous genes. The Multiple Mapper Resolution (MMR) tool, just released and available at https://github.com/ratschlab/mmr, aims to address this issue but provides no methodological information. Because multicellular eukaryotes have so many duplicated genes, any downstream analysis will be unreliable if RPFs are incorrectly assigned to paralogous genes. I'll describe the method used in the Tuxedo computer programme to assign RPFs to three or more paralogous genes. The assignment is quite straightforward when there are just two members of a gene family, thus it will not be covered in this article.

For the appropriate distribution of genome reads from a protein family that contains three or many alternatively spliced genomes, a phylogenetic tree is required. I use a genome that has three parts made up of the three alternatively spliced genes A, B, and C. in Fig. 1.4 to illustrate the allocation concept. The centre region of the three genes was the same for all three, and it had 23 matched reads (which must match all three paralogues equally well). The first portion of both genes B and C is the same, and 20 readings matched it. The initial section of Gene A differs from that of Genes B and C, and received four matched readings. The three genes also have a third diverged segment where paralogous genes A, B, and C all matched 3, 6, and 12 reads, respectively. The 23 reads that all three of us shared and the 20 reads that B and C shared must then be divided among the three paralogues.

TUXEDO employs a straightforward counting strategy using the following:

$$\begin{aligned} \mathrm{PA} &= \frac{3+4}{3+4+20+6+12} = 0.15556 \\ \mathrm{PB} &= (1-\mathrm{PA})\frac{6}{6+12} = 0.28148 \\ \mathrm{PC} &= (1-\mathrm{PA})\frac{12}{6+12} = 0.56296 \end{aligned} \tag{1.3}$$

Therefore, using PA, PB, and PC, we assign the 23 evenly matched reads to paralogous genes A, B, and C, respectively. We divide the 20 reads into 20*6/(6 + 12) for B and 20*12/(6 + 12) for C for 20 reads which exactly approximate B and C. The expected number of matches for each gene is provided as follows:

$$\begin{aligned} \mathrm{NA} &= 3+4+23\mathrm{PA} = 10.57778 \\ \mathrm{NB} &= 6+23\mathrm{PB}+20\left(\frac{6}{6+12}\right) = 19.14074 \\ \mathrm{NC} &= 12+23\mathrm{PC}+20\left(\frac{12}{6+12}\right) = 38.28148 \end{aligned} \tag{1.4}$$

1.7 Structural Biology and Drug Discovery

One should be able to: (1) predict the 3-D configuration of a peptide or RNA on the basis of physiological environment where it is interpreted or decoded, (2) "BLAST" a defined peptide or RNA molecule against datasets of peptide or RNA molecule to extract all peptide or RNA with alike structural features in order to simplify structural and usage interpretations and analysis of usable redundant systems of the target gene in the cell. The 3D configuration of a peptide or RNA predicated on the physiological environment where it is interpreted or decoded (Dykeman et al. 2013; Naveed et al. 2015), (3) locate and obtain all of a given query structure's potential binding partners to help evaluate the query's suitability as a drug target or drug candidate, including the effectiveness of the query and any negative effects resulting from which interacts physically with another cell elements, (4) using structural modelling and simulation instantly detect peptides and RNA which can make intricates and put those intricates together (like the ribosome and spliceosome), (5) offer novel structures that can physically interact with the query to activate or deactivate the query protein/RNA function in the cell. Steps 5 and 6 include predicting activity, whether alone or as part of the component of a network, of proteins and RNAs of defined structures. Despite not always being ideal, technologies and libraries for programming gathered at http://www.click2drug.org/ can perform almost all of these tasks.

The first thing to do when a protein piques one's curiosity is to see if the PDB already contains a structure for it (Rose et al. 2015; Westbrook et al. 2002). If not, then one can infer its structure using tools like homology modelling depending solely on a single or many structurally identified near homologs. SWISS-MODEL (Biasini et al. 2014), TASSER (Zhang et al. 2005), and their offspring are examples of such tools. Once the structure has been refined, it can be visualised UCSF Chimera (Pettersen et al. 2004), or PyMOL, and possible medication options which might affect the peptide of interest can be found using automated screening tools like SwissSimilarity (Zoete et al. 2016). The use of metabolic and ligand databases like ChEMBL (Gaulton et al. 2012) and SuperSite (Bauer et al. 2009) considerably improves this screening method.

To explore the physical interactions between proteins and tiny molecules, one can also utilise docking tools like SwissDock (Grosdidier et al. 2011) or SwissBioisostere (Wirth et al. 2013). Such structural research enhances the development of anti-HIV-1 protease medications (Heal et al. 2012). Each homodimer of the protease has 99 amino acids, and in order to interfere with the action of the enzyme, an inhibitor normally needs to wedge itself in between the two monomers (Broglia et al. 2008, Wlodawer and Erickson 1993, Wlodawer and Vondrasek 1998).

It makes sense to suppose that additional peptides to a related code or form may engage to that same reagent given a previously observed and well-documented protein–ligand interaction (Ding et al. 2014; Ekins et al. 2015). The software SwissTargetPrediction is conceptually built on this similarity-based method (Gfeller et al. 2014).

It is crucial to remember that a protein's structure behave differently depending also on cells surroundings and since the form obtained using X-ray or NMR only offers a glimpse of dynamic analysis. The characterisation of these changing engagements of peptides with their complexes is made easier by the software CHARMM (Brooks et al. 2009) and its derivatives. Such studies are aided by basic libraries of ligand–protein engagements (Hecker et al. 2012), specialist libraries describing peptide correlations in membrane surface including GPCR-ligand connections (Chan et al. 2015) or in tumour cells, or organism-specific libraries like as of *M. tuberculosis* (Kinnings et al. 2010).

1.8 Bioinformatics and Drug Resistance

After penicillin was discovered in 1928 and began to be used regularly in medicine in 1940, bacterial resistance to penicillin quickly became apparent (Abraham and Chain 1940; Abraham et al. 1941). Following the widespread use of artemisinin in Asian nations, multicellular microorganisms, like the plasmodium species, *Plasmodium falciparum*, are capable of developing alike resistance swiftly to the strongest anti-malarial medication (Noedl et al. 2008; Noedl et al. 2010; Noedl et al. 2009). Drug development is expensive, and drug resistance frequently renders expensively developed drugs useless (David et al. 2009; Drews and Ryser 1997). The quick

emergence of chemical tolerability in HIV-1 emphasises how crucial it is to comprehend drug resistance (Smyth et al. 2012; Smyth et al. 2014).

Extreme selectivity towards the microorganism is required for modern medication development against pathogens. Drug-mediated selection will only favour those with drug resistance in this specific population of bacteria if a medication is detrimental to that specific bacterium infection. Medication resistance may, however, arise in all of these species when a medication is toxic to the 100 additional non-parasitic bacteria, which frequently with subsequent transmission of drug resistance from a non-pathogenic species to a pathogenic one. Bioinformaticians have developed databases to make it easier to identify pathogenicity islands as therapeutic targets (Pundhir et al. 2008; Yoon et al. 2015). Pathogenicity islands (Gal-Mor and Finlay 2006; Hacker and Kaper 2000) are unique DNA segments found although not in its infectious counterparts, in a large variety of infectious bacterium diseases.

The speed at which microbial pathogens can evolve medication resistance has been clarified by contemporary bioinformatic analysis and creative experimentation. In one study (Belanger et al. 2002), random mutations were introduced into the *Streptococcus pneumonia* genes using error-prone PCR. Following the transformation of *S. pneumoniae* using these altered amplicons, some of the colonies that resulted showed resistance to the antibiotic fusidic acid. A single mutation in the fusA gene, identified through DNA sequence analysis, is responsible for the drug resistance. There are numerous instances when a single mutation in the HIV-1 protease can drastically alter how susceptible the protease is to inhibitors (Rhee et al. 2010; Young et al. 2010). These investigations we enable can determine the proportion of genetic drift that results in medication tolerance.

The rate of mutation, the size of the parasite population, and genetic diversity all have major roles in how quickly bacterial and eukaryotic infections can adapt to treatment resistance. Tolerance is probably due to a deficit of genomic diversity must develop from scratch, in where such situation the pace of variation becomes a substantial barrier wherein scenario the frequency of variation becomes important in the development of chemoresistance in microorganisms. Historically, the spontaneous mutation rate was assessed through laborious mutation accumulation experiments, which were typically conducted on a small number of rapidly replicating bacterial species and viruses (Drake 1964; Drake 1966). On determining the source of pseudogenes and evaluating their deviation from that of their operational versions, one can estimate the sporadic alteration frequency and alteration range (Smyth et al. 2012; Smyth et al. 2014; Pundhir et al. 2008). The chances of a pathogen becoming resistant to medications are increased by high-mutation rates and large population sizes.

Variations have been identified that have been long thought to exist independently from one another and represent distinct mutation events. Because UCU and AGU must undergo, it is indeed highly doubtful that the two serine codes in the conventional genetic sequence (UCU and AGU), which are generally exposed to severe purging choice, will ever evolve into one another. Nevertheless, omics analysis and modelling studies have demonstrated that multiple mutation events can take place in organisms in “clusters and showers” over the period of a single

generation (Schrider et al. 2011; Averof et al. 2000) as well as in viruses and bacteria species (Drake 2007a, 2007b, Drake et al. 2005).

Since genomic and morphological correlations usually indicate homology in complex formation, genomes in order to find conserved sequences or structures that can guide the creation of vaccines and ligands developed as blockers towards pathogenic microbial diseases, genome research and evolutionary biology have already been extensively explored (Manocheewa et al. 2015; O'Connell et al. 2014; Anisimova 2015).

Strongly conserved regions of a gene do not necessarily mean that mutations there will cause the gene to malfunction. The HIV-1 proteolytic enzyme has several amino acid residues that are shared by several variants of M group, suggesting that they belong to crucial for the enzyme's functionality. Protease inhibitors, on the other hand, rapidly because mutations at these highly conserved sites, decreasing the susceptibility to them (Rhee et al. 2010; Young et al. 2010). The emergence of antibiotics is analogous to how this adaptation to drug-induced selection operates. Plasmids in a bacterial species like *E. coli*, whether or not they carry antibiotic-resistant genes, represent a replication burden in the absence of antibiotics. As a result, they are promptly eliminated from *E. coli* cultures via selection. However, in the presence of antibiotics, the benefit of antibiotic resistance more than outweighs the expense of a replication burden, and the plasmids containing the antibiotic-resistant genes will proliferate.

The probability of a substance alteration will appear in the first population after drug administration, the likelihood that it won't appear until the second generation, or generally the likelihood that it won't appear until the Nth generation can all be estimated, when we are aware of the parasite's group density, its speed of random mutagenesis, as well as the ratio of drug-tolerant mutations among all random mutations in the bacterial population. The lots of decades are needed for the very first medication resistant variant to appear can also be calculated. This type of estimation falls under the purview of population genetics.

1.9 Bioinformatics Software and Database

The Swiss Institute of Bioinformatics maintains http://click2drug.org/, a comprehensive collection of tools, databases, and web services specifically relevant to drug discovery. The following are basically categorised: (1) databases, (2) chemical structure representations, (3) molecular modelling and simulation, (4) homology modelling to estimate the structure of a protein using a homologue of a known structure, (5) prediction of the binding site, (6) docking, (7) drug candidate screening, (8) prediction of the drug target, (9) design of the ligand, (10) calculation of the binding free energy, (11) QSAR, and (12) ADME-Toxicity.

Numerous software programmes are robust, available for free, and backed by well-known organisations. These include databases like ChEMBL (Gaulton et al. 2012) and SwissSidechain (Ekins et al. 2015), software programmes like UCSF

Chimera (Pettersen et al. 2004), a platform for structural biology-focused software developers, SwissSimilarity for virtual screening, SwissBioisostere for ligand design (Wirth et al. 2013), SwissTargetPrediction (Gfeller et al. 2014), SwissSideChain to facilitate experiments that expand the protein repertoire by introducing non-natural amino acids, and SwissDock for docking drug candidates. Even though certain software, such PyMOL (Schrödinger) and CHARMM (Brooks et al. 2009), is commercial, they often have free versions for students and teachers.

1.10 Conclusion

A data-driven field of study, bioinformatics, developed or modified many of its databases and algorithms in reaction to new kinds of data. This is the reason why high-throughput data gathering methods frequently outpace bioinformatic tools. But many molecular biologists, computer scientists, and mathematicians have devoted a great deal of their time and energy to creating new, potent software programmes and databases to expand our perspectives of nature, much like how microscopes and telescopes enable us to see patterns that we have never seen before. Examining this work by pharmaceutical scientists more closely could be very valuable for the bioinformatics research community as well as the pharmaceutical industry.

References

Abdel-Hameed EA, Ji H, Shata MT (2016) HIV-induced epigenetic alterations in host cells. Adv Exp Med Biol 879:27–38. https://doi.org/10.1007/978-3-319-24738-0_2

Abraham EP, Chain E (1940) An enzyme from bacteria able to destroy penicillin. Rev Infect Dis 10(3):677–678. https://doi.org/10.1038/146837a0

Abraham EP, Chain E, Fletcher CM, Florey HW, Gardner AD, Heatley NG, Jennings MA (1941) Further observations on penicillin. Lancet 238:177–189. https://doi.org/10.1016/S0140-6736(00)72122-2

Alderwick LJ, Seidel M, Sahm H, Besra GS, Eggeling L (2006) Identification of a novel arabinofuranosyltransferase (AftA) involved in cell wall arabinan biosynthesis in Mycobacterium tuberculosis. J Biol Chem 281(23):15653–15661. https://doi.org/10.1074/jbc.M600045200

Anisimova M (2015) Darwin and Fisher meet at biotech: on the potential of computational molecular evolution in industry. BMC Evol Biol 15(1):76. https://doi.org/10.1186/s12862-015-0352-y

Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D (2003) Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 100(7):3889–3894. https://doi.org/10.1073/pnas.0635171100

Arbibe L, Sansonetti PJ (2007) Epigenetic regulation of host response to LPS: causing tolerance while avoiding toll errancy. Cell Host Microbe 1(4):244–246. https://doi.org/10.1016/j.chom.2007.05.011

Arvaniti E, Moulos P, Vakrakou A, Chatziantoniou C, Chadjichristos C, Kavvadas P, Charonis A, Politis PK (2016) Whole transcriptome analysis of UUO mouse model of renal fibrosis reveals new molecular players in kidney diseases. Sci Rep 6:26235. https://doi.org/10.1038/srep26235

Averof M, Rokas A, Wolfe KH, Sharp PM (2000) Evidence for a high frequency of simultaneous double-nucleotide substitutions. Science 287(5456):1283–1286. https://doi.org/10.1126/science.287.5456.1283

Baird AL, Westwood S, Lovestone S (2015) Blood-based proteomic biomarkers of Alzheimer's disease pathology. Front Neurol 6:236. https://doi.org/10.3389/fneur.2015.00236

Bao J, Bedford MT (2016) Epigenetic regulation of the histone-toprotamine transition during spermiogenesis. Reproduction 151(5):R55–R70. https://doi.org/10.1530/REP-15-0562

Bastianelli G, Bouillon A, Nguyen C, Crublet E, Petres S, Gorgette O, Le-Nguyen D, Barale JC, Nilges M (2011) Computational reverse-engineering of a spider-venom derived peptide active against Plasmodium falciparum SUB1. PLoS One 6(7):e21812. https://doi.org/10.1371/journal.pone.0021812

Bauer RA, Gunther S, Jansen D, Heeger C, Thaben PF, Preissner R (2009) SuperSite: dictionary of metabolite and drug binding sites in proteins. Nucleic Acids Res 37(Database issue):D195–D200. https://doi.org/10.1093/nar/gkn618

Belanger AE, Lai A, Brackman MA, LeBlanc DJ (2002) PCR based ordered genomic libraries: a new approach to drug target identification for Streptococcus pneumoniae. Antimicrob Agents Chemother 46(8):2507–2512. https://doi.org/10.1128/AAC.46.8.2507-2512.2002

Bell D, Bell AH, Bondaruk J, Hanna EY, Weber RS (2016) In-depth characterization of the salivary adenoid cystic carcinoma transcriptome with emphasis on dominant cell type. Cancer 122(10):1513–1522. https://doi.org/10.1002/cncr.29959

Berger MF, Levin JZ, Vijayendran K, Sivachenko A, Adiconis X, Maguire J, Johnson LA, Robinson J, Verhaak RG, Sougnez C, Onofrio RC, Ziaugra L, Cibulskis K, Laine E, Barretina J, Winckler W, Fisher DE, Getz G, Meyerson M, Jaffe DB, Gabriel SB, Lander ES, Dummer R, Gnirke A, Nusbaum C, Garraway LA (2010) Integrative analysis of the melanoma transcriptome. Genome Res 20(4):413–427. https://doi.org/10.1101/gr.103697.109

Bergsten E, Uutela M, Li X, Pietras K, Ostman A, Heldin CH, Alitalo K, Eriksson U (2001) PDGF-D is a specific, protease activated ligand for the PDGF beta-receptor. Nat Cell Biol 3(5):512–516. https://doi.org/10.1038/35074588

Beverley SM, Owens KL, Showalter M, Griffith CL, Doering TL, Jones VC, McNeil MR (2005) Eukaryotic UDP-galactopyranose mutase (GLF gene) in microbial and metazoal pathogens. Eukaryot Cell 4(6):1147–1154. https://doi.org/10.1128/EC.4.6.1147-1154.2005

Bhatia B, Ponia SS, Solanki AK, Dixit A, Garg LC (2014) Identification of glutamate ABC-Transporter component in Clostridium perfringens as a putative drug target. Bioinformation 10(7):401–405. https://doi.org/10.6026/97320630010401

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42(Web Server issue):W252–W258. https://doi.org/10.1093/nar/gku340

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R (2011) High density DNA methylation array with single CpG site resolution. Genomics 98(4):288–295. https://doi.org/10.1016/j.ygeno.2011.07.007

Bierne H, Hamon M, Cossart P (2012) Epigenetics and bacterial infections. Cold Spring Harb Perspect Med 2(12):a010272. https://doi.org/10.1101/cshperspect.a010272

Bigaud E, Corrales FJ (2016) Methylthioadenosine (MTA) regulates liver cells proteome and methylproteome: implications in liver biology and disease. Mol Cell Proteomics 15(5):1498–1510. https://doi.org/10.1074/mcp.M115.055772

Birney E, Stamatoyannopoulos JA et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447(7146):799–816. https://doi.org/10.1038/nature05874

Boehringer D, Thermann R, Ostareck-Lederer A, Lewis JD, Stark H (2005) Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. Structure 13(11):1695–1706. https://doi.org/10.1016/j.str.2005.08.008

Bolden JE, Peart MJ, Johnstone RW (2006) Anticancer activities of histone deacetylase inhibitors. Nat Rev Drug Discov 5(9):769–784. https://doi.org/10.1038/nrd2133

Boxall AB, Rudd MA, Brooks BW, Caldwell DJ, Choi K, Hickmann S, Innes E, Ostapyk K, Staveley JP et al (2012) Pharmaceuticals and personal care products in the environment: what are the big questions? Environ Health Perspect 120(9):1221–1229. https://doi.org/10.1289/ehp.1104477

Broglia R, Levy Y, Tiana G (2008) HIV-1 protease folding and the design of drugs which do not create resistance. Curr Opin Struct Biol 18(1):60–66. https://doi.org/10.1016/j.sbi.2007.10.004

Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30(10):1545–1614. https://doi.org/10.1002/jcc.21287

Brucher BL, Jamall IS (2016) Somatic mutation theory - why it's wrong for most cancers. Cell Physiol Biochem 38(5):1663–1680. https://doi.org/10.1159/000443106

Cardon LR, Burge C, Clayton DA, Karlin S (1994) Pervasive CpG suppression in animal mitochondrial genomes. Proc Natl Acad Sci U S A 91(9):3799–3803. https://doi.org/10.1073/pnas.91.9.3799

Chan WK, Zhang H, Yang J, Brender JR, Hur J, Ozgur A, Zhang Y (2015) GLASS: a comprehensive database for experimentally validated GPCR-ligand associations. Bioinformatics 31(18):3035–3042. https://doi.org/10.1093/bioinformatics/btv302

Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, Feng GH, Peng H, Zhang X, Qian J, Duan E, Zhai Q, Zhou Q (2016) Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. Science 351(6271):397–400. https://doi.org/10.1126/science.aad7977

Chithambaram S, Prabhakaran R, Xia X (2014a) Differential codon adaptation between dsDNA and ssDNA Phages in *Escherichia coli*. Mol Biol Evol 31(6):1606–1617. https://doi.org/10.1093/molbev/msu087

Chithambaram S, Prabhakaran R, Xia X (2014b) The effect of mutation and selection on codon adaptation in Escherichia coli bacteriophage. Genetics 197(1):301–315. https://doi.org/10.1534/genetics.114.162842

Chu C, Qu K, Zhong FL, Artandi SE, Chang HY (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 44(4):667–678. https://doi.org/10.1016/j.molcel.2011.08.027

Chu C, Quinn J, Chang HY (2012) Chromatin isolation by RNA purification (ChIRP). J Vis Exp 61:3912. https://doi.org/10.3791/3912

Clark AT (2015) DNA methylation remodeling in vitro and in vivo. Curr Opin Genet Dev 34:82–87

Daily K, Patel VR, Rigor P, Xie X, Baldi P (2011) MotifMap: integrative genome-wide maps of regulatory motif sites for model species. BMC Bioinform 12:495

David E, Tramontin T, Zemmel R (2009) Pharmaceutical R & D: the road to positive returns. Nat Rev Drug Discov 8:609–610

Davies J, Davies D (2010) Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev 74:417–433

Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA (2012) Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell 149:233–1244

Deng Q, Ramskold D, Reinius B, Sandberg R (2014a) Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343:193–196

Deng W, Rupon JW, Krivega I, Breda L, Motta I, Jahn KS, Reik A, Gregory PD, Rivella S, Dean A, Blobel GA (2014b) Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell 158:849–860

Ding H, Takigawa I, Mamitsuka H, Zhu S (2014) Similarity-based machine learning methods for predicting drug-target interactions: a brief review. Brief Bioinform 15:734–747

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21

Doolittle RF, Hunkapiller MW, Hood LE, Devare SG, Robbins KC, Aaronson SA, Antoniades HN (1983) Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor. Science 221:275–277

Doudna JA, Sarnow P, Mathews MB, Sonenberg N, Hershey J (2007) Translation initiation by viral internal ribosome entry sites. In: Translational control in biology and medicine. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 129–154

Drake JW (1964) Studies on the induction of mutations in bacteriophage t4 by ultraviolet irradiation and by proflavin. J Cell Physiol 64(SUPPL 1):19–31

Drake JW (1966) Spontaneous mutations accumulating in bacteriophage T4 in the complete absence of DNA replication. Proc Natl Acad Sci U S A 55:738–743

Drake JW (2007a) Mutations in clusters and showers. Proc Natl Acad Sci U S A 104:8203–8204

Drake JW (2007b) Too many mutants with multiple mutations. Crit Rev Biochem Mol Biol 42: 247–258

Drake JW, Bebenek A, Kissling GE, Peddada S (2005) Clusters of mutations from transient hypermutability. Proc Natl Acad Sci U S A 102:12849–12854

Drews J, Ryser S (1997) The role of innovation in drug development. Nat Biotechnol 15:1318–1319

Dykeman EC, Stockley PG, Twarock R (2013) Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. J Mol Biol 425:3235–3249

Eckhardt F, Lewin J, Cortese R, Rakyan VK et al (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 38:1378–1385

Eder J, Sedrani R, Wiesmann C (2014) The discovery of first-in-class drugs: origins and evolution. Nat Rev Drug Discov 13:577–587

Ehnman M, Missiaglia E et al (2013) Distinct effects of ligand-induced PDGFR alpha and PDGFR beta signaling in the human rhabdomyosarcoma tumor cell and stroma cell compartments. Cancer Res 73:2139–2149

Ekins S, de Siqueira-Neto JL, McCall LI, Sarker M, Yadav M, Ponder EL, Kallel EA, Kellar D, Chen S, Arkin M, Bunin BA, McKerrow JH, Talcott C (2015) Machine learning models and pathway genome data base for *Trypanosoma cruzi* drug discovery. PLoS Negl Trop Dis 9: e0003878

Elroy-Stein O, Merrick W, Mathews MB, Sonenberg N, Hershey J (2007) Translation initiation via cellular internal ribosome entry sites. In: Translational control in biology and medicine. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 155–172

Fernandez-Pinar R, Lo Sciuto A, Rossi A, Ranucci S, Bragonzi A, Imperi F (2015) In vitro and in vivo screening for novel essential cell-envelope proteins in Pseudomonas aeruginosa. Sci Rep 5:7593

Forrester WC, Epner E, Driscoll MC, Enver T, Brice M, Papayannopoulou T, Groudine M (1990) A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. Genes Dev 4:1637–1649

Furukawa R, Hachiya T, Ohmomo H, Shiwa Y, Ono K, Suzuk S, Satoh M, Hitomi J, Sobue K, Shimizu A (2016) Intraindividual dynamics of transcriptome and genome-wide stability of DNA methylation. Sci Rep 6:26424

Gabrielsson J, Green AR (2009) Quantitative pharmacology or pharmacokinetic pharmacodynamic integration should be a vital component in integrative pharmacology. J Pharmacol Exp Ther 331:767–774

Gal-Mor O, Finlay BB (2006) Pathogenicity islands: a molecular toolbox for bacterial virulence. Cell Microbiol 8:1707–1719
Gapp K, Jawaid A, Sarkies P, Bohacek J, Pelczar P, Prados J, Farinelli L, Miska E, Mansuy IM (2014) Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. Nat Neurosci 17:667–669
Garraway LA, Lander ES (2013) Lessons from the cancer genome. Cell 153:17–37
Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, AlLazikani B, Overington JP (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40:D1100–D1107
Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S (2005) Bioinformatics and computational biology solutions using R and bioconductor. Springer, New York, p 473
Gfeller D, Grosdidier A, Wirth M DA, Michielin O, Zoete V (2014) SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res 42:W32–W38
Gibbs JB (2000) Mechanism-based target identification and drug discovery in cancer research. Science 287:1969–1973
Gilbert WV, Zhou K, Butler TK, Doudna JA (2007) Cap-independent translation is required for starvation-induced differentiation in yeast. Science 317:1224–1227
Goto M, Washio T, Tomita M (2000) Causal analysis of CpG suppression in the Mycoplasma genome. Microb Comp Genomics 5:51–58
Grigg G, Clark S (1994) Sequencing 5-methylcytosine residues in genomic DNA. BioEssays 16: 431–436
Grigg GW (1996) Sequencing 5-methylcytosine residues by the bisulfite method. DNA Seq 6:189–198
Grosdidier A, Zoete V, Michielin O (2011) SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res 39:W270–W277
Gruber TD, Borrok MJ, Westler WM, Forest KT, Kiessling LL (2009) Ligand binding and substrate discrimination by UDP-galactopyranose mutase. J Mol Biol 391:327–340
Hacker J, Blum-Oehler G, Muhldorfer I, Tschape H (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol Microbiol 23:1089–1097
Hacker J, Kaper JB (2000) Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol 54:641–679
Haustead DJ, Stevenson A, Saxena V, Marriage F, Firth M, Silla R, Martin L, Adcroft KF, Rea S, Day PJ, Melton P, Wood FM, Fear MW (2016) Transcriptome analysis of human ageing in male skin shows mid-life period of variability and central role of NF-kappaB. Sci Rep 6:26846
Heal JW, Jimenez-Roldan JE, Wells SA, Freedman RB, Romer RA (2012) Inhibition of HIV-1 protease: the rigidity perspective. Bioinformatics 28:350–357
Heath JR, Ribas A, Mischel PS (2016) Single-cell analysis tools for drug discovery and development. Nat Rev Drug Discov 15:204–216
Hecker N, Ahmed J, von Eichborn J, Dunkel M, Macha K, Eckert A, Gilson MK, Bourne PE, Preissner R (2012) SuperTarget goes quantitative: update on drug-target interactions. Nucleic Acids Res 40:D1113–D1117
Higgs PG, Attwood TK (2013) Bioinformatics and molecular evolution. Wiley, Chichester
Hofer A, Steverding D, Chabes A, Brun R, Thelander L (2001) Trypanosoma brucei CTP synthetase: a target for the treatment of African sleeping sickness. Proc Natl Acad Sci U S A 98:6412–6416
Holford NH, Sheiner LB (1981a) Pharmacokinetic and pharmacodynamic modeling in-vivo. Crit Rev Bioeng 5:273–322
Holford NH, Sheiner LB (1981b) Understanding the dose-effect relationship: clinical application of pharmacokinetic-pharmacodynamic models. Clin Pharmacokinet 6:429–453
Hou C, Zhao H, Tanimoto K, Dean A (2008) CTCF-dependent enhancer-blocking by alternative chromatin loop formation. Proc Natl Acad Sci U S A 105:20398–20403
Hua S, Sun Z (2001) Support vector machine approach for protein subcellular localization prediction. Bioinformatics 17:721–728

Huang HY, Chien CH, Jen KH, Huang HD (2006) RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. Nucleic Acids Res 34:W429–W434

Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJ, Jackson SE, Wills MR, Weissman JS (2014) Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. Cell Rep 8:1365–1379

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223

Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147:789–802

Ingrosso D, Cimmino A, Perna AF, Masella L, De Santo NG, De Bonis ML, Vacca M, D'Esposito M, D'Urso M, Galletti P, Zappia V (2003) Folate treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinaemia in patients with uraemia. Lancet 361:1693–1699

Ingrosso D, Perna AF (2009) Epigenetics in hyperhomocysteinemic states. A special focus on uremia. Biochim Biophys Acta 1790:892–899

Insinga A, Minucci S, Pelicci PG (2005a) Mechanisms of selective anticancer action of histone deacetylase inhibitors. Cell Cycle 4:741–743

Insinga A, Monestiroli S, Ronzoni S, Gelmetti V, Marchesi F, Viale A, Altucci L, Nervi C, Minucci S, Pelicci PG (2005b) Erratum: Inhibitors of histone deacetylases induce tumor-selective apoptosis through activation of the death receptor pathway. Nat Med 11:71–76

Jackson RJ, Hellen CU, Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol 11:113–127

Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3: 318–356

Jan E, Sarnow P (2002) Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. J Mol Biol 324:889–902

Jan E, Thompson SR, Wilson JE, Pestova TV, Hellen CU, Sarnow P (2001) Initiator Met-tRNA-independent translation mediated by an internal ribosome entry site element in cricket paralysis virus-like insect viruses. Cold Spring Harb Symp Quant Biol 66:285–292

Jiang YH, Bressler J, Beaudet AL (2004) Epigenetics and human disease. Annu Rev Genomics Hum Genet 5:479–510

Jin P, Alisch RS, Warren ST (2004) RNA and microRNAs in fragile X mental retardation. Nat Cell Biol 6:1048–1053

Kanehisa M (2013) Molecular network analysis of diseases and drugs in KEGG. Methods Mol Biol 939:263–275

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D642

Kincaid VA, London N, Wangkanont K, Wesener DA, Marcus SA, Heroux A, Nedyalkova L, Talaat AM, Forest KT, Shoichet BK, Kiessling LL (2015) Virtual screening for UDP-galactopyranose mutase ligands identifies a new class of antimycobacterial agents. ACS Chem Biol 10:2209–2218

Kinnings SL, Xie L, Fung KH, Jackson RM, Bourne PE (2010) The Mycobacterium tuberculosis drugome and its polypharmacological implications. PLoS Comput Biol 6:e1000976

Kioussis D, Vanin E, deLange T, Flavell RA, Grosveld FG (1983) Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. Nature 306:662–666

Komar AA, Hatzoglou M (2005) Internal ribosome entry sites in cellular mRNAs: mystery of their existence. J Biol Chem 280:23425–23428

Korenke GC, Fuchs S, Krasemann E, Doerr HG, Wilichowski E, Hunneman DH, Hanefeld F (1996) Cerebral adrenoleukodystrophy (ALD) in only one of monozygotic twins with an identical ALD genotype. Ann Neurol 40:254–257

Kotokorpi P, Venteclef N, Ellis E, Gustafsson JA, Mode A (2010) The human ADFP gene is a direct liver-X-receptor (LXR) target gene and differentially regulated by synthetic LXR ligands. Mol Pharmacol 77:79–86

Kozak M (1980) Evaluation of the "scanning model" for initiation of protein synthesis in eucaryotes. Cell 22:7–8

Krasemann EW, Meier V, Korenke GC, Hunneman DH, Hanefeld F (1996) Identification of mutations in the ALD-gene of 20 families with adrenoleukodystrophy/adrenomyeloneuropathy. Hum Genet 97:194–197

Kungulovski G, Jeltsch A (2016) Epigenome editing: state of the art, concepts, and perspectives. Trends Genet 32:101–113

Kutlar A (2007) Sickle cell disease: a multigenic perspective of a single gene disorder. Hemoglobin 31:209–224

Langmead B, Hansen KD, Leek JT (2010) Cloud-scale RNA sequencing differential expression analysis with Myrna. Genome Biol 11:R83

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293

Ling S, Hu Z, Yang Z, Yang F, Li Y, Lin P, Chen K, Dong L, Cao L, Tao Y, Hao L, Chen Q, Gong Q, Wu D, Li W, Zhao W, Tian X, Hao C, Hungate EA, Catenacci DV, Hudson RR, Li WH, Lu X, Wu CI (2015) Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. Proc Natl Acad Sci U S A 112:E6496–E6505

Liu X, Jiang H, Gu Z, Roberts JW (2015) High-resolution view of bacteriophage lambda gene expression by ribosome profiling. Proc Natl Acad Sci U S A 110:11928–11933

Ma P, Xia X (2011) Factors affecting splicing strength of yeast genes. Comp Funct Genomics 2011: 212146, 13 pages

MacKay VL, Li X, Flory MR, Turcott E, Law GL, Serikawa KA, Xu XL, Lee H, Goodlett DR, Aebersold R, Zhao LP, Morris DR (2004) Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone. Mol Cell Proteomics 3:478–489

Manocheewa S, Mittler JE, Samudrala R, Mullins JI (2015) Composite sequence-structure stability models as screening tools for identifying vulnerable targets for HIV drug and vaccine development. Viruses 7:5718–5735

Miller LH, Su X (2011) Artemisinin: discovery from the Chinese herbal garden. Cell 146:855–858

Mlera L, Lam J, Offerdahl DK, Martens C, Sturdevant D, Turner CV, Porcella SF, Bloom ME (2016) Transcriptome analysis reveals a signature profile for tick-borne flavivirus persistence in HEK 293T cells. MBio 7

Moffat JG, Rudolph J, Bailey D (2014) Phenotypic screening in cancer drug discovery - past, present and future. Nat Rev Drug Discov 13:588–602

Morita M, Shimozawa N, Kashiwayama Y, Suzuki Y, Imanaka T (2011) ABC subfamily D proteins and very long chain fatty acid metabolism as novel targets in adrenoleukodystrophy. Curr Drug Targets 12:694–706

Muller HJ, Altenburg E (1930) The frequency of translocations produced by X-rays in drosophila. Genetics 15:283–311

Muller PY, Milton MN (2012) The determination and interpretation of the therapeutic index in drug development. Nat Rev Drug Discov 11:751–761

Murphy J, Mahony J, Ainsworth S, Nauta A, van Sinderen D (2013) Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. Appl Environ Microbiol 79:7547–7555

Naveed H, Hameed US, Harrus D, Bourguet W, Arold ST, Gao X (2015) An integrated structure- and system-based framework to identify new targets of metabolites and known drugs. Bioinformatics 31:3922–3929

Noedl H, Se Y, Schaecher K, Smith BL, Socheat D, Fukuda MM (2008) Evidence of artemisinin-resistant malaria in western Cambodia. N Engl J Med 359:2619–2620

Noedl H, Se Y, Sriwichai S, Schaecher K, Teja-Isavadharm P, Smith B, Rutvisuttinunt W, Bethell D, Surasri S, Fukuda MM, Socheat D, Chan Thap L (2010) Artemisinin resistance in Cambodia: a clinical trial designed to address an emerging problem in Southeast Asia. Clin Infect Dis 51:e82–e89

Noedl H, Socheat D, Satimai W (2009) Artemisinin-resistant malaria in Asia. N Engl J Med 361: 540–541

O'Connell RJ, Kim JH, Excler JL (2014) The HIV-1 gp120 V1V2 loop: structure, function and importance for vaccine development. Expert Rev Vaccines 13:1489–1500

Ohta T, Gray TA, Rogan PK, Buiting K, Gabriel JM, Saitoh S, Muralidhar B, Bilienska B, Krajewska-Walasek M, Driscoll DJ, Horsthemke B, Butler MG, Nicholls RD (1999) Imprinting-mutation mechanisms in Prader-Willi syndrome. Am J Hum Genet 64:397–413

Ow GS, Ivshina AV, Fuentes G, Kuznetsov VA (2014) Identification of two poorly prognosed ovarian carcinoma subtypes associated with CHEK2 germ-line mutation and non-CHEK2 somatic mutation gene signatures. Cell Cycle 13:2262–2280

Ozsvari B, Lamb R, Lisanti MP (2016) Repurposing of FDA approved drugs against cancer—focus on metastasis. Aging (Albany NY) 8:567–568

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 32:232–246

Pauling L, Itano HA, Singer SJ, Wells IC (1949) Sickle cell anemia a molecular disease. Science 110:543–548

Pedersen LL, Turco SJ (2003) Galactofuranose metabolism: a potential target for antimicrobial chemotherapy. Cell Mol Life Sci 60:259–266

Pereira B, Chin SF, Rueda OM et al (2016) The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun 7:11479

Pestova TV, Lomakin IB, Hellen CU (2004) Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. EMBO Rep 5:906–913

Pestova TV, Shatsky IN, Fletcher SP, Jackson RJ, Hellen CU (1998) A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. Genes Dev 12:67–83

Petronis A (2004) The origin of schizophrenia: genetic thesis, epigenetic antithesis, and resolving synthesis. Biol Psychiatry 55:965–970

Petronis A (2006) Epigenetics and twins: three variations on the theme. Trends Genet 22:347–350

Petronis A, Gottesman II, Kan P, Kennedy JL, Basile VS, Paterson AD, Popendikyte V (2003) Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? Schizophr Bull 29:169–178

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612

Pietras K, Sjoblom T, Rubin K, Heldin CH, Ostman A (2003) PDGF receptors as cancer drug targets. Cancer Cell 3:439–443

Popa A, Lebrigand K, Barbry P, Waldmann R (2016) Pateamine A sensitive ribosome profiling reveals the scope of translation in mouse embryonic stem cells. BMC Genomics 17:52

Prabhakaran R, Chithambaram S, Xia X (2015) E. coli and Staphylococcus phages: effect of translation initiation efficiency on differential codon adaptation mediated by virulent and temperate lifestyles. J Gen Virol 96:1169. https://doi.org/10.1099/vir.0.000050

Pundhir S, Vijayvargiya H, Kumar A (2008) PredictBias: a server for the identification of genomic and pathogenicity islands in prokaryotes. In Silico Biol 8:223–234

Rhee SY, Taylor J, Fessel WJ, Kaufman D, Towner W, Troia P, Ruane P, Hellinger J, Shirvani V, Zolopa A, Shafer RW (2010) HIV-1 protease mutations and protease inhibitor cross-resistance. Antimicrob Agents Chemother 54:4253–42561

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129:1311–1323

Roberts A, Schaeffer L, Pachter L (2013) Updating RNA-Seq analyses after re-annotation. Bioinformatics 29:1631–1637

Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol 12:R22

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S (2007) Genome-wide profiles of STAT1 DNA association using chromatin immune-precipitation and massively parallel sequencing. Nat Methods 4:651–657

Rodgers AB, Morgan CP, Leu NA, Bale TL (2015) Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. Proc Natl Acad Sci U S A 112:13699–13704

Rose PW, Prlic A, Bi C, Bluhm WF, Christie CH, Dutta S, Green RK, Goodsell DS, Westbrook JD, Woo J, Young J, Zardecki C, Berman HM, Bourne PE, Burley SK (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res 43:D345–D356

Rouchka EC (1997) A brief overview of Gibbs sampling. IBC Statistics Study Group, Washington University, Institute for Biomedical Computing

Saadatpour A, Lai S, Guo G, Yuan GC (2015) Single-cell analysis in cancer genomics. Trends Genet 31:576–586

Schrider DR, Hourmozdi JN, Hahn MW (2011) Pervasive multi-nucleotide mutational events in eukaryotes. Curr Biol 21:1051–1054

Schuler M, Connell SR, Lescoute A, Giesebrecht J, Dabrowski M, Schroeer B, Mielke T, Penczek PA, Westhof E, Spahn CM (2006) Structure of the ribosome-bound cricket paralysis virus IRES RNA. Nat Struct Mol Biol 13:1092–1096

Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Belleannee C, Kucukural A, Serra RW, Sun F, Song L, Carone BR, Ricci EP, Li XZ, Fauquier L, Moore MJ, Sullivan R, Mello CC, Garber M, Rando OJ (2016) Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science 351:391–396

Shoemaker R, Deng J, Wang W, Zhang K (2010) Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Res 20:883–889

Shoemaker RH (2006) The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer 6:813–823

Singh VK, Chang HH, Kuo CC, Shiao HY, Hsieh HP, Coumar MS (2016) Drug repurposing for chronic myeloid leukemia: in silico and in vitro investigation of DrugBank database for allosteric Bcr-Abl inhibitors. J Biomol Struct Dyn 35:1833

Smircich P, Eastman G, Bispo S, Duhagon MA, Guerra-Slompo EP, Garat B, Goldenberg S, Munroe DJ, Dallagiovanna B, Holetz F, Sotelo-Silveira JR (2015) Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in Trypanosoma cruzi. BMC Genomics 16:443

Smyth RP, Davenport MP, Mak J (2012) The origin of genetic diversity in HIV-1. Virus Res 169:415–429

Smyth RP, Schlub TE, Grimm AJ, Waugh C, Ellenberg P, Chopra A, Mallal S, Cromer D, Mak J, Davenport MP (2014) Identifying recombination hot spots in the HIV-1 genome. J Virol 88:2891–2902

Snell TW, Johnston RK, Srinivasan B, Zhou H, Gao M, Skolnick J (2016) Repurposing FDA-approved drugs for anti-aging therapies. Biogerontology 17(5–6):907–920

Sonenberg N, Meerovitch K (1990) Translation of poliovirus mRNA. Enzyme 44:278–291

Song F, Amos CI, Lee JE, Lian CG et al (2014) Identification of a melanoma susceptibility locus and somatic mutation in TET2. Carcinogenesis 35:2097–2101
Steinberg MH, Rodgers GP (2001) Pathophysiology of sickle cell disease: role of cellular and genetic modifiers. Semin Hematol 38:299–306
Swinney DC (2009) The role of binding kinetics in therapeutically useful drug action. Curr Opin Drug Discov Devel 12:31–39
Swinney DC (2013) Phenotypic vs. target-based drug discovery for first-in-class medicines. Clin Pharmacol Ther 93:299–301
Swinney DC, Anthony J (2011) How were new medicines discovered? Nat Rev Drug Discov 10: 507–519
Swinney ZT, Haubrich BA, Xia S, Ramesha C, Gomez SR, Guyett P, Mensa-Wilmot K, Swinney DC (2016) A four-point screening method for assessing molecular mechanism of action (MMOA) identifies tideglusib as a time-dependent inhibitor of Trypanosoma brucei GSK3beta. PLoS Negl Trop Dis 10:e0004506
Tanabe M, Kanehisa M (2012) Using the KEGG database resource. Curr Protoc Bioinformatics Chapter 1:1.12.1. Unit 1:12
Taramelli R, Kioussis D, Vanin E, Bartram K, Groffen J, Hurst J, Grosveld FG (1986) Gamma delta beta-thalassaemias 1 and 2 are the result of a 100 kbp deletion in the human beta-globin cluster. Nucleic Acids Res 14:7017–7029
Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111
Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seqexperiments with TopHat and Cufflinks. Nat Protoc 7:562–578
Ulferts R, de Boer SM, van der Linden L, Bauer L, Lyoo HR, Mate MJ, Lichiere J, Canard B, Lelieveld D, Omta W, Egan D, Coutard B, van Kuppeveld FJ (2016) Screening of a library of FDA-approved drugs identifies several enterovirus replication inhibitors that target viral protein 2C. Antimicrob Agents Chemother 60:2627–2638
Vlasschaert C, Xia X, Coulombe J, Gray DA (2015) Evolution of the highly networked deubiquitinating enzymes USP4, USP15, and USP11. BMC Evol Biol 15:230
Vlasschaert C, Xia X, Gray DA (2016) Selection preserves Ubiquitin Specific Protease 4 alternative exon skipping in therian mammals. Sci Rep 6:20039
Voelter-Mahlknecht S (2016) Epigenetic associations in relation to cardiovascular prevention and therapeutics. Clin Epigenetics 8:4
Wade PA, Wolffe AP (2001) Recognizing methylated DNA. Nat Struct Biol 8:575–577
Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO, von Mering C (2012) PaxDb, a database of protein abundance averages across all three domains of life. Mol Cell Proteomics 11:492–500
Waterfield MD, Scrace GT, Whittle N, Stroobant P, Johnsson A, Wasteson A, Westermark B, Heldin CH, Huang JS, Deuel TF (1983) Platelet-derived growth factor is structurally related to the putative transforming protein p28sis of simian sarcoma virus. Nature 304:35–39
Wesener DA, May JF, Huffman EM, Kiessling LL (2013) UDP-galactopyranose mutase in nematodes. Biochemistry 52:4391–4398
Westbrook J, Feng Z, Jain S, Bhat TN, Thanki N, Ravichandran V, Gilliland GL, Bluhm W, Weissig H, Greer DS, Bourne PE, Berman HM (2002) The Protein Data Bank: unifying the archive. Nucleic Acids Res 30:245–248
Wirth M, Zoete V, Michielin O, Sauer WH (2013) SwissBioisostere: a database of molecular replacements for ligand design. Nucleic Acids Res 41:D1137–D1143
Wishart DS (2016a) Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov 15(7):473–484
Wishart DS (2016b) Introduction to cheminformatics. Curr Protoc Bioinformatics 53:1411–1421
Wlodawer A, Erickson JW (1993) Structure-based inhibitors of HIV-1 protease. Annu Rev Biochem 62:543–585

Wlodawer A, Vondrasek J (1998) Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. Annu Rev Biophys Biomol Struct 27:249–284
Wu J, Tzanakakis ES (2013) Deconstructing stem cell population heterogeneity: single-cell analysis and modeling approaches. Biotechnol Adv 31:1047–1062
Xia J, Psychogios N, Young N, Wishart DS (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res 37:W652–W660
Xia X (2003) DNA methylation and mycoplasma genomes. J Mol Evol 57:S21–S28
Xia X (2007) Bioinformatics and the cell: modern computational approaches in genomics, proteomics and transcriptomics. Springer, New York, p 349
Xia X (2012) Position Weight Matrix, Gibbs Sampler, and the associated significance tests in motif characterization and prediction. Scientifica 2012:917540, 15pp.
Xia X (2013) DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. Mol Biol Evol 30:1720–1728
Xia X (2016a) ARSDA: a Quick Start. XiaLab. http://dambe.bio.uottawa.ca/ARSDA/ARSDA.aspx.Ottawa2016. Accessed 3 Mar 2017
Xia X (2016b) PhyPA: phylogenetic method with pairwise sequence alignment outperforms likelihood methods in phylogenetics involving highly diverged sequences. Mol Phylogenet Evol 102:331–343
Xia X, MacKay V, Yao X, Wu J, Miura F, Ito T, Morris DR (2011) Translation initiation: a regulatory role for poly(A) tracts in front of the AUG codon in Saccharomyces cerevisiae. Genetics 189:469–478
Xia X, Xie Z (2001) AMADA: analysis of microarray data. Bioinfomatics 17:569–570
Xie X, Rigor P, Baldi P (2009) MotifMap: a human genome-wide map of candidate regulatory motif sites. Bioinformatics 25:167–174
Yoon JH, De S, Srikantan S et al (2014) PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. Nat Commun 5:5248
Yoon SH, Park YK, Kim JF (2015) PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. Nucleic Acids Res 43:D624–D630
Young TP, Parkin NT, Stawiski E, Pilot-Matias T, Trinh R, Kempf DJ, Norton M (2010) Prevalence, mutation patterns, and effects on protease inhibitor susceptibility of the L76V mutation in HIV-1 protease. Antimicrob Agents Chemother 54:4903–4906
Yu Y, Sweeney TR, Kafasla P, Jackson RJ, Pestova TV, Hellen CU (2011) The mechanism of translation initiation on Aichivirus RNA mediated by a novel type of picornavirus IRES. EMBO J J30:4423–4436
Zhang W, Tan AY, Blumenfeld J, Liu G, Michaeel A, Zhang T, Robinson BD, Salvatore SP, Kapur S, Donahue S, Bobb WO, Rennert H (2016) Papillary renal cell carcinoma with a somatic mutation in MET in a patient with autosomal dominant polycystic kidney disease. Cancer Genet 209:11–20
Zhang Y, Arakaki AK, Skolnick J (2005) TASSER: an automated method for the prediction of protein tertiary structures in CASP6. Proteins 61(Suppl 7):91–98
Zien A, Ratsch G, Mika S, Scholkopf B, Lengauer T, Muller KR (2000) Engineering support vector machine kernels that recognize translation initiation sites. Bioinformatics 16:799–807
Zoete V, Daina A, Bovigny C, Michielin O (2016) SwissSimilarity: a web tool for low to ultra high throughput ligand-based virtual screening. J Chem Inf Model 56:1399–1404
Zoghbi HY, Beaudet AL (2016) Epigenetics and human disease. Cold Spring Harb Perspect Biol 8: a019497

Chapter 2
Computational Modelling and Simulations in Drug Design

Akansha Agrwal

2.1 Introduction

Currently, computational chemistry, organic synthesis, chemical and structural biology, and pharmacology are all involved in the process of discovering new drugs. As a result, it is divided into numerous stages. In order to priorities and choose candidates with the best potential for development as a safe and effective medicine, both in vitro and in vivo investigations are carried out.

It is commonly acknowledged that the process of discovering and developing novel drugs is time-consuming, dangerous, and expensive. From concept to market, the typical drug discovery and development cycle lasts about 14 years and costs between $800 million and $1 billion USD (Myers and Baker 2001; Moses et al. 2005). Despite the fact that there has been a large increase in investment in the past few decades, the output is not positively correlated with the expenditure due to the low efficiency in addition to elevated failure rate of drug research. As a result, numerous strategies have been created to shorten the research cycle, lower costs, and eliminate failure risk in quest for new drugs. One of the most efficient ways to accomplish these objectives is through computer-aided drug design (CADD) (Ou-Yang et al. 2012; Prieto-Martínez et al. 2019). For storage, management, analysis, and modelling of chemicals, the term "CADD" is frequently used to refer to computational tools and sources. It encompasses a wide range of topics related to drug discovery, such as computer programmes for generating compounds, instruments for methodically evaluating possible lead candidates, and the creation of digital repositories for researching chemical interactions (Song et al. 2009; Baig et al. 2018). Leading pharmaceutical corporations and research organisations have

A. Agrwal (✉)
Department of Applied Science, KIET Group of Institutions, Ghaziabad, India
e-mail: akansha.agrwal@kiet.edu

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_2

accelerated the drug discovery and development process by using computer-aided drug discovery tools in preliminary investigations to reduce costs and failures in the final stage (Yu and Mackerell 2017). Drug development is increasingly relying on CADD approaches, which are essential for efficiently identifying viable drug candidates. These computational techniques help medicinal chemists and pharmacologists during the drug discovery process by reducing the usage of animal models in pharmacological research, assisting in the rational design of novel and safe drug candidates, and repositioning already-marketed medications (Brogi et al. 2020). In addition, structure–activity relationships (SARs) are created to identify relevant pharmacokinetic and pharmacodynamic characteristics that can be used to evaluate analogues that are synthesized. (a) Target identification entails the finding and isolation of individual targets in order to research their roles and connections to a particular disease. (b) Target validation is the process of establishing a connection between the therapeutic target and the desired disease as well as the target's ability to control biological processes in the body after attaching to a partner molecule. Numerous investigations are conducted to determine whether the target macromolecule and the sick condition are related. (c) The discovery of a synthetic chemical that demonstrates some potency and selectivity against a biological target and is presumed to contain the makings of a medication that can treat the intended (d) Lead optimisation entails enhancing potency and other crucial characteristics through repeated evaluations of the lead compound(s) and its analogues. (e) Research on drug production and formulation, in vivo tests on animals to determine toxicity and potency, and characterisation of mechanistic toxicity are all part of the preclinical stage. (f) Clinical studies consist of three phases that examine the proposed drug's pharmacokinetics and pharmacological properties as well as its safety, adverse side effects, dose, effectiveness, and safety in human volunteers.

CADD has been crucial in the identification of numerous pharmaceuticals that are currently on the market, have received FDA approval, and are available (Kitchen et al. 2004; Clark 2006; Talele et al. 2010). The discipline of CADD is developing quickly, and new techniques and technologies are constantly created. It holds great promise for the workflow of drug discovery.

The rising application of information technology in the discovery of novel molecular entities encourages the use of modern molecular modelling tools to assist in the teaching of critical aspects of drug design to undergraduate students of chemistry and pharmacy. When developing and optimizing a leading drug, statistical models like quantitative structure–activity relationships (QSAR)—often in its 3D QSAR variant—are frequently used (Herwig 2014; Ragno et al. 2020).

Molecular dynamics (MD) in drug designing is a scientific technique for investigating how atoms and molecules interact and move in accordance with Newtonian physics. The forces between interacting atoms are estimated using a force field, and the system's overall energy is computed when Newton's equations of motion are integrated during MD simulations; the resulting configurations of the evolving system produce trajectories that specify the positions and velocities of particles across time. Numerous properties, such as free energy, kinetics measurements, and other macroscopic values, which may be compared with experimental observables,

can be determined from these MD trajectories (De Vivo et al. 2016). Not only MD simulation demonstrates structural variation in response to environmental variables like pH, temperature, and residue mutations, but it can also show the dynamic process of protein or peptide misfolding with aggregation (Santini and Derreumaux 2004; Campos et al. 2010; Chen et al. 2014).

In this chapter, we will discuss the whole process, methods, and applications of computer-aided drug designing.

2.2 Computational Modelling and Methods in Drug Discovery

2.2.1 Structure-Based CADD

"Structure-based drug design" (SBDD) is generally effective and potent procedure in the whole drug finding concept (Batool et al. 2019). "De novo drug design (DNDD)"and virtual screening are both steps in the process known as "SBDD." Utilisation of computational resources (ADMET) speeds up the drug development procedure, which involves combinatorial chemistry, many selection techniques, and computation of possessions like "absorption, distribution, metabolism, excretion, and toxicity." SBDD is a process that goes all over several cycles and produces an improved drug contender ready for experimental testing. The drug finding procedure typically consists of four stages: the discovery phase, the development phase, the phases of clinical trials, and registries. Scientists and pharmaceutical businesses frequently employ the computational method known as SBDD. The market has a wide variety of medications that have been identified by SBDD. The primary success story of SBDD is the development of FDA-approved HIV-1 inhibitors (Wlodawer and Vondrasek 2003). Additional medications discovered using the SBDD method are the thymidylate synthase inhibitor raltitrexed and the antibiotic norfloxacin (Rutenber and Stroud 1996; Anderson 2003). Other success cases of SBDD are Raltitrexed (Anderson 2003), Amprenavir (Clark 2006), Isoniazid (Marrakchi et al. 2000), Pim-1KinaseInhibitors (Ren et al. 2011), Norfloxacin and Dorzolamide (Grover et al. 2006).

2.2.2 Ligand-Based CADD

In the need of an experimental 3D arrangement, "ligand-based drug design" techniques are helpful. Owed to the lack of an investigational arrangement, it is necessary to analyse the identified ligand molecules which bind to drug targets in order to comprehend the structural and physicochemical characteristics of the ligands which correspond with the intended pharmacological action of those ligands (Mason et al.

2005; Bernard et al. 2005; González et al. 2009; Guner et al. 2012). Pharmacophore modelling, QSARs", and artificial intelligence (AI), are a few methods frequently utilised in the ligand-based virtual screening methodology. A pharmacophore model reveals the spatial arrangement of chemical characteristics of ligands necessary for dealings through the object receptor (Schaller et al. 2020). Negatively charged ionisable groups, hydrophobic regions, positively charged ionisable groups, hydrogen bond donors, acceptors, and aromatic ring systems are some of the chemical characteristics employed in pharmacophore modelling (Schaller et al. 2020). Catalyst, ligand scout, PHASE, PharmMapper, and GALAHAD are a few regularly used applications that enable automatic generation of the pharmacophore model (Hecker et al. 2002; Wolber and Langer 2004). A computational technique called QSAR is used to measure the connection among a given chemical or biological process and the chemical structures of a group of chemicals. Equivalent structural or physiochemical qualities should result in similar activity, according to the fundamental theory of the QSAR approach (Akamatsu 2005; Verma and Hansch 2009). The foundation of QSAR investigations is the idea that differences in a compound's bioactivity can be linked to changes in its molecular structure. They are frequently utilised in the strike to direct identification or lead optimisation stage of the drug finding procedure. The correlation studies which are used to build a statistical model, and this resulting model be capable of forecasting the biological action of novel molecules (Melo-Filho et al. 2014). There must be an adequate figure of data sets through biological activities beginning frequent investigational protocol, the training and test set compounds should be carefully chosen, there must be no auto connection between the physiochemical property of the ligands that could lead to above appropriate of data, and the final model's applicability and predictability must be verified using both internal and external validates (Cherkasov et al. 2014). Depending on how the descriptors are created, there are six main forms of QSAR that can be distinguished: (i) 1D-QSAR, is the method which examines the connection between biological processes and some characteristics like logP and pKa; (ii) 2D-QSAR, is a method which investigates the connection between biological activity and structural pattern, including connectivity index and 2D-pharmacophores; and (iii) 3D-QSAR, is a method which investigates the connection between ligands' non-covalent interaction fields with biological activities, (iv) 4D-QSAR, is an expansion of 3D-QSAR which incorporates a group of ligand configurations, and (v) 5D-QSAR, which expands 4D-QSAR by using multiple induced-fit models (Patel et al. 2014). The HypoGen unit of Catalyst, PHASE (Dixon et al. 2006), comparative molecular field analysis (CoMFA) (Cramer et al. 1988), and comparative similarity indices analysis are some examples of 3D QSAR programmes (CoMSIA) (Klebe et al. 1994). In the chemical, biological, and engineering sciences, deterioration or categorisation models are used to provide quantitative structure–activity connection models. Regression methods like QSAR associate a variety of "predictive" variables with potency, just like other regression models. Followings are various online portals to find QSAR of test molecules.

[A] VEGA platform [https://www.vegahub.eu/portfolio-item/vega-qsar/].

We can access a variety of QSAR models using the VEGA platform for regulatory purposes or create your own for research purposes. A chemical compound's property can be predicted using QSAR models by using data from the compound's structure.

[B] DEMETRA [http://www.demetra-tox.net/].

An EU-funded initiative is DEMETRA. In particular, pesticide compounds, potential pesticides, and their derivatives were the focus of this effort, which aimed to create software and predictive models that provide a quantitative estimate of the toxicity of a chemical. The chemical makeup of the substance serves as the input. Software algorithms make use of QSARs. To assess the toxicity of pesticide molecules and related chemicals, utilise the DEMETRA software application. DEMETRA models can be found for free. To forecast the toxicity to trout, daphnia, quail, and bees, five models were created. The software is built on a homogeneous integration of the knowledge gained through the DEMETRA EU project, using the best algorithms discovered as the foundation for hybrid combination models to be utilised for prediction.

[C] "T.E.S.T [https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test]".

Utilizing the aforementioned QSAR approaches, users will be able to quickly estimate acute toxicity utilising the Toxicity Estimation Software Tool (T.E.S.T.).

[D] "OCHEM [https://ochem.eu/home/show.do]".

OCHEM is a modelling environment-integrated online database of experimental measurements. Build prognostic QSAR model intended for physical–chemical or biological virtues.

[E] "E-DRAGON [http://www.vcclab.org/lab/edragon/]".

It is an electronic remote version of the renowned programme DRAGON, which R. Todeschini developed to calculate molecular descriptors. These descriptors can be used for similarity analysis, molecular SAR, and high-throughput screening of the molecule database.

[F] SeeSAR [https://www.biosolveit.de/SeeSAR/].

A software device for visual, interactive composite prioritization and development is called SeeSAR. A multi-parameter optimisation to optimise the chance of achievement rather than just similarity is ideal for structure-based design work. The availability of the necessary parameters and real-time, 3D visual computer help are two advantages of SeeSAR.

[G] Dragon [https://chm.kode-solutions.net/products_dragon.php].

Dragon generates 5270 molecular descriptors, which cover the majority of the different theoretical stances. The basic atom types, functional groups, part counts, topological and geometric descriptors, three-dimensional descriptors, as well as a number of property prediction (like logP), drug- and lead-like stimuli (like Lipinski's) are all included in the list of descriptors (Alarm).

[H] PaDEL-Descriptor [http://www.yapcwsoft.com/dd/padeldescriptor/].

A programme that determines molecular fingerprints and identifications Currently, the software generates 12 different types of fingerprints and 1875 identifiers (including 1444 1D, 2D, and 431 3D identifiers) (16,092 bits total).

The Chemistry Development Kit is used to calculate additional identities and fingerprints, including ring numbers, the Crippen logP and MR, the extended topochemical atom (ETA) descriptors, the McGowan volume, and the descriptors for the molecular linear free energy relationship.

AI is a subset of machine intelligence that depends on computers' capacity to learn from data already in existence. Various computational modelling techniques have employed AI to forecast the biological activity and toxicity of pharmacological compounds (Patel et al. 2014). Additionally, AI has many uses in the drug discovery process, including de novo drug design, virtual screening, QSAR, protein–protein interaction prediction, and protein folding prediction (Wang et al. 2019). Machine learning (ML) and deep learning (DL) are two potent techniques that are frequently employed in rational drug design (Patel et al. 2020). Support vector machine (SVM) (Cortes et al. 1995), Random Forest (RF) (Breiman 2001), and Naive Bayesian (NB) ML algorithms have been widely used in drug discovery. Convolutional neural networks (CNN), deep neural networks (DNN), recurrent neural networks (RNN), auto encoders, and limited Boltzmann machines are a few examples of DL techniques (RBN) (Zhong et al. 2018).

2.3 ADMET Prediction in CADD

Successful drugs must have an appropriate ADMET profile in addition to biological activity to have a therapeutic effect. Although there are a number of ADMET screening techniques available for in vitro and in vivo research in the lab, it is still crucial to execute an in silico method for determining the pharmacokinetic activity of these medicinal medicines. In fact, the ability to predict the pharmacokinetic characteristics of drugs using a computational approach facilitates early ADMET research and reduces the need for laboratory animal testing (Villoutreix et al. 2013; Aguayo-Orozco et al. 2016). A chemically vigorous molecule must meet a number of requirements set forth by these pharmacokinetic characteristics in order to be considered as a potential therapeutic candidate. Therefore, it is crucial to determine a possible drug's ADMET qualities after in silico molecular docking has predicted and demonstrated that it has the ability to connect to object receptor. These preclinical candidates need to have a propensity for absorption into the object exterior, good distribution from the place of inclusion to the object site, normal hepatic metabolism, elimination through the excretory system following action, and they should not be lethal (Jorgensen and Duffy 2002). Additionally, for optimal drug-likeness and absorption, the possible medication candidates must adhere to the Lipinski rule of five (RO5) (Lipinski 2004). Christopher Lipinski based his drug-likeness on several physicochemical characteristics, including the quantity of

Table 2.1 Most widely used software in ADMET prediction

Software	Developer	Reference
ADMET predictor	Simulation plus	Liu (2022)
ADME suit	ACD labs	Yinghuang et al. (2022)
Tox suite	ACD labs	([CSL STYLE ERROR: Reference with no printed form.])
ADME works predictor	Fujitsu FQS	([CSL STYLE ERROR: Reference with no printed form.])
SwissADM	SIB	(Daina et al. 2017)
TopKat	BIOVIA software	([CSL STYLE ERROR: Reference with no printed form.])
OSIRIS	CRO/CMO services	Lipinski et al. (2001)
Metasite	Molecular discovery	Cruciani et al. (2005)

H-bond donors (must be 5) and acceptors (must be 10), molecular weight (must not exceed 500 Da), and partition coefficient represented by logp (must be 5). Poor ligands, phytochemicals, or bioactive substances (natural or manufactured) are those that break more than one of these guidelines (van de Waterbeemd and Gifford 2003). The development of new medications is said to depend heavily on ADMET data. Both in vitro and in vivo models offer information about the ADMET properties of pharmaceuticals, which can then be used to forecast how the drugs will behave once they have been taken. Drug candidates are advanced, held, or cancelled based on ADMET characteristics (Zhang et al. 2012). Since pharmacokinetic profiles can be inferred based on drug ADMET data, preclinical data on drug ADMET qualities are important in assessing how well medicines target after administration. When evaluating drug exposure in the intended site of action, variables such as drug absorption rate, deposition, and metabolism inside the embattled appendage are engaged into description (Zhuang and Lu 2016). The need to optimise physicochemical properties, together with their bioactivities and ADMET properties, has arisen in order to expand drugs by means of the fewest side effects, because drug developers today face serious challenges to develop more effective as well as cost-effective drugs compared to the existing therapies (Peach et al. 2012). The amount of doses and their frequency are additional characteristics that are considered while forecasting the behaviours of recently discovered medications. These characteristics include the medications' bioavailability, oral absorption, clearance, volume of distribution, and blood–brain barrier penetration (BBB) (van de Waterbeemd and Gifford 2003).

The most widely used software in ADMET prediction are given in Table 2.1.

2.4 Molecular Dynamics Simulation in Drug Discovery

2.4.1 Process of MD Simulation

"MD simulation" , a method which model the arrangement since a set of particle with the aim of interaction during classical mechanics, is a very effective computational scheme for achieving mechanistic knowledge. For those with a background in chemistry, it makes sense intuitively to assume that these particles symbolise atoms, by means of relations among the atoms generating the forces that control molecule structure and the intermolecular forces that organise relations among molecules. However, the particles can also be used to represent larger structures than just a single atom, such as groups of atoms, complete molecules, or even collections of molecules. These models, sometimes referred to as coarse-grained models, can provide information about the system on a longer time scale than an atomistic model does (De Vivo et al. 2016; Bunker and Róg 2020; Salo-Ahen et al. 2021). While its inception in the late 1970s, MD simulation has advanced from simulating a few hundred atoms to systems with biological significance, such as whole proteins in solution with explicit solvent representations, proteins embedded in membranes, or massive macromolecular complexes like nucleosomes (Sotriffer 2006; Roccatano et al. 2007; Tinoco and Wen 2009; Durrant and McCammon 2011). When the necessary computer resources are available, simulations of systems with 50,000–100,000 atoms are increasingly commonplace. Simulations of systems with 500,000 atoms are now quite popular. Due in significant part to the utilisation of high-performance computing (HPC) and the clarity of the fundamental MD algorithm, this amazing improvement has been achieved (Brandman et al. 2012; Le and Le 2012; Hernández-Rodríguez et al. 2016). Almost all types of macromolecules with biological or therapeutic significance are studied using MD simulations nowadays, including nucleic acids, proteins, and carbohydrates. In explicit, every thousand to millions of individual atoms move in a succession of brief (for example, 2 fs), discrete time steps. The forces on each atom are calculated at each step and updated in accordance with Newton's laws of motion along with the atomic position and velocity. The forces on each atom are derived from the "force field," a set of physics-based parameters that represent both bonded and non-bonded (for example, van der Waals) inter-atomic forces. Even while free energy can be computed using various computational techniques, it has been demonstrated that the one obtained using MD has low false positive rates and great performance. Trajectories simulation with either MM-PBSA or MM-GBSA is integrated into MD to calculate binding-free energy, which is the affinity of the ligand for the target. Because the degree of binding a ligand has to the target determines its capacity to provide a therapeutic effect, evaluating binding-free energy is important in the drug discovery process. A trustworthy MD tool for calculating the affinity of protein–ligand complexes has also developed and it is called the umbrella sampling technique. The difference between the highest and minimum values of the free energy change, calculated using the

potential of mean force, is used to measure the binding affinity (PMF) (Perez et al. 2014; Gao et al. 2018; Bao et al. 2019).

2.4.2 Software Used in MD Simulation

The conventional drug design process relies heavily on the protein–ligand interaction, and MD simulation's lower cost compared to its experimental counterpart minimises the expense and time involved (Arcon et al. 2017). The study of dynamics and conformational flexibility of drug-target complexes has made use of MD modelling as a key technique. In a computer simulation, MD simulation helps to simulate biological phenomena. It has revolutionised the field of drug development and has evolved into a standard computational tool for CADD. It offers a precise estimation of the thermodynamics and kinetics related to the interaction and binding of the drug with the target. The use of MD simulation among scientists working with CADD and in the biopharmaceutical business has increased as a result of the development of new techniques, software, and hardware. The simulation period needs to be long enough to yield accurate results. The simulation is more accurate and produces dependable findings when run at the millisecond and microsecond time scales. Even though this is not relevant for evaluating protein–ligand interactions, the study can be completed in a few nanoseconds of simulation time (Liu et al. 2017). To address the issue with molecular dynamic simulation time length, a special purpose machine named ANTON has been created. However, few scientists can afford this supercomputer due to its tremendous computing capability (Shaw et al. 2008). To accelerate the most expensive computational systems for MD simulation, other supercomputers with specialised hardware, such as MDGRAPE (Taiji et al. 2002), MD Engine (S. 1999), and FASTRUN (Fine et al. 1991), have been developed. A powerful Graphical Processing Unit (GPU)-equipped computer system was recently designed to enable running MD simulation at a low cost (Stone et al. 2016). The most widely used software in MD simulation is given in Table 2.2.

2.5 Case Studies

Finding new therapeutic possibilities has always been a component of research that has proven to be essential for enhancing human health (Taubenberger and Morens 2006). For many years, the use of MD modelling has revolutionised the process of finding new drugs. Using physics-dependent intermolecular interaction, simulation often aids in the timely and relevant prediction of atom mobility in a molecular system. However, in simulation, this difficult operation in the wet lab is simple, lowering the intricacy surrounding the wet tests in the smallest amount of time (Van Gunsteren and Mark 1998; Karplus and McCammon 2002; Hollingsworth and Dror 2018). Due to this potential, MD is now thought of as a computational microscope.

Table 2.2 Most widely used software in MD simulation

Software	Details	Reference
CHARMM	It stands for "Chemistry at HARvard Macromolecular Mechanics." It is a multi-scale MD modelling tool with broad applicability to many-particle systems.	Jo et al. (2008) and Brooks et al. (2009)
GROMACS	GROMACS is one of the most well-known open-source and free chemistry software applications. It is mainly used for dynamical simulations of biological molecules. It provides a wide variety of computing, preparation, and analysis tools.	Kumari et al. (2014) and Abraham et al. (2015)
NAMD	NAMD is an analogous MD programme which is intended and used for high-performance simulation of big biological system.	Phillips et al. (2005)
AMBER	The term "Amber" refers to a group of software applications that let users run and examine MD simulations, chiefly for nucleic acids, proteins, and carbohydrates.	Case et al. (2005)
DESMOND	Desmond is a piece of software used to carry out high-speed MD simulations of biological systems on traditional computer.	Gopinath and Kathiravan (2021)
TINKER	With certain unique capabilities for biopolymers, the Tinker molecular modelling programme is a comprehensive and all-encompassing solution for molecular mechanics and dynamics. Tinker can employ any of a number of widely used parameter sets, including Amber, CHARMM, and others.	Rackers et al. (2018)
DL_POLY	I.T. Todorov, W. Smith, A.M. Elena, and others created the general-purpose classical MD simulation programme known as DL POLY at Daresbury laboratory.	Smith et al. (2010)
ACEMD	It is a MD engine that supports the CHARMM and AMBER force fields is called ACEMD.	Harvey et al. (2009)
GENESIS	GENESIS stands for GENeralised-Ensemble SImulation System. GENESIS makes it feasible to efficiently simulate and model different biomolecular systems using MD.	Ito et al. (2022) and Oshima and Sugita (2022)
MDynaMix	A computer software programme called Molecular Dynamics of Mixes (MDynaMix) simulates mixtures of molecules interacting with AMBER- and CHARMM.	Lyubartsev and Laaksonen (2000)
Orac	Orac is a traditional MD programme, which used to simulate complicated molecular systems at the atomistic level.	([CSL STYLE ERROR: Reference with no printed form.])
YASARA	A computer application called Yet Another Scientific Artificial Reality Application (YASARA) is used to visualise, model, and simulate MD.	Krieger et al. (2002)

The enigma surrounding the disorders linked to protein misfolding and aggregation could also be revealed by MD simulation, which would then aid uncover small molecular modulators (Urbanc et al. 2010; Ye et al. 2013). In recent times, it was claimed that using "MD simulation," medications used to treat or control neurological illnesses might be targeted (Manglik et al. 2016; Spahn et al. 2017; McCorvy et al. 2017).

The spectrum of applications for MD simulation in biology is expanding swiftly, revealing distinctive characteristics of protein structure that are difficult to identify using only experimental techniques (Wu et al. 2022). Herotika et al. explained the role of MD simulation in virology research (Ode et al. 2012). Probable covalent and FDA-approved SARS-CoV-2 major protease 3CL inhibitors were discovered through MD simulation research (Alamri et al. 2021). Wadhwa et al. used MD simulations to estimate the membrane permeability of two naturally occurring medicines called withanolides (withaferin-A and withanone), which are structurally similar but have strikingly different lethal properties (Wadhwa et al. 2021). Liu et al. assessed the uses of MD simulations for drug detection, including virtual screening, drug–target interaction mechanisms, and the pathogenic mechanisms of illnesses induced by amyloidosis (Liu et al. 2018). As discussed by Durrant et al., atomistic computer simulations of macromolecular (such as protein) receptors and the small molecules that bind to them can be used to identify cryptic or allosteric binding sites, improve conventional virtual-screening techniques, and directly predict the binding energies of small molecules (Durrant and McCammon 2011). A 2013 study geared toward the detection of novel "aldose reductase inhibitors" (ARI) that was in print with the "Journal of Chemical Information and Modelling" employed MD as part of its methodology. It is challenging to identify a specific obligatory place to be worn meant for this enzyme because of the active site architecture's sampling of diverse conformations while various ligands are bound to it. For example, when ligand A binds, the residues it interacts with are vastly different from those of ligands B and C. Three average conformations were obtained from a quick MD simulation of this protein target (aldose reductase), and these conformations were then utilised to dock a variety of substances. Later, MD was tried to assess the communication among the lead chemical and the target protein (complex dynamics). Compared to Epalrestat, which are being marketed, pre-clinical and clinical testing of this lead chemical showed it to have high-biological activity and a decreased off-target effect (Wang et al. 2014). Another team of Filipino researchers examined MD applications in tuberculosis and came to the conclusion that MD has accelerated the pace of drug discovery for this condition (Macalino et al. 2020). In a study to identify possible inhibitors for "HMG-CoA reductase," the rate-limiting enzyme in cholesterol manufacture, bio-active chemicals from 10 plants were virtually screened. Rutin emerge, and the constancy of the "Rutin-HMG-CoA reductase" complex was assessed using MD. Comparing the complex to the unbound apoprotein, a close examination of its "RMSD, RMSF, ROG, and H-bond" spectra revealed that it was more stable (Suganya et al. 2017). Another study was focused on the "γ-aminobutyric acid (GABA)" receptor, a common target for anaesthetics. Corresponding data from MD analysis (RMSD) assured the researchers that the lead 5 ligand was a potent

GABA receptor inhibitor and should be directed to the pre-clinical stage of the drug discovery process. This group in China was the first to use in silico study to search for novel anaesthetic compounds (Peng et al. 2014). Permeability is a crucial factor that is optimised in the drug discovery process. Drugs having intracellular targets must be able to penetrate membranes in order to be effective, as this is dependent on their bioavailability. A number of techniques are already available to determine a potential drug candidate's capacity to infuse membranes, including the parallel artificial membrane permeability test (PAMPA) and the "immobilised artificial membrane" (IAM) (Tsopelas et al. 2016). Martini force field coarse-grained MD, a technique that reduces the lengthy time that would have been invested in all atomistic MD by grouping atoms jointly into super-particles or bead, can currently be used to evaluate the penetration of a drug candidate (Hoffmann et al. 2020). In the lead optimisation process, MD modelling is helpful. Scientists always assume a desire to introduce a medicine with high potencies into clinics after the lead identification stage, which occurs before pre-clinical trial in the drug discovery scheme. To achieve a stronger therapeutic impact, this may, however, necessitate the optimisation of the lead chemical at the target's active site. MD has been used to support the idea that comprehending a ligand's ensuing communication in the dynamic pocket of the target is a vibrant procedure that involves adding or removing groups (Carnero 2006; Dror et al. 2011). It is now abundantly obvious that the utility of MD simulation in drug development extends commencing the earliest stage of drug finding to direct optimisation.

2.6 Conclusion and Future Perspectives

In the protracted process of drug discovery and development, Computer-Aided Drug Designing is now a vital instrument. Additionally, it offers alternatives for comprehending chemical systems in various methods, giving data that are difficult to collect in laboratory research and requiring significantly less time and money than trials. As is common in the early phases of practically any new technology or research, CADD initially had a rough reputation in the field of drug development and may have overstated its claims. The two disciplines of computer-aided drug discovery that play a vital part in the design and identification of therapeutic molecules in a quicker and more affordable manner are structure- and ligand-based drug design. We think that there is now a lot of room for growth in drug development and other fields by carefully combining MD simulations with complementary experimental techniques. This possibility will further expand as simulations get more precise, swifter, less expensive, and broadly available. Applying simulations to molecular biology and drug development successfully necessitates careful consideration of the experimental and computational data at hand and as a result, benefits from a wide range of skills and interdisciplinary partnerships. At the moment, conformational changes or ligand binding can be accurately modelled when routine simulations are reaching the microsecond scale. We can transition

from the analysis of single structures, the foundation of molecular modelling as we know it, to the analysis of conformational ensembles thanks to advancements made in computational hardware, particularly the use of GPUs, and optimisation of MD algorithms, including coarse-grained ones. In conclusion, CADD is useful for pharmaceutical development in the areas of 3D structure prediction, compound design, druggability prediction, and *in silico* ADMET prediction; however, it must be understood that computational predictions need to be combined with experimental methods for effective drug discovery and development.

References

Abraham MJ, Murtola T, Schulz R 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

Aguayo-Orozco A, Audouze K, Brunak S, Taboureau O (2016) In silico systems pharmacology to assess drug's therapeutic and toxic effects. Curr Pharm Des 22:6895–6902. https://doi.org/10.2174/1381612822666160907093215

Akamatsu M (2005) Current state and perspectives of 3D-QSAR. Curr Top Med Chem 2:1381–1394. https://doi.org/10.2174/1568026023392887

Alamri MA, Tahir Ul Qamar M, Mirza MU et al (2021) Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CLpro. J Biomol Struct Dyn 39:4936–4948. https://doi.org/10.1080/07391102.2020.1782768/SUPPL_FILE/TBSD_A_1782768_SM9175.DOCX

Anderson AC (2003) The process of structure-based drug design. Chem Biol 10:787–797. https://doi.org/10.1016/j.chembiol.2003.09.002

Arcon JP, Defelipe LA, Modenutti CP et al (2017) Molecular dynamics in mixed solvents reveals protein-ligand interactions, improves docking, and allows accurate binding free energy predictions. J Chem Inf Model 57:846–863. https://doi.org/10.1021/ACS.JCIM.6B00678/SUPPL_FILE/CI6B00678_SI_001.PDF

Baig MH, Ahmad K, Rabbani G et al (2018) Computer aided drug design and its application to the development of potential drugs for neurodegenerative disorders. Curr Neuropharmacol 16:740. https://doi.org/10.2174/1570159X15666171016163510

Bao Y, Zhou L, Dai D et al (2019) Discover potential inhibitors for PFKFB3 using 3D-QSAR, virtual screening, molecular docking and molecular dynamics simulation. J Recept Signal Transduct 38:413–431. https://doi.org/10.1080/10799893.2018.1564150

Batool M, Ahmad B, Choi S (2019) A structure-based drug discovery paradigm. Int J Mol Sci 20:2783

Bernard D, Coop A, MacKerell AD (2005) Conformationally sampled pharmacophore for peptidic δ opioid ligands. J Med Chem 48:7773–7780. https://doi.org/10.1021/JM050785P/SUPPL_FILE/JM050785PSI20050921_043101.PDF

Brandman R, Brandman Y, Pande VS (2012) A-site residues move independently from P-site residues in all-atom molecular dynamics simulations of the 70S bacterial ribosome. PLoS One 7:29377. https://doi.org/10.1371/JOURNAL.PONE.0029377

Breiman L (2001) Random forests. Mach Learn 451(45):5–32. https://doi.org/10.1023/A:1010933404324

Brogi S, Ramalho TC, Kuca K et al (2020) Editorial: In silico methods for drug design and discovery. Front Chem 8:612

Brooks BR, Brooks CL, Mackerell AD et al (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30:1545. https://doi.org/10.1002/JCC.21287

Bunker A, Róg T (2020) Mechanistic understanding from molecular dynamics simulation in pharmaceutical research 1: drug delivery. Front Mol Biosci 7:604770

Campos SRR, MacHuqueiro M, Baptista AM (2010) Constant-pH molecular dynamics simulations reveal a β-rich form of the human prion protein. J Phys Chem B 114:12692–12700. https://doi.org/10.1021/JP104753T/SUPPL_FILE/JP104753T_SI_001.PDF

Carnero A (2006) High throughput screening in drug discovery. Clin Transl 87(8):482–490. https://doi.org/10.1007/S12094-006-0048-2

Case DA, Cheatham TE, Darden T et al (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668. https://doi.org/10.1002/JCC.20290

Chen W, Van Der Kamp MW, Daggett V (2014) Structural and dynamic properties of the human prion protein. Biophys J 106:1152. https://doi.org/10.1016/J.BPJ.2013.12.053

Cherkasov A, Muratov EN, Fourches D et al (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57:4977–5010. https://doi.org/10.1021/JM4004285/ASSET/IMAGES/MEDIUM/JM-2013-004285_0009.GIF

Clark DE (2006) What has computer-aided molecular design ever done for drug discovery? Expert Opin Drug Discov 1:103–110. https://doi.org/10.1517/17460441.1.2.103

Cortes C, Vapnik V, Saitta L (1995) Support-vector networks. Mach Learn 203(20):273–297. https://doi.org/10.1007/BF00994018

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

Cruciani G, Carosati E, De Boeck B et al (2005) MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. J Med Chem 48:6970–6979. https://doi.org/10.1021/JM050529C/SUPPL_FILE/JM050529CSI20050823_063632.PDF

Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 71(7):1–13. https://doi.org/10.1038/srep42717

De Vivo M, Masetti M, Bottegoni G, Cavalli A (2016) Role of molecular dynamics and related methods in drug discovery. J Med Chem 59:4035–4061

Dixon SL, Smondyrev AM, Rao SN (2006) PHASE: a novel approach to pharmacophore modeling and 3D database searching. Chem Biol Drug Des 67:370–372. https://doi.org/10.1111/J.1747-0285.2006.00384.X

Dror RO, Pan AC, Arlow DH et al (2011) Pathway and mechanism of drug binding to G-protein-coupled receptors. Proc Natl Acad Sci U S A 108:13118–13123. https://doi.org/10.1073/PNAS.1104614108/SUPPL_FILE/SM01.AVI

Durrant JD, McCammon JA (2011) Molecular dynamics simulations and drug discovery. BMC Biol 9:1–9. https://doi.org/10.1186/1741-7007-9-71/FIGURES/4

Fine R, Dimmler G, Levinthal C (1991) FASTRUN: a special purpose, hardwired computer for molecular simulation. Proteins Struct Funct Bioinforma 11:242–253. https://doi.org/10.1002/PROT.340110403

Gao J, Zhang Y, Chen H et al (2018) Computational insights into the interaction mechanism of transcription cofactor vestigial-like protein 4 binding to TEA domain transcription factor 4 by molecular dynamics simulation and molecular mechanics generalized Born/surface area calculation. J Biomol Struct Dyn 37:2538–2545. https://doi.org/10.1080/07391102.2018.1491889

González PM, Acharya C, MacKerell AD, Polli JE (2009) Inhibition requirements of the human apical sodium-dependent bile acid transporter (hASBT) using aminopiperidine conjugates of glutamyl-bile acids. Pharm Res 267(26):1665–1678. https://doi.org/10.1007/S11095-009-9877-3

Gopinath P, Kathiravan MK (2021) Docking studies and molecular dynamics simulation of triazole benzene sulfonamide derivatives with human carbonic anhydrase IX inhibition activity. RSC Adv 11:38079–38093. https://doi.org/10.1039/D1RA07377J

Grover S, Apushkin MA, Fishman GA (2006) Topical dorzolamide for the treatment of cystoid macular edema in patients with retinitis pigmentosa. Am J Ophthalmol 141:850–858. https://doi.org/10.1016/j.ajo.2005.12.030

Guner O, Clement O, Kurogi Y (2012) Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances. Curr Med Chem 11:2991–3005. https://doi.org/10.2174/0929867043364036

Harvey MJ, Giupponi G, De Fabritiis G (2009) ACEMD: accelerating biomolecular dynamics in the microsecond time scale. J Chem Theory Comput 5:1632–1639. https://doi.org/10.1021/CT9000685

Hecker EA, Duraiswami C, Andrea TA, Diller DJ (2002) Use of catalyst pharmacophore models for screening of large combinatorial libraries. J Chem Inf Comput Sci 42:1204–1211. https://doi.org/10.1021/CI020368A

Hernández-Rodríguez M, Rosales-Hernández MC, Mendieta-Wejebe JE et al (2016) Current tools and methods in molecular dynamics (MD) simulations for drug design. Curr Med Chem 23:3909–3924. https://doi.org/10.2174/0929867323666160530144742

Herwig R (2014) Computational modeling of drug response with applications to neuroscience. Dialogues Clin Neurosci 16:465. https://doi.org/10.31887/DCNS.2014.16.4/RHERWIG

Hoffmann C, Centi A, Menichetti R, Bereau T (2020) Molecular dynamics trajectories for 630 coarse-grained drug-membrane permeations. Sci Data 71(7):1–7. https://doi.org/10.1038/s41597-020-0391-0

Hollingsworth SA, Dror RO (2018) Molecular dynamics simulation for all. Neuron 99:1129. https://doi.org/10.1016/J.NEURON.2018.08.011

Ito S, Yagi K, Sugita Y (2022) Computational analysis on the allostery of tryptophan synthase: relationship between α/β-ligand binding and distal domain closure. J Phys Chem B 126:3300–3308. https://doi.org/10.1021/ACS.JPCB.2C01556/SUPPL_FILE/JP2C01556_SI_002.PDF

Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 29:1859–1865. https://doi.org/10.1002/JCC.20945

Jorgensen WL, Duffy EM (2002) Prediction of drug solubility from structure. Adv Drug Deliv Rev 54:355–366. https://doi.org/10.1016/S0169-409X(02)00008-X

Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 99(9):646–652. https://doi.org/10.1038/nsb0902-646

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

Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 37:4130–4146. https://doi.org/10.1021/JM00050A010/ASSET/JM00050A010.FP.PNG_V03

Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. Proteins Struct Funct Bioinforma 47:393–402. https://doi.org/10.1002/PROT.10104

Kumari R, Kumar R, Lynn A (2014) G-mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model 54:1951–1962. https://doi.org/10.1021/CI500020M/SUPPL_FILE/CI500020M_SI_001.PDF

Le L, Le L (2012) Incorporating molecular dynamics simulations into rational drug design: a case study on influenza a neuraminidases. Bioinformatics. https://doi.org/10.5772/52642

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

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

Liu X, Shi D, Zhou S et al (2017) Molecular dynamics simulations and novel drug discovery. Expert Opin Drug Discov 13:23–37. https://doi.org/10.1080/17460441.2018.1403419

Liu X, Shi D, Zhou S et al (2018) Molecular dynamics simulations and novel drug discovery. Expert Opin Drug Discov 13:23–37

Liu Y (2022) Use in silico and in vitro methods to screen hepatotoxic chemicals and CYP450 enzyme inhibitors. Methods Mol Biol 2474:189–198. https://doi.org/10.1007/978-1-0716-2213-1_17

Lyubartsev AP, Laaksonen A (2000) M.DynaMix – a scalable portable parallel MD simulation package for arbitrary molecular mixtures. Comput Phys Commun 128:565–589. https://doi.org/10.1016/S0010-4655(99)00529-9

Macalino SJY, Billones JB, Organo VG, Carrillo MCO (2020) In silico strategies in tuberculosis drug discovery. Molecules 25:665. https://doi.org/10.3390/MOLECULES25030665

Manglik A, Lin H, Aryal DK et al (2016) Structure-based discovery of opioid analgesics with reduced side effects. Nature 537(7619):185–190. https://doi.org/10.1038/nature19112

Marrakchi H, Lanéelle G, Quémard A (2000) InhA, a target of the antituberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS-II. Microbiology 146:289–296. https://doi.org/10.1099/00221287-146-2-289/CITE/REFWORKS

Mason J, Good A, Martin E (2005) 3-D pharmacophores in drug discovery. Curr Pharm Des 7:567–597. https://doi.org/10.2174/1381612013397843

McCorvy JD, Butler KV, Kelly B et al (2017) Structure-inspired design of β-arrestin-biased ligands for aminergic GPCRs. Nat Chem Biol 142(14):126–134. https://doi.org/10.1038/nchembio.2527

Melo-Filho C, Braga R, Andrade C (2014) 3D-QSAR approaches in drug design: perspectives to generate reliable CoMFA models. Curr Comput Aided-Drug Des 10:148–159. https://doi.org/10.2174/1573409910666140410111043

Moses H, Dorsey ER, Matheson DHM, Thier SO (2005) Financial anatomy of biomedical research. JAMA 294:1333–1342. https://doi.org/10.1001/JAMA.294.11.1333

Myers S, Baker A (2001) Drug discovery—an operating model for a new era. Nat Biotechnol 198(19):727–730. https://doi.org/10.1038/90765

Ode H, Nakashima M, Kitamura S et al (2012) Molecular dynamics simulation in virus research. Front Microbiol 3:258. https://doi.org/10.3389/FMICB.2012.00258/BIBTEX

Oshima H, Sugita Y (2022) Modified Hamiltonian in FEP calculations for reducing the computational cost of electrostatic interactions. J Chem Inf Model 62:2846–2856. https://doi.org/10.1021/ACS.JCIM.1C01532/SUPPL_FILE/CI1C01532_SI_001.PDF

Ou-Yang SS, Lu JY, Kong XQ et al (2012) Computational drug discovery. Acta Pharmacol Sin 33:1131–1140

Patel HM, Noolvi MN, Sharma P et al (2014) Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. Med Chem Res 23:4991–5007. https://doi.org/10.1007/S00044-014-1072-3

Patel L, Shukla T, Huang X et al (2020) Machine learning methods in drug discovery. Molecules 25(22):5277. https://doi.org/10.3390/MOLECULES25225277

Peach ML, Zakharov AV, Liu R et al (2012) Computational tools and resources for metabolism-related property predictions. 1. Overview of publicly available (free and commercial) databases and software. Future Med Chem 4:1907–1932. https://doi.org/10.4155/FMC.12.150

Peng QX, Guan XH, Yi ZG, Su YP (2014) Insilico approaches in anesthetic drug development: computer aided drug designing. Drug Res 8:587–591. https://doi.org/10.1055/S-0034-1395564/ID/R2014-08-0840-0029

Perez C, Faust B, Mehdipour AR et al (2014) Substrate-bound outward-open state of the betaine transporter BetP provides insights into Na+ coupling. Nat Commun 51(5):1–11. https://doi.org/10.1038/ncomms5231

Phillips JC, Braun R, Wang W et al (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781. https://doi.org/10.1002/JCC.20289

Prieto-Martínez FD, López-López E, Eurídice Juárez-Mercado K, Medina-Franco JL (2019) Computational drug design methods—current and future perspectives. In: In silico drug design:

repurposing techniques and methodologies. Academic Press, London, pp 19–44. https://doi.org/10.1016/B978-0-12-816125-8.00002-X

Rackers JA, Wang Z, Lu C et al (2018) Tinker 8: software tools for molecular design. J Chem Theory Comput 14:5273. https://doi.org/10.1021/ACS.JCTC.8B00529

Ragno R, Esposito V, Di Mario M et al (2020) Teaching and learning computational drug design: student investigations of 3D quantitative structure-activity relationships through web applications. J Chem Educ 97:1922–1930. https://doi.org/10.1021/acs.jchemed.0c00117

Ren JX, Li LL, Zheng RL et al (2011) Discovery of novel Pim-1 kinase inhibitors by a hierarchical multistage virtual screening approach based on SVM model, pharmacophore, and molecular docking. J Chem Inf Model 51:1364–1375. https://doi.org/10.1021/CI100464B/SUPPL_FILE/CI100464B_SI_001.PDF

Roccatano D, Barthel A, Zacharias M (2007) Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation. Biopolymers 85:407–421. https://doi.org/10.1002/BIP.20690

Rutenber EE, Stroud RM (1996) Binding of the anticancer drug ZD1694 to E. coli thymidylate synthase: assessing specificity and affinity. Structure 4:1317–1324. https://doi.org/10.1016/S0969-2126(96)00139-6

Salo-Ahen OMH, Alanko I, Bhadane R et al (2021) Molecular dynamics simulations in drug discovery and pharmaceutical development. Processes 9:71. https://doi.org/10.3390/PR9010071

Santini S, Derreumaux P (2004) Helix H1 of the prion protein is rather stable against environmental perturbations: molecular dynamics of mutation and deletion variants of PrP(90–231). Cell Mol Life Sci CMLS 61:951–960. https://doi.org/10.1007/S00018-003-3455-3

Schaller D, Šribar D, Noonan T et al (2020) Next generation 3D pharmacophore modeling. Wiley Interdiscip Rev Comput Mol Sci 10:e1468. https://doi.org/10.1002/WCMS.1468

Shaw DE, Deneroff MM, Dror RO et al (2008) Anton, a special-purpose machine for molecular dynamics simulation. Commun ACM 51:91–97. https://doi.org/10.1145/1364782.1364802

Smith W, Yong CW, Rodger PM (2010) DL_POLY: application to molecular simulation. Mol Simul 28:385–471. https://doi.org/10.1080/08927020290018769

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

Sotriffer CA (2006) Molecular dynamics simulations in drug design. In: Encyclopedic reference of genomics and proteomics in molecular medicine. Springer, Berlin, pp 1153–1160. https://doi.org/10.1007/3-540-29623-9_0820

Spahn V, Del Vecchio G, Labuz D et al (2017) A nontoxic pain killer designed by modeling of pathological receptor conformations. Science 355:966–969. https://doi.org/10.1126/SCIENCE.AAI8636/SUPPL_FILE/AAI8636_SPAHN_SM.PDF

Stone JE, Hallock MJ, Phillips JC, et al (2016) Evaluation of emerging energy-efficient heterogeneous computing platforms for biomolecular and cellular simulation workloads. In *2016 IEEE 30th International Parallel and Distributed Processing Symposium IPDPS*, Chicago, IL, USA, pp 89–100. https://doi.org/10.1109/IPDPSW.2016.130

Suganya S, Nandagopal B, Anbarasu A (2017) Natural inhibitors of HMG-CoA reductase—an Insilico approach through molecular docking and simulation studies. J Cell Biochem 118:52–57. https://doi.org/10.1002/JCB.25608

Taiji M, Futatsugi N, Narumi T et al (2002) Protein explorer: a petaflops special-purpose computer for molecular dynamics simulations. Genome Inform 13:461–462. https://doi.org/10.11234/GI1990.13.461

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

Taubenberger JK, Morens DM (2006) 1918 Influenza: the mother of all pandemics. Emerg Infect Dis 12:15–22. https://doi.org/10.3201/EID1201.050979

Tinoco I, Wen J-D (2009) Simulation and analysis of single-ribosome translation. Phys Biol 6: 025006. https://doi.org/10.1088/1478-3975/6/2/025006

Tsopelas F, Vallianatou T, Tsantili-Kakoulidou A (2016) Advances in immobilized artificial membrane (IAM) chromatography for novel drug discovery. Expert Opin Drug Discov 11: 473–488. https://doi.org/10.1517/17460441.2016.1160886

Urbanc B, Betnel M, Cruz L et al (2010) Elucidation of amyloid β-protein oligomerization mechanisms: discrete molecular dynamics study. J Am Chem Soc 132:4266–4280. https://doi.org/10.1021/JA9096303/SUPPL_FILE/JA9096303_SI_001.PDF

Van Gunsteren WF, Mark AE (1998) Validation of molecular dynamics simulation. J Chem Phys 108:6109. https://doi.org/10.1063/1.476021

Verma RP, Hansch C (2009) Camptothecins: a SAR/QSAR study. Chem Rev 109:213–235. https://doi.org/10.1021/CR0780210/ASSET/CR0780210.FP.PNG_V03

Villoutreix BO, Lagorce D, Labbé CM et al (2013) One hundred thousand mouse clicks down the road: selected online resources supporting drug discovery collected over a decade. Drug Discov Today 18:1081–1089. https://doi.org/10.1016/J.DRUDIS.2013.06.013

Wadhwa R, Yadav NS, Katiyar SP et al (2021) Molecular dynamics simulations and experimental studies reveal differential permeability of withaferin-A and withanone across the model cell membrane. Sci Rep 11:2352. https://doi.org/10.1038/s41598-021-81729-z

Wang B, Buchman CD, Li L et al (2014) Enrichment of chemical libraries docked to protein conformational ensembles and application to aldehyde dehydrogenase 2. J Chem Inf Model 54: 2105–2116. https://doi.org/10.1021/CI5002026/SUPPL_FILE/CI5002026_SI_001.PDF

Wang L, Ding J, Pan L et al (2019) Artificial intelligence facilitates drug design in the big data era. Chemom Intell Lab Syst 194:103850. https://doi.org/10.1016/J.CHEMOLAB.2019.103850

van de Waterbeemd H, Gifford E (2003) ADMET in silico modelling: towards prediction paradise? Nat Rev Drug Discov 2:192–204. https://doi.org/10.1038/NRD1032

Wlodawer A, Vondrasek J (2003) Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. Annu Rev Biophys Biomol Struct 27:249–284. https://doi.org/10.1146/ANNUREV.BIOPHYS.27.1.249

Wolber G, Langer T (2004) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model 45:160–169. https://doi.org/10.1021/CI049885E

Wu X, Xu L-Y, Li E-M, Dong G (2022) Application of molecular dynamics simulation in biomedicine. Chem Biol Drug Des 99:789–800. https://doi.org/10.1111/CBDD.14038

Ye W, Wang W, Jiang C et al (2013) Molecular dynamics simulations of amyloid fibrils: an in silico approach. Acta Biochim Biophys Sin Shanghai 45:503–508. https://doi.org/10.1093/abbs/gmt026

Yinghuang GSS, Gong ZH et al (2022) Mechanism of Sanhua decoction in the treatment of ischemic stroke based on network pharmacology methods and experimental verification. Biomed Res Int 2022:7759402. https://doi.org/10.1155/2022/7759402

Yu W, Mackerell AD (2017) Computer-aided drug design methods. Methods Mol Biol 1520:85. https://doi.org/10.1007/978-1-4939-6634-9_5

Zhang D, Luo G, Ding X, Lu C (2012) Preclinical experimental models of drug metabolism and disposition in drug discovery and development. Acta Pharm Sin B 2:549–561. https://doi.org/10.1016/J.APSB.2012.10.004

Zhong F, Xing J, Li X et al (2018) Artificial intelligence in drug design. Sci China Life Sci 61(10): 1191–1204. https://doi.org/10.1007/S11427-018-9342-2

Zhuang X, Lu C (2016) PBPK modeling and simulation in drug research and development. Acta Pharm Sin B 6:430–440. https://doi.org/10.1016/J.APSB.2016.04.004

Chapter 3
Informatics: Tools and Databases in Drug Discovery

Jurnal Reang, Vivek Yadav, Vinita, Jaseela Majeed, Prabodh Chander Sharma, Rajiv Kumar Tonk, and Kalicharan Sharma

3.1 Introduction

Modern day drug development or drug discovery research is based on the information or data gathered over many years correlated to studies on therapeutic contender. These information comprises data linked to chemical, generic, bio-chemical, pharmacological, and physiological properties. Nowadays accessing and operating enormous magnitudes of information are essential for the drug discovery process. The field informatics or information technology is providing the systematic organization, storage, and retrieval of data that required during the process of drug discovery. Information technology allows the researcher to easy access and analysis of data to yield profound understanding. Cheminformatics and bioinformatics are instances of growing acknowledgment of informatics in drug discovery (Gunjal 2003).

3.1.1 Databases

There is an inevitability of information ranging from basic to scientific in our daily life. To maintain the information, print and electronic files (books and journals) were established. Databases are organized depositories of information in the form of text, numerical, graphical, and structural details. Databases enable easy access of relevant information with the ease of up-to-date management. In the late 60s, database arise due to an emergent mandate among users of different organization for additional information to run the day-to-day function and also for future planning. The

J. Reang (✉) · V. Yadav · Vinita · J. Majeed · P. C. Sharma · R. K. Tonk · K. Sharma
Pharmaceutical Chemistry Department, Delhi Pharmaceutical Sciences and Research University, New Delhi, India

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_3

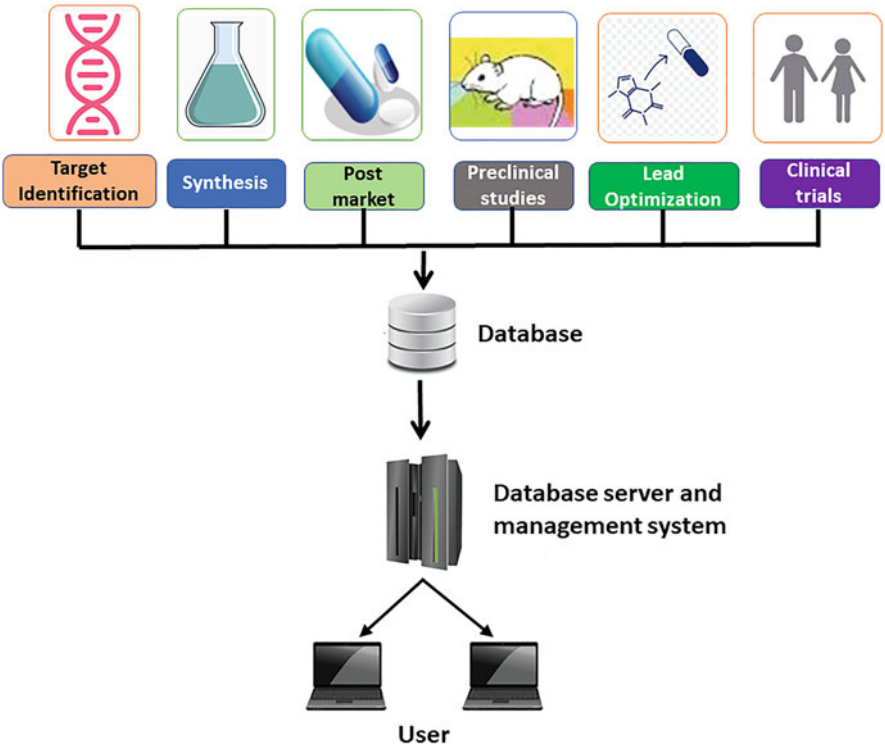


Fig. 3.1 The modern computer-based database

traditional database system is comprised of paper-based records, which necessitate hefty storage space with the risk of data repetition, astray or impairment of data, and also involving time intense pursuit. The modern computer-based databases were technologically advanced to tackle the issues linked traditional database system. It offers data integrity, data sharing, and data migration between systems, and security restriction with lessen data inconsistencies. Modern databases employ algorithms for the mining of information from the large pool (Fig. 3.1). It has become vital part of public health care, as it takes account of patient condition recording, observation of patient situation, and therapy which all cumulatively plays a vital role in drug design and discovery-linked procedures. MEDLINE, DrugBank, Protein Data Bank (PDB) and Binding Database (BD) are some of the examples of databases (Gunjal 2003).

Functions of Database:

- User-friendly: Database delivers easy to learn and use the information.
- Data independence: It permits modifications at any one level of the database without disturbing the further levels.
- Economy: Additional data at small budget as database provides protection and transformation of data economically.

- Data Sharing: A database permits distribution of data under its supervision to several number of users.
- Accuracy and Integrity: Centralized control of the database helps in avoiding incorrect data as the unified controls distinguish data incorrectness where they arise. The correctness of a database safeguards that data feature and content stay persistent.
- Privacy and Security: Complete jurisdictions over the operational data prevent unauthorized access of data to sustain privacy.
- Concurrency control: Database allows rapid access to a database, while shielding information integrity.
- Performance: It stresses on retort time of investigations, making fit to use the information based on the type of user-database dialogue.
- Enforcement of Standards: In order to aid data interchange between systems, the concept of database facilitates regulation of stored data in particular desirable formats.
- Support: Database support complex file assembly and access route.

3.1.2 Database Structure

Data are organized as per the data model, where data are divided into a component based on the elements of data. For example, book information is a data structure comprising the data components such as author name, title, publisher's name, ISBN, and year (Gunjal 2003).

There are numerous diverse methods to evaluate the rational configuration of information in complicated databases. But there are three varieties of data structure,

(a) List Structure
(b) Tree/Hierarchical Structure
(c) Network Structure

(a) List Structure

In this type of database structure, data are recorded basically based on the position of Nth record where the Nth is associated $(N - 1)$ and $(N - 2)$. This brings one-to-one correlation (Gunjal 2003).

(b) Tree or Hierarchical Structure

It is a multilevel and non-linear structure of data in which every node may be linked to N-nodes at some level lower it but to simply one node overhead it. The record is as of the topmost, and the course of examination is momentary down. Data traced at diverse stages alongside a specific division from the root are called the node. The foundation of a data tree or hierarchy is the root. This type of structure is useful for storing data in the form of parent–child connection (Gunjal 2003).

(c) Network Structure

It is a different arrangement of hierarchical structure with broader arrangement than a hierarchy model. Similar to the hierarchy method, the data are characterized by accounts and associations. This structure allows relationships among entities. It is a structure of database consists of a sum of distinct incidents record in which a certain node may have several quantities of underlings' nodes. It is paralleled to a chart assembly and conveys many-to-many connection (Gunjal 2003).

3.1.3 Database Management System (DBMS)

Creating, modernizing, and erasing information are vital part of a database which help in gathering and administration of pertinent data in a database. The set of programs or software tools that implement these tasks is entitled as database management system (DBMS). DBMS offers a convenient and efficient setting for the retrieval and storing of information in database. Large amounts of data can be managed through DBMS. DBMS assists the physical accumulation of logically related records (data) and the retrieval procedure. It makes an effort to prevent physical data from being redundant and upholds data integrity and independence. DBMS offers conveniences for institution to retrieve and regulate (Gunjal 2003).

Objectives of DBMS
Main goal of DBMS is to offer a convenient setting for retrieving and storing information in database. It is supported for both single and multi-user scenario (Gunjal 2003).

- To organize the mass storing of pertinent information.
- Make it simple for consumers to access the information.
- Retort quickly to consumer demands for information.
- Remove the unnecessary data.
- Permit several users to be active at once.
- Guard the information from physical damage and unlawful access.
- Control over the accuracy, reliability, security, and consistency of the data.
- Allow the database system to expand.
- To avail most recent updates in database instantly accessible.

Functions of DBMS
Some of the major functions of DBMS are as follows (Gunjal 2003):

- **Data Storage, Retrieval, and Update**: Since numerous users may share a database, the DBMS must support several consumer observations and facilitate effective storing, recovery, and modernizing for entire consumers.
- **Transaction Integrity**: A transaction is a succession of activities that account a specific type of viable action. The DBMS must include tools for the consumer or application software to create operation restrictions, i.e., the rational start and

expiration of trades, in order to assurance operation reliability. The DBMS should then approve alterations for trades that are successful and deny variations for those that fail.

- **Recovery Services**: In the incident of a system breakdown, the DBMS should be capable to recover the database. Operator error, disc head accidents, and program faults are all potential sources of system failure.
- **Concurrency Control**: As a database is shared by several consumers, it is possible for numerous users to crack to obtain the similar information at once.
- **Security Mechanisms**: Information should be safeguarded contrary to unauthorized or deliberate usage or diversion. The DBMS offers tools for limiting user access to data and outlining the activities they are permitted to perform.
- **Data Communication Interface**: Users frequently use remote terminals in a telecommunications network to access databases. The course of dealings and the distant terminations is processed by a telecommunication monitor. In order for the system to support the user rather than be a burden, the DBMS must have an interface with one or added telecommunication observers. This will ensure that all indispensable tasks are carried out.
- **Integrity Services**: Facilities that help users ensure the integrity of their data must be provided by the DBMS. The DBMS and its software interfaces can be configured with a variability to manage checks and reliability constraints. The data dictionary is usually used to administer these tests.

3.1.4 Bioinformatics

Bioinformatics is a branch of research that combines computer science, biology, and information technology. It includes all of the many approaches and techniques for examining and gleaning physiologically pertinent data from the exponentially expanding biological and crucial sequence databases. A bioinformatician's prime possessions are the internet and computer tools. Anyone with internet connection and access to appropriate websites can now easily learn the makeup of biological fragments like nucleic acids and proteins through simple bioinformatics tools, from doctors to molecular biologists. Bioinformatics employ a selection of information bases, comprising raw DNA arrangements, protein arrangements, macromolecular assemblies, and genome sequencing (Bayat 2002; Diniz and Canduri 2017).

Functions of Bioinformatics

1. To examine gene expression and variation.
2. To analyze and forecast the structure and functionality of genes and proteins.
3. Estimation and discovery of gene directive systems.
4. Providing model settings of whole-cell modeling.

5. Intricate displaying of gene directing undercurrents and systems.
6. Employed to visualize and study of biological pathways to comprehend gene–disease relationships.

3.1.5 Cheminformatics

A subject that organizes and coordinates the use of computers in chemistry is known as "cheminformatics," a phrase that lexically combines the words "chemistry" and "informatics" (computer science). In the modern drug discovery process, cheminformatics has become an essential part (Gasteiger 2016). The functions of cheminformatics are (Gasteiger 2016):

1. Databases

   Undoubtedly, one of cheminformatics' most well-known functions is its ability to give users access to databases of chemical data on a scale that is unfeasible to reach by manually sifting through the chemical literature. Without databases, it would simply be unmanageable to get a summary of the identified chemistry given that there are currently 90 million known chemicals. Additionally, chemists are able to connect with databases using their native tongue of graphical representations such as structure diagrams and reaction equations. Present chemical study would not be conceivable without the progressions in cheminformatics combined into database.
2. Property Prediction

   For the property prediction of a chemical compound, numerous techniques have been established. In all of these techniques for a particular structure, descriptors are derived from a cheminformatic dataset, followed by data analysis or model building technique to establish the connection between the structural signifiers and the studied characteristic. Thus, cheminformatics are important for the property estimation of a particular structure.
3. Drug Design

   The field of drug design has seen the vast majority of cheminformatics applications. It is used in lead discovery, lead optimization, and modeling of ADMET properties. Cheminformatics has significantly aided in the creation of numerous novel medications. All major pharmaceutical corporations now have cheminformatics departments, and virtually every medicine that has recently been produced has used cheminformatics procedures at some stage.
4. Analytical Chemistry

   The majority of analytical chemistry tasks include a categorization issue, such placing a sample in a particular category. Early on, it was realized that computational approaches may substantially aid in the classification of analytical samples. This caused in the progression of chemometrics arena; which is a subfield of

cheminformatics. In chemometrics, maximum chemical information was acquired by analyzing chemical data from cheminformatics.

3.2 Drug Discovery Informatics

Information on biology, chemistry, and literature are integrated with the aid of informatics (computer-based storage and retrieval of information). The design of molecules for therapeutic intervention is made easier by this technical progress. The processes for drug research and development are accelerated and strengthened by integrated technology (informatics). When medical researchers conduct investigation into treatments for definite ailments or for certain patient demographics, massive quantities of data are generated. Discovery informatics implicates the establishment of systems that can work more efficiently with such massive collections of data. In future, these technological advancements are anticipated to surge in both consistency and opportunity. Thus, the integration of new informatics with pharmaceutical sciences is becoming a crucial part of the drug discovery process. In future, these technical expansions are projected to nurture both in terms of their dependability and opportunity. Thus, the developing informatics combined with pharmaceutical sciences is a crucial element of drug discovery (Yadav et al. 2020).

The component of drug discovery informatics can be classified into two categories (Fig. 3.2):

- Information resources
- Software tool

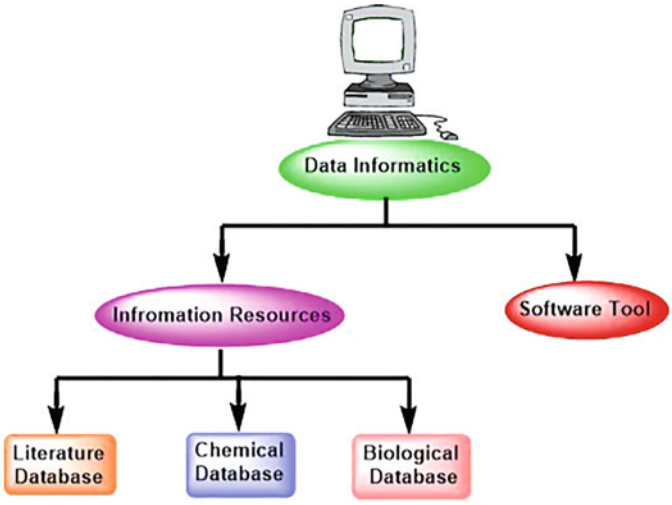


Fig. 3.2 The different component in classification of drug discovery informatics

3.2.1 Information Resources

Drug development and discovery is an extremely complicated process that generates enormous volumes of data and information. Information resources include all the available data for an existing molecule including structural, graphical, physicochemical, pharmacodynamics, and pharmacokinetics data. Information resources are classified into following categories;

- Literature databases
- Chemical databases
- Biological databases

3.2.1.1 Literature Databases

Past print and electronic indices continue to exert govern over the remarkable increase in the amount of circulated data (journals and books). Electronic files were created to help with information storage, retrieval, and dissemination. Database search engines can now be used to perform the literary searches and information retrieval that was hitherto the domain of librarians. These engines are used for information retrieval by researchers from academic research centers and research labs. Many literature databases (databases for information) are available on the internet, and these databases are helpful to both biological and pharmaceutical experts.

Following are some of the example of literature databases.

1. MEDLARS

   MEDLARS or Medical Literature Analysis and Retrieval System is a bibliographic database. It is created by the National Library of Medicine (NLM), a division of the National Center for Biotechnology Information (NCBI) in the United States. MEDLARS revolutionized the process of searching for literature. It retains a medical science article index. Every day, the general public, scientists, librarians, and health professionals conduct about 350,000 MEDLINE searches over the World Wide Web (Dee 2007).
2. MEDLINE

   MEDLINE is the name of the MEDLARS online version (MEDLARS online). It is a well-known bio-medical database that contains information on more than 25 million citations. Medical data are accessible through this database (clinical and therapeutic topics). Literature published between 1966 and the present is included in MEDLINE, as is some older literature. PubMed is the exploration contraption presented for pursuit against MEDLINE. The Literature Selection Technical Review Committee (LSRTC) of NCBI recommends that the vast majority of journals be added to MEDLINE. MEDLINE gives preference to the fields of biology, behavioral sciences, chemical sciences, bioengineering, clinical

care, public health, and health policy development. Additionally, it discusses biophysics, marine biology, plant and animal science, and biology (Cross 2006).

3. EMBASE

   EMBASE or Excerpta Medica Database is a bibliographic database consists of biological and pharmacological published literature. EMBASE is created by Elsevier, extending coverage to publications from Europe and works authored in languages other than English. It offers access to journals as old as going back to 1947. It includes MEDLINE titles and encompasses more than 32 million records. This database helps researchers understand drugs–disease association and drug–drug relations. It is highly helpful in finding pharmacological side effects (Lefebvre et al. 2008).

4. International Pharmaceutical Abstract (IPA)

   The International Pharmaceutical Abstracts (IPA) database was molded by the American Society of Health-System Pharmacists (ASHP). It gives users access to a vast database of data on drug use. It includes the full range of drug therapy and pharmaceutical information, making it simple for researchers, toxicologists, cosmetic manufacturers, medical librarians, and healthcare professionals to find solutions to any drug-related issues they run into.

5. Chemical Abstract Service (CAS)

   The Chemical Abstracts Service (CAS) is an arm of the American Chemical Society. For more than a century, it has offered the most complete record of research in chemistry and related fields. It includes scholarly publications such as books, conference proceedings, technical reports, patents, and dissertations. This database is highly helpful in the process of finding and developing new drugs (Efremenkova and Krukovskaya 2007).

6. TOXLINE

   TOXLINE is a toxicity database and subset of the National Library of Medicine's (NLM), which offered bibliographic data on the biochemical, pharmacological, physiological, and toxicological properties of medications and other compounds. References to TOXLINE were collected from a variety of sources and categorized into component sub files (Schultheisz 1981).

3.2.1.2 Chemical Databases

Chemical databases include encoded chemical configurations as well as molecular and atomic information. Most chemical databases contain two- (2D) and three-dimensional (3D) essential simulations in various domains. The large number of fragments in chemical databases, as well as the supplementary information, necessitates refined information structures. Present drug discovery processes demand the use of systems that can access and control enormous amounts of chemical data. Computer technologies improve the progression of storing and retrieving chemical information. It includes the design, organization, storage, management, retrieval, analysis, dissemination, visualization, and application of chemical data. Chemical database contains an extensive variety of data, which have vital input in the course of

rational drug design. Following are some of the important chemical database (Yadav et al. 2020).

1. ChemBank

   It was initially established by the Harvard Institute of Chemistry and Cell Biology as an open, web-based informatics. This database is currently kept up-to-date by the Chemical Biology Program and the Chemical Biology Platform at Broad Institute of MIT and Harvard. This database is openly accessible and uses Daylight Chemical Information Systems. It contains information about biochemical assays as well as molecular characteristics of small molecules. This database supports both text- and structure-based searches. The pursuit can be narrowed down to specific subsections (e.g., natural products; FDA permitted medications) (Seiler et al. 2008).
2. Chemical Entities of Biological Interest (ChEBI)

   Chemical Entities of Biological Interest (ChEBI) is an open molecular object dictionary aiming on trivial chemical objects. It is a component of the European Bioinformatics Institute's Open Biomedical Ontologies initiative. ChEBI includes an ontological arrangement and elucidates the connections among molecular objects and their parent or children. The molecular objects comprise atom, molecule, ion, radical, complex, and conformer (de Matos et al. 2010).
3. Developmental Therapeutics Program (DTP)

   To aid in the finding and improvement of novel anticancer medicines, the National Cancer Institute's (NCI's) Developmental Therapeutics Program (DTP) offers support facilities and resources to academic and commercial research communities around the globe. The DTP maintains a stockpile of pure natural and synthetic chemicals that are being investigated as possible anticancer drugs. DTP assisted in the development of anti-cancer medicines such as paclitaxel, romidepsin, eribulin, sipuleucel-T, and dinutuximab (Monga and Sausville 2002).
4. Comparative Toxicogenomics Database (CTD)

   This database was created by the National Institutes of Environmental Health Sciences (NIEHS) at North Carolina State University (NCSU). It sheds light on intricate networks of protein and chemical interactions. It explains the molecular processes underpinning disease susceptibility variation and environmental influences (Davis et al. 2021).
5. DrugBank

   It is a web-based free database contains information on targets and drugs that is chemical, pharmaceutical, medical, and molecular biological. It includes chemical structures and more than 200 data fields for each drug. It includes FDA permitted small molecule medications and peptide, as well as nutraceuticals and experimental medications. DrugBank can be dug using both text- and structure-based approaches. There are presently 14,940 drug entries in DrugBank, including 2729 approved small molecule drugs. DrugBank data can help with target identification, biological activity screening, and drug breakdown estimations (Wishart et al. 2008).

6. Carcinogenic Potency Database (CPDB)

   Lawrence Berkeley Laboratory and the University of California collaborated to create this database. It offers outcomes of persistent and longstanding animal cancer examinations that have been documented in published works (Gold et al. 1991).
7. Non-Redundant Database of Small Molecules (NRDBSM)

   It is suitable for virtual high-throughput screening of small molecules. Particular attention is paid to physicochemical characteristics and Lipinski's rule of five (RO5). IIT Delhi's Supercomputing Facility for Bioinformatics & Computational Biology is the source of this information.
8. Cambridge Structural Database (CSD)

   A thorough and carefully curated chemical database. It comprises coordination compounds and small organic molecules with empirically established molecular structures that have been documented in the literature. Additionally, it also includes data that were avail straight through the CSD and are not accessible elsewhere (Groom et al. 2016).
9. CrossFire Beilstein database

   The Beilstein Institute created the first database, which was based on the Beilstein handbook of organic chemistry. This database includes compounds prior to 1960. It includes arrangements, Beilstein and CAS registry records, titles, formulas, natural product separations, chemical analogues, and reference information dating back to 1771. It comprises Gmelin database as well as Elsevier's patent chemistry database. It also offers spectral, thermodynamic, biological, and toxic property information, as well as their applications (Vanco 2003).
10. Therapeutic Target Database (TTD)

    It contains details on the prime targets and mechanism of action of the medications'. This database offers information on metabolic pathway connections, signal transduction, and metabolic response disruption. This database can be used to research the drug's molecule's affinity for the macromolecule (Zhu et al. 2009).
11. Kyoto Encyclopedia of Genes and Genomes (KEGG)

    For the majority of species and chemical compounds, it covers genomic, chemical, and network/pathway data. This offers metabolic paths, gene nodding, protein connections, and metabolite structure information relevant to an organism (Kanehisa and Goto 2000).
12. ZINC

    ZINC is an open database of small molecule that can be found on the market for use in virtual screening. The chemical in the database fulfills the Lipinski rule of five (RO5). ZINC provides capabilities for docking, examining substructures, and procurement aspect of chemicals. The drug-like characteristics of molecular structures are interpreted in the ZINC. Molecular mass, log P, numeral of rotatable bonds, hydrogen bond donors and acceptors, chiral centers, desolvation energy, net charge, rigid fragments, and molecular function information can all be used to search for molecules (Irwin and Shoichet 2005).

13. PubChem

    The National Institutes of Health (NIH) created and maintain this open chemical database. It contains information on the small molecules structures, identifiers, chemical and physical characteristics, biological actions, patents, safety, health, and harmfulness ratings. It consists of nucleotides, carbohydrates, lipids, peptides, and macromolecules that have undergone chemical modification. It ranks database molecules against a query compound using an 881-bit finger point (Kim et al. 2016).

3.2.1.3 Biological Databases

Biological databases house well-ordered and determined biological information. These databases are collections of organized biological data, such as DNA sequences, protein sequences, and molecular structures. The growing pool of biological information necessitated the creation of biological databases. The most significant biological datasets are generally categorized into:

1. Sequence databases
2. Structure databases

Sequence Databases

Each gene or protein is represented by its nucleotide and amino acid arrangement. The sequencing of genes and proteins has been aided by developments in molecular biology and related sciences. The following list includes some of the most important sequencing databases.

1. GenBank

    GenBank is a genetic sequence database retained by the National Institutes of Health (NIH). It is a vast public database of protein and nucleotide sequences. The most rapidly expanding database of known genomic sequences is GenBank. GenBank files contain references, phylogenetic categorization, accession numbers, and gene names. More than 20 million distinct sequences make up the more than 27 billion nucleotide bases in GenBank. The GenBank database is anticipated to deliver and boost access to the most recent and broad DNA arrangement data in the scientific circle (Clark et al. 2016).
2. European Molecular Biology Laboratory (EMBL)

    The European Molecular Biology Laboratory (EMBL) is a large database that contains DNA and RNA sequences. These information are gathered from patent applications and scientific literature. The crucial component of EMBL is the data exchange between GenBank and the DNA Database of Japan (DDBJ). About 110 separate research and service teams working in the fields of molecular biology and bioinformatics undertake research at the EMBL (Patel et al. 2022).

3. DNA Data Bank of Japan (DDBJ)

   The DNA Data Bank of Japan (DDBJ) was founded by the Japanese National Institute of Genetics. It interchanges molecular data with GenBank at NCBI and EMBL of EBI. DDBJ generally takes data from Japanese researchers but also from other nations. For every entry, accession number is allocated by DDBJ (Mashima et al. 2017).
4. SwissProt

   SwissProt is a curated protein sequence database, set up by the Swiss Institute of Bioinformatics (SIB), a section of the University of Geneva's department of Medical Biochemistry, and the European Molecular Biology Laboratory (EMBL). The curated protein sequence database is called SwissProt. It offers extensive annotations, including those on roles, domain structures, and post-transitional alterations (PTMs). The three distinctive criteria that set the SWISS-PROT database apart from other protein sequence databases are (i) annotations, (ii) low redundancy, and (iii) assimilation with other databases (Bairoch and Apweiler 2000).
5. Protein Information Resource (PIR)

   Protein Information Resource (PIR) was found by the National Biochemical Research Foundation (NBRF). It aids in the study of computational biology, functional genomics, and molecular evolution. PIR was in association with Munich Information Center for Protein Sequences (MIPS) and Japan Information Database formed International Protein Sequence Database (PIR-PSD). GeneBank of NCBI, EMBL of EBI, DDBJ translations, published publications, and uninterrupted submissions to PIR are the main data sources for PIR-PSD (Wu et al. 2003).
6. PFAM

   PFAM is an EMBL centered database of protein families that contain various sequence alignments and annotations created using Hidden Markov Models (HMM). With extensive protein coverage, it offers a thorough and accurate classification of protein families and domains. For enhanced homology detection, this database offers sequence arrangement for protein domains and preserved protein regions (Finn et al. 2016). It is divided into two sections:

   (a) PFAM-A: It comprises curated families with corresponding profile HMMs.
   (b) PFAM-B: It contains the group of sequence sections that PFAM-A does not.

7. InterPro

   InterPro is a database for protein sequence functional enquiry that is based on the EMBL. Proteins are categorized at the superfamily, family, and subfamily stages. It provides estimated investigation of the existence of repeats and functional domain (Blum et al. 2021).
8. PROSITE

   PROSITE is a repository of biologically important sites, patterns, and profiles. It is responsible for defining the role of proteins transcribed from genomic or cDNA sequences. It contains a brief account of the protein family/domain as well as a summary of the pattern or profile's development (Sigrist et al. 2010).

9. iProClass

   The iProClass database is an integrated classification database designed to serve as a central repository for annotated protein family information. It provides access to above 175 biological databases. Protein families, functions and pathways, interactions, structures and structural classifications, genes and genomes, ontologies, literature, and taxonomy are all covered. It aids in the annotation of protein sequences, as well as genomic and proteomic research (Huang et al. 2003).

Structure Databases

The macromolecular characteristics cannot be fully explained by the 2D structures. The molecular framework in 3D is crucial for describing their roles. The 3D structures can provide a greater understanding of the structure. The 3D structures of macromolecules have been created using molecular biology and biophysical approaches such as X-ray crystallography and NMR spectroscopy. These 3D structures are stored in repositories known as structure databases. The list and description of the most significant structural databases are provided below.

1. Protein Data Bank (PDB)

   Brookhaven National Laboratories created the Protein Data Bank (PDB), which is overseen by the Research Collaboratory for Structural Bioinformatics (RCSB). It manages the databases of ligands and macromolecules with their 3D experimentally confirmed structures. Access to sequence information, atomic coordinates, crystallization circumstances, 3D structure neighbors, calculate methods, geometric data, structural factor, 3D photos, are all made available in this database. Each protein is given a special PDB-ID or PDB code, which consists of four alphanumeric characters. By entering text-based search terms (keywords), the PDB-ID, and the author name, you can examine against 3D structure of the interest protein. Options for advanced searches are offered for more focused and effective search. PDB provides the following molecular sequence analysis tools: alignments of sequence and structure, protein symmetry, structural quality, and protein location in the genome (Berman et al. 2000).
2. CATH

   The CATH database is an openly accessible internet tool that offers details on the interactions between protein domains across time. Protein domain structures are categorized hierarchically at four levels using the CATH system. The term CATH of the database stands for Class (C), Architecture (A), Topology (T), and Homologous superfamily (H). And these are the four primary levels in a hierarchical classification system,

   (a) Class: Proteins can be grouped together according to how much secondary structure they share. It speaks to the content of secondary structure.
   (b) Architecture: Regardless of the topological connection type, it refers to the common layout of the secondary structural components.

(c) Topology: It describes the topological connectivity and shape of structural parts.
(d) Homologous Super Family: This shows how the various proteins have changed throughout time.

CATH is an important tool for scholars because proteins with even minor arrangement resemblances are often organizationally and functionally correlated (Dawson et al. 2017).

3. Structural Classification of Proteins (SCOP)

The crucial element in determining the evolutionary process of sequences is understanding the relationships between structural similarities. To define the structural and evolutionary link among the proteins, whose 3D assemblies have previously been determined, the structural grouping of proteins (SCOP) database was created. A variety of automatically generated visualization-focused software tools are used to build the database. It includes experimentally determined biomolecular structures of proteins, RNA, and DNA that are deposited in the Protein Data Bank (PDB). It provides bonds, ligands, literature, chemical graphs, 3D domains, and explicit (Lo Conte et al. 2000).

4. Molecular Modeling Database (MMDB)

The National Center for Biotechnology Information (NCBI) hosts the Molecular Modeling Database (MMDB), a collection of 3D bimolecular structures that have been identified through research. It is an essential component of the NCBI Entre information retrieval system. By entering a key word, accession number, author, and journal name, a protein structure can be searched in the MMDB. More than 28,000 structures can be found in MMDB, which is connected to the other NCBI databases for sequences, bibliographic references, taxonomic groupings, and sequences and structures neighbors (Madej et al. 2012).

5. DNA Databank

This database was created by the National Institutes of Health (NIH). DNA clones of many tissues, including liver, muscle, skin, hair, and saliva, are kept in the DNA Data Bank. Information from DNA databanks is helpful in genetic genealogy, paternity testing, criminal investigations (genetic fingerprinting), and illness screening for genetic disorders(Johnson et al. 2003).

3.2.2 Software Tools Used in Drug Discovery

There are numerous computer programs or tools with a variety of applications, accuracy, and approaches accessible globally for the drug development process. These computational tools or programs are used in the design and discovery of new medications for molecular modeling, docking, and protein conformation, ADMET predictions, pharmacophore mapping, visualization of docking postures, computation of force fields, homology modeling, generation of 3D structures, and pharmacophore searches. Computational techniques are crucial in the current drug

discovery system for lead optimization, lead identification, and pre-clinical in vitro assessment criteria for entering clinical development (Patel et al. 2022). Following are some of the widely used software tools used for the different purpose in drug development process.

3.2.2.1 Chemical Drawing Tools

1. ChemDraw
   A molecular editor was developed by the cheminformatics company CambridgeSoft. The first iteration of the chemical editor, named ChemDraw, was developed by David A. Evans and Stewart Rubenstein in 1985. PerkinElmer bought the business in 2011. ChemDraw, Chem3D, and ChemFinder are all part of the ChemOffice software suite, which is accessible on both the Macintosh and Microsoft Windows operating systems. The recommended XML-based CDXML format as well as the binary CDX file format are the native file formats for ChemDraw. Additionally, ChemDraw supports MOL, SDF, and SKC chemical file formats for import and export. It is used to design chemical structures, swap chemical IUPAC names for structures, replicate NMR and mass spectral data, tidy up structures, convert 2D to 3D structures, and construct chemical structures using many different international journal-style templates (Patel et al. 2022).
2. ChemSketch
   The ACD/Labs-developed molecule editor named ChemSketch. The tool is used for 2D and 3D structure drawing, optimization, and viewing. ChemSketch empowers molecules and molecular models presented in two and three dimensions, to apprehend the arrangement of chemical bonds and the type of the functional groups. It also calculates molecular properties such as molecular weight, density, and molar refractivity. ChemSketch is suitable only for Windows operating systems (Patel et al. 2022).

3.2.2.2 Molecular Modeling Tools

1. CHARMM
   CHARMM stands for Chemistry at Harvard Macromolecular Mechanic, a program for simulating molecules with a wide range of applications to many-particle systems, including support for multi-scale techniques like QM/MM (Quantum mechanics/Molecular mechanics), MM/CG (molecular mechanics/coarse-grained), and a variety of implicit solvent models. CHARMM also finds extensive uses for inorganic materials in material design. A big group of developers led by Martin Karplus regularly maintains CHARMM (Brooks et al. 2009).
2. Amber
   Assisted Model Building with Energy Refinement, sometimes known as Amber, is a group of biomolecular simulation tools. Two things are stated by the label "Amber." It is first a primary pool of molecular mechanical force fields

for the simulation of biomolecules. Second, it is a collection of applications for molecular simulation that contains source code and demonstrations (Patel et al. 2022).

3.2.2.3 Homology Modeling Tools

1. I-TASSER
   A hierarchical method to protein arrangement estimation and structure-based role annotation is I-TASSER (Iterative Threading ASSEmbly Refinement). It initiates by using the multiple threading tactic LOMETS to detect structural models from the PDB, and then iterative template-centered fragment assembly models are used to construct full-length atomic prototypes. By re-threading the 3D prototypes via the protein function database BioLiP, function insights of the target are obtained. The server is being actively developed with the aim of offering the most precise predictions of protein structure and function using cutting-edge algorithms. The server is generally available for non-commercial usage (Yang et al. 2015).
2. SWISS-MODEL
   SWISS-MODEL is a web-based cohesive facility devoted to homology modeling of protein structural. It offers instructions for creating protein homology models at various levels of intricacy. The first completely automated protein homology modeling service was delivered by SWISS-MODEL. The modeling of homo and heteromeric complexes using the amino acid arrangements of the interaction partners as a starting point has recently been added to its modeling functionality. Other newly added qualities comprise the incorporation of a novel modeling engine, ProMod3, with improved precision of the created models and better local model quality estimation methodology. It is one of the extensively engaged structure modeling servers globally (Waterhouse et al. 2018).

3.2.2.4 Binding Site Prediction Tools

1. MED-SuMo
   The Protein Data Bank is an exclusive open cradle of macromolecular structures often with co-crystallized ligand(s). MED-SuMo software relates and superimposes any 3D interaction on molecule surface over the PDB. It offers medicinal chemists, molecular modelers, and crystallographers fresh possibilities for structure-based drug discovery. All PDB assemblies that are sharing a 3D system of intermolecular interactions, comprising as charges, H-bonds, hydrophobic and aromatic stacking, are recovered by MED-SuMo and sited onto binding spot of concern or a identified protein whole surface (Doppelt-Azeroual et al. 2009).
2. CASTp
   An online tool entitled the Computed Atlas of Surface Topography of Proteins (CASTp) can be used to find, define, and measure hollow surface areas on 3D protein assemblies. These consist of cavities hidden inside proteins and compartments on protein sides. The measurement takes into account the molecular

surface model (Connolly's surface) and the solvent reachable surface model (Richards' surface), both of which may be determined analytically, to determine the extent and dimensions of the pocket or void. Protein surface characteristics and functional areas can be studied using CASTp. A graphical user interface, configurable cooperative visualization, and on-the-fly calculation for handler uploaded assemblies are all features of CASTp (Tian et al. 2018).

3.2.2.5 Docking Tools

1. AutoDock
   AutoDock is a useful docking software that calculates the affinity of a known 3D structure docked with a target protein. This docking score provides information on how a new molecule is fixed to the target's active site. This can reveal a molecule's predictive activity prior to actual production, which can readily save time, money, and resources. The AutoDock has found use in X-ray crystallography, structure-based drug design, lead optimization, virtual screening (HTS), combinatorial library design, protein–protein docking, chemical mechanism investigations, and other fields (Patel et al. 2022).
2. GEMDOCK
   GEMDOCK is a molecular docking software that uses a generic evolutionary algorithm and scoring function. GEMDOCK reduces the search space of ligand structure conformations by using an empirical scoring function similar to the AMBER-based energy function and a generic evolutionary strategy based on a novel rotamer-based mutation operator. GEMDOCK is an automatic system that creates all docking variables such as atom formal charge, atom type, and a protein's ligand binding spot. The program was created by Jinn-Moon Yang, a professor at Institute of Bioinformatics, National Chiao Tung University (Patel et al. 2022).

3.2.2.6 Pharmacophore Finding Tools

1. AnchorQuery
   A program called AnchorQuery is utilized for interactive virtual screening of pharmacophore searches. It is an online tool for the rational plan of protein–protein interaction (PPI) inhibitors based on structural analysis. Rapid library screening for pharmacophore search of over 31 million synthesizable compounds are directed by customized plan to favorably target PPIs. Every chemical in the library has an anchor motif that is bioisosteric to an amino acid residue and is reachable over one-step multi-component reaction (MCR) chemistry. For consumers to conduct online collaborative virtual screens of millions of chemicals, including pharmacophore explanation and exploration, as well as improvement analysis, AnchorQuery offers all the tools required (Koes et al. 2018).

2. Ligand Scout

   Ligand Scout is a computer program used to generate 3D pharmacophore simulations from the structural information of macromolecule-ligand complexes or from sets of organic compounds used for training and testing. It includes a thorough characterization of the 3D chemical properties that characterize the interaction between a bound small organic molecule (ligand) and the macromolecule's surrounding binding site, such as lipophilic regions, hydrogen bond donors, and acceptors (Wolber and Langer 2005).

3.2.2.7 Screening Tools

1. MedChem

   MedChem Studio is a comprehensive cheminformatics software suite that includes all of the capabilities needed for high-throughput screening analysis, lead discovery and prioritization, de novo design, scaffold hopping, and lead optimization. A license is not required for MedChem Studio's "VIEWER" mode. It promotes association between scientists with various sets of expertise by enabling an unrestricted number of users to import structure files, including the exclusive file format of MedChem Studio. For instance, a medicinal chemist can see the outcomes of an analysis conducted by a computational chemist. A free molecular sketch tool called MedChem Designer is accessible through MedChem Studio. It can be employed to input or manage structures, explain structure disputes, envision metabolites, and much more (Patel et al. 2022).
2. PLANTS

   PLANTS software is employed for lead optimization and virtual screening. This software is based on ant colony optimization (ACO) algorithm, which uses an artificial ant colony to identify the lowest energy conformation of the ligand in the protein's binding region. ACO-based search engine, two scoring functions (PLANTSCHEMPLP and PLANTSPLP), rigid-body docking of multi conformer collections into rigid and flexible receptors, constraint system, docking with certain explicit, displaceable water molecules, fully automatic ligand setup, virtual screening, and rescoring capability are some of the features of this software (Patel et al. 2022).

3.2.2.8 Target Prediction Tools

1. Swiss Target Prediction

   SwissTarget Prediction is a web service that has been available since 2014 and attempts to forecast the most likely protein targets of small compounds. Calculations are made using reverse screening and the similarity principle. The calculations are made by looking for similar molecules in 2D and 3D among a bigger gathering of 376,342 compounds that have been experimentally shown to be active on a larger set of 3068 macromolecular targets. The ability to interoperate allows for the simple offer of any input or output molecule to other online computer-aided drug design tools established by the Swiss Institute of

Bioinformatics (SIB). The new SwissTarget Prediction is completely free (Patel et al. 2022).

2. SEA

   SEA (Similarity Ensemble Approach) is an online search tool that identifies proteins created on their ligands' set-wise chemical resemblance. It can be employed to scan big compound databases quickly and create cross-target similarity maps. It employs a pool of over 65,000 ligands annotated for pharmacological objects, with the majority of annotations containing hundreds of ligands. SEA shows both predicted and unexpected commonalities, which can be tested by investigating the ligands' "off-target" activities. Only ligand chemistry was employed to generate these links, and such grouping is a promising aspect of this approach.

3.2.2.9 Ligand Design Tools

1. AutoT&T

   The Automatic Tailoring and Transplanting (AutoT&T) approach was created as a multipurpose computational tool for both automatic lead optimization and lead discovery in molecular-targeted drug discovery. This approach finds appropriate fragments on reference molecules and then transplants them onto the specified lead compound to create new ligand molecules. Binding affinities, synthesis feasibility, and drug-likeness features are also assessed in order to identify interesting contenders for additional attention (Li et al. 2016).

2. Mcule

   Mcule is a drug discovery software that operates online. By offering the best downloadable compound databases and molecular modeling tools, it provides a distinctive solution for the pharmaceutical and biotechnology industries. A ligand-based strategy can be used to create novel scaffolds by drawing a reference structure followed by picturing the relationship between the demand and the discovered scaffold (Odhar et al. 2019).

The following are some of the lead optimization tools employed by Mcule:

1. 1-Click Docking
2. 1-Click Scaffold Hop
3. Property calculator

3.2.2.10 Binding-Free Energy Estimation

1. Lead Finder

   The free energy of binding is accurately calculated using Lead Finder software concurrently with ligand docking calculations. To determine the binding energy of protein–ligand interactions, a different scoring function known as the dG-score function is employed. In Lead Finder, an extensive library of 330 experimentally described protein–ligand complexes with a wide range of ligand binding affinities

and other physicochemical characteristics have been used to validate the accuracy of binding energy prediction. For a preset protein–ligand structure derived from experimental data or other molecular modeling research, binding energy calculations can be carried out either concurrently with ligand docking or separately. Lead Finder's capability to precisely anticipate the free energy of ligand binding can be highly helpful in drug discovery investigations, simulating ADMET characteristics, researching enzyme specificity, and rationally designing enzymes (Patel et al. 2022).

2. BFEE (Binding-free energy estimator)

   Binding-free energy estimator (BFEE) is python-based software that automates total BFEE via molecular dynamics simulations using either the alchemical or geometric path. It employs generalized, best-fit-rotation-based geometric variables, making it theoretically accessible to several protein–ligand composite. The geometric course investigates the grades of liberty one by one using 1D free-energy assessments. To safeguard the merging of the simulations, a thermodynamic cycle in which the ligand and geometric restrictions are dissociated separately (Fu et al. 2018).

3.2.2.11 QSAR Prediction Tools

1. CDK

   A java-based open-source program called CDK (Chemistry Development Kit) is used to compute molecular descriptors for chemical structure. Structure–activity relationships and QSAR descriptor calculations both use these chemical descriptors. Coordinate creation and rendering, canonical identifiers for quick accurate searching, substructure and SMARTS pattern searching, ECFP, Daylight, MACCS, and other fingerprint approaches for resemblance exploration are only a few of the important features it has. SMILES, SDF, InChI, Mol2, and CML are some of the file types that this program can read and write (Patel et al. 2022).

2. QSAR-Co

   The abbreviation QSAR-Co stands for "QSAR with conditions," an open source standalone tool made for building reliable classification-based QSAR models. This software's prime purpose is to easily create classification-based QSAR models that can handle both types of instances, that is, situations with and without conditions. The software has two components, namely the Screen/Predict module and the Model Development module. For the building of models and screening queries on chemical databases, these modules offer a number of crucial functions (Patel et al. 2022).

3.2.2.12 ADME Toxicity Prediction Tools

1. SwissADME

   The SwissADME web tool is openly available, designed for comprehensible offer and simple result inquiry, even for CADD novices. In comparison to the most

advanced free web-based tools for ADME and pharmacokinetics (such as pkCSM and admetSAR) and aside from exclusive access to effective techniques (such as iLOGP16 or the BOILED-Egg and Bioavailability Radar), SwissADME's solid facts comprise, but are not restricted to: a variety of input methods, computation for numerous molecules, and the capability to display, save, and share outcomes either per specific molecule or over global intuitive and interactive displays. Established and managed by the Molecular Modeling Group of the Swiss Institute of Bioinformatics (Daina et al. 2017).

2. pkCSM

   pkCSM is a free online tool that utilize graph-based marks to forecast pharmacokinetic features. The improvement of predictive regression and classification simulations by the pkCSM signatures was successful for each of the five key types of pharmacokinetic features. The 30 predictors built by the pkCSM, which is organized into five primary classes: absorption (seven predictors), distribution (four predictors), metabolism (seven predictors), excretion (two predictors), and toxicity (ten predictors), predict their pharmacokinetic and toxicological features. The pkCSM platform for predicting ADMET characteristics can be split into two categories of great analytical simulations: (a) 14 regression simulations that seek to calculate a numeric quantification of the pharmacokinetic or toxicological characteristics and (b) 16 groups simulations that divide the result into two classes (Pires et al. 2015).

3.3 Conclusion and Future Perspectives

Preclinical drug discovery research is a highly data demanding progression that can be significantly accelerated by easy retrieve to knowledge that has already been gathered by the public. Information technologists may easily take use of advancements in database technologies made by a variety of disciplines as they create data management plans for the development of new drugs. Through the encoding of chemical cartridges that facilitate and normalize the storing and use of molecular data, chemical informaticians have been instrumental in this advancement.

Given key new achievements in articulating molecular demonstrations that back much added effective and smart database enquiries exclusively in the space of chemical resemblance pursuits that are possibly essential for analoging and novel lead discovery, one can visualize the boundless aids that can be resulting from a conjugal between the remarkable supremacy of present databases and the abundant intellect being programed into novel chemical informatics approaches.

This new effective and efficient method of building and using the understanding pyramid can be a crucial component of a developing formula for feat in the extremely difficult and competitive task of finding fresh leads with the ability to endure in the pharmaceutical industry. The structure, dynamics, surface characteristics, and thermodynamics of inorganic, biological, and polymeric systems are increasingly often studied using computational software tools. A crucial element

of the roadmap for drug discovery involves computational software tools. It is frequently employed in the processes of structure-based drug design and rational drug design.

Through the assistance of software, we may create novel medications (ligands) by examining the receptor-drug binding affinity of 3D structures (protein, receptor, and enzymes). The creation of a novel chemical entity requires the design and discovery of drugs. Numerous computational tools are accessible for this procedure on a worldwide scale; these computational software tools are quick, free, and available online in abundance, and there are some computational tools which are paid, but more powerful and effective.

Informatics tools and databases are slowly turning into the pillar of foundation in modern drug discovery research, as they can be easily, freely accessible and non-expert users can explore them and build complex queries. Shorter time to market for new treatments as a result of the use of these informatics tools and databases has had a huge impact on society, especially in times of national and international crises like the Zika and Wuhan virus epidemic. In the days to come with exposure to modern technologies, informatics tools and databases will grow exponentially and they will be integral part of public health care in finding new remedies.

References

Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28(1):45–48

Bayat A (2002) Science, medicine, and the future: bioinformatics. BMJ 324(7344):1018–1022

Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. Nucleic Acids Res 28(1): 235–242

Blum M, Chang HY, Chuguransky S et al (2021) The InterPro protein families and domains database: 20 years on. Nucleic Acids Res 49(D1):D344–D354

Brooks BR, Brooks CL, Mackerell ADJ et al (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30(10):1545–1614

Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2016) GenBank. Nucleic Acids Res 44(D1):D67–D72

Cross J (2006) MEDLINE, PubMed, PubMed Central, and the NLM. Editors' Bulletin 2(1):1–5

Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7(1):42717

Davis AP, Grondin CJ, Johnson RJ et al (2021) Comparative Toxicogenomics Database (CTD): update 2021. Nucleic Acids Res 49(D1):D1138–D1143

Dawson NL, Lewis TE, Das S et al (2017) CATH: an expanded resource to predict protein function through structure and sequence. Nucleic Acids Res 45(D1):D289–D295

Dee CR (2007) The development of the Medical Literature Analysis and Retrieval System (MEDLARS). J Med Libr Assoc 95(4):416–425

Diniz WJ, Canduri F (2017) REVIEW-ARTICLE bioinformatics: an overview and its applications. Genet Mol Res 16(1):gmr16019645. https://doi.org/10.4238/gmr16019645

Doppelt-Azeroual O, Moriaud F, Adcock S, Delfaud F (2009) A review of MED-SuMo applications. Infect Disord Drug Targets 9(3):344–357

Efremenkova VM, Krukovskaya NV (2007) Chemical abstracts service centennial: facts and figures. Sci Tech Inf Proc 34(6):328–334

Finn RD, Coggill P, Eberhardt RY et al (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44(D1):D279–D285

Fu H, Gumbart JC, Chen H, Shao X, Cai W, Chipot C (2018) BFEE: a user-friendly graphical interface facilitating absolute binding free-energy calculations. J Chem Inf Model 58(3): 556–560

Gasteiger J (2016) Chemoinformatics: achievements and challenges, a personal view. Molecules 21(2):151

Gold LS, Slone TH, Manley NB et al (1991) The Carcinogenic Potency Database: analyses of 4000 chronic animal cancer experiments published in the general literature and by the U.S. National Cancer Institute/National Toxicology Program. Environ Health Perspect 96:11–15

Groom CR, Bruno IJ, Lightfoot MP, Ward SC (2016) The Cambridge structural database. Acta Crystallogr B Struct 72(2):171–179

Gunjal B (2003) Database system: concepts and design. Proceedings of 24th IASLIC–SIG-2003, Dehradun, Survey of India (15–18 December 2003)

Huang H, Barker WC, Chen Y, Wu CH (2003) iProClass: an integrated database of protein family, function and structure information. Nucleic Acids Res 31(1):390–392

Irwin JJ, Shoichet BK (2005) ZINC—a free database of commercially available compounds for virtual screening. J Chem Inf Model 45(1):177–182

Johnson P, Williams R, Martin P (2003) Genetics and forensics: making the national DNA database. Sci Stud 16(2):22–37

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30

Kim S, Thiessen PA, Bolton EE et al (2016) PubChem substance and compound databases. Nucleic Acids Res 44(D1):D1202–D1213

Koes DR, Dömling A, Camacho CJ (2018) AnchorQuery: rapid online virtual screening for small-molecule protein-protein interaction inhibitors. Protein Sci 27(1):229–232

Lefebvre C, Eisinga A, McDonald S, Paul N (2008) Enhancing access to reports of randomized trials published world-wide—the contribution of EMBASE records to the Cochrane Central Register of Controlled Trials (CENTRAL) in The Cochrane Library. Emerg Themes Epidemiol 5:13

Li Y, Zhao Z, Liu Z, Su M, Wang R (2016) AutoT&T v.2: an efficient and versatile tool for lead structure generation and optimization. J Chem Inf Model 56(2):435–453

Lo Conte L, Ailey B, Hubbard TJ et al (2000) SCOP: a structural classification of proteins database. Nucleic Acids Res 28(1):257–259

Madej T, Addess KJ, Fong JH et al (2012) MMDB: 3D structures and macromolecular interactions. Nucleic Acids Res 40(D1):D461–D464

Mashima J, Kodama Y, Fujisawa T et al (2017) DNA Data Bank of Japan. Nucleic Acids Res 45 (D1):D25–D31

de Matos P, Alcántara R, Dekker A et al (2010) Chemical entities of biological interest: an update. Nucleic Acids Res 38(Database issue):D249–D254

Monga M, Sausville EA (2002) Developmental therapeutics program at the NCI: molecular target and drug discovery process. Leukemia 16(4):520–526

Odhar HA, Rayshan AM, Ahjel SW, Hashim AA, Albeer AAMA (2019) Molecular docking enabled updated screening of the matrix protein VP40 from Ebola virus with millions of compounds in the MCULE database for potential inhibitors. Bioinformation 15(9):627–632

Patel J, Joshi H, Shah U, Jayvadan P (2022) A review on computational software tools for drug design and discovery. Indo Global J Pharm Sci 12:53–81

Pires DE, Blundell TL, Ascher DB (2015) pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. J Med Chem 58(9):4066–4072

Schultheisz RJ (1981) TOXLINE: evolution of an online interactive bibliographic database. J Am Soc Inf Sci 32(6):421–429

Seiler KP, George GA, Happ MP et al (2008) ChemBank: a small-molecule screening and cheminformatics resource database. Nucleic Acids Res 36(suppl_1):D351–D359

Sigrist CJ, Cerutti L, de Castro E et al (2010) PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res 38(Database issue):D161–D166

Tian W, Chen C, Lei X, Zhao J, Liang J (2018) CASTp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Res 46(W1):W363–W367

Vanco J (2003) The Beilstein CrossFire Information System and its use in pharmaceutical chemistry. Ceska Slov Farm 52(2):68–72

Waterhouse A, Bertoni M, Bienert S et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46(W1):W296–W303

Wishart DS, Knox C, Guo AC et al (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 36(Database issue):D901–D906

Wolber G, Langer T (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model 45(1):160–169

Wu CH, Yeh LS, Huang H et al (2003) The protein information resource. Nucleic Acids Res 31(1): 345–347

Yadav V, Tonk RK, Khatri R (2020) Molecular docking, 3D-QSAR, fingerprint-based 2D-QSAR, analysis of pyrimidine, and analogs of ALK (anaplastic lymphoma kinase) inhibitors as an anticancer agent. Lett Drug Des Discov 18(5):509–521

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. Nat Methods 12(1):7–8

Zhu F, Han B, Kumar P et al (2009) Update of TTD: therapeutic target database. Nucleic Acids Res 38(Database issue):D787–D791

Chapter 4
Multi-Omics Approaches in Drug Discovery

Gourav Rakshit, Komal, Pankaj Dagur, and Venkatesan Jayaprakash

4.1 Introduction

4.1.1 Definition of Multi-Omics

Understanding human health and illnesses thoroughly requires assessing molecular diversity and complexity at various levels, including the genomes, epigenomes, transcriptomes, proteomes, and metabolomes (Manzoni et al. 2018). Since the development of sequencing technology, molecular biology has depended more and more on data produced at these levels, referred to as "multi-omics" data. The advent of multi-omics datasets has transformed the ways biology and medicine are studied by making integrated system-level approaches possible (Manzoni et al. 2018).

In order to gain useful insights into how cells work, integrating clinical data with the analysis of multi-omics has become of prime importance to researchers (Subramanian et al. 2020). An integrative network of multi-omics data that provide insight into biomolecules from several layers appears to hold promise for a systematic and comprehensive understanding of complicated biology. To comprehend how molecules interact, integrated techniques incorporate various omics data in a concurrent or coherent way (Subramanian et al. 2020). They aid in evaluating the information flow from one omics level to the next and, in doing so, assist in bridging the morphophysiological gap. Since integrative approaches may assess biochemical pathways comprehensively, they have the potential to enhance prognostics and accuracy of the proposed disease phenotypes, which could eventually lead to better

G. Rakshit (✉) · Komal · P. Dagur · V. Jayaprakash (✉)
Department of Pharmaceutical Sciences and Technology, Birla Institute of Technology, Ranchi, Jharkhand, India
e-mail: venkatesanj@bitmesra.ac.in

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_4

Table 4.1 Different softwares used for multi-omics analysis

Sl. No.	Software	Utility of the software
1	Mothur	Marker gene data analysis pipeline
2	QIIME	Marker gene data analysis pipeline
3	RDPipeline	Marker gene data analysis pipeline
4	CloVR	Marker gene data analysis pipeline
5	PyroNoise	Flowgram noise correction
6	PyroTagger	Flowgram noise correction
7	Denoiser	Flowgram noise correction
8	AmpliconNoise	Flowgram noise correction
9	FlowgramFixer	Flowgram noise correction
10	Decipher	Chimeric sequence detection
11	ChimeraSlayer	Chimeric sequence detection
12	Perseus	Chimeric sequence detection
13	Uchime	Chimeric sequence detection
14	RDP classifier	Taxonomic identification
15	RTA X	Taxonomic identification
16	Muscle	Multiple sequence alignment
17	MAFFT	Multiple sequence alignment
18	Infernal	Multiple sequence alignment
19	PyNAST	Multiple sequence alignment
20	FastTree	Multiple sequence alignment

therapy and prevention (Subramanian et al. 2020). The different softwares used for multi-omics analysis are listed in Table 4.1 given below (Singh et al. 2017).

4.1.2 Different Omic Strategies of Multi-Omics Studies

The multi-omics approach combines all omic fields as illustrated in Fig. 4.1. Complicated syndromes and illnesses are a result of atmospheric change, climate change, and human evolution. Analysis of a single type of omics cannot provide a solution or cure for such disorders. Multi-omics strategy emerges as a hero in such situations (Hasin et al. 2017). The goal of multi-omics research is to combine several omic approaches to discover and treat a variety of complex diseases and allergies. The multi-omics approach makes use of all omic domains and aids in comprehending an organism's original and altered states (Hasin et al. 2017).

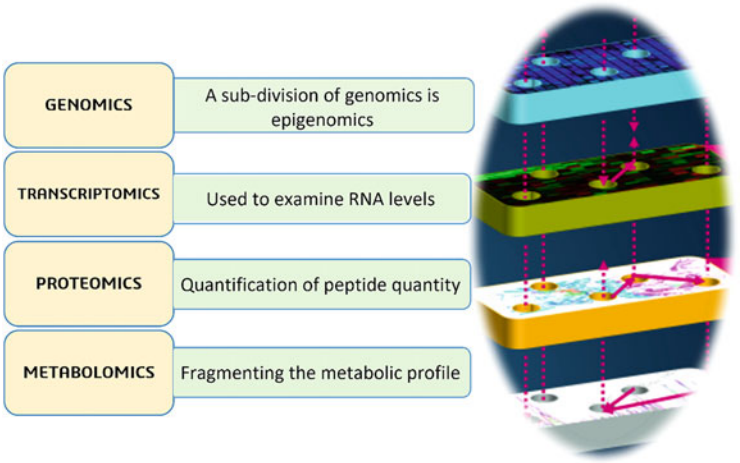


Fig. 4.1 A representation of different multi-omics fields

4.2 Omics Data Types and Repositories

4.2.1 Genomics

The omics field with the most experience in genomics. Genomic research in medicine focuses on finding genetic variations linked to illness, treatment outcomes, or patient prognosis in the future. The GWAS approach has been effectively applied across multiple human populations to identify hundreds of genetic changes connected to complex illnesses. In these studies, hundreds of subjects have dozens of biomarkers genotyped, and significant variations in recessive allele frequencies between cases and controls are utilized as evidence of a relationship (Hasin et al. 2017). Technologies & Methods Used Frequently: (i) Whole-genome, exome, and targeted sequencing are examples of NGS, and (ii) Microarrays.

4.2.2 Epigenomics

Through the study of epigenomics, the complete genome is examined for reversible changes to DNA or proteins associated with DNA, such as DNA methylation or histone acetylation. Modifications of DNA (covalent) and histones are important for the regulation of gene transcription and, subsequently, of cellular fate (Hasin et al. 2017). These changes can be long-lasting, and occasionally heritable, and they can be impacted by both genetic and environmental influences. Numerous published epigenome-wide association studies demonstrate the significance of epigenetic alterations in biological processes and the development of disease, despite the fact that

their role as mediators of transgenerational environmental impacts is still debatable (Weinhold 2006). Differentially methylated DNA regions could be used as markers of the disease state for metabolic syndromes, cardiovascular diseases, varied cancers, and many other pathophysiological disorders.

4.2.3 Transcriptomics

Transcriptomics is used to examine RNA levels both qualitatively and numerically across the entire gene sequence. The fundamental idea of biology holds that RNA serves as a biochemical link between DNA and proteins, which are considered to be the main operating interpreters of DNA (Hasin et al. 2017). Numerous times, it has been assumed that RNA's structural and regulatory functions, such as those in ribosome complexes and ChrX inactivation, are anomalous deviations from the normal. Large-scale transcriptome investigations that have emerged in the last ten years have demonstrated that up to 80% of the genome is transcribed, even though only around 3% of it encodes proteins (Hasin et al. 2017). Studies using RNA-Seq technology discovered thousands of novel isoforms and revealed the protein-coding transcriptome to be more complicated than was previously thought (Kukurba and Montgomery 2015). However, the advancement of the noncoding RNA field was these studies' even more significant contribution (Hasin et al. 2017).

4.2.4 Proteomics

Quantifying peptide quantity, modification, and interaction is done using proteomics. MS-based approaches have revolutionized analysis and quantification of proteins, and these techniques have recently been improved to allow for high-throughput examinations of tens of thousands of proteins in cells or bodily fluids (Tuli and Ressom 2009). Traditional unbiased techniques like yeast two-hybrid experiments and phage display can be used to find protein interactions. A single molecule is isolated using an immunoglobulin or a chromosomal identifier while employing affinity purification procedures. Any related proteins are then found using MS (Hasin et al. 2017). These affinity techniques have been used to study the overall interactions involving both the nucleic acids and proteins, sometimes in conjunction with chemical crosslinking (e.g., ChIP-Seq) (Tuli and Ressom 2009).

4.2.5 Metabolomics

The accepted definition of the emerging area of metabolomics is the comprehensive characterization of all substrates and low-weight molecules in a biological sample.

Because metabolomics tries to quantify molecules with different physical properties than genomic and proteomic approaches, it poses a considerable analytical difficulty (Eicher et al. 2020). As a result, as illustrated in Fig. 4.2, holistic metagenomic technology platforms typically employ the approach of fragmenting the metabolic profile into clusters of intermediates based on composite molecule polarization, and shared structural features to develop customized sample preparation and analytical techniques that are optimized for each. This leads to the quantification of the microbiome as a mosaic of results from several analysis tools. The methods used in metabolomics are constantly evolving and progressing as a field that is still in its infancy. This is, at least in part, due to the continual advancement of analytical equipments that have advanced capabilities every year (Clish 2015).

4.3 Multi-Omics Data Repositories

Data from the epigenome, the transcriptome, the metabolome, and the proteome are all included in the category of "multi-omics data" in general (Martorell-Marugán et al. 2021). Other biological information like the lipidome, phosphoproteome, and glycol-proteome can be added to the omics spectrum. Multi-omics information obtained for the identical assortment of samples, further insights into the transmission of biomedical data at various levels may be provided, which can aid in the identification of the mechanisms behind the biological state of relevance (Subramanian et al. 2020). Several publicly accessible databases that offer patient-diverse omics data sets are included in Table 4.2.

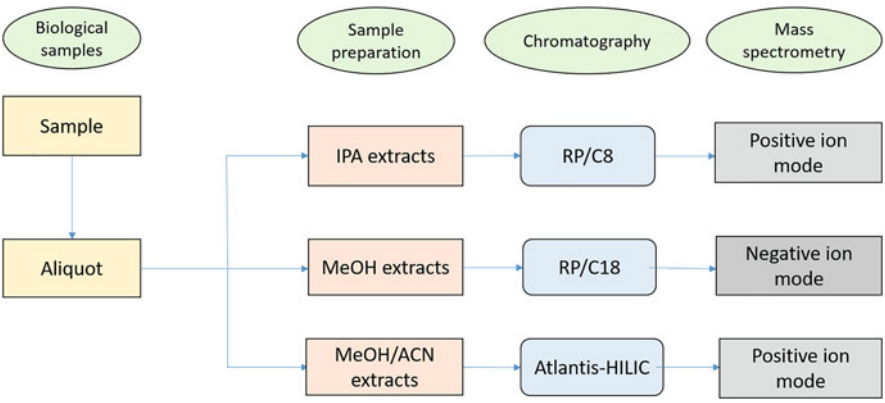


Fig. 4.2 Analytical techniques optimized for each metabolomics cluster

Table 4.2 Publicly accessible databases that offer patient-diverse omics data sets (Martorell-Marugán et al. 2021)

Repository	Disease targeted	Types of omics datasets available
The Cancer Genome Atlas (TCGA)	Cancer	RNA-Seq, DNA-Seq, miRNA-Seq, SNV, CNV, DNA methylation, and RPPA
Clinical Proteomic Tumor Analysis Consortium (CPTAC)	Cancer	Proteomics data corresponding to TCGA cohorts
International Cancer Genomics Consortium (ICGC)	Cancer	Whole-genome sequencing, genomic variations data
Cancer Cell Line Encyclopedia (CCLE)	Cancer cell line	Gene expression, copy number, and sequencing data
Omics Discovery Index	Consolidated data sets	Genomics, transcriptomics, proteomics, and metabolomics

4.4 Strategies Toward Multi-Omics Studies

4.4.1 Design of Omics Studies

Researchers' understanding of the information dissemination, from the fundamental cause of disease (genomic, ecological, or evolutionary) to the operational ramifications or relevant linkages, can be improved by multi-omics. The basic experimental conditions that should be taken into account while designing an omics investigation are covered in this section (Altmäe et al. 2014).

4.4.1.1 Complexity of Diseases

The type of the illness is a crucial factor in the design of a multi-omic investigation. Few etiological variables are involved in the formation of simple diseases, which result from single-gene mutations; nevertheless, "augmentation genes" or outside elements can affect the severity or development of many illnesses (Hasin et al. 2017). For instance, a single chloride channel mutation is the most common factor that causes cystic fibrosis, allowing research on the illness to concentrate on one gene's functionality. It is therefore expected that concentrated omics initiatives at certain durations, focusing on the instant molecular modifications brought on by the causal factor, will yield adequate information to increase knowledge of potential treatment methods (Hasin et al. 2017).

4.4.1.2 Power, Sample Sizes, and Subsequent Analysis

Omics techniques generate results to provide molecular insight based on mathematical interpretation from frequently large databases. As a result, effect size, background noise heterogeneity, and sample size all have a significant impact on the

ability to discover relationships or the flow of information, with the latter factor frequently being the only one that can be controlled by researchers (Hasin et al. 2017). Because of this, the effectiveness of omics techniques to shed light on human illnesses is heavily reliant on the sizes of the samples that are available. An imbalanced research is frequently more prone to produce false-positive outcomes in addition to being a stab in the dark that overlooks significant indications (Hasin et al. 2017).

4.4.1.3 Human Study and Animal Model of Disease

Omics research on both people and animal models offers crucial information about disease. Since humans are the primary target population for medical research, results from human studies have a higher potential for translation than those from studies using animal models. Many human-centric consortia, such as the Roadmap Epigenomics Project, have generated a sizable volume of epigenomics and transcriptomics data in many of the tissues. Even while it provides useful information, human omics research has a number of limitations that, if the appropriate animal model of the illness is used, can only be addressed in animal experiments. Primary human cell lines have been broadly applied to analyze specific particular mechanistic approaches, supporting the claim that they are a good platform for disease exploration without the usage of animal models. However, their usage is constrained by the complexity and convergence of several cell types that underlie the majority of complicated illnesses. Reproducibility, environmental factor control, employing animal models have advantages such as availability of relevant tissues, precise phenotyping, availability of almost limitless numbers of identical biological duplicates, and the opportunity to experimentally verify theories (Hasin et al. 2017).

4.4.2 Analysis and Network Methods

Network techniques have been created and successfully used for any form of data or analytical technique. Table 4.3 lists the tools in alphabetical order by type (Yan et al. 2018). However, most research to date focuses on a single omics layer and ignores the linkages across other omics levels. Despite the fact that route and network studies incorporate a variety of omics data types, their conclusions still predominantly depend on a single omics layer without truly merging them. But even with big discoveries that can be replicated, science cannot advance fully as none of the separate components can give sufficient context or information to provide a full analysis of a biological system, without merging the multi-omics data into a paradigm. Another exciting area of integrated omics is the use of computational approaches to understand the molecular relationships between various omic layers (Picard et al. 2021).

Table 4.3 Tools used for any forms of data or analytical techniques

Epistasis	Features
PLINK	Regression-based
InterSNP	Regression-based
Parallelized PLINK (FastEpistasis)	Regression-based
BOOST	Regression-based; high-performance tool
SNPHarvester	Regression-based; high-performance tool
SIXPAC	Contrast test-based
EPIQ	Contrast test-based
MDR	Data mining-based
BEAM	Bayesian-based

Despite these advancements, the majority of studies still employ simple network methodologies and computational approaches in integrated omics are currently significantly explored less. Although simple network techniques are straightforward to use, care should be given because they have limitations that could generate bias, especially as the number of markers increase. Several leading organizations have been creating increasingly advanced computer approaches for integrative assessment in these systems since it is substantially easier to gather multi-omics data from straightforward model organisms (Yan et al. 2018).

To extract the most important information from the multi-omics data sets, the need for integrated research tools is equally critical in addition to network approaches. One of the early methods for integrative analysis was PARADIGM that stands for Pathway Recognition Algorithm using Data Integration on Genomic Model (Dellinger et al. 2014). It can calculate the degree of pathway activation for each sample by merging gene expression data with copy number, which accounts for the use of a stochastic graphical for varied interactions across paths.

4.4.3 Leveraging Multi-Omics Data for Actionable Insights

To carry out intricate biological operations, genes, transcripts, proteins, metabolites, and other macro/micro molecules work in concert. Numerous studies have demonstrated how the integration of multi-omics data sets can aid in the discovery of the underlying mechanisms at various omics levels. Here, we go into great detail about the tools and approaches that enable the fusion of multi-omics data sets to address the different issues surrounding disease and its mechanisms. The tools are arranged according to how well they can answer relevant biological questions (Subramanian et al. 2020).

Each case study's tools or methods can be broadly categorized into one or more namely: networks, Bayesian, fusions, similarity-based, correlation-based, and other multivariate methods. The integrative tools and methodologies are schematically depicted in Fig. 4.3 and are organized into groups based on the approaches used

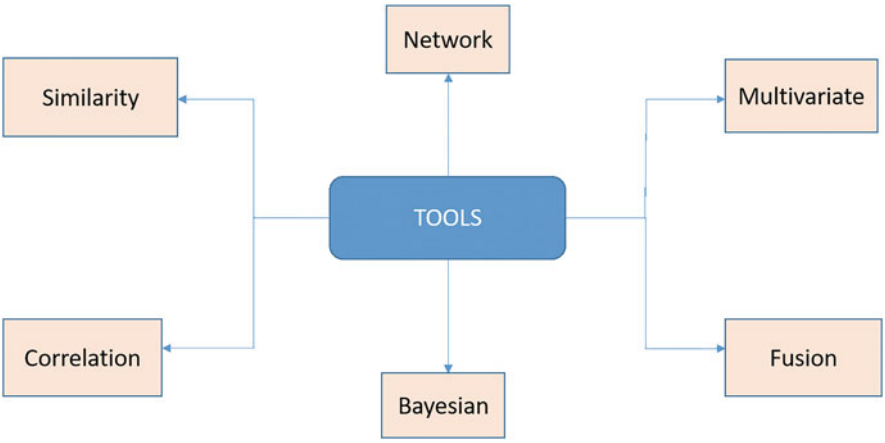


Fig. 4.3 The tools and approaches that enable the fusion of multi-omics data sets

(Edmondson et al. 2014). The combination of these approach types is used by a select few tools, such as PARADIGM, similarity network fusion (SNF), and others, as shown in Fig. 4.3. Each case study's tools and techniques are organized into approach categories (Subramanian et al. 2020).

The data integration technologies must deal with the unavoidable existence of missing values in multi-omics data (Voillet et al. 2016). A small number of the tools discussed in this article can handle missing data using imputation techniques, but others need missing values to be handled or removed during preprocessing processes (Subramanian et al. 2020).

4.5 Multi-Omics Approaches in Drug Discovery

4.5.1 Target Identification

In pharmaceutical research, the identification of new targets has been a key application of omics technologies. Early on in the postgenomic age, differentially regulated genes received a lot of attention with the goal of discovering new targets (Yan et al. 2015). This focus, however, failed to take into consideration the relatively poor correlation between gene and protein expression as well as the fact that many promising therapeutic targets do not exhibit differential expression. For the purpose of identifying critical nodes controlling significant disease pathways based on network topology, more recent methodologies aggregate gene expression data together with other information into networks (Hasin et al. 2017).

A large-scale protein was assembled and DNA interaction network incorporating gene expression data, expression quantitative trait loci (eQTL) analysis, and molecular interaction information were employed to identify potentially causal genes and

dysregulated pathways in an effort to discover novel targets for glioblastoma (van Dam et al. 2018).

To comprehend targets and disease pathways from a more comprehensive angle, a different strategy was adopted. In the Gene Expression Omnibus (GEO) database, it was discovered that illness-regulated genes by are associated with disease concepts and integrating this information with protein-protein interaction data. Known drug target genes were preferentially located in modules that were dysregulated in numerous diseases after functional modules, pathways, and complexes were assessed using disease-specific transcription data sets (Hasin et al. 2017).

Despite these excellent efforts to identify new targets, one of the challenges confronting researchers studying drug development is that the majority of the targets they identify lack easily druggable active regions that can be inhibited by small molecule medicines. They may function as transcription factors, structural elements of important cellular complexes, or have unknown functions. This circumstance has rekindled interest in more straightforward drug development methods that involve screening chemical libraries using phenotypic tests.

Systems biology methods, which were created to better understand cell signaling and pathway mechanisms, can now be used to help with the deconvolution of complex modes of action, in contrast to the pregenomic age (Chen and Thorner 2005).

4.5.2 Mechanism of Action and Cell Systems Biology

Characterizing the mechanism of action is important for many pharmaceutical development decisions. Typically, the target through which a medicine exerts a pharmacologic effect is referred to as the mechanism of action (Mast et al. 2014).

There are numerous instances where systems biology and omics methods have proven useful for identifying mechanisms of action. A compilation of gene expression profiles derived from a panel of yeast mutants was used in one of the earliest large-scale transcriptomics studies in yeast—to discover erg2, a yeast homolog of the sigma receptor, as a possible target of the topical anesthetic dyclonine. This study demonstrated the ability of transcriptomics analysis to produce testable hypotheses, despite the fact that it is doubtful that most pharmacological targets have yeast homologs (Mast et al. 2014).

A particularly fascinating field of research is the analysis of drug combinations for both research and discovery unique combinations and new methods. There was also a demonstration of the coupled gene expression research and whole-genome methylation profiling to look into the drug combination's mechanism of action of the histone deacetylase and the multitarget flavonoid genistein Vorinostat, an inhibitor of HDAC.

Cell systems biology is a phenotypic drug discovery strategy that blends combinatorial design and the depth of human disease biology into the development of assays. Primary human cell-based assays known as BioMAP1 systems are created in

an efficient in vitro manner to represent complex human disease. Combinations of pathway activators are used to stimulate primary human cell types and cocultures in order to produce cell signaling networks that are more pertinent to human disease. By assessing the quantities of secreted and cell surface proteins and mediators, these cell culture systems are examined. A database that can be searched to find for functional similarities has different changes in protein readout levels caused by medication actions. In this method, it is discovered that inhibitors or activators of particular targets change the levels of numerous endpoints reproducibly, frequently in a predetermined pattern, allowing the resulting signatures to be connected to particular mechanisms of action (Butcher et al. 2004).

4.5.3 Phenotypic Drug Discovery

Before the 1990s, the majority of drug discovery research was based on phenotypic drug discovery, which, in the absence of a clearly identified biological target, finds drugs using animal and cellular models of illness. But for a variety of reasons, targeted medication development has taken center stage in contemporary pharmaceutical research with the human genome project's culmination. To investigate and comprehend the target function and pharmacological mechanisms of action at the molecular level, gene knockouts and transgenic systems can be used (Yan et al. 2018).

Despite the fact that target-based drug discovery has been successful in many fields, including biologics, targeted chemotherapeutics, and second-generation medications, it is debatable whether it has yet to result in the expected rise in the number of new medications entering the market. In fact, attrition of compounds in late clinical development has actually risen in recent years.

In addition to target-based initiatives, these worries have rekindled interest in phenotypic drug discovery. In fact, growing investment in phenotypic drug discovery initiatives is being driven by systems biology advancements to better understand disease pathways, new tools for the deconvolution of target pathways and processes, and past achievements (Yan et al. 2018).

4.6 Application of Multi-Omics Technologies in Tubercular Drug Discovery

Multi-omics approaches are critical in the hunt for novel antituberculosis therapeutics. To begin, in a discovery biology strategy, novel targets in druggable pathways are identified for target-based inquiry, progressing from target to lead molecule. Second, in a discovery chemistry method, determine the mode of action of lead compounds produced from high-throughput screens as they advance from molecule

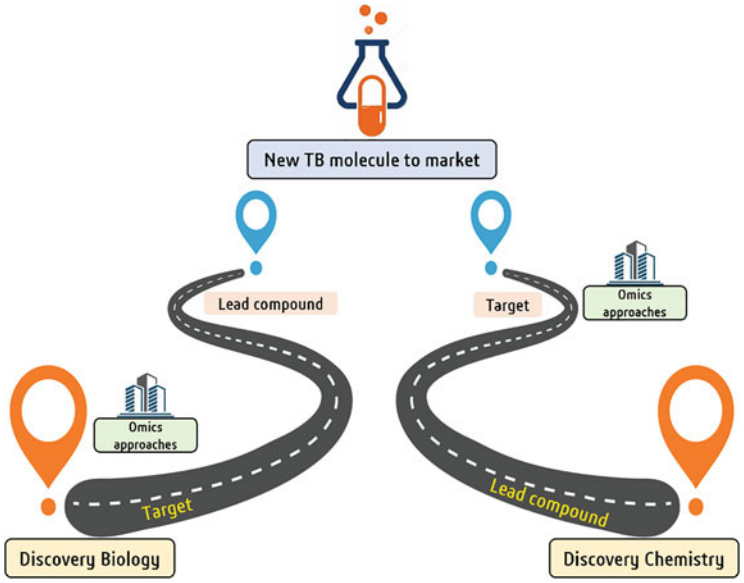


Fig. 4.4 Pathway for the use of omics technologies in the discovery of novel drugs for TB

to target. The benefit of multi-omics techniques in both of these contexts is that they are unregulated and unbiased to a preconceived preconception, making them helpful tools for confirming therapeutic action, revealing new insights into compound activity, and discovering new lines of investigation as shown in Fig. 4.4.

4.6.1 Genomics

4.6.1.1 Target Identification

The *Mycobacterium tuberculosis* (Mtb) H37Rv genome sequence represents the starting point for antitubercular drug discovery. As a result, the application of genomes to Mtb has offered a framework of prospective therapeutic targets. Manipulation of gene function via gene inactivation techniques has contributed in the identification of pathways critical to Mtb in various microenvironments, identifying potential targets for drug discovery programs. The application of genomes to Mtb has offered a framework of prospective therapeutic targets. Manipulation of gene function via gene inactivation techniques has contributed in the identification of pathways critical to Mtb in various microenvironments, identifying potential for drug discovery programs (Hasin et al. 2017).

Methods for generating unmarked single-gene knockout mutants have been used alongside global procedures that use transposons (Tn) to inactivate gene function. Mtb's capacity to survive and multiply in phagocytes, evading phagosome-lysosome

fusion and adapting to an intracellular lifestyle, is a critical pathogenic characteristic (Sassetti et al. 2003). To establish intramacrophage survival pathways, Barczak et al. used high content imaging and multiplexed cytokine analyses on macrophages infected with Mtb Tn mutant libraries to map genes required for intracellular growth (Barczak et al. 2017).

CRISPRi (Clustered Regularly Interspaced Short Palindromic Repeats Interference) has the potential to transform the field by allowing precise gene silencing to identify and validate therapeutic targets (Rock et al. 2017). A single-guide RNA directs the activity of a nuclease, dCas9, which has two mutations that eliminate its nuclease activity (sgRNA). When the dCas9-sgRNA combination binds to the target site, the DNA duplex destabilizes and inhibits gene transcription by preventing RNA polymerase promoter access. The gene silencing of folate metabolism revealed the potential of CRISPRi in target-based drug development (Tsai et al. 2015; Haeussler et al. 2016).

4.6.1.2 Mode of Action

Andries et al. used whole-genome sequencing (WGS) of M. smegmatis and Mtb bedaquiline-resistant mutants cultivated in vitro to show that this diarylquinoline targets the product of atpE, a subunit of the mycobacterial ATP synthase anchored in the mycobacterial membrane, in the discovery of bedaquiline (Andries et al. 2005). Kundu et al. refined this mechanism of action by demonstrating that bedaquiline binds to the epsilon subunit (Kundu et al. 2016). In addition, WGS has been employed to determine the mechanism of action of repurposed licensed medicines in Mtb. Rybniker et al. discovered lansoprazole from the Prestwick library of 1280 FDA-approved medications to be protective to lung fibroblasts in an M tb intracellular infection model. Through the large-scale sequencing of clinical isolates, WGS is also changing our understanding of pharmacological action and treatment resistance (Rybniker et al. 2019). Drug-resistance conferring mutation mapping consortiums have discovered novel mechanisms of resistance and probable novel modes of action of existing anti-Mtb medicines in patients. Melief et al. created a library of Mtb strains overexpressing single genes that could be examined in a high-throughput format to increase the approaches available for understanding drug mechanism of action. Each gene in the library was cloned downstream of a tetracycline-inducible promoter. The drug's target Alr was overexpressed, resulting in a sevenfold rise in the lowest inhibitory concentration of D-cycloserine (Melief et al. 2018).

4.6.2 Transcriptomics

4.6.2.1 Target Identification

In order to identify druggable pathways and bacterial responses to drug exposure, it is necessary to have an understanding of the mycobacterial transcriptome. In Mtb

reproducing intracellularly in macrophages and in expectorated Mtb in patient sputa, transcriptomics in a discovery biology context has revealed the stimulation of potentially druggable pathways involved in oxidation of fatty acids, the glyoxylate shunt, and cholesterol metabolism (Schnappinger et al. 2003; Rienksma et al. 2015). Cultivating the possibility of cidal therapeutic action by targeting pathways active in vivo, RNA profiles from animal models of TB infection and human tissue provide valuable information on the expression of targets in human disease (Rachman et al. 2006). Given that the bactericidal or bacteriostatic suppression of a crucial target in vitro does not always predict in vivo therapeutic efficacy, this is useful evidence for drug discovery decision making. Integrating data on gene essentiality with transcriptome analysis provides a multi-omics approach to determining which pathways should be studied in depth first.

4.6.2.2 Mode of Action

Since the mechanism of Mtb death is unknown, this unsupervised method is very helpful for interpreting the results of high-throughput screens' lead compounds. Drugs now in phase I/II research or the clinic have had transcriptomics applied to it to determine their mechanism of action. Drug mode of action-specific reporters were chosen by Boot et al. by analyzing RNA-seq data for responses of Mtb and Mycobacterium marinum to subinhibitory doses of ciprofloxacin, ethambutol, isoniazid, streptomycin, and rifampicin (Boot et al. 2018). Screening a library of 196 antimycobacterial compounds with the MMAR 4645-ciprofloxacin reporter and iniBAC-isoniazid reporter provided proof of concept that these drug reporters could hasten TB drug discovery by identifying the mode of action of hit compounds. The screening identified one molecule with a mode of action similar to ciprofloxacin, which could block DNA replication, and two others with a mode of action similar to isoniazid, both of which presumably target the mycobacterial cell wall.

4.6.3 Proteomics

The proteomics field has not yet expanded to the depths of genomes and transcriptomics (Bespyatykh et al. 2017).

4.6.3.1 Target Identification

Proteomics offers a unique and valuable insight of Mtb physiological responses and target expression, as seen by the discrepancy between protein abundance and related mRNA abundance. Several reports have confirmed the expression of proteins that could be druggable targets in Mtb, providing new perspectives on the intracellular and in vivo expression of this pathogen (Vogel and Marcotte 2012). There are two

main types of proteomics approaches; the first is a top-down technique, in which proteins are first separated from a biological sample and then sorted according to their physical and chemical properties using gel electrophoresis, before being identified using mass spectrometry (MS). This mapped the expression of efflux systems that might affect drug efficacy and brought to light the expression of potentially druggable proteins linked with transmembrane transport, supplementing genomics and transcriptomics techniques to validate existence of target protein (Bespyatykh et al. 2017). The second approach is the bottom-up method that involves proteolytic cleavage of a complete set of proteins into peptides, then high-performance liquid chromatography, and tandem mass spectrometry (LC-MS/MS) analysis. Although computational resolution of profiles and insufficient sensitivity currently limit proteomics findings, this technique has the potential to measure many more proteins and follow a chosen group of proteins to better sensitivity (Bespyatykh et al. 2017). The field of proteomics has been used in in vitro infection models to identify novel protein targets for medication development.

4.6.3.2 Mode of Action

Mtb reactions to drug exposure have been deconvoluted using proteomics, shedding light on the mechanisms of drug action and resistance (Sharma et al. 2018). Proteomics was recently used by Meneguello et al. to investigate the metabolic pathways involved in rifampicin's action. Proteomics was used by Sarkar et al. to chart Mycobacterium tuberculosis's response to sulfamethoxazole, and they found the drug-induced oxidative stress and electron transport chain pathways (Sarkar et al. 2018). Proteomics has the potential to be used in the analysis of medication resistance. To determine which proteins are secreted by Mtb strains that are resistant to isoniazid and rifampicin, Putim et al. used a shotgun proteomics approach. Proteins in bacterial culture filtrates were collected using low-binding protein-cellulose acetate membranes, separated using SDS-PAGE, digested in gel, and analyzed using liquid chromatography mass spectrometry (Putim et al. 2018).

4.6.4 Metabolomics

4.6.4.1 Target Identification

The examination of the metabolite network inside a biological system is known as metabolomics, and it is a crucial omics tool for drug development since it reveals which cellular processes could be targeted by therapeutics. Through the use of metabolomics in a discovery biology context, the metabolic pathways utilized by Mtb in various niches have been identified. Similarly, Serafini et al. employed this method to determine how Mtb assimilates pyruvate and lactate. The authors showed a novel function for the methylcitrate cycle by highlighting that it could be reversed

for the biosynthesis of propionyl-CoA and the metabolism of pyruvate and lactate, which identifies new targets for drug discovery efforts despite the well-established fact that lipids are important carbon sources for Mtb during infection (Serafini et al. 2019). This study is exemplary of a multi-omics strategy because it utilized transposon-directed insertion site sequencing in conjunction with RNAseq transcriptomics, proteomics, and metabolomics to provide a comprehensive functional picture of the carbon metabolic network in Mtb. Amino acids are used by Mtb as a nitrogen source, and Agapova et al. combined stable isotope tracing of labeled amino acids with mass spectrometry to learn more about this. The scientists demonstrated that utilizing several different amino acids as nitrogen sources did not result in enhanced development over using only one. The possibility for targeting specific pathways within the Mtb nitrogen metabolic network was elucidated by metabolomics (Agapova et al. 2019).

4.6.4.2 Mode of Action

Studies elucidating the mechanism of action of new drugs show the value of metabolomics. The high-throughput metabolomics method used by Zampieri et al. involved analyzing the mass spectra of supernatants from Mtb cultures of *M. smegmatis* to characterize a collection of 212 antimycobacterial chemicals with unknown mode of action (Zampieri et al. 2018). Before evaluating the similarities between the reference and test compound profiles, the metabolomic signatures were created for 62 reference compounds with 17 identified targets. More than 70% of the 212 compounds had a clear mechanism of action, and only 16% had metabolomic profiles that were significantly different from the reference compounds. Six of these 16 substances showed a metabolomic response indicating they inhibited lipid and trehalose metabolism in a similar fashion. This method identified previously undetected druggable pathways in Mtb and, more crucially, enables drug development programs to diversify target pathways, eliminating compounds that are expected to block targets of existing medicines. The *M. smegmatis* cultures treated with ampicillin, ethambutol, ethionamide, isoniazid, kanamycin, linezolid, rifampicin, and streptomycin had a different metabolite profile than those treated with pretomanid. Pathway analysis of differentially abundant metabolites revealed a potential role for the buildup of the toxic metabolite methylglyoxal in pretomanid's antibacterial action (Baptista et al. 2018; Wang et al. 2019).

4.6.5 Lipidomics

The exploration of this networking of cellular lipids within a biological system is widely classified as lipidomics, a branch of metabolomics that investigates lipid species present and how they interact with other lipids, metabolites, and proteins in a cell.

4.6.5.1 Target Identification

Lipidomics is based on mass spectroscopy, which measures the mass-to-charge ratio and abundance of gas-phase ions. It is further classified as gas chromatography (GC)-MS, liquid chromatography (LC)-MS, and direct infusion-MS (Wu et al. 2014). Lipidomics has been used to identify potentially druggable lipid production pathways based on Mtb's sensitivity to environmental changes. Raghunandanan et al. studied the pattern of Mtb lipid alterations during hypoxia-induced dormancy and resuscitation, discovering that lipid concentration dropped dramatically during dormancy and progressively increased after reaeration (Raghunandanan et al. 2019). Lipidomics is also a useful approach for identifying lipid biosynthesis pathway targets. The fatty acid synthase FAS-II multienzyme system is required for mycolic acid production and proper cell wall function in mycobacteria (Lefebvre et al. 2018). Because it is the target of various antimycobacterial agents, including isoniazid, it has further therapeutic promise.

4.6.5.2 Mode of Action

By mapping the impact of the natural antimycobacterial component vanillin in M. smegmatis, Sharma et al. proved the relevance of lipidomics in mechanism of action research. Vanillin, the authors discovered, alters the makeup of fatty acids, glycolipids, glycerophospholipids, and saccharolipids, disrupting cell membrane homeostasis (Sharma et al. 2020). Howard et al. found that rifampicin-resistant Mtb isolates with rpoB mutations had different lipid profiles depending on where the single-nucleotide polymorphisms (SNPs) were located (Howard et al. 2018). Lipidomics, coupled with genomes and transcriptomics, were used to determine the mode of action of HC2091, a new drug that appears to target MmpL3. SNPs in the mmpL3 gene were found in Mtb HC2091-resistant mutants, potentially conferring drug resistance (Li et al. 2016). As a result, lipidomics, which is frequently used in conjunction with other omics techniques, is an excellent tool, particularly in the investigation of cell wall biosynthesis pathways, which are a rich source of druggable Mtb targets.

4.7 Conclusion and Future Perspectives

The expanding application of omics technologies and the incorporation of these tools in the drug development process have brought to light a number of important concerns that can direct future study. Some of these are technological, such as the well-known computational difficulties brought on by the enormous datasets with numerous features but few samples. Missing data are a concern as well, and

published literature offers more details in well-known study fields where systemic bias is introduced (Paananen and Fortino 2020).

Another crucial challenge is comprehending and controlling the data's fluctuation. Inaccurate models can be produced by using inconsistent approaches and paying insufficient attention to experimental sources of variability. One significant drawback of network models is their tendency to disguise the ambiguities in the underlying data, which can lead to an illusion of comprehension. Both biological and technical factors can cause variation. Cost considerations frequently prevent experimental repeats in large omics collections. Media components, cell passage number, and other culture variables can result in contradictory data sets with cell culture samples. Results for patient samples can be impacted by experimental variances, such as sample gathering methods, storage time, and temperature, in addition to usual variables like genetics, gender, age, stage of disease, and treatment history.

The need for stronger terminology and more practical ontologies has been brought to light by efforts to connect omics data sets with other datasets. There has been some progress thanks to initiatives like the National Institutes of Health's BioAssay Research Database (BARD) project and OpenBEL, a computational but human readable, semantically rich language for representing and describing causal relations between biological and scientific findings.

Future research will be conducted in a number of areas, including better methods and procedures for data integration as well as the creation and sharing of networks. There are still new technological advancements happening. These include cutting-edge proteomic techniques that provide maps of proteins within the cell that are spatially and temporally detailed. It will also be necessary to develop new computational techniques for assessing and integrating multiscale data. Longer timescale approaches to integrate computational disease models with cell signaling networks will be of special relevance (Butcher et al. 2004).

References

Agapova A, Serafini A, Petridis M et al (2019) Flexible nitrogen utilisation by the metabolic generalist pathogen Mycobacterium tuberculosis. elife 8:e41129

Altmäe S, Esteban FJ, Stavreus-Evers A et al (2014) Guidelines for the design, analysis and interpretation of 'omics' data: focus on human endometrium. Hum Reprod Update 20:12–28

Andries K, Verhasselt P, Guillemont J et al (2005) A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. Science 307:223–227

Baptista R, Fazakerley DM, Beckmann M et al (2018) Untargeted metabolomics reveals a new mode of action of pretomanid (PA-824). Sci Rep 8:1–7

Barczak AK, Avraham R, Singh S et al (2017) Systematic, multiparametric analysis of Mycobacterium tuberculosis intracellular infection offers insight into coordinated virulence. PLoS Pathog 13:e1006363

Bespyatykh JA, Shitikov EA, Ilina EN (2017) Proteomics for the investigation of mycobacteria. Acta Naturae 9:15–25

Boot M, Commandeur S, Subudhi AK et al (2018) Accelerating early antituberculosis drug discovery by creating mycobacterial indicator strains that predict mode of action. Antimicrob Agents Chemother 62:e00083–e00018

Butcher EC, Berg EL, Kunkel EJ (2004) Systems biology in drug discovery. Nat Biotechnol 22: 1253–1259

Chen RE, Thorner J (2005) Systems biology approaches in cell signaling research. Genome Biol 6: 1–5

Clish CB (2015) Metabolomics: an emerging but powerful tool for precision medicine. Mol Case Studies 1:a000588

van Dam S, Vosa U, van der Graaf A et al (2018) Gene co-expression analysis for functional classification and gene–disease predictions. Brief Bioinform 19:575–592

Dellinger AE, Nixon AB, Pang H (2014) Integrative pathway analysis using graph-based learning with applications to TCGA colon and ovarian data: supplement issue: array platform modeling and analysis (A). Cancer Inform 13:1

Edmondson R, Broglie JJ, Adcock AF, Yang L (2014) Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. Assay Drug Dev Technol 12: 207–218

Eicher T, Kinnebrew G, Patt A et al (2020) Metabolomics and multi-omics integration: a survey of computational methods and resources. Metabolites 10:202

Haeussler M, Schönig K, Eckert H et al (2016) Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol 17:1–12

Hasin Y, Seldin M, Lusis A (2017) Multi-omics approaches to disease. Genome Biol 18:1–15

Howard NC, Marin ND, Ahmed M et al (2018) Mycobacterium tuberculosis carrying a rifampicin drug resistance mutation reprograms macrophage metabolism through cell wall lipid changes. Nat Microbiol 3:1099–1108

Kukurba KR, Montgomery SB (2015) RNA sequencing and analysis. Cold Spring Harb Protoc 2015:951

Kundu S, Biukovic G, Grüber G, Dick T (2016) Bedaquiline targets the ε subunit of mycobacterial F-ATP synthase. Antimicrob Agents Chemother 60:6977–6979

Lefebvre C, Boulon R, Ducoux M et al (2018) HadD, a novel fatty acid synthase type II protein, is essential for alpha-and epoxy-mycolic acid biosynthesis and mycobacterial fitness. Sci Rep 8:1–15

Li W, Obregón-Henao A, Wallach JB et al (2016) Therapeutic potential of the Mycobacterium tuberculosis mycolic acid transporter, MmpL3. Antimicrob Agents Chemother 60:5198–5207

Manzoni C, Kia DA, Vandrovcova J et al (2018) Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief Bioinform 19:286–302

Martorell-Marugán J, López-Domínguez R, García-Moreno A et al (2021) A comprehensive database for integrated analysis of omics data in autoimmune diseases. BMC Bioinform 22:1–14

Mast FD, Ratushny AV, Aitchison JD (2014) Systems cell biology. J Cell Biol 206:695–706

Melief E, Kokoczka R, Files M et al (2018) Construction of an overexpression library for Mycobacterium tuberculosis. Biol Methods Protoc 3:bpy009

Paananen J, Fortino V (2020) An omics perspective on drug target discovery platforms. Brief Bioinform 21:1937–1953

Picard M, Scott-Boyer M-P, Bodein A et al (2021) Integration strategies of multi-omics data for machine learning analysis. Comput Struct Biotechnol J 19:3735–3746

Putim C, Phaonakrop N, Jaresitthikunchai J et al (2018) Secretome profile analysis of multidrug-resistant, monodrug-resistant and drug-susceptible Mycobacterium tuberculosis. Arch Microbiol 200:299–309

Rachman H, Strong M, Ulrichs T et al (2006) Unique transcriptome signature of Mycobacterium tuberculosis in pulmonary tuberculosis. Infect Immun 74:1233–1242

Raghunandanan S, Jose L, Gopinath V, Kumar RA (2019) Comparative label-free lipidomic analysis of Mycobacterium tuberculosis during dormancy and reactivation. Sci Rep 9:1–12

Rienksma RA, Suarez-Diez M, Mollenkopf H-J et al (2015) Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. BMC Genomics 16:1–15

Rock JM, Hopkins FF, Chavez A et al (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. Nat Microbiol 2:1–9

Rybniker J, Kohl TA, Barilar I, Niemann S (2019) No evidence for acquired mutations associated with cytochrome bc 1 inhibitor resistance in 13,559 clinical mycobacterium tuberculosis complex isolates. Antimicrob Agents Chemother 63:e01317–e01318

Sarkar R, Mdladla C, Macingwana L et al (2018) Proteomic analysis reveals that sulfamethoxazole induces oxidative stress in M. tuberculosis. Tuberculosis 111:78–85

Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 48:77–84

Schnappinger D, Ehrt S, Voskuil MI et al (2003) Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: insights into the phagosomal environment. J Exp Med 198: 693–704

Serafini A, Tan L, Horswell S et al (2019) Mycobacterium tuberculosis requires glyoxylate shunt and reverse methylcitrate cycle for lactate and pyruvate metabolism. Mol Microbiol 112:1284–1307

Sharma D, Bisht D, Khan AU (2018) Potential alternative strategy against drug-resistant tuberculosis: a proteomics prospect. Proteomes 6:26

Sharma S, Hameed S, Fatima Z (2020) Lipidomic insights to understand membrane dynamics in response to vanillin in Mycobacterium smegmatis. Int Microbiol 23:263–276

Singh B, Crippen TL, Tomberlin JK (2017) An introduction to metagenomic data generation, analysis, visualization, and interpretation. In: Forensic Microbiology. Wiley, Hoboken, p 94

Subramanian I, Verma S, Kumar S et al (2020) Multi-omics data integration, interpretation, and its application. Bioinform Biol Insights 14:1177932219899051

Tsai SQ, Zheng Z, Nguyen NT et al (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol 33:187–197

Tuli L, Ressom HW (2009) LC–MS based detection of differential protein expression. J Proteomics Bioinform 2:416

Vogel C, Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 13:227–232

Voillet V, Besse P, Liaubet L et al (2016) Handling missing rows in multi-omics data integration: multiple imputation in multiple factor analysis framework. BMC Bioinform 17:1–16

Wang Z, Soni V, Marriner G et al (2019) Mode-of-action profiling reveals glutamine synthetase as a collateral metabolic vulnerability of M. tuberculosis to bedaquiline. Proc Natl Acad Sci 116: 19646–19651

Weinhold B (2006) Epigenetics: the science of change. Environ Health Perspect 114(3):A160–A167

Wu Z, Shon JC, Liu K-H (2014) Mass spectrometry-based lipidomics and its application to biomedical research. J Lifestyle Med 4:17

Yan J, Risacher SL, Shen L, Saykin AJ (2018) Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. Brief Bioinform 19:1370–1381

Yan S-K, Liu R-H, Jin H-Z et al (2015) "Omics" in pharmaceutical research: overview, applications, challenges, and future perspectives. Chin J Nat Med 13:3–21

Zampieri M, Szappanos B, Buchieri MV et al (2018) High-throughput metabolomic analysis predicts mode of action of uncharacterized antimicrobial compounds. Sci Transl Med 10: eaal3973

Chapter 5
Computational Methods in Natural Products-Based Drug Discovery

Pankaj Dagur, Shreya, Rahul Ghosh, Gaurav Rakshit, Abanish Biswas, and Manik Ghosh

5.1 Introduction

Drug discovery is an extensive, costing big-budget, time-consuming process with the low rate of success. The development of a drug from scratch to market value, maintaining its efficacy, takes around 13–15 years and costs billions of dollars on average and still counting. In comparison to that, the rate of the launching novel drugs in the market is less. It is estimated that about more than half of all the drugs approved in the last three decades were either NPs (Natural products) or their semisynthetic derivatives (Newman and Cragg 2016) (Patridge et al. 2016).

The reason is their diversity in species and utilization for medicinal purposes since ancient civilizations. NPs possess comparatively greater molecular mass and a number of sp^3 carbon atoms, H-bond acceptors and donors, more hydrophilic nature, and molecular rigidness than that of nonnatural compounds' libraries (Atanasov et al. 2015) (Feher and Schmidt 2003). The structural upper hand can be advantageous while tackling protein-protein interaction owing to the greater rigidity of NPs (Lawson et al. 2017). Despite not adhering to Lipinski's rule of five, NPs are still a class that is used for therapeutic purposes, owing to their high molecular mass.

Natural products, despite being an inspirational source for NP-based drug discovery, pose disadvantages for the pipeline. NPs have diverse and complex molecular structures which means a challenge for generating 3D molecular structures and their analogs while considering stereochemistry, force fields, and algorithm for predicting protein-bound conformations (Friedrich et al. 2019). Dereplication tools are required to circumvent the rediscovery of known compounds. Other challenges

P. Dagur (✉) · Shreya · R. Ghosh · G. Rakshit · A. Biswas · M. Ghosh (✉)
Department of Pharmaceutical Sciences and Technology, Birla Institute of Technology, Ranchi, Jharkhand, India
e-mail: manik@bitmesra.ac.in

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_5

include procuring the materials, extraction, detection, and isolation of bioactive compounds and generating activity profiles are time-consuming and the success rate is less. Considering the facts, the prior prediction of activity using in silico methods can aid in simplifying the process.

The capital needed for in silico experiments is comparatively less than the expenses (for example- scikit-learn, CDK) associated with experimental procedures of which software licensing costs alone, continue to be a significant cost component and have been steadily rising in recent years. Moreover, on site efficient computing center is no longer necessary as calculations can be performed affordably in the cloud at very large scales, with a low degree of complexity. Computational-based drug discovery has well-established techniques equipped with cheminformatics for easing the process, reducing the loss and comparatively less time-consuming. These techniques involve data mining on large data, dereplication, chemical space analysis, visualization and comparison, prediction of bioactivity, ADME and safety profiles' natural products-inspired de novo design, and prediction of natural products prone to cause interference with biological assays (Chen and Kirchmair 2020).

5.2 Natural Products' Collections

The definition of "natural products" is not universally agreed upon, with some authors limiting the term to small molecule secondary metabolites while others broadly accept that chemical substance produced by a living organism as NP. The latter one holds more diversity and hence the line separating the subclasses remains ill-defined. The therapeutic class of NP as per the definition can be classified into phytochemicals, fungal metabolites, toxins, antibodies, and NPs with limited activity. The NPs collection can also be categorized as physical and virtual collections for in silico technology.

5.2.1 Physical Collection

The importance of NPs in ailment curing can be dated back to ancient civilizations. In earlier decades, natural compounds and their structural analogs have significantly added to the therapeutic arsenal for curing numerous diseases, including cancer and infectious disorders. According to a survey, only 6% of the estimated four lakh plant species have undergone activity studies, while less than 20% have undergone phytochemical investigations (Fabricant and Farnsworth 2001). Phytochemicals being antioxidants and a source for many life-saving medicines form a broad class of NPs including polyphenols, terpenoids, and alkaloids. The fungal metabolites have been explored for their use as antidiabetic, antibacterial, antioxidant, antitumor, and even insecticidal agents (Daley et al. 2017). In most cases, chemotherapy medications are made from naturally occurring poisons produced by large clades

of organisms, such as plants, fungi, and bacteria. The next important therapeutic class is antibiotics with more than 60% of drugs approved and more than 500 in the developmental stage as per the survey in 2016 (Cragg et al. 1997). The common mechanism of action includes receptor blocking or downregulation and induction of target cell signaling which can be exploited for rheumatoid arthritis, non-Hodgkin lymphoma, multiple myeloma, and various other diseases (Carter and Lazar 2018). NPs such as biopolymers, spider silk are known for their activity in drug delivery systems rather than therapeutic value.

For virtual screening of NPs for in silico studies, the majority of compound suppliers across the world now freely offer information related to the structures and some other features of the compounds. According to a survey, of the total known NP compounds in virtual databases, only about 10% of them are available for experimental procedures (Chen et al. 2017). This lack of availability of NPs physically serves as a blockage in the path of drug discovery. However, the readily available ones have favorable physicochemical properties for the drug discovery pipeline. Moreover, more than half of them have a molecular weight of less than 300 Da hence, providing many prospects for optimization (Chen et al. 2017). There are more than 100 commercial suppliers of purified NPs in the world, but only few of them supply more than 5000 NPs.

The fact that the (abovementioned) 25 k easily accessible NPs encompass more than 5700 Murcko scaffolds is noteworthy in this respect. Additionally, these NPs including alkaloids, steroids, and flavonoids, provide a fair representation of all of the major NP classes (Chen et al. 2018).

5.2.2 Virtual Collection

The rapidly growing attention of NPs has led to steep growth in NP-based databases. The virtual collection (or databases) of NPs can be categorized into (i) the generalized NP-based databases, (ii) databases of traditional NPs, and (iii) specialized databases (Chen et al. 2019a). The second category includes databases for traditionally used NP-based drugs whereas the third category includes databases focusing on some exclusive organisms belonging to a specific habitats, biological activities, or specific NP classes. A survey reported that since the 2000s, approximately 120 different databases and collections have been released and used in context with NPs (Sorokina and Steinbeck 2020). And of them, approximately 50 are open access, whereas 98 are still in some way accessible. These open -access databases include NP-based database collections published as supplementary material in scientific publications as well as those available in the ZINC database (Sterling and Irwin 2015). The collection of NPs on ZINC database provides information about their structure and their origin but no other additional information. The databases could be open access or commercially available. Amongst free NP databases is Super Natural II, consisting of more than 325 k NPs (Banerjee et al. 2015). A chemistry-aware online interface can be used to query the database, although the bulk download is not

Table 5.1 Examples of some active databases

Databases	Size
TCM database@Taiwan(Chen 2011)	>60 k
Natural Product Atlas (Van Santen et al. 2019)	>25 k
Collective Molecular Activities of Useful Plants (CMAUP) (Zeng et al. 2019)	47 k NPs
Marine Natural Library (Bugni et al. 2008)	14 k

officially supported. Universal Natural Products Database (UNPD) is another free database with more than 200 k NPs and downloadable resources (Gu et al. 2013). Unfortunately, UNPD database appears to be nonfunctional. These virtual databases are either specific to a particular geographical region (like databases only for Chinese herbals), or particular section of NPs (like database for only marine-based NPs), or could be generalized (COCONUT) (Sorokina et al. 2021). Some examples of functional databases are listed below in Table 5.1.

Some other examples include NuBBEDB (Pilon et al. 2017), KnapSack (Nakamura et al. 2013),CMAUP (Zeng et al. 2019), and smaller databases like FooDB. On the contrary, the data available on the therapeutic efficacy and protein-bound conformations of NPs suffer from scarcity. Amongst the most relevant ones, the Marine Natural Library has special mention, as it allows the download of the full dataset of more than 14 k marine NPs (Bugni et al. 2008). NPs seem to have a slight upper hand over synthetic compounds, as their "libraries" already exist in nature. The generalized databases of chemical compounds (Li et al. 2010; Leach 2017) (such as PubChem and ChEMBL) also include databases related to NPs that are annotated by their class, while, more specific ones (such as ArachnoServer, VenomKB, and the Dictionary of Marine Natural Products) provide even more granular annotations for aggregating NP libraries with various characteristics of interest (Dona et al. 2017; Romano et al. 2018).

5.3 Cheminformatics and Computational Approaches for NP-Based Drug Discovery

5.3.1 Computational-Based Approaches

Computer-based approaches being the broader term encloses within cheminformatics technology. Cheminformatics is the application of computational approaches to facilitate collection, storage, analysis of large databases addressing the major concern, drug discovery. Along with cheminformatics, other informatic approaches such as bioinformatics, semantic methods have also been reviewed (Romano and Tatonetti 2019). Computational techniques have long been regarded as an important part of drug development and discovery procedures. The various approaches it offers for drug discovery purpose are structural elucidation, analysis of the physicochemical and structural properties, in determining macromolecular

targets, prediction of ADME properties and safety profiles. Computational methods can be broadly classified into: structure-based and ligand-based for the abovementioned approaches (Podlogar et al. 2001). This classification is revolving around the level of structural information available in context with target to support the computational calculations. Structure-based methods operate on the availability of info regarding three-dimensional (3D) molecular target of interest, typically obtained from X-ray crystallography, nuclear magnetic resonance, or homology modeling (Cerqueira et al. 2015). Whereas ligand-based approaches focus on the availability of information in context with active ligands (and inactive compounds, when available) (Lill 2007). With the increasing need for prior virtual screening of NPs and maintaining of databases, cheminformatics has made its way through drug discovery process. The methods are generally classified as direct and indirect approaches, based on the type of properties they exploit. Direct approaches deal with chemical activity, their constants, reactive groups, ADME profiling, whereas indirect ones deal with structural specifications, compound category or other observations (Romano and Tatonetti 2019).

5.3.2 Cheminformatics and NP-Based Pipeline

So far, cheminformatics and other related informatics approaches have been reviewed in drug discovery pipeline. Cheminformatics and other approaches have played important part in curating NP-based fragmented databases and analyzing the result. Cheminformatics and computational approaches share an important linkage, basically cheminformatics is the application of computational approaches as shown in Fig. 5.1. Cheminformatics techniques exclusive to NP-based drug discovery are NP-based QSAR analysis, Molecular Docking and Dynamics, Computational Mutagenesis, and Library Construction. Numerous classes of NPs have been studied using QSAR, and the chemical descriptors used tend to be dictated by the particular classes (Huang et al. 2016). For example, small-molecule NPs include categorical variables suggesting their specific category of classification, species of origin. Similarly, in case of molecular docking, the specific classes of NPs decide the interaction of target and ligand. For example, if a macromolecular NP (belonging to specific class) is suspected of showing interactions with small-molecule metabolites, docking simulations can be used for mining which metabolites could bind to that NP (Pithayanukul et al. 2009). Other aspects of molecular docking include protein preparation and flexibility, pose scoring in context with binding affinity. The generation of extensive libraries of compounds and its screening aids in prediction of potential drug candidates along with awareness of encountering small fraction of hits (Terrett et al. 1995). In case of NPs, their databases exist in nature way before synthetics. In this chapter, we are going to discuss different analytical methods used in computational approaches for NPs. Antibodies, despite of their large molecular weight, are relatively easy to screen for large numbers via docking, indicating their specificity in structural and binding properties that eventually reduces computational

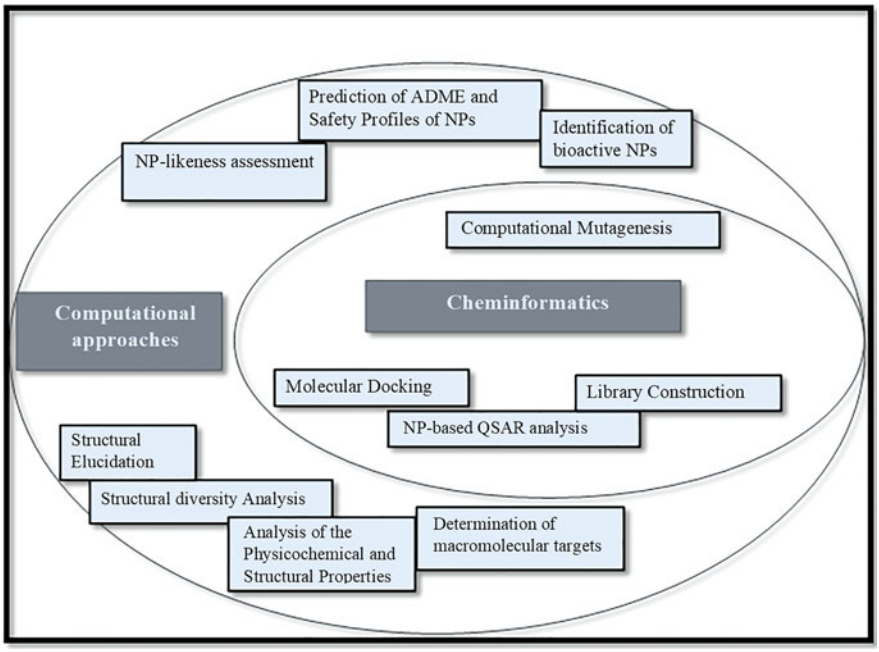


Fig. 5.1 Amalgamation of cheminformatics and computational approaches

complexity for simulations (Mann 2002). Additionally, noteworthy success stories have emerged from screening smaller NP-based databases against specific drug targets. For example, the compound ellagic acid, known to have both antiproliferative and antioxidants properties, was identified by Moro et al. by screening a proprietary database of 2000 NPs against the oncoprotein casein kinase 2 (Cozza et al. 2006).

5.4 Computational Approaches Related to Natural Products

5.4.1 Structural Elucidation

For the extraction and isolation of NPs, the source of materials area is going to be highly priced and long-time taking and when everyone gained knowledge about the NPs, the discovering of novel compounds is decreasing. Of order to make the most of the available experimental resources, it is necessary to integrate analytical and computational approaches for early detection of both favorable and negative features in NPs (Pereira and Aires-de-Sousa 2018). Databases that provide measurable analytical statistics (such as bioactivities, chromatographic data, MS, NMR

spectroscopy, and FTIR data) for known NPs and their interrogation using computational methods play a crucial role in this interaction of technologies. However, even the biggest of these databases only include a small subset of the NPs that are considered. This is why NMR and MS fragmentation predictions are increasingly being made using computational methods, often in tandem with structure generators (Pereira and Aires-de-Sousa 2018).

In recent days, for the virtual screening of natural product (NP) candidates in both small datasets of isolated chemicals and huge databases, structure-based (SB) and ligand-based (LB) cheminformatics techniques have become indispensable tools. Quantitative structure-activity relationships (QSAR), assessment of drug similarity, forecasting surface assimilation, distribution, metabolism, excretion prediction, similarity detection, and pharmacophore identification are the most often used LB approaches. Similar techniques used in SB methods include molecular dynamics, docking, and binding cavity analysis (Pereira and Aires-de-Sousa 2018).

The potential of re-isolating well-known molecules has recently, however, put a hold on the drug development process from natural products. The process of dereplication, which automates the quick identification of previously isolated compounds, directs researchers to fresh discoveries and cuts down on the time and effort needed to develop innovative medication leads. Dereplication uses processed experimental data to identify compounds by comparing it to data from known compounds, hence it requires a variety of computing tools and resources to process and analyze compound data. The combination of analytical data analysis and multivariate data analysis is a key technique for computer-assisted dereplication (Chanana et al. 2017). Dimensionality reduction methods like principal component analysis (PCA), cluster analysis, and/or discrimination assessment may be used to isolate interesting NPs from complicated mixtures, such as NPs in extracts that are specific to a certain organism of interest (Chanana et al. 2017; Abdelmohsen et al. 2014).

By analyzing spectroscopic data, computer-assisted shape elucidation (CASE) systems aim to identify the optimal shape for an active molecule. Structures that are in agreement with experimental (spectroscopic) data are listed and ranked by CASE systems for greater precision. CASE structures ideally operate at low mistake rates and in a fully computerized form. The assignment of stereochemical attributes to NP structures can be done using sophisticated CASE systems because they also take stereospecific NMR data and/or calculations based on DFT (density functional theory) into account (Burns et al. 2019).

NP dereplication is a topic that machine learning techniques find very appealing. Using 13C NMR spectroscopic data, for instance, a recent study once investigated the possibility of machine learning algorithms to assign NPs to eight NP classes (such as chromans) (Martínez-Treviño et al. 2020). It is used to take an XGBoost classifier to achieve the remarkable overall performance. More than 80% of a test set's compounds were correctly assigned for the majority of NP classes. For the quick identification of novel NPs from a filamentous marine cyanobacterium, another discovery successfully applied a convolutional neural network-based method (Reher et al. 2020).

One of the most up-to-date resources for managing MS/MS spectra and sharing the results of such analyses is the Global Natural Product Social Molecular Networking (GNPS). It enables researchers to investigate a dataset and compare its results to anything else that is publicly available. Online dereplication is made possible by GNPS's usage of automated molecular networking analysis (Wang et al. 2016).

5.4.2 Analysis of Physicochemical and Structural Properties

By utilizing the physicochemical and structural characteristics of NPs, NPs have been characterized in a significant way by cheminformatics. The chemical space that NPs occupy is substantially larger than that of synthetic compounds, and they also occupy regions of the chemical space that are often inaccessible to synthetic molecules (Ertl and Schuffenhauer 2008) (Singh and Culberson 2009).

Compared to synthetic pharmaceuticals and synthetic, drug-like substances, NPs are generally heavier and more hydrophobic (Chen et al. 2019b). In addition, their structural complexity is usually higher, particularly when it comes to stereochemistry (often measured by the number of chiral centers and the number of bridgehead atoms in ring systems) and three-dimensional molecular form (Henkel et al. 1999) (Lucas et al. 2015).

The vast variety of ring systems displayed by NPs, particularly in aliphatic systems, is astounding (Ertl and Schuffenhauer 2008) (Grabowski and Schneider 2007). Researchers found that commercially available screening databases lacked core ring scaffolds for 83% of NPs. The two characteristics of NPs that set them apart from synthetic compounds in terms of atom composition are their low variety of nitrogen atoms and their large number of oxygen atoms (Feher and Schmidt 2003; Wetzel et al. 2007; López-Vallejo et al. 2012). However, the vast majority of known NPs and, even more so, those found in actual NP libraries have pharmacological properties (Chen et al. 2018).

Physicochemical and structural characteristics vary across NPs from various kingdoms. For instance, marine species are more likely to have macrocycle-containing NPs or lengthy aliphatic chains than terrestrial species (El-Elimat et al. 2012) (Muigg et al. 2013) (Saldivar-Gonzalez et al. 2018). Their NPs are distinguished by an excessive number of heteroatoms and, in conjunction with this, a wide range of functional groups (Pilkington 2019) (Shang et al. 2018) (Ertl and Schuhmann 2020) (Ertl and Schuhmann 2019).

5.4.3 Structural Diversity Analysis

In terms of structural variety, NPs are incomparable, and this is something that is also evident at the fragment level (Tran et al. 2020). Using the concept of molecular

scaffolds, some research compares natural products (NPs) to synthetic ones in order to evaluate the structural diversity of NPs (Bemis and Murcko 1996). Recent research contrasts the scaffolds that are unique to natural products (NPs) with those of synthetic chemicals and presents an intuitive depiction of them (Ertl and Schuhmann 2020). This then allows us to compare the scaffolds often seen in NPs derived from bacteria, plants, fungi, or mammals (Chen et al. 2018).

Scaffold Hunter is a potent java-based application for the intuitive, visual study of the structural variety of a set of chemicals (Schäfer et al. 2017; Lachance et al. 2012). The concept of molecular scaffolds being represented and categorized hierarchically forms the foundation of Scaffold Hunter. An early version of this tool was used to develop the structural categorization of NPs (SCONP), a technique for mapping the chemical space of NPs (Koch et al. 2005).

Principal component analysis (PCA) is a common technique for mapping the chemical space since it transforms high-dimensional data into a low-dimensional space with little loss of information. The most useful result of principal components analysis (PCA) is the PCA scatter plot, which shows how the data points are distributed in a low-dimensional space (Saldívar-González et al. 2019; Shen et al. 2012).

A method called ChemGPS was created and updated for usage with NPs under the name ChemGPS-NP in order to prevent the need for the principal components to be recalculated as new compounds are added to the datasets. For mapping the chemical space of tiny compounds, predicting modes of action, and analyzing structure-activity connections, ChemGPS-NP has been employed in a number of research (Frédérick et al. 2012; Korinek et al. 2017; Muigg et al. 2013).

The recently developed UMAP for Dimension Reduction method and t-SNE are two more reliable methods for dimensionality reduction. When various items are modeled by distant points and the same objects are generally close together, t-SNE creates plots. Although UMAP is quicker, it delivers results conceptually comparable to those of t-SNE (Van der Maaten and Hinton 2008) (Burton 2020).

Researchers recently developed Statistical-Based Database Fingerprint (SB-DFP), which is a new technique for representing the chemical space of compound databases by a single fingerprint. In theory, any chemical fingerprint and any reference set might be used to derive the SB-DFP, which has a wide range of applicability. By contrasting the binomial distributions of the preferred molecular fingerprint features among the compounds in an interest dataset with those in a reference dataset, the SB-DFP is created (Sánchez-Cruz and Medina-Franco 2018).

5.4.4 Natural Product-Likeness Assessment

The NP-likeness of compounds can be quantified using computational techniques, which can also distinguish NPs and NP-like substances from manufactured compounds with high accuracy. As a result, they are often used in the development of new compounds, the construction of libraries, the selection of NPs (and NP

derivatives and analogs) from collections of mixed compounds, and the prioritizing of compounds (Chen et al. 2022) (Yu 2011).

The NP-Likeness Score is one of the most well-known strategies (Ertl and Schuffenhauer 2008). This score assesses the NP-likeness of compounds using Bayesian statistics, mostly based on how similar their fragments are to those of recognized NPs. With certain changes, the NP-Likeness Score has been modified in several programs and platforms (Jayaseelan and Steinbeck 2014; Vanii Jayaseelan et al. 2012; Sorokina and Steinbeck 2019). Additionally, a rule-based strategy and a theoretically related method using extended connectivity fingerprints (ECFPs) are other options (Zaid et al. 2010). A more recent method for locating NPs and NP-like substances in vast sets of molecules is called NP-Scout (Chen et al. 2019b).

In order to properly characterize the structural properties of NPs, a novel method known as the Natural Compound Molecular Fingerprint (NC-MFP) has been developed (Seo et al. 2020).

5.4.5 Identification of Bioactive Natural Products

With regard to identifying bioactive NPs, computational approaches have demonstrated their effectiveness. For NP research, the full spectrum of virtual screening methods has been used, from straightforward, quick methods based solely on 2D molecular fingerprint similarity to more sophisticated, 3D methods largely based on similarity in molecular structure, pharmacophore models, molecular interaction fields, or docking. Machine learning techniques have recently become a cornerstone in virtual screening for bioactive NPs (Kirchweger and Rollinger 2018).

The sparseness of the structural information that is now available will make it extremely difficult to attach NPs to the structures of macromolecules. This is due to the fact that docking algorithms and scoring criteria are particularly sensitive to even very small changes in 3D form, as those frequently brought on by ligand binding (including solvent effects). The careful employment of homology modeling techniques, induced fit docking methods, and molecular dynamics simulations, however, can also aid to overcome this challenge. Docking toward a variety of representative protein structures may be an effective strategy when dealing with highly adaptable proteins (for binding mode prediction as well as virtual screening) (Amaro et al. 2018; Grienke et al. 2010).

In terms of binding mode prediction, docking algorithms frequently produce accurate results as opposed to virtual screening. It is possible to generate a sufficiently accurate binding pose that offers crucial insights for the development of optimization techniques if the target NP is no longer excessively large or flexible, the ligand binding site is well-defined (i.e., not too shallow, not solvent-exposed), and the interaction between the binding companions consists of two or more directed interactions (Chen and Kirchmair 2020). Binding posture prediction is more practical than virtual screening because it completely ignores the most difficult part of docking—scoring compounds according to their binding affinity—and permits

researchers to focus their efforts on a single ligand-target combination. Importantly, docking makes it possible to clarify the stereoselectivity of ligand binding, especially in the context of NP research (and different processes, such as metabolism). It is impossible to exaggerate how important it is to employ the proper stereochemical data when using 3D techniques, particularly docking (Warren et al. 2006).

5.4.6 Determination of Macromolecular Targets

When one, few, or even many compounds are tested against the broadest range of macromolecules, it may be said that *in silico* target prediction is a large-scale use of virtual screening (Grisoni et al. 2019). Numerous techniques including models have been described in recent years, and they are now recognized as crucial resources in the early stages of drug development. The majority of target prediction algorithms are ligand-based due to the difficulties associated with docking and structure-based approaches in general (specifically, the restricted representation of macromolecules through the available structural data) (Cereto-Massagué et al. 2015; Ezzat et al. 2019; Sam and Athri 2019; Chaudhari et al. 2017).

Ligand-based approaches span the whole spectrum, from simple similarity-based methods to sophisticated machine learning and network-based methods. Unexpectedly, despite the wide variety of computer approaches available today for target prediction, we still have a limited understanding of the importance of these tactics in practical situations. This is especially true given the (generally) expensive expenses associated in experimentally evaluating such models in a systematic, prospective manner. However, it is also true given the common use of partially inadequate, cursory retrospective validation techniques (Mathai et al. 2020; Mathai and Kirchmair 2020). To the best of our knowledge, the Similarity Ensemble Approach (SEA) is the only computational strategy for which consistent experimental validation has been documented (Keiser et al. 2007) (Keiser et al. 2009)(Lounkine et al. 2012).

In recent research comparing the effectiveness and scope of a similarity-based strategy and a machine learning technique toward determining the targets of small molecules, it was discovered that the structural similarity between both the compounds of interest and the compounds reflected in the training set is a key factor in both methods' predictability (or knowledge base). Given that target prediction models are essentially created for and trained on experimental measurements for synthetic chemicals, it is important to take this fact into consideration while working with NPs (Mathai et al. 2020).

Surprisingly, in the same research, the similarity-based technique beat the machine learning strategy for the data at hand. The results imply that the basic similarity-based strategy is a realistic choice, in particular when taking into consideration model interpretability. However, a direct comparison in between two approaches should be approached with extreme caution for a number of reasons.

Additionally, this is demonstrated by the successful operation of several well-known, similarity-based approaches like SwissTargetPrediction (Gfeller et al. 2014).

In addition to 3D similarity-based methods, 3D pharmacophore-based approaches are extensively utilized in the field of NP research for target prediction. A profiling investigation, for example, evaluated secondary metabolites extracted from the medicinal plant Ruta graveolens against a battery of over 2000 pharmacophore models covering over 280 targets (Rollinger et al. 2009). Arborinine was discovered to be an inhibitor of acetylcholinesterase (estimated IC50 = 35 M) as a result of this in silico search, among other potential bioactive NPs and interactions.

Machine learning-based methods have undoubtedly sparked the most interest in NP target prediction in recent years. SPiDER, TIGER, and STarFish are a few notable examples (Reker et al. 2014b) (Schneider and Schneider 2017a) (Cockroft et al. 2019).

With the use of "fuzzy" molecular descriptors, SPiDER employs self-organizing maps in an acronym that enables NPs to utilize it (Rodrigues et al. 2016b; Merk et al. 2018). The mannequin helped identify the targets of the macrolide PPAR, archazolid A (Reker et al. 2014a), including 5-lipoxygenase, FXR, glucocorticoid receptor, as well as, prostaglandin E2 synthase 1. It also successfully predicted the target of the 16-membered depsipeptide doliculide, which is prostanoid receptor 3 (Schneider et al. 2016). Numerous fragment-like NPs were also successfully recognized by SPIDER, including (i) sparteine, whose targets include the nicotinic receptors, muscarinic, p38 mitogen-activated protein kinase, and kappa opioid receptor (Rodrigues et al. 2016a), (ii) DL-goitrin, whose targets include the muscarinic M1 receptor and the pregnane X receptor, (iii) Isomacroin, whose targets were experimentally verified to be the adenosine A3 receptor and the platelet-derived growth factor receptor, and (iv) graveolinine, whose objectives were scientifically proven to be cyclooxygenase-2 and the serotonin 5-HT2B receptor (Rodrigues et al. 2015).

SPiDER and TIGER have a similar conceptual framework. The projected targets are scored using a new methodology and updated CATS descriptions (taking into account ensemble similarity). The marine NP (+−)-marinopyrrole A (Schneider and Schneider 2017a) has been effectively discovered by TIGER as a target of cholecystokinin receptor, the orexin receptor, and glucocorticoid receptor. The model correctly identified the estrogen receptors and as targets of the stilbenoid resveratrol, among other proteins (Schneider and Schneider 2017b).

A stacked ensemble target prediction approach called STarFish was developed using synthetic chemical data (Cockroft et al. 2019).

Most recently, medical indication information was used to train multitask deep neural networks and use them to identify privileged chemical scaffolds in NPs (in this instance, scaffolds are used for which many NPs built within the same scaffold are active inside the same indication). A privileged scaffold dataset was created for 100 indications based on the predictions of these models, which may be used as the starting point for NP-based drug development (Lai et al. 2020).

5.4.7 *Prediction of ADME and Safety Profiles of NPs*

ADME and safety profiling has a major say in drug discovery. ADME failures contribute to around 40% of all the drug failures (Bhhatarai et al. 2019). So far, the *in silico* ADME techniques have seen significant progress as shown in Table 5.2. Drug toxicity is still a major concern despite the fact that pharmacokinetics (PK) failures have decreased as a result of preclinical ADME investigations. These failures at late phases of drug discovery pipeline causes huge loss of time and capital. The *in silico* models provide a prior prediction for optimization. Another concern is drug–drug interactions (DDI) which can result in toxicity and severe ADR, obscuring the whole process. Established and broadly applicable computational filters will serve the best for screening and synthesizing and optimizing the drug product (Ekins et al. 2000). In the 1960s, the early phase of ADME models was developed using Hansch's conventional QSAR methods. As a result, comparative molecular field analysis (CoMFA), a type of molecular modeling software, was developed, in such a way that three-dimensional visualization became an important direction for QSAR.

The different ADME properties that can be evaluated by computational approach are solubility, permeability, clearance, metabolic stability, drug–drug interactions, blood–brain barrier, and cardiotoxicity.

The different software available for predicting ADME properties are MolCode toolbox, preADMET, MolCode toolbox, Discovery Studio,volsurfC, QikProp, ADMEWORKS Predictor C Chembench, and admetSAR (Shin et al. 2017).

The major challenges addressed by NPs related to ADME profiling are off-target receptors such as—hERG channel, cytochrome P450 enzymes (suspected for drug-drug interactions, and toxicity), and the P-glycoprotein (suspected for drug resistance). A plethora of such models based on statistical, machine learning, pharmacophore address these and many other off-targets. Another major concern is most of the computational models are validated by synthetic origin drug product. Computational models such as FAME 3 have reportedly known to for their effectiveness even when majority of compounds in the training set are again of synthetic origin (Šícho et al. 2019).

Table 5.2 Progress in *in silico* ADME (Bhhatarai et al. 2019)

Phase	Progress
1960s	Classical QSAR methods with small datasets developed by Hansch (1972), introduction of use of octanol ±water log *P*
1980s	CoMFA was developed along with other membrane permeability and intestinal absorption models—CYP 3D-QSAR and 4D-QSAR modeling
2010s	More than 100,000 data for in vitro ADME properties in big pharma, open access data in thousands, growth of open projects (for example, eTOX, OpenTox, Tox21, ToxCast). wide variety of ML algorithms (RF, SVM, KNN, NB, DNN)

5.4.8 Case Study

Scientists have shown that five tropical plants—M. charantia, B. javanica, E. longifolia, T. divaricata, and G. mangostana—exhibit inhibitory effect against H5N1 neuraminidase. For the purposes of bioassays, different plant parts (leaves, roots, and fruits) were extracted, chromatographed, and fractionated. The anti-H5N1 neuraminidase activity of the plant fractions and extracts ranged from excellent to moderate. At 250 g/ml, G. mangostana showed the maximum inhibition (82.95 percent). Following this, pure chemicals were extracted from the five plants. The IC50 values of rubraxanthone, mangostin, and garcinone C ranged from 89.71 to 95.49 M, making them stand out (Ikram et al. 2015). This process is depicted below (Fig. 5.2) and the docking results of the abovementioned plant derivative are mentioned in Fig. 5.3.

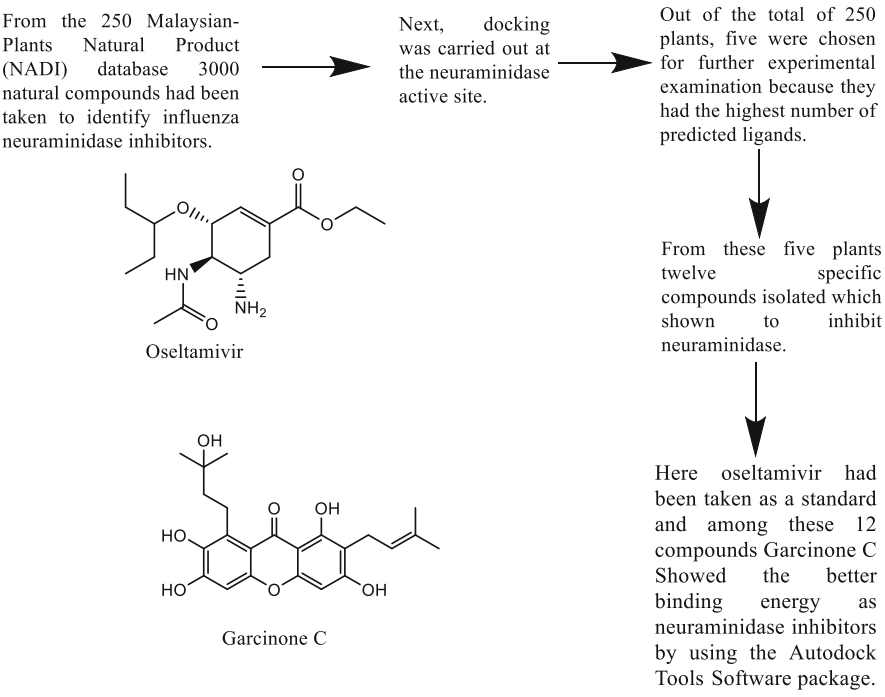


Fig. 5.2 Strategies for novel neuraminidase inhibitors discovery of natural product (Ikram et al. 2015)

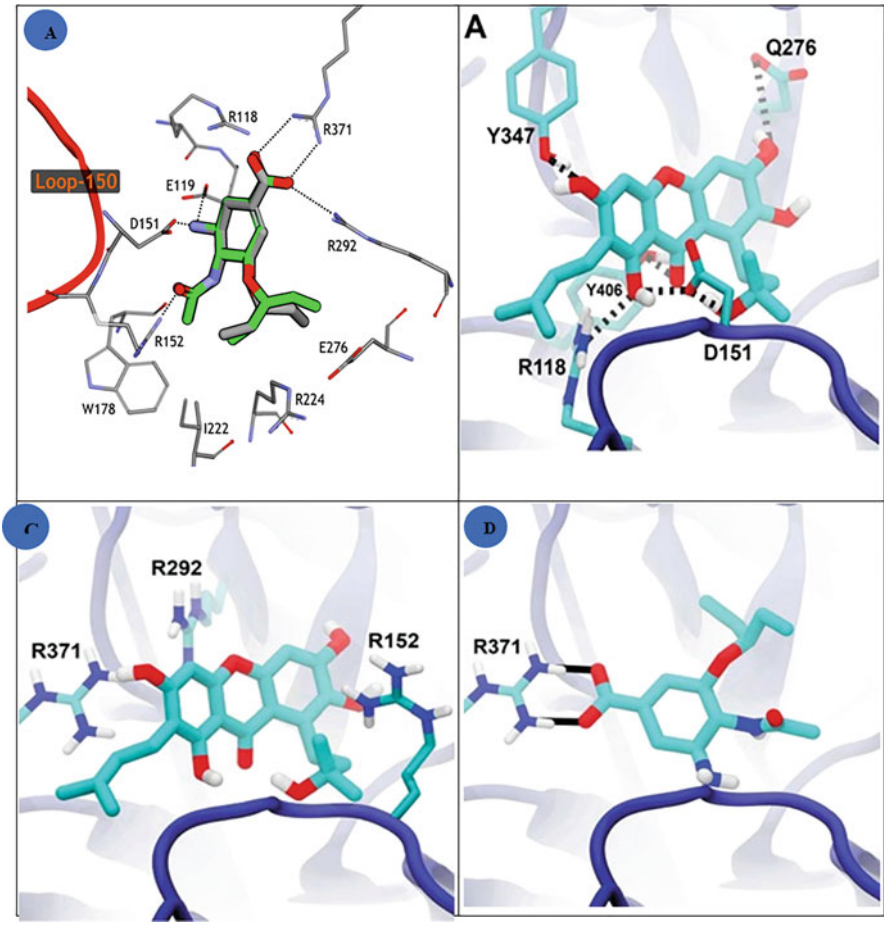


Fig. 5.3 (**a**) The superimposition of the docked and crystallographic oseltamivir poses (green and blue, respectively). The RMSD was 0.84 Å. (**b**) Predicted hydrogen bonds of Garcinone C in the active site of neuraminidase inhibitors. (**c**) Predicted cation-π interactions between R371, R292, R152, and the xanthone moiety of Garcinone C in the active site of neuraminidase inhibitors. (**d**) The crystallographic pose of oseltamivir, a potent inhibitor, shown for reference (PDB ID: 2HU4) (Ikram et al. 2015)

5.5 Challenges to Computational Approaches

The major challenges for NP-based drug discovery is management and representation of the data. Although ArachnoServer and ConoServer are rich and highly descriptive NP databases, but reserved only to specific clade of species producing toxins (Kaas et al. 2012). A partial solution for this is Tox-Prot manual annotation program within UniProtKB/Swiss-Prot which provides a more generalized and improved representation of databases for NPs (Jungo et al. 2012). However, this

Table 5.3 List of databases discontinued in 2019

Database	Type of NPs	Size
3DMET (Maeda and Kondo 2013)	General	18,248
AfroDB (Ntie-Kang et al. 2013b)	tm, plants, Africa	954
CamMedNP (Ntie-Kang et al. 2013a)	tm, plants, Africa	>2500
Traditional Chinese Medicine Systems Pharmacology (TCMSP) (Ru et al. 2014)	Chinese herbs	499

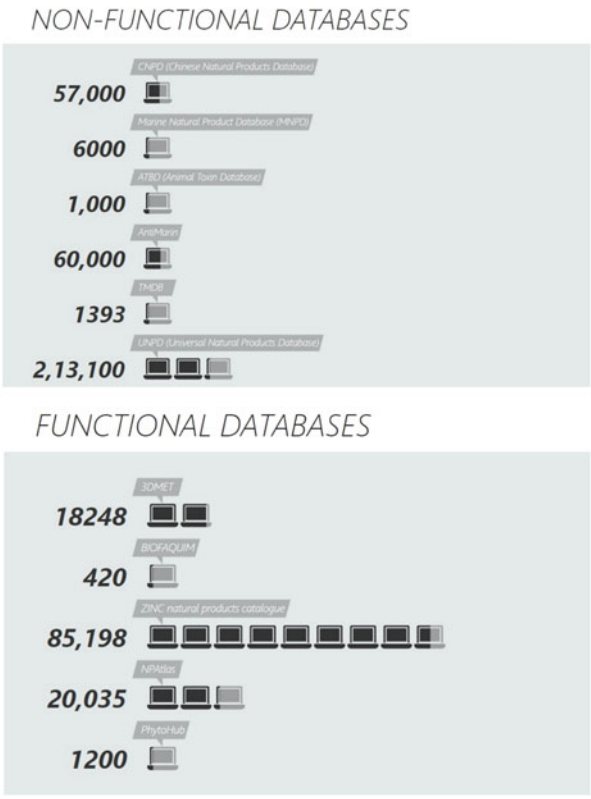


Fig. 5.4 Pictorial representation of functional and defunct databases

does not seem to be the complete solution. Another concern associated with NPs is fragmentation of databases which means more scattered form of data to be maintained by smaller or larger organizations. The added difficulty is shortage of funding required for maintaining those databases which leads to mismanagement of data, ultimately disabling the function of that database. Examples of such databases include as follows (Table 5.3):

To have a clear view, a comparative data of functional and defunct databases have been depicted in Fig. 5.4. A fundamental obstacle to the experimental screening of

NPs is their propensity to interact with biological tests. This could be explained with the example of quercetin which has reportedly shown active in more than about 800 unique bioassays. The most common mechanism followed for interference is aggregate formation, covalent binding, membrane disruption, metal chelation, interference with assay spectroscopy, and buffer decomposition buffers (Baell and Holloway 2010). These problems could be overcome by specific set of rules following statistical approach known as pan-assay interference compounds (PAINS) rule set (Baell and Nissink 2018).

5.6 Conclusion and Future Perspectives

Between the 1980s and the 2010s, two-thirds of the medications were either featured NP pharmacophores (35%) or were analogs of NPs (5%). Modern computational techniques discussed above can significantly expedite and reduce the risk of NP-based drug development. The integration of computational approaches with cheminformatics and other informatics methods has led to ease the management, storage, and representation of vast NP-based databases. Computational tools offer assistance in structural elucidation of bioactive NPs, in prior prediction of various properties of NPs as discussed above which eases the procedure for drug discovery pipeline. However, the major challenge being availability of descriptive database, fragmented databases, and its maintenance along with physical availability of the particular NP. These challenges have been resolved partially with introduction of databases like COlleCtion of Open Natural prodUcTs (COCONUT) which provides a web interface to browse and download elucidated and predicted NPs collected from open sources. On a larger parameter, machine learning (ML) has been using computational methods in drug discovery. For instance, clustering techniques have enabled de novo molecular design, projected protein target druggability, and segmented cell type imaging. The computational approach for NP-based drug discovery holds great future for NP-based drug discovery. The amalgamation of computational methods with advanced technologies in analytical domains can improvise the drug discovery pipeline for NPs. The advancement of higher-field NMR instruments and probe technology has made it possible to determine the structure of NPs from extremely small amounts hence, less wastage of hardly obtained product. Pauli and associates suggested conducting early, relatively sophisticated purity analyses on lead nanoparticles using quantitative NMR and LC–MS to avoid pointless downstream initiatives. Further advancement of metabolomics, genome mining, microbial culturing technique has added to the future scope of NP-based drug pipeline. In addition, antivirulence strategies may represent an alternative method for combating infections, for which NPs that target bacterial quorum sensing may be of interest. *In silico* Medicine, an American company, created an AI system called GENTRL (Generative Tensorial Reinforcement Learning) in 2019 that, in just 46 days, successfully created six kinase inhibitors of the discoidin domain receptor 1 linked to lung fibrosis. Cheminformatics, bioinformatics, and other related fields

have made significant contributions to NP-based drug discovery over the years. Recently, reviews of their successful applications and limitations were conducted.

References

Abdelmohsen UR, Cheng C, Viegelmann C et al (2014) Dereplication strategies for targeted isolation of new antitrypanosomal actinosporins A and B from a marine sponge associated-Actinokineospora sp. EG49. Mar Drugs 12:1220–1244

Amaro RE, Baudry J, Chodera J et al (2018) Ensemble docking in drug discovery. Biophys J 114: 2271–2278

Atanasov AG, Waltenberger B, Pferschy-Wenzig EM et al (2015) Discovery and resupply of pharmacologically active plant-derived natural products: a review. Biotechnol Adv 33:1582–1614. https://doi.org/10.1016/j.biotechadv.2015.08.001

Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem 53:2719–2740

Baell JB, Nissink JWM (2018) Seven-year itch: pan-assay interference compounds (PAINS) in 2017—utility and limitations. ACS Chem Biol 13:36–44

Banerjee P, Erehman J, Gohlke B-O et al (2015) Super Natural II—a database of natural products. Nucleic Acids Res 43:D935–D939

Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. J Med Chem 39:2887–2893

Bhhatarai B, Walters WP, Hop CECA et al (2019) Opportunities and challenges using artificial intelligence in ADME/Tox. Nat Mater 18:418–422

Bugni TS, Richards B, Bhoite L et al (2008) Marine natural product libraries for high-throughput screening and rapid drug discovery. J Nat Prod 71:1095–1098. https://doi.org/10.1021/np800184g

Burns DC, Mazzola EP, Reynolds WF (2019) The role of computer-assisted structure elucidation (CASE) programs in the structure elucidation of complex natural products. Nat Prod Rep 36: 919–933

Burton R (2020) Unsupervised learning techniques for malware characterization: understanding certain DNS-based DDoS attacks. Digit Threat Res Pract 1:1–26

Carter PJ, Lazar GA (2018) Next generation antibody drugs: pursuit of the 'high-hanging fruit'. Nat Rev Drug Discov 17:197–223

Cereto-Massagué A, Ojeda MJ, Valls C et al (2015) Tools for in silico target fishing. Methods 71: 98–103

Cerqueira NM, Gesto D, Oliveira EF et al (2015) Receptor-based virtual screening protocol for drug discovery. Arch Biochem Biophys 582:56–67

Chanana S, Thomas CS, Braun DR et al (2017) Natural product discovery using planes of principal component analysis in R (PoPCAR). Meta 7:34

Chaudhari R, Tan Z, Huang B, Zhang S (2017) Computational polypharmacology: a new paradigm for drug discovery. Expert Opin Drug Discov 12:279–291. https://doi.org/10.1080/17460441.2017.1280024

Chen CY-C (2011) TCM database@ Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. PLoS One 6:e15939

Chen Y, de Bruyn KC, Kirchmair J (2017) Data resources for the computer-guided discovery of bioactive natural products. J Chem Inf Model 57:2099–2111

Chen Y, de Bruyn Kops C, Kirchmair J (2019a) Resources for chemical, biological, and structural data on natural products. Prog Chem Org Nat Prod 110:37–71

Chen Y, Garcia de Lomana M, Friedrich N-O, Kirchmair J (2018) Characterization of the chemical space of known and readily obtainable natural products. J Chem Inf Model 58:1518–1532
Chen Y, Kirchmair J (2020) Cheminformatics in natural product-based drug discovery. Mol Inform 39:e2000171
Chen Y, Rosenkranz C, Hirte S, Kirchmair J (2022) Ring systems in natural products: structural diversity, physicochemical properties, and coverage by synthetic compounds. Nat Prod Rep 39: 1544–1556
Chen Y, Stork C, Hirte S, Kirchmair J (2019b) NP-Scout: machine learning approach for the quantification and visualization of the natural product-likeness of small molecules. Biomol Ther 9:43
Cockroft NT, Cheng X, Fuchs JR (2019) STarFish: a stacked ensemble target fishing approach and its application to natural products. J Chem Inf Model 59:4906–4920
Cozza G, Bonvini P, Zorzi E et al (2006) Identification of ellagic acid as potent inhibitor of protein kinase CK2: a successful example of a virtual screening application. J Med Chem 49:2363–2366
Cragg GM, Newman DJ, Snader KM (1997) Natural products in drug discovery and development. J Nat Prod 60:52–60
Daley DK, Brown KJ, Badal S (2017) Fungal metabolites. In: Pharmacognosy: fundamentals, applications and strategy. Elsevier, London
Dona MSI, Prendergast LA, Mathivanan S et al (2017) Powerful differential expression analysis incorporating network topology for next-generation sequencing data. Bioinformatics 33:1505–1513
Ekins S, Waller CL, Swaan PW et al (2000) Progress in predicting human ADME parameters in silico. J Pharmacol Toxicol Methods 44:251–272
El-Elimat T, Zhang X, Jarjoura D et al (2012) Chemical diversity of metabolites from fungi, cyanobacteria, and plants relative to FDA-approved anticancer agents. ACS Med Chem Lett 3:645–649
Ertl P, Schuffenhauer A (2008) Cheminformatics analysis of natural products: lessons from nature inspiring the design of new drugs. Prog drug Res 66(217):219–235. https://doi.org/10.1007/978-3-7643-8595-8_4
Ertl P, Schuhmann T (2019) A systematic cheminformatics analysis of functional groups occurring in natural products. J Nat Prod 82:1258–1263
Ertl P, Schuhmann T (2020) Cheminformatics analysis of natural product scaffolds: comparison of scaffolds produced by animals, plants, fungi and bacteria. Mol Inform 39:2000017
Ezzat A, Wu M, Li X-L, Kwoh C-K (2019) Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. Brief Bioinform 20:1337–1357
Fabricant DS, Farnsworth NR (2001) The value of plants used in traditional medicine for drug discovery. Environ Health Perspect 109:69–75. https://doi.org/10.1289/ehp.01109s169
Feher M, Schmidt JM (2003) Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. J Chem Inf Comput Sci 43:218–227
Frédérick R, Bruyère C, Vancraeynest C et al (2012) Novel trisubstituted harmine derivatives with original in vitro anticancer activity. J Med Chem 55:6489–6501
Friedrich N-O, Flachsenberg F, Meyder A et al (2019) Conformator: a novel method for the generation of conformer ensembles. J Chem Inf Model 59:731–742
Gfeller D, Grosdidier A, Wirth M et al (2014) SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res 42:W32–W38. https://doi.org/10.1093/nar/gku293
Grabowski K, Schneider G (2007) Properties and architecture of drugs and natural products revisited. Curr Chem Biol 1:115–127
Grienke U, Schmidtke M, Kirchmair J et al (2010) Antiviral potential and molecular insight into neuraminidase inhibiting diarylheptanoids from Alpinia katsumadai. J Med Chem 53:778–786
Grisoni F, Merk D, Friedrich L, Schneider G (2019) Design of natural-product-inspired multitarget ligands by machine learning. ChemMedChem 14:1129–1134

Gu J, Gui Y, Chen L et al (2013) Use of natural products as chemical library for drug discovery and network pharmacology. PLoS One 8:e62839
Henkel T, Brunne RM, Müller H, Reichel F (1999) Statistical investigation into the structural complementarity of natural products and synthetic compounds. Angew Chem Int Ed 38:643–647
Huang P-S, Boyken SE, Baker D (2016) The coming of age of de novo protein design. Nature 537:320–327
Ikram NKK, Durrant JD, Muchtaridi M et al (2015) A virtual screening approach for identifying plants with anti-H5N1 neuraminidase activity. J Chem Inf Model 55:308–316
Jayaseelan KV, Steinbeck C (2014) Building blocks for automated elucidation of metabolites: natural product-likeness for candidate ranking. BMC Bioinform 15:1–9
Jungo F, Bougueleret L, Xenarios I, Poux S (2012) The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data. Toxicon 60:551–557
Kaas Q, Yu R, Jin A-H et al (2012) ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. Nucleic Acids Res 40:D325–D330
Keiser MJ, Roth BL, Armbruster BN et al (2007) Relating protein pharmacology by ligand chemistry. Nat Biotechnol 25:197–206. https://doi.org/10.1038/nbt1284
Keiser MJ, Setola V, Irwin JJ et al (2009) Predicting new molecular targets for known drugs. Nature 462:175–181
Kirchweger B, Rollinger JM (2018) Virtual screening for the discovery of active principles from natural products. In: Natural products as source of molecules with therapeutic potential. Springer, Cham, pp 333–364
Koch MA, Schuffenhauer A, Scheck M et al (2005) Charting biologically relevant chemical space: a structural classification of natural products (SCONP). Proc Natl Acad Sci 102:17272–17277
Korinek M, Tsai Y-H, El-Shazly M et al (2017) Anti-allergic hydroxy fatty acids from Typhonium blumei explored through ChemGPS-NP. Front Pharmacol 8:356
Lachance H, Wetzel S, Kumar K, Waldmann H (2012) Charting, navigating, and populating natural product chemical space for drug discovery. J Med Chem 55:5989–6001
Lai J, Hu J, Wang Y et al (2020) Privileged scaffold analysis of natural products with deep learning-based indication prediction model. Mol Inform 39:2000057. https://doi.org/10.1002/minf.202000057
Lawson ADG, MacCoss M, Heer JP (2017) Importance of rigidity in designing small molecule drugs to tackle protein–protein interactions (PPIs) through stabilization of desired conformers: miniperspective. J Med Chem 61:4283–4289
Leach AR (2017) The ChEMBL database in. Nucleic Acids Res 45:D945–D954
Li Q, Cheng T, Wang Y, Bryant SH (2010) PubChem as a public resource for drug discovery. Drug Discov Today 15:1052–1057
Lill MA (2007) Multi-dimensional QSAR in drug discovery. Drug Discov Today 12:1013–1017
López-Vallejo F, Giulianotti MA, Houghten RA, Medina-Franco JL (2012) Expanding the medicinally relevant chemical space with compound libraries. Drug Discov Today 17:718–726
Lounkine E, Keiser MJ, Whitebread S et al (2012) Large-scale prediction and testing of drug activity on side-effect targets. Nature 486:361–367
Lucas X, Grüning BA, Bleher S, Günther S (2015) The purchasable chemical space: a detailed picture. J Chem Inf Model 55:915–924
Maeda MH, Kondo K (2013) Three-dimensional structure database of natural metabolites (3DMET): a novel database of curated 3D structures. J Chem Inf Model 53:527–533
Mann J (2002) Natural products in cancer chemotherapy: past, present and future. Nat Rev Cancer 2:143–148
Martínez-Treviño SH, Uc-Cetina V, Fernández-Herrera MA, Merino G (2020) Prediction of natural product classes using machine learning and 13C NMR spectroscopic data. J Chem Inf Model 60:3376–3386
Mathai N, Chen Y, Kirchmair J (2020) Validation strategies for target prediction methods. Brief Bioinform 21:791–802. https://doi.org/10.1093/bib/bbz026

Mathai N, Kirchmair J (2020) Similarity-based methods and machine learning approaches for target prediction in early drug discovery: performance and scope. Int J Mol Sci 21:3585
Merk D, Grisoni F, Friedrich L et al (2018) Computer-assisted discovery of retinoid X receptor modulating natural products and isofunctional mimetics. J Med Chem 61:5442–5447
Muigg P, Rosén J, Bohlin L, Backlund A (2013) In silico comparison of marine, terrestrial and synthetic compounds using ChemGPS-NP for navigating chemical space. Phytochem Rev 12: 449–457
Nakamura K, Shimura N, Otabe Y et al (2013) KNApSAcK-3D: a three-dimensional structure database of plant metabolites. Plant Cell Physiol 54:e4–e4
Newman DJ, Cragg GM (2016) Natural products as sources of new drugs from 1981 to 2014. J Nat Prod 79:629–661. https://doi.org/10.1021/acs.jnatprod.5b01055
Ntie-Kang F, Mbah JA, Mbaze LM et al (2013a) CamMedNP: building the Cameroonian 3D structural natural products database for virtual screening. BMC Complement Altern Med 13:1–10
Ntie-Kang F, Zofou D, Babiaka SB et al (2013b) AfroDb: a select highly potent and diverse natural product library from African medicinal plants. PLoS One 8:e78085
Patridge E, Gareiss P, Kinch MS, Hoyer D (2016) An analysis of FDA-approved drugs: natural products and their derivatives. Drug Discov Today 21:204–207
Pereira F, Aires-de-Sousa J (2018) Computational methodologies in the exploration of marine natural product leads. Mar Drugs 16:236
Pilkington LI (2019) A Chemometric analysis of deep-sea natural products. Molecules 24:3942
Pilon AC, Valli M, Dametto AC et al (2017) NuBBEDB: an updated database to uncover chemical and biological information from Brazilian biodiversity. Sci Rep 7:1–12
Pithayanukul P, Leanpolchareanchai J, Saparpakorn P (2009) Molecular docking studies and anti–snake venom metalloproteinase activity of Thai mango seed kernel extract. Molecules 14:3198–3213
Podlogar BL, Muegge I, Brice LJ (2001) Computational methods to estimate drug development parameters. Curr Opin Drug Discov Devel 4:102–109
Reher R, Kim HW, Zhang C et al (2020) A convolutional neural network-based approach for the rapid annotation of molecularly diverse natural products. J Am Chem Soc 142:4114–4120
Reker D, Perna AM, Rodrigues T et al (2014a) Revealing the macromolecular targets of complex natural products. Nat Chem 6:1072–1078
Reker D, Rodrigues T, Schneider P, Schneider G (2014b) Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. Proc Natl Acad Sci 111:4067–4072
Rodrigues T, Reker D, Kunze J et al (2015) Revealing the macromolecular targets of fragment-like natural products. Angew Chem Int Ed 54:10516–10520
Rodrigues T, Reker D, Schneider P, Schneider G (2016a) Counting on natural products for drug design. Nat Chem 8:531–541. https://doi.org/10.1038/nchem.2479
Rodrigues T, Sieglitz F, Somovilla VJ et al (2016b) Unveiling (−)-Englerin A as a modulator of L-type calcium channels. Angew Chem Int Ed 55:11077–11081
Rollinger JM, Schuster D, Danzl B et al (2009) In silico target fishing for rationalized ligand discovery exemplified on constituents of Ruta graveolens. Planta Med 75:195–204
Romano JD, Nwankwo V, Tatonetti NP (2018) VenomKB v2. 0: a knowledge repository for computational toxinology. Sci Data 2:150065
Romano JD, Tatonetti NP (2019) Informatics and computational methods in natural product drug discovery: a review and perspectives. Front Genet 10:368
Ru J, Li P, Wang J et al (2014) TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. J Cheminform 6:1–6
Saldívar-González FI, Angélica Pilón-Jiménez B, Medina-Franco JL (2019) Phys Sci Rev 4: 20180103
Saldivar-Gonzalez FI, Valli M, Andricopulo AD et al (2018) Chemical space and diversity of the NuBBE database: a chemoinformatic characterization. J Chem Inf Model 59:74–85

Sam E, Athri P (2019) Web-based drug repurposing tools: a survey. Brief Bioinform 20:299–316
Sánchez-Cruz N, Medina-Franco JL (2018) Statistical-based database fingerprint: chemical space-dependent representation of compound databases. J Cheminform 10:1–13
Schäfer T, Kriege N, Humbeck L et al (2017) Scaffold Hunter: a comprehensive visual analytics framework for drug discovery. J Cheminform 9:1–18
Schneider G, Reker D, Chen T et al (2016) Deorphaning the macromolecular targets of the natural anticancer compound doliculide. Angew Chem Int Ed 55:12408–12411
Schneider P, Schneider G (2017a) De-orphaning the marine natural product (±)-marinopyrrole A by computational target prediction and biochemical validation. Chem Commun 53:2272–2274
Schneider P, Schneider G (2017b) A computational method for unveiling the target promiscuity of pharmacologically active compounds. Angew Chem Int Ed 56:11520–11524
Seo M, Shin HK, Myung Y et al (2020) Development of natural compound molecular fingerprint (NC-MFP) with the dictionary of natural products (DNP) for natural product-based drug development. J Cheminform 12:6. https://doi.org/10.1186/s13321-020-0410-3
Shang J, Hu B, Wang J et al (2018) Cheminformatic insight into the differences between terrestrial and marine originated natural products. J Chem Inf Model 58:1182–1193
Shen M, Tian S, Li Y et al (2012) Drug-likeness analysis of traditional Chinese medicines: 1. Property distributions of drug-like compounds, non-drug-like compounds and natural compounds from traditional Chinese medicines. J Cheminform 4:1–13
Shin HK, Kang Y-M, No KT (2017) Predicting ADME properties of chemicals. In: Handbook of computational chemistry, vol 59. Springer, Cham, pp 2265–2301
Šícho M, Stork C, Mazzolari A et al (2019) FAME 3: predicting the sites of metabolism in synthetic compounds and natural products for phase 1 and phase 2 metabolic enzymes. J Chem Inf Model 59:3400–3412
Singh SB, Culberson JC (2009) Chapter 2: Chemical space and the difference between natural products and synthetics. In: Natural product chemistry for drug discovery. The Royal Society of Chemistry, Cambridge, pp 28–43
Sorokina M, Merseburger P, Rajan K et al (2021) COCONUT online: collection of open natural products database. J Cheminform 13:1–13
Sorokina M, Steinbeck C (2019) NaPLeS: a natural products likeness scorer—web application and database. J Cheminform 11:1–7
Sorokina M, Steinbeck C (2020) Review on natural products databases: where to find data in 2020. J Cheminform 12:1–51
Sterling T, Irwin JJ (2015) ZINC 15–ligand discovery for everyone. J Chem Inf Model 55:2324–2337
Terrett NK, Gardner M, Gordon DW et al (1995) Combinatorial synthesis—the design of compound libraries and their application to drug discovery. Tetrahedron 51:8135–8173
Tran TD, Ogbourne SM, Brooks PR et al (2020) Lessons from exploring chemical space and chemical diversity of propolis components. Int J Mol Sci 21:4988
Van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579
Van Santen JA, Jacob G, Singh AL et al (2019) The natural products atlas: an open access knowledge base for microbial natural products discovery. ACS Cent Sci 5:1824–1833
Vanii Jayaseelan K, Moreno P, Truszkowski A et al (2012) Natural product-likeness score revisited: an open-source, open-data implementation. BMC Bioinform 13:1–6
Wang M, Carver JJ, Phelan VV et al (2016) Sharing and community curation of mass spectrometry data with global natural products social molecular networking. Nat Biotechnol 34:828–837
Warren GL, Andrews CW, Capelli A-M et al (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49:5912–5931

Wetzel S, Schuffenhauer A, Roggo S et al (2007) Cheminformatic analysis of natural products and their chemical space. Chim Int J Chem 61:355–360

Yu MJ (2011) Natural product-like virtual libraries: recursive atom-based enumeration. J Chem Inf Model 51:541–557

Zaid H, Raiyn J, Nasser A et al (2010) Physicochemical properties of natural based products versus synthetic chemicals. Open Nutraceuticals J 3:194

Zeng X, Zhang P, Wang Y et al (2019) CMAUP: a database of collective molecular activities of useful plants. Nucleic Acids Res 47:D1118–D1127

Chapter 6
Virtual Screening in Lead Discovery

Vinita, Jurnal Reang, Vivek Yadav, Jaseela Majeed, Prabodh Chander Sharma, Kaalicharan Sharma, and Rajiv Kumar Tonk

6.1 Introduction

Drug discovery continues to be a major concern in biomedicine. The initial phase in the drug discovery process is typically identifying targets for a condition of interest. The next step is to undertake high-throughput screening (HTS) studies to find successes in the synthesized compound library or molecules with promising bioactivity. The next phase is to refine the hit compounds to create molecules with more potent properties and other desired traits, such as solubility or the disappearance of negative side effects. After finishing the preclinical research, possible drug candidates must successfully complete a chain of clinical studies to become licensed pharmaceuticals. A single drug takes between 10 and 15 years and more than 2 billion dollars to create. Although HTS investigations are very effective, they are nonetheless time- and money-consuming since they need a huge amount of protein supply, hundreds of produced compounds, and established procedures for assessing bioactivity in the lab. Computational approaches have been extensively cast-off in the design process during the past three decades to rationalize and expedite drug development. A popular technique that has a strong chance of binding to an important target is Virtual Screening (VS) (Berdigaliyev and Aljofan 2020). When using low-cost platforms like ZINC or MolPORT, VS techniques can quickly scan millions of (commercially) accessible chemicals and prioritize which ones should be put through testing, internally synthesized, or purchased from outside vendors. Additionally, virtual compound libraries can be employed for screening (VS), which broadens the chemical space and is used to prioritize compounds from (ultra) huge compound libraries and a database comprising about two billion

Vinita (✉) · J. Reang · V. Yadav · J. Majeed · P. C. Sharma · K. Sharma · R. K. Tonk
Pharmaceutical Chemistry Department, Delhi Pharmaceutical Sciences and Research University, New Delhi, India

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_6

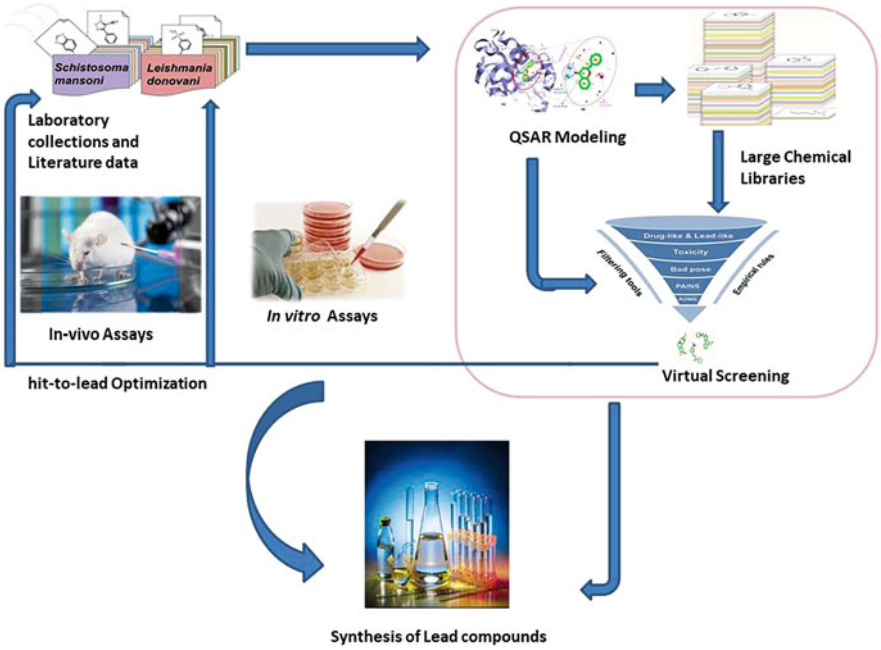


Fig. 6.1 Schematic representation of virtual screening

drug-like chemicals. Enamine REAL, which has over 17 billion make-on-demand compounds, is one of these libraries. The search area can be narrowed down to a few hundred compounds with the necessary characteristics for further research even though VS methods cannot always pinpoint the molecule that is the most active (Scior et al. 2012).

Technically, VS is any statistical or computational filtering tool that is used to choose molecules from a significant database. The obvious next step is to procure these chemicals and perform the experimental testing depicted in Fig. 6.1 after that. An operational definition of VS, which asserts that it is the exercise of ranking molecules by descending order of likelihood of relevant biological activity, encapsulates the essence of VS regardless of the method employed to rank molecules (Hawkins et al. 2007). Usually, the ranking algorithm is chosen based on the target's known qualities, the substances active in the relevant biological assay, how distinct the desired ligands must be from well-known bioactive molecules, and what percentage of the ranked database would be picked for experimental testing. The ranking algorithm must be more effective to produce successful hits from VS when there are fewer compounds to test (Hert et al. 2005).

The benefit of using computational methods is that they can deliver new medication candidates more quickly and inexpensively. In many stages of the discovery process, using complementing experimental and informatics methodologies raises the likelihood of success (Yadav et al. 2021).

Some considerations are necessary for the practical deployment of virtual screening applications (Warren et al. 2006).

- The amount of time needed for computing can be very high, especially when dealing with big databases containing millions of chemical components.
- Particularly when using computationally straightforward procedures, the accuracy of the employed attributes is crucial.
- The techniques used for data analysis are crucial, especially in light of the enormous amount of data that high-performance modeling produces.
- For timely results, the gear and software being used must be appropriate.

6.2 Virtual Screening Methods

The methods of Virtual Screening can be largely categorized into following methods:

1. Approach that ranks compounds according to their degree of resemblance to known actives, depending on the molecule's 2D or 3D structure is known as Ligand-based Virtual Screening (LBVS) (Yadav et al. 2022)
2. Technique that determines a pharmacophore which is a 3D arrangement of characteristics that enhance or inhibit binding and search for it in the database being searched is known as pharmacophore-based Virtual Screening (PBVS). This methodology focuses on the characteristics like hydrogen bond donors, acceptors, acidic or basic units, and hydrophobic fragments and makes it possible to find unexpected scaffolds with the ideal properties.
3. Technique that makes use of the target's structural information, which is often determined by protein crystallography, to find compounds that match the "binding site" through advantageous protein-ligand interactions is known as structure-based Virtual Screening (SBVS).
4. Technique that asks for the target's 3D structure to be present is known as receptor-based Virtual Screening (RBVS).

Depending on the facts at hand, the approach chosen may be aided or restricted. More than one active small compound is identified in the absence of target structural knowledge, LBVS or PHBVS are viable options. SBVS can be taken into account if there are no known active chemicals but there is an investigational or computational model of the protein assembly. Several suitable techniques can be used, or multiple techniques can be combined, if both active chemicals and the target structure are available.

(a) Bench marking
(b) Data base creation
(c) Data base filtering

Drug discovery now includes VS as a crucial component because in order to select potentially active chemicals, a hierarchical workflow is typically implemented that combines various techniques (either sequentially or concurrently) as filters as shown in Fig. 6.1 (Tang and Marshall 2011).

6.2.1 Structure-Based Virtual Screening

Target-based virtual screening (TBVS), often referred to as structure-based virtual screening (SBVS), seeks to predict the most efficient manner in which ligands would interact with a molecular target to produce a complex hence, the ligands are listed in ascending order of target affinity, with the most promising compounds at the top. The target protein's 3D structure must be known for SBVS approaches to predict *in silico,* the interactions between the target protein and each chemical molecule. According to how strongly they bind to the receptor site, the chemicals are selected from a database and organized into groups in this technique. Molecular docking stands out among SBVS techniques due to its less computational expense and successful outcomes (Meng Zhang et al. 2011).

This approach was developed and tested for the first time by developing a series of algorithms to analyze the geometrically possible arrangements of a ligand and target. Despite the method's promise, it was not until the 1990s that this technique was popular, thanks to improvements in the techniques used, a jump in computing power, and simpler right to use to the basic data of the mark molecules (Kuntz et al. 1982). Based on how strongly they attach to the receptor site, the evaluated molecules are sorted throughout the SBVS procedure. As a result, it is possible to discover ligands that are more expected to act together pharmacologically with the molecular target. Score jobs are cast-off to confirm the possibility of a binding site affinity between the ligand and target; hence, a trustworthy scoring function is essential to the docking process in this approach (Leelananda and Lindert 2016).

There are benefits and drawbacks to using SBVS. The following are a few benefits (Lionta et al. 2014):

- The time and money required for screening millions of tiny compounds have decreased.
- As the molecule need not physically exist, it can be computationally examined even before it is created.
- Various tools are available to help SBVS as mentioned in Table 6.1.

The following are some examples of the drawbacks:

- Some tools are more effective in certain situations but not in others.
- The complicated nature of ligand-receptor binding interactions makes it difficult to parameterize as a result which the proper binding position and classification of molecules are not accurately forecast.
- It may provide erroneous results, both positive and negative.

Table 6.1 Key features of widely used docking software

Sl. No.	Software	Author Institution and website	Tutorial	Free trial	Parallel computing	Operating system
1	AutoDock	Molecular Graphics Laboratory, The Scripps Research Institute, CA, USA http://autodock.scripps.edu	Available	Free noncommercial l use	Not available	Linux Windows SGI IRIX Solaris Mac OS X
2	DOCK	Kuntz Lab, University of California, San Francisco, USA http://dock.compbio.ucsf.edu	Available	Free for academic institutions	MPICH library	Linux Windows Mac OS X
3	FlexX	BioSolveIT GmbH Sankt Augustin, Germany http://www.biosolveit.de/FlexX	Available	Six-week evaluation license	Virtual high-throughput screening on cluster platforms	Windows Linux
4	Gold	Cambridge Crystallographic Data Centre (CCDC), UK (in collaboration with University of Sheffield and GlaxoSmithKline) http://www.ccdc.cam.ac.uk/ products/life_sciences/gold	Available	Interactive Web Trial	Grid Computing	Windows Linux IRIX

Despite the flaws mentioned above, several research involving SBVS have been produced recently, proving that despite these disadvantages, SBVS is still regularly because of time and money savings for developing new drugs (Nunes et al. 2019).

Researchers can perform VS procedures more easily by using a variety of VS software tools that use different docking techniques because they are not required to have a thorough understanding of computer science to design the algorithms. VS software reduces the expenses by acting as filter to select compounds from thousands of databases that are more likely to show biological activity against the selected target.

6.2.2 Ligand-Based Virtual Screening

There are a few ligand-based approaches that are rather sophisticated scientifically, such quantitative structure-activity relationship (QSAR) modeling, molecular similarity search, and ligand-based pharmacophores. As templates, LBVS starts with

known active substances. The component of LBVS techniques such as computational descriptors of molecular structure, features, or pharmacophore aspects examines the correlations between active database or test compounds in a distinct chemical descriptor space (Bajorath 2001). The sophistication and complexity of these descriptors might differ significantly. If there are any "descriptor solutions" at all, finding them universally applicable or desirable is difficult. This is due to the fact that it might be difficult to accurately depict molecules using mathematical models and descriptors for LBVS calculations. Multiple descriptor types are frequently required by different search issues and methods, and different descriptor combinations frequently produce results that are either extremely similar or quite divergent. Depending on specific methods and applications, descriptors of variable dimensions may also work (Bradley et al. 2000). It has also been found that, in some cases, even simple 2D descriptions of molecules can faithfully reproduce search results that were previously thought to be dependent on the usage of complex 3D descriptors (Xue et al. 2001). To enable objective and automatic choice of descriptors for definite applications, machine learning techniques or information theoretic methodologies need to be applied. Different methods generally produce divergent results depending on the features of the quest problem under investigation and multifaceted classes that are evaluated, which also reflects the large algorithmic range of approaches exploited for VS applications (Martin et al. 2002).

Sometimes, it is challenging to recognize level of estimated chemical similarity that correlates to biological similarity of test molecules, which complicates compound selection techniques. Even with these challenges and possible downsides, LBVS has produced several prominent success stories, both in sovereign case studies and when used in combination with HTS. For instance, it has been confirmed in various autonomous studies that pairing LBVS with HTS in repetitive way can raise the success proportions by one to a factor of two. Despite the innate limits of computational procedures, it is acceptable to undertake that successful applications of VS are considerably more frequent than those published so far given the high-class flora of the majority of discovery activities. Despite the fact that compound filtering has received a lot of attention, the most common (and challenging) use of LBVS may be the hunt for compounds that structurally differ from patterns but exhibit equivalent activity (Bajorath 2002).

6.2.3 Pharmacophore-Based Virtual Screening

The term "pharmacophore" has an extended and fruitful history in medicinal chemistry. Before crystal structures enabled for the observation of protein-ligand interaction, chemists employed within a certain sequence would identify the portions of the fragment most involved with a desired biological function by trial and error (Van Drie 2007). Changes to the molecule's other components could alter activity as long as the pharmacophore remained stable, but they frequently guaranteed that potency was maintained, with the exception of cases where extra molecular segments caused

significant disturbance. The two different kinds of compounds may have comparable biological processes if a pharmacophore is pleased by supplementary functional assemblies, analogous collections, or atoms organized in a spatially equivalent way on another scaffold. This idea can be applied more broadly. Even while 2D topology may not reveal a joint design of characteristics, the presence of necessary pharmacophoric features in the anticipated three-dimensional geometry is sufficient to produce meaningful biological action. These concepts were then established in order to quest a database of 3D assemblies for ligands that matched 3D pharmacophores. These techniques fall under the umbrella of "pharmacophore-based VS" (Kurogi and Guner 2012). A "pharmacophore" is any representation of pharmacophore properties that eliminate 3D geometry and instead make use of a number of atoms or functional groups. The 3D pharmacophore's most fundamental form is the geometric presence of several significant elements, frequently selected from aromatic rings, hydrophobic groups, hydrogen bond donors, and acceptors. In the lack of the three-dimensional (3D) structures of receptors complexed to ligands, the major biologically relevant metric that connected molecular structure to biological activity was thought to be the pharmacophore. However, as can be shown as in Table 6.2, a set of tools that capture the unique characteristics, the charges, the hydrophobic nature, and the shape can easily characterize a 3D pharmacophore in greater detail. These descriptors were used into modeling and design as part of 3D QSAR, and VS tests were conducted utilizing a range of various methodologies, from multidimensional QSAR to a simple grouping of pharmacophoric-binding components.

6.2.4 Receptor Structure-Based Methods

The availability of a target's 3D structure is necessary for receptor-based virtual screening techniques, which are also known as structure-based techniques. These techniques require the obvious molecular docking of each ligand keen on the target's binding site. This results in a projected binding mode for each chemical in the database and a measurement of the compound's degree of fit in the target-binding site. The ligands that bind to the target protein strongly are then separated from those that do not use this information. In comparison to ligand-based techniques, receptor-based approaches are more significant, particularly because target proteins' 3D structures are revealed and made available which leads to more dependable and accurate outcomes. The Receptor-based Virtual Screening process consists of following computational steps, including:

(a) Target selection and database preparation
(b) Ligand selection
(c) Docking
(d) Postprocessing stage

Table 6.2 Tools available for pharmacophore identification and ligand-based design (Cheeseright et al. 2008)

Sl. No.	Tool	Pharmacophore identification programs and resources
1	GALAHALD	Genetic algorithm with linear assignment for hypermolecular alignment of datasets (GALAHALD), Tripos, Inc. http:/www.tripos.com/data/SYBL/ GALAHALD_9-7-05.pdf
2	DISCO and DISCO tech	DIStance Comparison for multiple pharmacophores generations. Based on clique detection. The conformational search is separated. Tripos, Inc.www.tripos.com/data/SYBL/DISCOTech-072505.pdf
3	GASP	Genetic algorithm Similarity Program. A flexible genetic algorithm. The pharmacophoric features are defined by the SYBYL package. Tripos, Inc.www.tripos.com/data/SBYL/GASP-072505.pdf
4	ROCS method	A shape-based method developed by Open Eye (www.eyesopen.com) involves converting a single molecule in a potentially bioactive conformation into a series of Gaussian grid functions that represent shape or atomic character. This information is then compared to similar data from a precomputed database of stored conformations using a scaled similarity function that measures shape overlap or atomic character similarity. This approach stands out for its quickness, logical command-line interface, parallelization, and reliable performance with various ligand classes.
5	Catalyst	An integrated setting for managing databases and running queries for medicine distribution. It essentially functions as a VS tool against a database that has a precomputed conformational expansion for each ligand. Every chemical has several stored conformations, each of which can be used to create a 3D pharmacophore using hydrophobic, hydrogen bonding, and potentially positively and negatively ionizable functional groups. Software from Chemical Computing Group and Schrodinger for computational chemistry also offers useful variations of Catalyst-like capabilities. www.accelrys.com/products/catalyst/catalystproducts/cathypo.html Accelrys, Inc.
6	Surflex-Sim method	An array of "observer" points that describe the local nature of the surface and any potential interactions surround each molecule. There will be a shared subset of comparable observer points between two related molecules. When the disparities in pharmacophore character and molecular surface between two molecules that may be deduced from observer points are minimized, this is known as an ideal alignment. To speed up the algorithm, large molecules can be fragmented into parts which are then compared, and then tested for consistency. This feature also makes the method capable of identifying alignments when one molecule is much smaller than the other.

(continued)

Table 6.2 (continued)

Sl. No.	Tool	Pharmacophore identification programs and resources
7	UNITY package from Tripos Inc	Additionally, the user must be able to recognize the spatial arrangement and pharmacophore properties. This can be used to concentrate on a small number of features or to omit a certain volume from the molecule when there are several chemicals present and biological activity is known. The database's chemicals are then compared to the results of the Lead Discovery query, using a versatile-guided tweak method with the Virtual Screening 99 pharmacophore. To get a decent recall of actives, it is frequently required in practice to tune tolerances and characteristics. Before doing a thorough database search, validation using known actives against a limited, diverse background of inactives is frequently advised.
8	Ligand Scout from Inte:Ligand,	An additional technique that extracts pharmacophores from a protein crystallographic complex (www.inteligand.com). Hydrophobes, normals to aromatic rings, hydrogen bond donors and acceptors (including extension points), and other limited pharmacophore characteristics are included in this method. In actual reality, the suspected or known binding sites have been transformed into pharmacophore search queries, and then the pharmacophore data have been transmitted to software like catalyst or MOE. It successfully reproduces pertinent binding mechanisms in validation investigations.

Figure 6.2 displays a typical schematic representation of receptor-based virtual screening. A variety of computational methods that will be thoroughly covered in the following sections must all be implemented correctly for this workflow's many stages to function properly. (Budzik et al. 2010).

(a) Target Selection and Database Preparation

The virtual screening campaign's first phase, target identification, is crucial for the success of the drug development process. Polysaccharides, lipids, nucleic acids, and proteins are the macromolecules which are targeted with small-molecule compounds. Proteins, and within them enzymes, are always the preferable one as they possess excellent binding pocket properties, which enable high specificity, potency, and low toxicity. The Protein Data Bank is the foremost source for experimentally verified 3D structures of large biological molecules. So, the first step in acquiring a protein 3D structure for a VS operation is to use this database. Once a protein's 3D structure has been established, its druggability score can be calculated. Druggability is the capacity of a receptor to bind substances with drug-like properties. The determining factor in this process is the ability of the molecule to interact well with a particular pocket or cleft in the protein. Finding these binding sites is straightforward when a ligand and the target protein have been explicitly co-crystallized. If this kind of information is unavailable, it may be challenging to pinpoint the binding site's

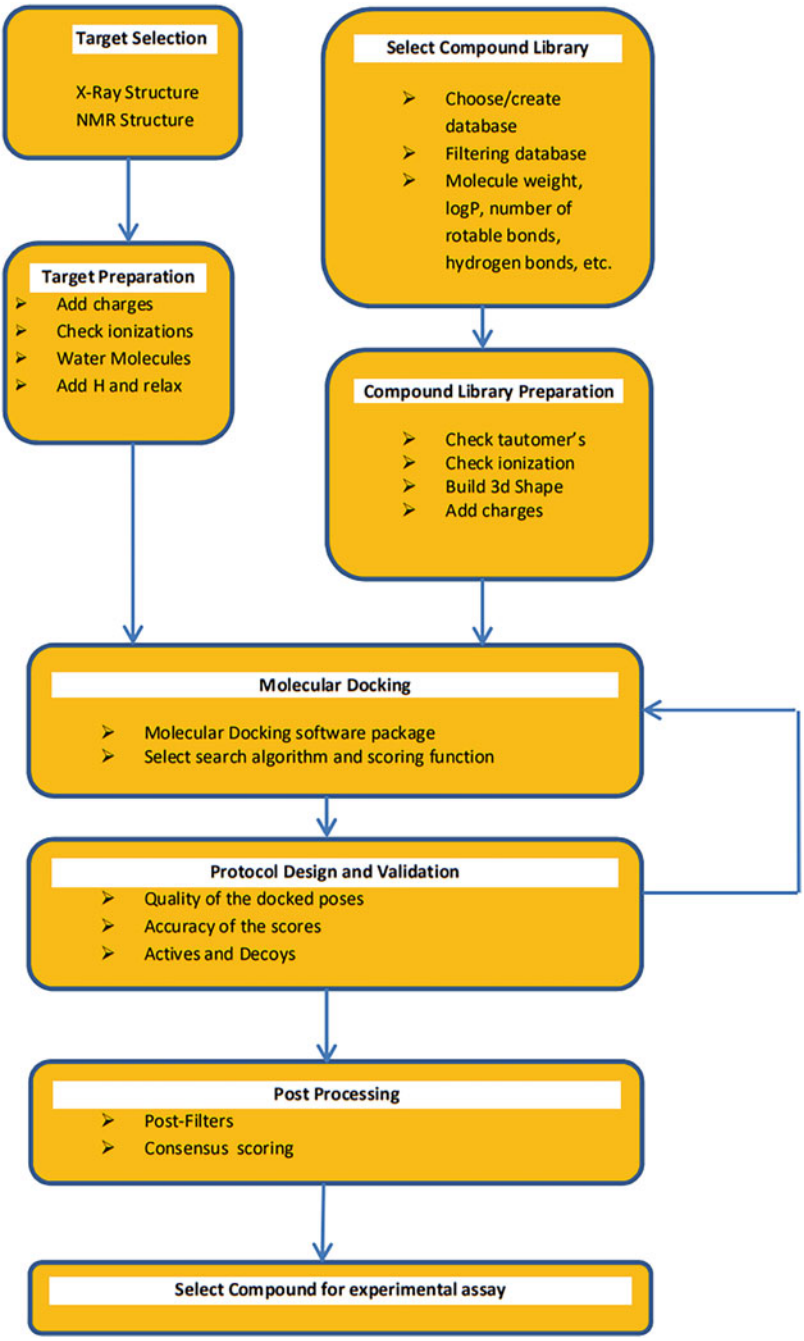


Fig. 6.2 Schematic representation of receptor-based drug designing

position. Under these circumstances, it is possible to categorize potential binding locations using computational approaches. In order to find binding pockets, several computational methods, such as POCKET, LIGSITE, SURFNET, SPHGEN, FPOCKET, etc., primarily rely on geometric features. These algorithms detect and rank probable binding sites based on the energy of probes interacting with those sites. Geometry-based algorithms are usually preferred because they are rapid and trustworthy in managing structural changes or missing atoms or residues in the input structure (Schmidtke and Barril 2010). Energy-based algorithms, on the other hand, are frequently more accurate and sensitive. Once the target has been identified and the most druggable binding site has been chosen, the target must be prepared for docking. Target preparation often involves withdrawing solvent and ligand molecules, introducing hydrogen atoms, establishing bond ordering and formal charges, capping chain termini, and defining the protonation states of amino acids (atom types). It might also be necessary to improve the crystallographic structure and identify the binding site's flexible regions. Although it is commonly overlooked, target preparation may have a big impact on virtual screening enrichment (Kortagere and Ekins 2010).

(b) Ligand Selection

A large number of compounds can be tested in a shorter amount of time because of virtual screening. The number of compounds that must be evaluated must be limited in some way because it is obviously impossible to quickly screen the entire chemical space for a single target. We need to filter out some molecules beforehand in order to create a workable library, which may be done using a variety of techniques.

Assembling all the structures we want to evaluate is required before beginning a new virtual screening campaign. Typically, a library of possible ligands must be constructed before docking; this library may have thousands of different compounds that will eventually be evaluated. Many databases of chemical structures have been created recently. These databases not only contain information about these compounds' structures, but also a plethora of pertinent chemical and biological data. ZINC is an example of such database which is free containing approximately 35 million buyable structures, including more than 4.5 million clean leads, is one of the most widely used compound databases. Other databases include PubChem, ChemDB, and ChemSpider, which together contain more than 32 million chemicals (63 million unique structures). Additionally, all significant pharmaceutical firms have internal corporate libraries that contain millions of chemicals (Abdo et al. 2010).

Scientists have developed the idea of drug-likeness to determine whether a new molecule is a viable candidate to become a drug or not which means that a new molecule must possess some of the traits shared by the vast majority of medications in practice, traits that are connected to the compound's bioavailability after injection. Right now, there are a number of techniques available for determining how drug-like a certain molecule is, allowing us to narrow down our database. There are functional group filters and basic counting techniques among them.

6.2.5 *Counting Methods*

Counting techniques make it simple to narrow the search field. The partition coefficient (log P), molecular weight, and hydrogen bonding groups, which are all associated with bioavailability, are properties that are taken into account by counting techniques. The candidates which fulfill the above criteria have greater chances to pass clinical trials. One of the most famous counting methods is Rule of Five which represents a set of rules that can be used to assess the drug-likeliness of a compound, which are:

- Hydrogen bond donors ≥ 5
- Hydrogen bond acceptors ≥ 10
- Molecular weight > 500 Da
- Log $P > 5$

The fact that all numbers are multiples of 5 is where the name "Rule of Five" comes from. The majority of these exceptions involve antibiotics, vitamins, antifungals, and cardiac glycosides, though there are seem to be quite a few circumstances in which they do not seem to apply. Veber et al. further suggested that to improve oral bioavailability, the number of rotatable 360 bonds should be less than 7 (Veber et al. 2002).

6.3 Functional Group Filters

Functional group filters are grounded on the notion that some functional assemblies are inappropriate for use in drugs because they are either poisonous to organisms or extremely reactive. Functional groups that are known to harm the organism, such as those that have teratogenic or mutagenesis potential, are frequently eliminated since the molecules they produce are typically also toxic. Furthermore, particular reactive groups, such as metals, alkyl-bromides, etc., can result in false positives. By concentrating just on compounds that are expected to be good leads, we can save our valuable time.

6.4 Docking

The next step of RBVS, i.e., molecular docking demands the most processing resources and time in the VS that is why it is observed as the core of most of virtual screening campaign. Molecular docking can begin once the target protein and a database of substances have been chosen. In this stage of the computational technique, it is possible to foretell the preferred posture and conformation of one molecule (a ligand) in relation to another when their binding results in the

development of a stable complex. The binding affinity between the receptor and ligand can then be predicted using the molecule's preferred orientation in reference to the receptor.

The number of molecular docking applications available for use in virtual screening campaigns are now fairly considerable and growing. Each one of them involves looking for the ligand's preferred alignments with the receptor. The search algorithm and the scoring function are two different sorts of algorithms that can be used to achieve these poses/conformations. The search method generates the many configurations and orientations that could be used to fit the ligand into the receptor's binding pocket. The scoring function assigns a score to each of the many ligand postures and positions that the search algorithm produces. The best-scoring results correspond to a real-binding conformation which should be practically near to value which is observed experimentally.

Postprocessing Stage This stage consists of choosing which compounds from the library database should move on to the experimental testing after they have all docked into the binding pocket of the drug target. The simplest method is to simply rank the compounds according to the values of the scoring function that is straight integrated into the docking algorithm, and then select the compounds with the highest scores for experimental testing (Cerqueira et al. 2015).

6.5 Other Tools

6.5.1 Fragment-Based Virtual Screening (FBVS)

Many pharmaceutical researchers watched at a rival's patent chemical and considered for leads that preserve the rival's molecule's action while still being sufficiently different not to overstep on the rival's patent. It is common practice in these situations to substitute isosteric equivalents for molecular fragments. Only fragments and substructures are included in the scope of this FBVS discussion; fragments with a molecular weight of five atoms or less are excluded. Due to the demanding need of research and development in pharmaceutical world leads resulting from more than one chemical class for a given target, researchers sometimes put efforts to imitate their own compounds with suitably divergent scaffolds to assist as a standby in case of unpredicted letdown of the lead entrant in the clinic trials. Fragment-based drug development is not new because computational chemistry frequently involves computing the properties of molecules using the attributes of their constituents. Strong evidence has emerged from recent studies supporting the notion that strong, ligand-efficient binding fragments can be used to build larger, highly affine ligands from weaker, ligand-inefficient ones, as long as the additional fragments are carefully chosen to prevent noticeably impairing ligand efficiency (Congreve et al. 2008).

6.5.2 Text-Mining Techniques

All information about molecules can be discovered in publications that are mostly text-based, but for now, even the most sophisticated text-mining algorithm cannot find anything for us unless the molecular query is a basic phrase like glucose or pyrrolidine. This would seem to be an impossibility to a modern scientist if one did not stop to reflect that the question would have been difficult for the average layman to understand just 10 years ago. Text mining and natural language processing (NLP) were not what they are now 10 years ago. The bits and bytes used to hold molecular information adopt a format that is significantly different from regular language and cannot be quickly deciphered without appropriate notice, despite the fact that to a computer scientist VS appears to be just another type of text mining. Modern texts do contain some nonnative language content, but it is properly indicated and does not interfere with the NLP program's functionality. The material to be searched, the algorithms used, and the retrieval procedures are all specialized for structure perception and manipulation in the case of chemical structures, even though the information is still saved and processed as bits and bytes. This restriction results from the fact that molecular information is difficult to understand when expressed in natural language, and that when it is, as in patents, it is so cryptic that very few people attempt to read and decipher the chemical structure or composition by reading the IUPAC name detailed in a patent. Nowadays, everyone searches for the best software translator that can change a name into a recognized molecular structure.

6.5.3 Other Techniques

There are more methods for determining the interactions between a protein and its ligand; these methods use algebraic topology and are recorded in MathDL (Imrie et al. 2018) and TopologyNet (Cang and Wei 2017). MathDL uses sophisticated mathematical techniques (such as geometry, topology, and/or graph theory) to encode the representations of the physicochemical interactions into lower-dimensional rotational and translational invariant representations. DeepBindRG (Meng Zhang, H. X., Mezei, M., & Cui, M. 2011) and DeepVS (Pereira et al. 2016) focus on the complex's interacting atom environments using atom pair and atom context encodings, with R between 0.5 and 0.6 and an RMSE for a particular protein ranging from 1.6 to 1.8.

6.6 Role of Virtual Screening in Drug Discovery

6.6.1 Case Study 1

The study by Agrawal and colleagues illustrates crucial elements in carrying out SBVS. They investigated for E. coli DNA primase inhibitors. Three potential-binding sites were found with the help of the GRID program. To extract the desired molecules from the database, a number of filters were applied, yielding roughly 500,000 molecules. The top 2500 compounds were obtained using glide docking. There were 79 inhibitors on the short list, 68 of which could be bought. Of these, four inhibitors had an IC50 of less than 100 mM for inhibiting primase (Agrawal et al. 2007).

6.6.2 Case Study 2

A study highlights the benefit of employing a pharmacophore isolated from a protein complex's binding site. Tervo and colleagues developed two pharmacophore hypotheses based on the docking of three well-known sirtuin-2 histone deacetylase inhibitors using the UNITY software. The libraries of Maybridge and LeadQuest were flexible searched, yielding 34 compounds and by using the Volsurf permeability model, these were condensed to 32 compounds. Eleven molecules were purchased after more investigation followed by *in vitro* investigation. Out of 11 molecules, four molecules showed IC50 inhibition of less than 200 μM (Tervo et al. 2006).

6.7 Conclusion

It is widely acknowledged that finding new drugs and developing them need time and resources. To speed up drug discovery, design, development, and optimization, there is a rising push to apply computational power to the combined chemical and biological domain. CADD is being used in the pharmaceutical business to speed up hit identification, hit-to-lead selection, improve the absorption, distribution, metabolism, excretion, and toxicity profile, and prevent safety concerns. It is no longer debatable how computational techniques like VS contribute to the development of new drugs. Computational design tools are used by all of the major pharmaceutical and biotechnology businesses in the world. CADD is currently predicted to make up 10% of pharmaceutical R & D spending and will reach 20% by 2016. There are numerous successful studies where CADD, and especially VS, helped develop new drugs. The usefulness and influence of VS are still limited by significant scientific problems that have not yet been fully overcome, despite the commonly presented

extremely positive picture. First of all, there are problems with erroneous posture rating, binding energy estimations, and similarity-based compound rankings, all of which need for a time-consuming follow-up analysis in order to select possible leads based on knowledge or intuition. To make VS the main stage in drug development, it is still necessary to show that the components on which it depends are precise and reproducible. This can be done either by developing new virtual screening methods or by carefully validating existing methods in experiments. The interplay between computational modeling and experimental research is therefore a pivotal phase where the contributions from each of these domains are essential for their mutual progress. Despite these limitations, VS is still the most time- and resource-consuming, cost-effective method for exploring a wide range of chemical possibilities since it allows access to a large number of possible ligands, the bulk of which are easily accessible for purchase and subsequent testing. Because of the increasing number of targets identified by genome and proteomics as well as improved technique that can anticipate higher hit rates and better forecasts of geometries, VS methods will soon play an even more dominating role in drug design.

References

Abdo A, Chen B, Mueller C, Salim N, Willett P (2010) Ligand-based virtual screening using Bayesian networks. J Chem Inf Model 50(6):1012–1020. https://doi.org/10.1021/ci100090p

Agrawal H, Kumar A, Bal NC, Siddiqi MI, Arora A (2007) Ligand-based virtual screening and biological evaluation of inhibitors of chorismate mutase (Rv1885c) from Mycobacterium tuberculosis H37Rv. Bioorg Med Chem Lett 17(1F1):3053–3058. https://doi.org/10.1016/j.bmcl.2007.03.053

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

Bajorath J (2002) Virtual screening in drug discovery: methods, expectations and reality. Curr Drug Discov: 24–28

Berdigaliyev N, Aljofan M (2020) An overview of drug discovery and development. Future Med Chem 12(10):939–947. https://doi.org/10.4155/fmc-2019-0307

Bradley EK, Beroza P, Penzotti JE, Grootenhuis PDJ, Spellmeyer DC, Miller JL (2000) A rapid computational method for lead evolution: description and application to α1-adrenergic antagonists. J Med Chem 43(14):2770–2774. https://doi.org/10.1021/jm990578n

Budzik B, Garzya V, Shi D, Walker G, Woolley-Roberts M, Pardoe J, Lucas A, Tehan B, Rivero RA, Langmead CJ, Watson J, Wu Z, Forbes IT, Jin J (2010) Novel N-substituted benzimidazolones as potent, selective, CNS-penetrant, and orally active M1 mAChR agonists. ACS Med Chem Lett 1:244–248

Cang Z, Wei G (2017) TopologyNet: topology-based deep convolutional and multi-task neural networks for biomolecular property predictions. PLoS Comput Biol 13(7):1–27. https://doi.org/10.1371/journal.pcbi.1005690

Cerqueira NMFSA, Gesto D, Oliveira EF, Santos-Martins D, Brás NF, Sousa SF, Fernandes PA, Ramos MJ (2015) Receptor-based virtual screening protocol for drug discovery. Arch Biochem Biophys 582(June):56–67. https://doi.org/10.1016/j.abb.2015.05.011

Cheeseright T, Mackey M, Rose S, Vinter A (2008) Rapid discovery of new leads for difficult targets: application to CCK2 and 11beta-HSD1. Chem Cent J 2(S1):10–11. https://doi.org/10.1186/1752-153x-2-s1-p32

Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent developments in fragment-based drug discovery. J Med Chem 51(13):3661–3680. https://doi.org/10.1021/jm8000373

Hawkins PCD, Skillman AG, Nicholls A (2007) Comparison of shape-matching and docking as virtual screening tools. J Med Chem 50(1):74–82. https://doi.org/10.1021/jm0603365

Hert J, Willett P, Wilton DJ, Acklin P, Azzaoui K, Jacoby E, Schuffenhauer A (2005) Enhancing the effectiveness of similarity-based virtual screening using nearest-neighbor information. J Med Chem 48(22):7049–7054. https://doi.org/10.1021/jm050316n

Imrie F, Bradley AR, van der Schaar M, Deane CM (2018) Protein family-specific models using deep neural networks and transfer learning improve virtual screening and highlight the need for more data. J Chem Inf Model 58(11):2319–2330. https://doi.org/10.1021/acs.jcim.8b00350

Kortagere S, Ekins S (2010) Troubleshooting computational methods in drug discovery. J Pharmacol Toxicol Methods 61(2):67–75. https://doi.org/10.1016/j.vascn.2010.02.005

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161(2):269–288. https://doi.org/10.1016/0022-2836(82)90153-X

Kurogi Y, Guner O (2012) Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. Curr Med Chem 8(9):1035–1055. https://doi.org/10.2174/0929867013372481

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

Lionta E, Spyrou G, Vassilatis D, Cournia Z (2014) Structure-based virtual screening for drug discovery: principles, applications and recent advances. Curr Top Med Chem 14(16):1923–1938. https://doi.org/10.2174/1568026614666140929124445

Martin YC, Kofron JL, Traphagen LM (2002) Do structurally similar molecules have similar biological activity? J Med Chem 45(19):4350–4358. https://doi.org/10.1021/jm020155c

Meng Zhang HX, Mezei M, Cui MXY (2011) Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des 7(2):146–157

Nunes RR, Da Fonseca AL, Pinto ACDS, Maia EHB, Da Silva AM, Varotti FDP, Taranto AG (2019) Brazilian malaria molecular targets (BraMMT): selected receptors for virtual high-throughput screening experiments. Mem Inst Oswaldo Cruz 114(2):1–10. https://doi.org/10.1590/0074-02760180465

Pereira JC, Caffarena ER, Dos Santos CN (2016) Boosting docking-based virtual screening with deep learning. J Chem Inf Model 56(12):2495–2506. https://doi.org/10.1021/acs.jcim.6b00355

Schmidtke P, Barril X (2010) Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. J Med Chem 53:5858–5867. https://doi.org/10.1021/jm100574m

Scior T, Bender A, Tresadern G, Medina-Franco JL, Martínez-Mayorga K, Langer T, Cuanalo-Contreras K, Agrafiotis DK (2012) Recognizing pitfalls in virtual screening: a critical review. J Chem Inf Model 52(4):867–881. https://doi.org/10.1021/ci200528d

Tang YT, Marshall GR (2011) In: Satyanarayanajois SD (ed) Virtual screening for lead discovery BT - drug design and discovery: methods and protocols. Humana Press, Totowa, NJ, pp 1–22

Tervo AJ, Suuronen T, Kyrylenko S, Kuusisto E, Kiviranta PH, Salminen A, Leppänen J, Poso A (2006) Discovering inhibitors of human sirtuin type 2: novel structural scaffolds. J Med Chem 49(24):7239–7241. https://doi.org/10.1021/jm060686r

Van Drie JH (2007) Computer-aided drug design: the next 20 years. J Comput Aided Mol Des 21(10):591–601. https://doi.org/10.1007/s10822-007-9142-y

Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45(12):2615–2623. https://doi.org/10.1021/jm020017n

Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49(20):5912–5931. https://doi.org/10.1021/jm050362n

Xue L, Stahura FL, Godden JW, Bajorath J (2001) Mini-fingerprints detect similar activity of receptor ligands previously recognized only by three-dimensional pharmacophore-based methods. J Chem Inf Comput Sci 41(2):394–401. https://doi.org/10.1021/ci000305x

Yadav V, Reang J, Vinita, Tonk RK (2022) Chapter 3: Ligand-based drug design (LBDD). In: Rudrapal M, Egbuna CBT-CADD (CADD): FL-BM to S-BA (eds) Drug discovery update. Elsevier, Amsterdam, pp 57–99

Yadav V, Tonk KR, Khatri R (2021) Molecular docking, 3D-QSAR, fingerprint-based 2D-QSAR, analysis of pyrimidine, and analogs of ALK (anaplastic lymphoma kinase) inhibitors as an anticancer agent. Lett Drug Des Discov 18:509–521

Chapter 7
Target-Based Screening for Lead Discovery

Monalisa Kesh and Sachin Goel

7.1 Introduction

7.1.1 Target-Based Drug Discovery Strategy

Target-based drug discovery, also known as TBDD, has emerged as the primary strategy used by several pharmaceutical industries over the past few decades as a result of the tremendous advancements in molecular biology, recombinant technology, and genomics. Despite growing investment from pharmaceutical firms, the strategy of target-based drug discovery has not resulted in any simultaneous rise in the number of novel molecular components and discovery of certain biological products, despite being beneficial in throughput and costs. A gene or a gene product or a molecular process that was able to be identified through genetic research or other biological research or investigations serves as the initial molecular target for TBDD (Hughes et al. 2011). It is customary to use genetic and molecular biology techniques to pinpoint the genes responsible for specific diseases, and with the advent of more affordable and rapid sequencing technologies, findings from wide-ranging enterprises have also been extensively used to increase the pool of the putative molecular targets. After that, genes are expressed via recombinant technology in less complex organisms like yeast, allowing high-throughput screening against massive chemical libraries to find "hits"—a small compound and/or a biologic that can have interactions with the desired target. Target identification and selection, target verification, and assay development, followed by hit detection, lead optimization, and preclinical

M. Kesh
Amity Institute of Biotechnology, Amity University Uttar Pradesh, Sector 125, Noida, India

S. Goel (✉)
Department of Biological Sciences and Engineering, Netaji Subhas University of Technology, New Delhi, India

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_7

and clinical development, are all part of the process (Clark et al. 2010; Lindsay 2003). It is crucial to find a target, which is "druggable." Any target is said to be so if a therapy, either it is a small-molecule therapeutic or a biologic, may change its activity, function, or behavior. Some examples of biological targets include both proteins and nucleic acids. If a target is known to affect how a disease manifests or is implicated in its etiology, it can be regarded as a prospective pharmacological target. The 3D framework of a such target should be accessible to evaluate druggability, and also, the target's expression should not be uniformly distributed throughout the body, but its toxicity profile should appear encouraging. Because the target is readily "assayable," high-throughput screening is possible. Lastly, in application to pharmaceutical firms, the status of the proposed target's intellectual property (IP) should be good (Gashaw et al. 2011). Target-based drug discovery strategy inputs include the target biomolecules' 3D structural data, interactions with ligands, inhibitors, and/or modulators, as well as the nature, characteristic properties, and shape and size of their binding pockets. The target protein, such as an enzyme, may be any number of biomolecules, such as cell surface or intracellular receptors, ion channels, pumps, and transcriptional regulators, each with a clearly defined role in the biological milieu. Designing novel compounds can benefit from the spatial information of different intermolecular interactions. Here, the end result of testing an idea in various procedures put together to suit the new needs of particular well-defined requirements of a mandate is a lead or potential therapeutic molecule. An unsolvable structure, a need for fresh proof-of-concept treatments, the exploration of new material and idea frontiers, or any other conundrum may be the mandate. The target-based drug discovery methods can actually be thought of as a collection of computational techniques, including visualization tools, put together to create decision support systems for the drug designing and discovery process (Anderson 2003). They are considered direct approaches because they include inputs from the target protein structure into the drug design/discovery process. The focus point or central idea of the adopted approaches in these investigations involves the 3D framework of the target. Numerous pharmaceutical research units have included them in their regular operating processes due to their resource and cost-effectiveness. The approach to be taken to accomplish the goal is determined by the task at hand and the amount of information available to the target/ligand. In light of this, suitable modeling and simulation tools are chosen to produce a result, such as a proposed protein or small-molecule chemical structure, which might be used as an input to the subsequent stage of operation. For instance, a drug development process that begins with a study of disease genesis will go on to identify a target, determine that target's 3D structure, find that target's ligand-binding site, etc. Additionally, there are a ton of leads and potential medications from the virtual routes waiting to enter wet labs. Without bias, a lot of the tasks in target-based drug discovery can be accomplished even with low resources.

7.1.2 Types of Drug Discovery Approaches

The identification of more molecular targets has been made easier by developments in combinatorial chemistry and molecular biology, which has necessitated the creation of innovative screening techniques (Berg 2010). The following are a few of the several screening methods applied in the drug discovery procedure. It is crucial to consider the techniques' key advancements and difficulties.

7.1.2.1 High-Throughput Screening and High-Content Screening

In the initial phases of drug development and designing, high-throughput screening or HTS is widely used. HTS is used to locate hit compounds displaying action averse to a chosen target, from large libraries of compounds that can comprise thousands of molecules (Hevener et al. 2018; Martis 2010). A hit molecule is confirmed and developed into a lead chemical compound with greater selectivity and potency when it is found. From there, more research can be done to identify a potential therapeutic candidate for preclinical testing. To locate, track, and measure the events in HTS, robotics, liquid/microplate handling devices, and microplate readers are utilized. In addition, specialized software is required for instrumentation control and data processing (Fernandes et al. 2009). HTS is an important tool; however, it may be difficult to evaluate drug attributes like toxicity and bioavailability. HTS is mostly used to support lead optimization; it can be thought of as a brief scan of biological organisms that enables the rapid exclusion of candidates with weak or negligible effects (Armstrong 1999). High-content screening (HCS), a method that was first created to support HTS, has become quite popular in recent years. HCS uses cellular imaging and high-throughput methods to efficiently gather quantitative data from intricate biological systems. The term "high-content screening" was first used to emphasize the complicated subcellular morphological and intensity-based readings that allow studying variations in a cell population as opposed to a single population-averaged readout per perturbation in the 1990s, when high-throughput tests using multi-tier plates and automated fluorescence microscopy came together. Multiple characteristics of distinct cells or species can be explored simultaneously with HCS. Data are extracted from cell populations using a combination of automated microscopy, image processing, and visualization technologies. HCS typically involves high-throughput fluorescence imaging of the samples and generates quantitative data on a variety of details, including the geographical distribution of targets and the morphology of individual cells and organelles. Preclinical drug discovery throughput is increased through high-content screening. In contrast to biochemical assays, screening small molecules, various natural products, and genetic or authorized medication libraries in a monolayer cell culture format enable the testing of hundreds of perturbations in a single experimental study while maintaining physiological relevance. Currently, traditional or two-dimensional (2D) tissue culturing is used for the majority of HCS. However, research into 3D models of cell culture is

also ongoing. 3D models are poised to revolutionize the HCS space by boosting physiological relevance. Spheroids and organoids are used in 3D high-content screening, which enhances the physiological significance and frequently seeks to lower the high rate of attrition in drug development experiments. It is actually possible to screen up to approximately more than 9000 compounds every day by the use of HTS procedures.

7.1.2.2 Fragment-Based Drug Discovery

Another well-known method for finding new drugs is called fragment-based drug discovery strategy or FBDD. This strategy of drug discovery makes use of smaller pools with hundreds of low-complexity compounds, or "fragments," as opposed to HTS campaigns, which use larger libraries of complex compounds to be screened (Chen et al. 2017). Fragment-based drug discovery takes less money for research than HTS. By employing the fragments, the complexity of the substances being screened is decreased, enabling a deeper examination of the target's binding site. A fantastic place to start when designing lead compounds with higher ligand efficiency is with a fragment-based drug discovery strategy. Thus, FBDD offers a bottom-up approach that enables the exploration of a greater area and the production of lead compounds with higher affinity and higher specificity. A variety of biophysical methods are used in FBDD for screening. The next step is to structurally characterize fragment binding using X-ray crystallography or even NMR spectroscopy. A high-throughput process that uses X-ray crystallography to simultaneously screen individually soaked fragments is the most recent advancement in the FBDD approach. It is believed that the use of AI, like that of deep learning, will speed up the increment of fragment hits. Artificial intelligence can optimize fragment hits while taking into account crucial elements like ADMET characteristics, solubility parameters, biological activity, and synthetic viability. Different studies utilizing fragment-based screening have highlighted its relevance and importance (Lu et al. 2021).

7.1.2.3 Virtual Screening

A method used *in silico* to find potential bioactive drug candidates is called virtual screening or VS. Virtual screening methodologies employ computational techniques to automatically scan massive datasets of recognized 3D structures (Patrick Walters et al. 1998). Two basic strategies are used when using VS tool: One is based on protein and ligand interactions, and the other is centered on the molecular homology principle. For their applications, all that is required is a virtual compound database and a starting point, the protein structure of the receptor or target molecule and for the latter, at least one known active chemical. Because the system's data are limited in these early stages of drug discovery, they are ideal instruments. According to reports, virtual screening procedures are a superior alternative to HTS because they increase the likelihood of locating the best outcome from a sizable virtual database.

Additionally, VS is seen as a more affordable way to identify compounds than physical alternatives that require screening enormous libraries of compounds. This is because it is a computer-based screening methodology. Only the best promising molecules are created after using virtual screening to discover the most promising hits that can bind to the target. VS can also be used to find hazardous substances or those with negative pharmacological and pharmacokinetic characteristics. Recent years have observed a significant spike in the number of new methods and software programs that can be used in this strategy. The most notable developments in this area are influenced by technological advancements. One would single out the techniques that mix structure and ligand-based approaches as being among the most significant innovations. A reasonable computational cost is added to the search process as a result. Additionally, given the explosive growth of big data, it is important to combine machine learning techniques with tried-and-true VS tactics. They have frequently demonstrated an effective method for dealing with crucial biological features from numerous compound databases for medication designers. Their findings must still be carefully examined though (Bhunia et al. 2021; Kim 2016).

7.2 Structure-Based Drug Designing (SBDD)

Drug design, often referred to as "rational drug design" or just rational design, is the process of developing innovative medications based on the knowledge of a target organism (Klebe 2000). One such example of a biomolecule whose activity is induced or hindered by the medication and assists the patient therapeutically is a protein. Designing compounds that engage with and attach to biomolecular targets complementing one another in terms of configuration and charge is the essence of drug design. Drug design often uses computer simulation techniques. It is also known as computer-aided drug design or CADD. Last but not least, drug design that depends on understanding the 3D framework of the desired macromolecular target is actually SBDD.

Based on the above discussions, the steps involved in structure-based drug designing (SBDD) are as follows:

1. Target structure preparation: The protein structure was produced by the addition of certain polar hydrogen atoms and the Kollman charge to it after water and heteroatoms were removed (Fig. 7.1).
2. Recognition of the active site/binding site of the ligand.
3. Preparation of the library of compounds.
4. Molecular docking and scoring functions.
5. Molecular dynamic (MD) simulations.
6. Calculation of binding free energies.

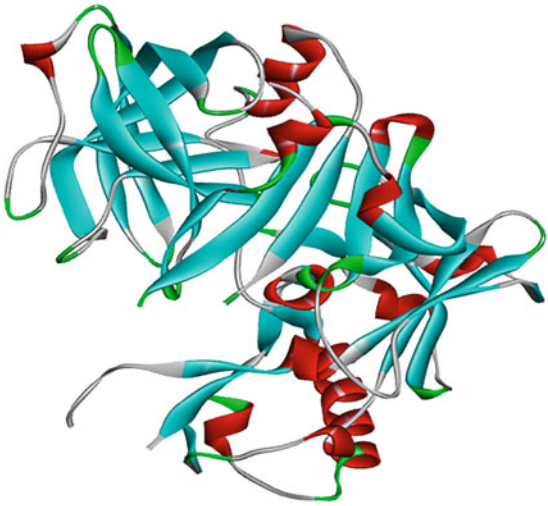


Fig. 7.1 3 D structure–BACE 1, having PDB ID: 6EJ2

7.2.1 Homology Modeling

This modeling technique is a computer approach for foretelling the 3D formation of a protein from its amino acid sequence (Cavasotto and Phatak 2009; Muhammed and Aki-Yalcin 2019) and is mostly accurate in doing so. It consists of several straightforward and basic stages. There are several software packages and services available that are utilized for homology modeling. A single modeling program or server cannot be said to be superior in every manner to others. Improving the grade of homology modeling is essential since the functioning of the model relies on the caliber of the created protein 3D structure.

Numerous uses for homology modeling exist, including drug searching. Drug development depends on the 3D determination of protein structures since medications interact with receptors, which are primarily made up of proteins in their structure. As a result, using 3D structures of proteins created using homology modeling, protein interactions have been clarified. This aids in the discovery of fresh medication prospects. Homology modeling is crucial for accelerating, simplifying, lowering the cost, and improving the utility of drug development (França 2015). The following are the steps in homology modeling:

1. Template recognition and initial alignment.
2. Alignment correction/multiple sequence alignment.
3. Generation of backbone.
4. Loop modeling.
5. Side-chain modeling.
6. Optimization of the model.
7. Validation of the model.

7.2.2 Rational Drug Design (Role of SBDD)

The technique of preparing innovative medications based on the knowledge of a target organism is referred to as rational drug design (e.g., protein and nucleic acid) (Reddy and Parrill 1999). By relying on a preexisting understanding of the structure, function, and mechanism of the target, it skips screening thousands of compounds at random (Mandal et al. 2009). CADD is crucial for supporting pharmaceutical chemists in the entire phases of developing a therapeutic candidate (Ramírez 2016).

7.2.3 Approaches of SBDD

SBDD depends on using methods like X-ray crystallography or NMR spectroscopy to ascertain the three-dimensional geometry of the biological target. If the experimental framework of the target is unknown, one can still construct a comparative model of the desired target based on the theoretical protein structure that is similar to the target. Prospective medications that are anticipated to interact with the biological target with close connection along with selectivity can be made using associated resolution and perception. Numerous computer strategies are alternative approaches for generating prospective innovative medicine options.

There are essentially three major categories that can be used to categorize current approaches to structure-based drug creation. The first technique involves finding novel ligands for a particular receptor by searching through massive databases of small-molecule 3D structures for those that match the binding pocket of the receptor using quick approximation docking algorithms. This method is known as virtual screening. A second category includes new ligands that are created from scratch. By piecemeal assembly, ligand molecules are generated in this technique inside the confines of the binding pocket. These fragments could be single atoms or bits of molecules. The main benefit of such a strategy is the ability to suggest novel structures that are not found in any database. The third technique involves improving existing ligands by testing potential analogs inside the binding cavity. Structure-based drug design (SBDD) can be differentiated roughly into two types (Wilson and Lill 2011):

1. Ligand-based drug design or database searching
2. Receptor-based drug design

7.2.3.1 Ligand-Based Drug Design

Understanding the compounds attaching to the chosen targets has been necessary for indirect drug design, sometimes referred to as ligand-based drug design (Badalà et al. 2008). With the help of these extra molecules, one pharmacophore model that details the essential structural conditions for such a molecule to attach to the target may be

developed. In the absence of 3D knowledge about the receptor, a method known as ligand-based drug design—which focuses on understanding molecules that interact with the specific biological target—is applied. Pharmacophore modeling and 3D quantitative structure–activity interactions are the most significant and often utilized techniques in ligand-based drug design (3D QSAR). They can offer forecasting analytics that is appropriate for lead optimization and lead discovery.

Quantitative Structure–Activity Relationship (QSAR) Models

Quantitative structure–activity relationship models or QSAR models, often known as regression or classification models, are used in the domains of chemical, biological, and engineering (Roy et al. 2015). While classification QSAR models connect a set of "predictor" variables (X) to the categorical value of the response variable (Y), QSAR regression approaches, like other regression models, relate a set of "predictor" variables (X) to the intensity of the response variable (Y).

The biological activity of a chemical might be the QSAR response variable; for QSAR modeling, the predictor is the substance's physical–chemical characteristics or hypothetical molecular descriptors. First off, QSAR models summarize a chemical data set's purported link between chemical compositions and biological activity. Additionally, QSAR models project unique compound characteristics.

For instance, the quantity of a chemical needed to produce a specific biological response might be used to express biological activity. Additionally, when physicochemical properties or structures are expressed in numerical form, a mathematical relationship—also known as a quantitative structure–activity relationship—can be discovered. The mathematical equation can then be used to predict the behavior of new chemical structures if it is well-validated.

A QSAR has the form of a mathematical model:

$$\text{Activity} = f(\text{physicochemical properties and/or structural properties}) + \text{error}.$$

The error includes the model error (bias) and the observational variability, which is, even with an appropriate model, the observations that can tend to vary.

Pharmacophore Modeling

An abstracted description of a molecular property necessary for a biological macromolecule to identify a ligand is called a pharmacophore. According to IUPAC, a pharmacophore is "an assemblage of steric and electronic characteristics necessary to permit the optimal supramolecular interactions with such a given biological target and to activate (or obstruct) its biological response." The ability of ligands with varied structural characteristics to adhere to a given receptor site is demonstrated using a pharmacophore model. Pharmacophore models may also be utilized to discover new ligands that will attach to the same receptor.

7.2.3.2 Receptor-Based Drug Designing

Constructing ligands fall within the topic of structure-based drug designing techniques, which is known as RBDD. Here, ligands are generated via piecemeal assembly inside the boundaries of the binding pocket. These fragments might be single atoms or bits of molecules. The main benefit of such a strategy is the ability to suggest unique structures that are not found in any database.

Docking

Bringing two or even more molecules altogether, such as a drug and an enzyme or protein, is the subject of research of molecular docking. In order to predict how an enzyme or a protein will interact with small molecules, a technique known as docking uses molecular modeling (ligands). We used AutoDock4.2 program (Rizvi et al. 2013) for our work, with the following the below steps (Fig. 7.2):

1. Retrieving the Target.pdb files from the major protein databases.
2. Retrieving the Ligand.pdb files from the major ligand databases.
3. Preparation of the Target.pdbqt file.
4. Preparation of the Ligand.pdbqt file.
5. Preparation of the Grid Parameter File.
6. Preparation of the Docking Parameter File.

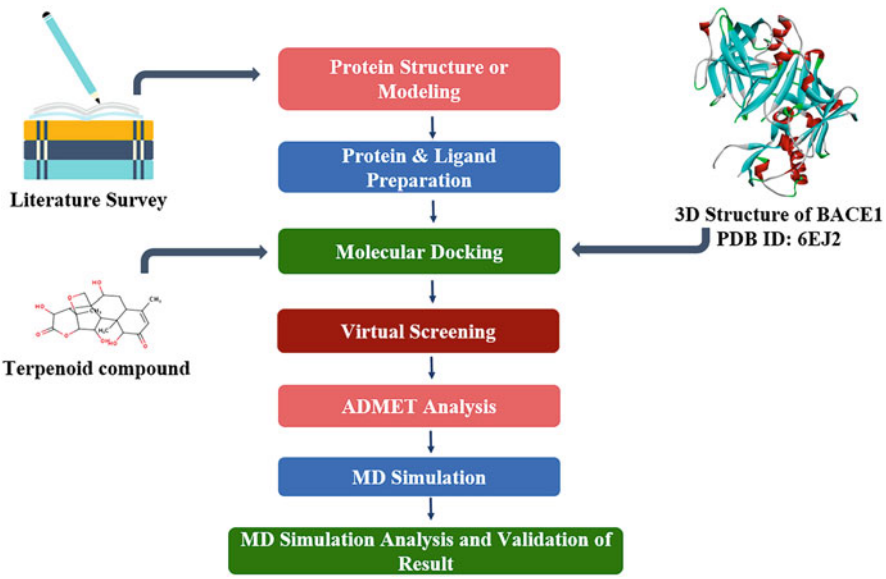


Fig. 7.2 Workflow diagram of drug discovery process

7. Using Cygwin for Molecular Docking.
8. Analyzing the results and retrieving the Ligand-Enzyme interaction complex.pdb.

7.3 Advantages, Applications, and Challenges

7.3.1 Advantages of Target-Based Approaches

One benefit of target-based strategies is that they are often quicker, simpler, and less expensive to design and implement than other methods (Zheng et al. 2013). They also commonly work with the understanding of a drug's molecular processes from an earlier stage. Drug discovery efforts can employ crystallography, computer modeling, biochemistry, binding molecular pharmacology, kinetics, genomics, and mutational analysis to illustrate how a drug interacts with a target once the target has been discovered. This paves the way for the development of effective biomarkers and the identification of new drug classes that act at the target in the future. Target-based drug development has, by some standards, been exceedingly successful. 70% of the 113 first-in-class medications approved by the US FDA between 1999 and 2013 were found via target-based drug discovery (Eder et al. 2014). The strategy is not just for finding small-molecule drugs. Target-based techniques are typically used to find antibody medicines and other protein biologics and gene therapy and also nucleic acid-based treatments are intrinsically target-based, as known. A target-based strategy has proven to be particularly effective with some target classes. Protein kinases are increasingly being targeted by a growing number of medications, particularly in cancer (Moffat et al. 2014), despite once being considered to be extremely difficult to execute. Over 30% of licensed medications target G-protein-coupled receptors or GPCRs, making them the most popular class of pharmacological target. Older GPCR-targeting medications were frequently phenotypically found based on pharmacological reactions, while target-based GPCR drug development made possible by genetics and recombinant technologies have produced better next-generation therapies. For instance, the chemical genomic characterization of previous neuropsychiatric drugs revealed the 5-HT2a serotonin receptor to be a prominent target implicated in psychosis. After that, the FDA in 2016 authorized pimavanserin, a 5-HT2a receptor inverse agonist, for the target-based therapy of Parkinson's disease psychosis (Vanover et al. 2006). The majority of peptide therapies have well-established targets, like the GLP-1 receptor, which is the target of exenatide, or the V1a vasopressin receptor, which is the target of selepressin in sepsis (Laporte et al. 2011). Phenotypic studies in tissues and animals can help assure translation to the clinic and identify unanticipated pharmacology, such as the discovery in target-based assays that selepressin reduces vascular leak in addition to vasoconstriction (Maybauer et al. 2014). Human genetics is one of the most effective methods for locating therapeutic targets. A medicine that targets a human gene or a gene product has a good likelihood of being successful when that gene or the gene product is

strongly associated with a disease. Enzyme replacement therapy has been successful in treating rare hereditary illnesses including lysosomal storage diseases, which occasionally have molecular targets discovered through genetics. Additionally, pharmacological targets for more widespread illnesses may be revealed by rare genetic conditions. Humans with gain-of-function mutations in the PCSK9 gene have unusually elevated blood LDL cholesterol and elevated chance of evolving heart disorders (Abifadel et al. 2003). From there, PCSK9-blocking therapeutic antibodies were discovered, and the FDA authorized alirocumab and evolucumab in 2015.

7.3.2 Challenges of Target-Based Approaches

Target-based drug discovery has certain downsides, one of which is the overuse of simple testing. Artificial recombinant cell-based assays frequently fail to represent the nuances of the physiological milieu of a whole organism. In reality, due to major improvements in phenotypic screening technologies including induced pluripotent stem cells, organ-on-a-chip systems, organoids, various high-content imaging techniques, and CRISPR-Cas, TBDD has experienced a setback in the pharmaceutical sector. Even when a pharmacological target is known and the mechanism is outlined, target-based drug discovery faces a hurdle since sometimes the full picture is not revealed. Drugs may target several targets, and the observed efficacy may not necessarily be attributable to the expected mechanisms. Selective serotonin reuptake inhibitors or SSRIs appear to have a more sophisticated mechanism of action than, for instance, simply increasing serotonin in synapses (Walker 2013). The clinical reaction to Sildenafil in erectile dysfunction instead of its impact on cardiovascular disease is an example of an unanticipated pharmacological effect that can occur with drugs. The complexity of life can astound one. Recombinant systems have the potential to mislead scientists into working on projects and compounds that do not produce positive clinical outcomes, which is another issue with target-based strategies. Because targets created for streamlined cell-based assays may not always act the same way as in the complex environment of entire organisms, findings from gene engineering in model animals might not even be applicable to the patients. The biology of intact animals and people is extremely complicated, and a recombinant system that is overly simplified may be unable to represent this complexity. Phenotypic assays examine substances in undamaged biological systems like cells, tissues, or animals with the goal of enhancing the translation of the medication discovery to the clinic by only causing the presence of disease-relevant traits. A biological system's ability to forecast how medications will affect human disease will, ideally, increase with reduced perturbation. The fact that medicines may affect several targets and that the reported treatment success may not necessarily be attributable to the recognized molecular mechanism presents a real difficulty for target-based drug discovery. For instance, by controlling inflammation, selective serotonin reuptake inhibitors may potentially be used to treat depression. Another rather

well-known example of the same is Sildenafil or Viagra, initially developed for heart illness but unintentionally developed pharmacological properties that also helped to treat erectile dysfunction. It will be difficult to pinpoint a single biological target for widespread disorders with significant socioeconomic repercussions like obesity and depression because they are frequently complex. The therapeutic advantage of reducing only one in several risk variables might not offset the adverse impact in situations where there may be unfavorable side effects.

7.3.3 Promising Examples of TBDD

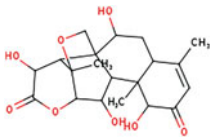
The key advantage of this strategy over the phenotypic strategy is the ease of implementation. It is usually quicker and less expensive as well. Furthermore, after a molecular target has been discovered, drug discovery researchers may completely comprehend how a drug interacts with its target using cutting-edge methods like crystallography, computer modeling, genomics, and mutational analysis. They may be able to increase their medications' biodistribution or optimize the structural–activity connection as a result. Dr. Cassandra Kennedy and her colleagues at the Francis Crick Institute have created a new photoaffinity alkyne-tagged probe to bind to MCC950, a strong small chemical inhibitor of the NLR family inflammasome pyrin domain containing 3 (NLRP3) protein that has been found to decrease inflammation in the animal models. They published their work in 2021 (Kennedy et al. 2021). According to this research, MCC950 could bind to targets that have not yet been discovered in addition to those for which it was designed to attach. If unintended disruptions to other biological pathways occur as a result of these off-target interactions, harm may result. The research team used photoaffinity labeling (PAL) and thermal protein profiling (TPP) with proteomics to identify and confirm carbonic anhydrase 2 or CA2 as a non-target protein for MCC950 at biologically relevant doses. MCC950 is a strong NLRP3 inhibitor. Through this study, the team emphasized the usefulness of proteomics and molecular biology methods in identifying and validating target molecules. The ability of these inhibitors to attach with human hedgehog acyltransferase (HHAT), a representative of the mammalian membrane-bound O-acyl-transferase (MBOAT) superfamily thought to be involved in diseases like cancer and obesity, is another promising example of a target-based drug discovery approach. A photochemical probe was recently created by Dr. Lanyon-Hogg, a principal scientist at the University of Oxford, and his colleagues. Their research was published in 2021 (Lanyon-Hogg et al. 2021). To create biologically active small compounds, the Lanyon-Hogg group uses synthetic chemistry and biology techniques. Small-molecule HHAT inhibitors are still poorly understood in terms of their chemical processes and binding location. The most effective HHAT inhibitor known to date, IMP-1575, was shown to be a single-enantiomer inhibitor in this study by making use of medicinal chemistry and a new assay for the HHAT lipid transferase activity (Acyl-cLIP), according to Lanyon-Hogg. Since HHAT has a small-molecule binding domain and its cryo-EM structure

has been solved, it is now actually possible to use structure-guided methods to speed up the development of inhibitors. As these molecules are further improved, there may be a lot of follow-up research in medicinal chemistry investigations of HHAT inhibitors. By assisting in the creation and development of tools and molecules for target validation in the *in vivo* models, structure-guided medicinal chemistry may be a key step in validating HHAT as a therapeutic target in cancer, says the principal scientist Lanyon-Hogg. The abovementioned probes are examples of recently developed techniques that offer a reliable way to confirm drug–target interactions. Understanding this interaction allows for further tuning of the chemical using medicinal chemistry methods to boost specificity and affinities to the drug target. A drawback of the phenotypic screening strategy is that it can be difficult to further enhance the drug's properties in the absence of knowledge about a drug–target interaction. In other words, by improving medication design, target-based drug discovery enables the researcher team to identify an issue and effectively solve it. Additionally, TBDD is particularly helpful when a gene and a disease have a well-established, strong association; this is valid for both uncommon illnesses and monogenic diseases. For instance, it was shown that a gain-of-function mutation in the PCSK9 gene raised blood levels of low-density lipoproteins (LDLs), commonly referred to as "bad cholesterol." Due to this finding in 2015, the FDA finally authorized the drugs alirocumab and evolucumab. Both medications act by inhibiting PCSK9, which lowers LDL cholesterol and lowers the risk of developing heart diseases (Duan et al. 2022; Guo et al. 2014; Singh et al. 2022)**.** Phenotypic test technology is frequently cutting edge, although the methodology itself is not. There were not many alternatives available before the development of cloned molecular targets. The effectiveness of a target-based strategy grows along with our understanding of a disease's molecular basis. When a popular drug target idea fails, for example, a new strategy utilizing phenotypic drug discovery may be necessary.

7.3.4 Molecular Docking

Putting several molecules together is the subject of research of molecular docking (Meng Zhang et al. 2011). In order to predict how much an enzyme or protein will interact with small molecules (ligands), a technique known as docking is performed (Berry et al. 2015). The capacity of a protein (or enzyme) and nucleic acid to bind with small molecules to create a supramolecular complex has a significant impact on the dynamics of a protein, which may either enhance or impede its biological activity. Small molecules' behavior in target proteins' binding sites can be explained by molecular docking. The technique seeks to determine the proper positions of ligands in a protein's binding pocket and to forecast their affinities (Ferreira et al. 2015). Based on the types of ligands, docking can be classified as follows:

Table 7.1 Molecular docking analysis of anticancer compounds against BACE 1

Sl. No.	Compound	2D structure	Binding energy (kcal/mol)	Molecular interactions
1.	Cedronolactone D		−9.5	Conventional Hydrogen bond THR481, TRP485 Pi-Alkyl PHE517 Pi-Sigma TYR480 Van der Waals ASP441, GLY443, SER444, VAL478, ILE527, ARG537, TYR607, ASP637, GLY639, THR640, THR738

1. Protein–small-molecule (ligand) type docking.
2. Protein–nucleic acid-type docking.
3. Protein–protein type docking.

The example of protein–small-molecule (ligand) docking is discussed below in Tables 7.1 and 7.2 and Fig. 7.3.

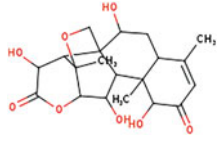
7.3.4.1 Molecular Docking Approaches

Some techniques were specifically beloved by the docking society (Dar and Mir 2017). One of those uses a match process to represent the protein and the ligand as harmonizing surfaces. The other one computes paired interaction effects between the two, while simulating that actual docking process. Both approaches offer several advantages and some disadvantages.

Shape Complementarity Approach

The ligand and target's surface structure properties are employed in this strategy to determine how their molecules will interact. In this instance, the molecular interface of the ligand is portrayed in relation to the target's accessible surface area of the solvent. The complementarity between two surfaces based on form matching illustration makes it simpler to search the complementary channel for ligand on the target surface. For instance, the hydrophobicity of the target molecules can be analyzed by the bend numbers in the main chain. This approach rapidly scans hundreds of ligands swiftly to assess their putative binding capacity on the target surface.

Table 7.2 ADME properties of terpenoid compound (cedronolactone D)

Sl. No.	Compound	Molecular formula	ADME properties (Lipinski's rule of five)		Structure	Drug likeliness
			Properties	Values		
1.	Cedronolactone D	$C_{20}H_{26}O_8$	Molecular Weight (≤500 Da)	394.4		Yes
			LogP (≤5)	−1.4		
			H-Bond Donor (≤5)	4		
			H-Bond Acceptor (≤10)	8		
			Violations	0		

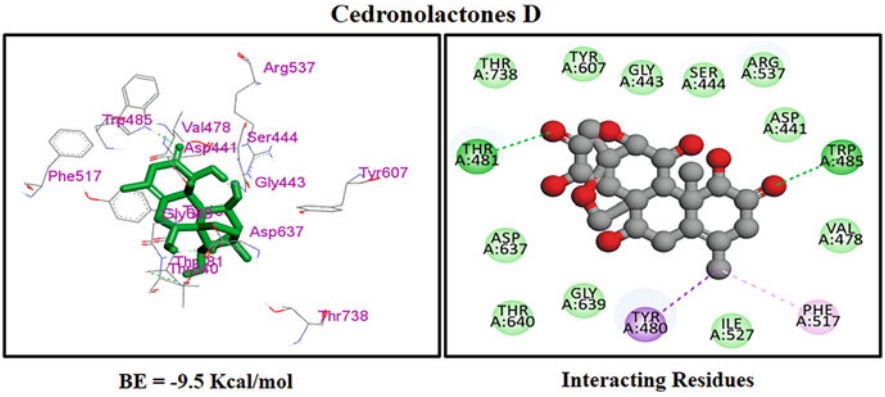


Fig. 7.3 2-D interaction of cedronolactone D with BACE 1

Simulation Approach

Here, after "specified durations of movements" in the target's conformational space, the ligand is allowed to attach to the groove after the target and ligand have been physically separated. The movements entail either internal (rotations of the torsional angle) or exterior (changes to the structure of the ligand) alterations (rotations and translations). "Total energy" is generated each time the ligand moves inside the conformational limit. This tactic is better since it is more willing to accommodate ligand flexibility. Evaluating the molecular interactions between both the ligand and the target is also more precise. The best docked crystallographic prediction using this technique takes longer since each conformation involves a large energy loss. This disadvantage has recently been significantly modernized to make the simulation technique more user-friendly via quick optimization methods and grid-based tools.

7.3.4.2 Docking—Its Mechanisms

Before performing docking, the target protein's structure is required. The structure has traditionally been determined using a biophysical technique like X-ray crystallography, NMR spectroscopy, or cryo-electron microscopy (cryo-EM); however, homology modeling construction can also be employed to do so. This protein structure together with a database of potential ligands serves as inputs to a docking algorithm. Two elements that affect a docking program's success are the algorithm-based and the scoring scheme.

7.3.5 Molecular Dynamics

The real atomic and molecular motions are examined using any computer simulation method known as molecular dynamics (MD) (Hollingsworth and Dror 2018). The molecules and atoms communicate for some time period, revealing information on dynamic "evolution" of the system. The most popular approach is computationally resolving Newton's motion equations for a system of interacting particles. Forces between the particles and their potential energies are typically computed using interatomic potentials for molecular mechanic force fields. By using this technique, atoms' and molecules' paths may be determined. The technique is mostly employed in chemical physics, materials science, and biophysics.

Drug development is a particularly intriguing example of an area where simulations may have an impact on research. Recent structural biology findings have identified several crucial targets for the creation of novel neural medicines (e.g., GPCRs, ion channels, and transporters). To fully make use of the promise of structure-based drug design for this and other targets, it is necessary to take into account the dynamic properties of these proteins.

MD modeling is very helpful when modifying a ligand to improve its effectiveness or other aspects, such as other qualities and lead optimization. On a qualitative level, simulations can provide a range of information to guide the ligand optimization process, including the ability to identify the key interactions a ligand has with the ligand-binding pocket, predict how a ligand will reorganize the binding pocket, and test and enhance potential ligand poses. The binding location and stance of a ligand can occasionally be seen in simulations of the full ligand-binding process. When compared to other computational techniques like docking, simulation-based approaches provide far more accurate estimates of ligand-binding constants (free energies). Some methods require a lot of computation and are only precise when determining the relative binding energies of ligands that share a scaffold. While using MD simulation, the molecular mechanics/generalized born surface area or MM/GBSA and molecular mechanics/Poisson–Boltzmann surface area or MM/PBSA approaches are both noticeably quicker but less accurate because they rely on continuum solvent models rather than an explicit description of water.

MD may also be used in virtual screening, which selects a beginning set of ligands predicted to bind a target. Docking tools and a particular target protein structure are used for conventional virtual screening. A subset of binders may only be found by docking to a particular structure since the binding pocket may really be somewhat flexible. The variety of binding ligands discovered can be increased by taking into account multiple potential structures discovered through simulation.

7.4 Target-Based Screening Versus Phenotypic Screening

7.4.1 What Does Each Screening Approach Involve?

Empirical drug discovery has experienced a resurgence, and it has been formalized under the name "phenotypic drug discovery" or PDD, following an analysis of the discovery processes for novel molecular entities authorized by the US Food and Drug Administration (FDA) between the years 1999 and 2008 (Swinney and Anthony 2011). This statistic shows that most first-in-class small-molecule drugs were discovered by experimental means, but the vast majority of those that came after were found through target-based drug discovery or TDD. This work came to the conclusion that it is frequently difficult to find a roadmap for the discovery of first-in-class medications using the mechanistic information that is available at the time a program is started. This requires understanding not just the anticipated drug target but also how that target corresponds to a particular, therapeutically advantageous phenotype, the molecular mechanism of action (Gilbert 2013; Kaminski et al. 2017; Swinney 2013). In biological research, drug designing discovery, and development, phenotypic screening is a screening strategy used to find certain compounds including small molecules, proteins, peptides, or RNAi that has the potential to change or alter the phenotypic characteristic of a cell or a certain organism in a desirable way (Childers et al. 2020; Kotz 2012). These compounds can be identified by analyzing both cell-based tests and animal model studies. A molecule has frequently been examined to see if it has the intended impact in cells, isolated tissues, and organs and even in animals. Often and sometimes for many long years, the precise mechanism by which the medicine exerts its effects, that is, its target, was not discovered. Target-based screening has historically produced more best-in-class medications, but on the other hand, phenotypic screening had a minor edge in discovering first-in-class treatments (Swinney 2013). This has been ascribed to the lack of prejudice when determining the mechanism of action of a medicine. However, phenotypic tests are also more suited to detecting chemicals that are active in cells, according to Professor Elizabeth Sharlow, a researcher at the University of Virginia School of Medicine in Virginia, USA. She explains this in simple words. According to her, due to the requirement for frequently complex downstream target deconvolution procedures, phenotypic tests are often difficult and challenging. Additionally, they can take longer to deploy in some cases, which could eventually affect the performance. Furthermore, the entire screening paradigm and screening strategy cost need to consider all of this into account. For many years, drug companies and academia have adopted a "target-first" strategy, in which a molecule known to be significant in a disease procedure is often implemented to screen large compound libraries in search of a "hit"—a candidate drug (Frearson and Wyatt 2010; Herrera-Acevedo et al. 2022). This lack of throughput, coupled with the revolution in the area of genomics studies, has led to this strategy's adoption. One benefit of this strategy is that one may screen millions of molecules that resemble drugs while being certain that, in the event of a hit, he already has a candidate that has the makings of a viable

medication (Verma and Prabhakar 2015). Target-based assays can be implemented more quickly in general, but they can occasionally run into problems with common readouts like enzyme activity, as noted by several researchers. Additionally, there are more comparatively complex target-based assays, such as those based on protein–protein interactions, which are much more difficult to execute despite being more physiologically intriguing. Both screening procedures are valuable and required for the identification of the chemical probes and/or the drugs, and typically, the screening strategy that one chooses will ultimately rely on the resources at his disposal. Both of these strategies have their proponents and opponents as well as their merits and shortcomings. Although phenotypic approaches employ semi-empirical techniques that do not need a comprehension of the mechanism, they do call upon their biology knowledge for human disease-related certain markers to be identified. Furthermore, it might be challenging to take the risk of integrating a chemical into the development without a working knowledge of the mechanism that would guide the evaluation of dose–response relationships. The effectiveness in a translational phenotypic test reduces the early risk in the phenotypic method. Of course, both strategies need to have their prediction power for human biology investigated and confirmed. The development of the therapeutic candidate may be slowed by the absence of mechanism knowledge because further research will need to be empirical. Target-based techniques should actually make it possible to quickly and methodically go for clinical trials, albeit which would be essential for testing a number of putative targets and the molecular mechanisms of action before selecting the best one. Because several hypotheses must be evaluated, it is probable that a TBDD strategy might increase the expenditure of designing. It is interesting to note that the amount of mechanistic understanding necessary to advance a molecule is the key aspect of any method in talks about contemporary drug research, although regulatory clearance does not need to be understood from a mechanistic perspective. However, at this point of time, it can be stated that phenotypic screening is making a minor comeback from the perspective of drug development.

7.4.2 Combination Approaches

For bridging the gap in between target-based approaches, phenotypic approaches, and other relevant drug discovery approaches, the gray area between the same has to be marked. Integrating the greatest qualities from different drug discovery methodologies is the perfect way for successfully developing a medication. In actuality, the majority of effective drug development initiatives combine an understanding of a molecular target with physiologically pertinent cellular experiments. This is beneficial since there is a higher likelihood of successfully marking a medicine when employing an integrated procedure as technologies for various drug discovery methodologies advance. Drug discovery techniques that were established in chemical biology include phenotypic and target-based approaches. Between proponents of phenotypic and target-focused screening, a heated debate had developed over

which strategy offers the best chance of successful drug development. After phenotypic screenings, effective target deconvolution (TD) offers a chance to harmonize these two methods. However, despite the variety of *in vitro* TD techniques that are currently accessible, it is still difficult to match a phenotypically active drug with a biomolecular target (Heilker et al. 2019). In the discipline of toxicology, they are also used as an adverse outcome pathway framework (Allen et al. 2014; Ankley et al. 2010). The phenotypic approach is an experimental method for assessing the phenotypic responses of cells or tissues to chemical substances. It uses techniques like cell-based and *in vivo* assays and tests. These tests are used in the drug development process to look for active substances that cause phenotypic responses that help cells or tissues in diseased states. Multi-parameter molecular profiling has proved invaluable in this situation. A system-level knowledge of biological processes and how they respond to small-molecule treatments are provided by small-molecule multi-parameter phenotypic profiling. As a result, it warrants more attention in the preliminary phase of the process of drug discovery. The mRNA-, protein-, and imaging-based multi-parameter profiling methods are being employed for phenotypic profiling in the context of medication development. Early phenotypic profiling technology integration together with effective experimental and in silico target identification approaches can boost the success rates of lead selection and optimization in the drug development process (Feng et al. 2009; Moffat et al. 2017). The target-based approach, on the other hand, is a logical method for evaluating drug candidates that target a disease-causing biomolecule (Croston 2017). With the development of high-throughput experimental technology in recent years, the target-based strategy has seen a significant increase in usage for drug discovery (Zheng et al. 2013). In this topic, Sams-Dodd has tried to analyze target-based drug discovery in detail. He discussed the pharmaceutical industry's consistent productivity reduction during the last 10 years. A startling finding in this context is that the emergence of target-based medication discovery corresponded with this reduction. He highlighted how the procedure of target validation is difficult and fraught with ambiguity. In light of this paradigm's supremacy and its ability to create rational drug discovery programs and its screening capabilities, he carefully examined these aspects to see if any flaws that might account for why it has not helped to spike yield over the conventional *in vivo* approach (Sams-Dodd 2005). Many other studies have tried to explain this phenomenon on different approaches of drug designing and development (Hajduk and Greer 2007; Rubin et al. 2006). Target deconvolution and polypharmacology, two significant issues in drug discovery, cannot be resolved by phenotypic or target-based methods. The multi-targeted activity of substances is referred to as polypharmacology. The use of many disease-relevant targets in polypharmacology can improve treatment effectiveness, prevent drug resistance, or reduce adverse effects associated with therapeutic targets. Unintentional polypharmacology can have negative effects. The interrelated features of polypharmacology are crucial here. It is important to consider the importance of polypharmacology for drug safety, risk mitigation, and methods for identifying polypharmacological molecules early in the drug discovery procedure. It is also noteworthy to consider the benefits of polypharmacology in treating multigenic

diseases and infections, as well as possibilities for drug discovery, and its development, followed by repurposing (Antolin et al. 2016; Anighoro et al. 2014; Peters 2013; Ravikumar and Aittokallio 2018; Zhang et al. 2016). To pinpoint a target biomolecule in charge of a phenotypic response, target deconvolution is performed. In the phenotypic approach, it is impossible to determine the desired target molecule on which the active drug directly works and the relationship here between the target biomolecule and the phenotype, even if we are successful in identifying an active compound that changes a target phenotype. The target-based method, which is used to hunt for and scan for medications, cannot be employed if the target biomolecule is uncertain since it focuses precisely and operates effectively on a known target biomolecule. Target deconvolution is thus one of the main obstacles to drug development in phenotypic and target-based techniques. Advanced molecular and chemical genetics have created a number of experimental strategies to address this problem. Swimney et al. discussed the development of novel pharmaceuticals (Swinney and Anthony 2011). To identify potential therapeutic candidates, several preclinical techniques are utilized. The team examined the molecular mechanisms of action (MMOA) and discovery strategies for new molecules and biologics that were endorsed by the FDA between 1999 and 2008 to determine whether some approaches were more effective compared to others, in the development of novel drugs. Out of the 259 compounds that received approval, 75 were first-in-class medications with novel MMOAs, 50 (about 67%) of these were small molecules, and about 25 (about 33%) were biologics. The results also show that, with 28 and 17 of these medications, respectively, phenotypic screening significantly contributed more to the creation of first-in-class small-molecule therapeutics than target-based techniques during a period when target-based approaches garnered the majority of the focus. The authors hypothesized that a target-centric approach for first-in-class pharmaceuticals without consideration of an ideal MMOA may be to blame for the existing high turnover rates and poor performance in pharmaceutical research and development. An important argument in favor of an integrated approach is the notion that fundamental human disease biology is still poorly known (therefore, molecular targets) and that existing disease models do not effectively represent actual illnesses (therefore, insufficiency of the phenotypic drug discovery approach). The majority of animal preclinical models are unable to adequately simulate a complex tissue microenvironment, such as that encountered in late-stage solid tumors, where there is high patient heterogeneity in cell content and mutations. The lack of quantitative test outputs that mechanistically match a causative disease biomarker and accurately estimate animal models strengthens the need for an integrated method.

7.4.3 How Can These Screenings Be Taken to the Next Level?

From the time of product development to the time of product approval and commercialization, medication development typically takes 12 years. Target-based drug designing has been demonstrated to be a potent method to select and validate a

therapeutic target by taking advantage of understanding of particular chemical pathways, despite its limitations. A better mechanistic understanding of the biology of human diseases will be made possible by ongoing advancements in molecular biology and genomics, and machine learning can further improve techniques such as high-content imaging and different computational modeling strategies to enhance lead optimization. In fact, phenotypic and target-based drug discovery methods may work together, despite frequently being described as opposing approaches, and combining them may increase the likelihood of successfully generating a medicine. Integration of different screening techniques will ultimately ease the path of successful design and discovery of a drug. Various studies and research have been conducted till date on combination approaches to strengthen and pave the future of various drug designing and discovery processes. A work recently published in 2020 by Kawamura et al. talks about how the identification of targets after the identification of small-molecule elicitors of various phenotypic modifications typically leads to new knowledge of cellular processes (Kawamura et al. 2020). The study team used comprehensive profiling and modifications for screening the compounds. By doing so, they were able to discover an indane derivative, namely NPD9055 that differed mechanistically from the reference substances with pre-established mechanisms of action. Later, they used a chemical proteomics technique to show that NPD9055 binds parts of the heterotrimeric G-protein Gi. An *in vitro* study revealed that NPD9055 reduced GDP/GTP exchange on a Gi subunit activated by a G-protein-coupled receptor protagonist, but not on a different G-protein from the Gs family [35S]GTP-S-binding experiment. Following its separation from Gi, NPD9055 increased intracellular ERK/MAP and Ca2+ level K phosphorylation in intact HeLa cells, both of which are controlled by G. Based on the study findings of the group, it was suggested that NPD9055 modulates G-dependent cellular functions by targeting Gi, most likely via triggering the separation of G from Gi. Another application of combination therapy was brought into limelight recently by Ye et al. in 2021 when they published their original research work on "ScaffComb" (Ye et al. 2021). Combination therapy has long been utilized in the treatment of cancer to overcome medication resistance brought on by monotherapy as well as the swift development of deep learning methods and the increase in pharmacological data have made it possible to build models that can predict and evaluate drug pairings. The problem arises when the drug libraries can only include a few hundred to thousands of chemicals, nevertheless. Here, in this article, the ScaffComb framework has been suggested by the research team as a way to fill in the gaps in large-scale databases' drug pair scannings. In ScaffComb, phenotypic information is incorporated into molecular scaffolds that can be utilized to screen the drug library and identify potent drug combinations. ScaffComb was created as a motivation for phenotype-based drug design. First, some known pharmacological combinations are successfully reidentified by using US Food and Drug Administration information to validate ScaffComb. Following that, the ZINC and ChEMBL databases are then screened using ScaffComb, yielding unique medication combinations, and demonstrating the capability to find new synergistic pathways. This research team indicates that ScaffComb is the first method to use phenotype-based virtual screening of drug

combinations in huge chemical data sets. In another study, recently in 2021, Wilke et al. published a study report highlighting an example of combination approaches in drug discovery (Wilke et al. 2021). To identify and determine the appropriate molecular targets, affinity-based chemical proteomics is frequently paired with a phenotypic screening strategy for bioactive small compounds. Although target identification frequently necessitates chemical derivatization of the identified compound, such assays and the studied bioactivity are skewed toward the observed phenotype. When a drug is perturbed, unbiased cellular profiling tools, on the other hand, record about hundreds of parameters in order to map the bioactivity in comparatively a larger biological context. These approaches may also be able to connect a profile to a target molecule or its mode of action. In this study, the research team presented "Cell-Painting" assessment in conjunction with thermal proteome profiling to identify the diaminopyrimidine DP68 as a sigma 1 (σ1) receptor antagonist. Their findings demonstrate how the combination of complementary profiling techniques may help identify small chemical targets and detect bioactivity. A review article published by Isgut et al. in 2017 (Isgut et al. 2018) portrayed the benefits and application of combination approaches for the identification and discovery of certain combination drugs based on natural products. These authors explained how the foundation for a greater emphasis on natural products in drug discovery and development is laid by studies on the advantages of medication combinations. Natural products are made up of a wide range of components that might interact with a wide range of bodily targets to cause pharmacodynamic reactions that could together result in an additive or a synergistic therapeutic effect. Natural compounds can serve as a beginning point for the development of powerful combination therapies even though they cannot be patentable. Phenotypic screening can be used to discover the ideal combination of bioactive components in natural goods. Due to the low success rates and rising development costs of modern drug discovery efforts pharmaceutical scientists have been looking for novel ways to focus on drug discovery, network pharmacology or NP was considered to be a solution to the issue, which focuses on many targets and pharmacological combinations for the treatment of diseases. With the development of the disciplines of systems biology and metabolomics, it has just lately begun to emerge and the fact that combinatorial drug screening is gaining popularity and importance is emphasized through this article. In another study published by Malyutina et al. in 2019 (Malyutina et al. 2019), the authors discussed how finding new therapeutic combinations for oncology can be facilitated by high-throughput drug screening. A thorough matrix design has been applied in several recent research studies to describe how different medication combinations affect cancer cells. However, the full matrix layout may not be the optimum choice because a drug pairing must be linked at various doses in a factorial design. Additionally, a lot of computational methods simply evaluate the synergy of drug combinations rather than their sensitivity, which could result in misleading positive findings. In order to more efficiently and synchronously examine the sensitivity and synergy of medication combinations, the study team devised a drug combination sensitivity score, or CSS, to measure the sensitivity of a drug pair. They also proposed a unique cross-design. In order to validate its usage as a reliable metric, they also showed that

the combination sensitivity score is highly repeatable between replicates. The researchers also showed how CSS might be forecasted using machine learning methods that determined the most crucial pharmaco-features for classifying cancer cell lines according to their medication combination sensitivity patterns and behaviors. A S-synergy score, which was used to assess the degree of drug interactions that use the cross-design, was created by comparing the dose–response profiles for the medicine combination and the individual drug. They also showed that the S score has an accuracy-level equivalent to the complete matrix design for identifying actual antagonistic and synergistic drug combinations. Another article written by Swamidass et al. and published in 2014 discusses the combined analysis of target-based and phenotypic drug screening methods (Swamidass et al. 2014). The research group carried out a thorough analysis of the small-molecule data in drug development by constructing test networks and linking experimental studies if they had involved non-promiscuous chemicals. Such a network incorporates innovative polypharmacology, recapitulates current biology, and identifies different screening types. Target-based screens that link phenotypic and biochemical information can provide the facility to repurpose biologic and small-molecule medicines. ALOX15 is associated with efforts to find drugs that can halt cell death brought on by a mutant version of superoxide dismutase. The prospect that ALOX15 inhibitors might be utilized medically to address amyotrophic lateral sclerosis is suggested by this connection. They also emphasized the interactive version of this network's website. In this context, a number of additional studies on various drug discovery screening techniques merit mentioning include those by Clark et al. 2015, Forsberg et al. 2014, Kaur et al. 2016, Matlock et al. 2013, O'Reilly et al. 2014, Sidders et al. 2018, and Tran-Nguyen and Rognan 2020.

7.5 Case Studies and Examples

7.5.1 Case Study 1

Global fatality rates have grown since SARS-CoV-2's rapid spread (Choudhury et al. 2022; Satyam et al. 2020). The unregulated increase in COVID-19 infection cases, which affected different communities of several countries, prompted concerns about world health. Another study recently examined the virus's current state of knowledge. SARS is recognized as the severe acute respiratory syndrome associated with coronavirus. These are all members of the order Nidovirales and family Coronaviridae. An RNA virus contains positive strands, and the SARS coronavirus is contagious. According to reports, coronaviruses have the largest genomes of all other RNA viruses, having average sizes between 27 and 32 kilobytes.

Molecular Docking The molecular docking approach to identify therapeutic targets is one of the most used methods for ligand-based computer-aided drug development (LB-CADD). With the use of this technique, enormous amounts of data from

drug catalogs may be swiftly analyzed and annotated, reducing the amount of effort, time, and money spent on CADD. Finding appropriate therapeutic targets is crucial since the COVID-19 virus presently lacks effective therapeutics. We employed a molecular modeling technique incorporating molecular docking and MD simulation to identify putative phytochemicals effective against the Mpro protein of COVID-19 (Gurung et al. 2021). These observed organic compounds may pave the way for the development of COVID-19 antiviral drugs. Based on AutoDock binding affinity, carvacrol, oleanolic acid, and ursolic acid have all shown sufficient associations with active site residues. According to research, the binding energies of these substances are 4.0 kcal/mol, 6.0 kcal/mol, and 5.9 kcal/mol, accordingly (Kumar et al. 2021).

MD Simulation MD simulation is one of the tried-and-true in silico methods for collecting data in real time with atomic spatial precision and picosecond or greater temporal resolution. Carvacrol, ursolic acid, and oleanolic acid are the principal phytochemical compounds docked with proteases. Simulation research was conducted over a simulation time of 50 ns to examine the durability of these compounds in the binding domain of Mpro.

Summary of the Case Study

1. It has been established that a key and extremely effective target for the suppression of new COVID-19 is the Mpro protein. Three natural substances—ursolic acid, carvacrol, and oleanolic acid—have been identified by this investigation as possible inhibitors of Mpro.
2. According to molecular docking research, carvacrol has lower binding energy than oleanolic acid and ursolic acid.
3. It was discovered that the binding form of interaction was fairly effective. According to MD simulations, all three docking complexes exhibited stability at about 50 ns. These inhibitors further meet Lipinski's rule of five and the ADME requirements.
4. For additional *in vivo/in vitro* validations, each of the presented substances is both natural and commercially available. Future research on further phytochemical-based therapies against COVID-19 may make use of the knowledge obtained from this work.

7.6 Future Roadmap

Any progress in technology and science is promptly put to use in the fields of medical, pharmacy, and drug development. The more effectively a specific drug candidate is created during the experimental stage, the less probable it is that the medicine will fail in the late phases, where the experimental studies are more expensive, particularly in clinical trials. The COVID-19 epidemic made us reevaluate ways to quicken the processes of medication and vaccine discovery and development. There has been a significant amount of interest in artificial intelligence

(AI) in recent years as a way to accelerate early-stage drug discovery and lower the cost of bringing novel medications to market. New AI-driven biotech startups have received a number of sizable funding. Most of these businesses use artificial intelligence (AI) to help make sense of the massive body of scientific literature and the expanding patient genomic, proteomic, and transcriptomic datasets, synthesizing these data to provide new therapeutic targets (Burki 2020). A recent instance of this was a partnership between Imperial College London and Benevolent AI to quickly find repositionable medications to assist in treating COVID-19. Artificial intelligence has the ability to offer new, efficient, and more affordable approaches to drug discovery (Richardson et al. 2020). In a short amount of time, AI can acquire and analyze massive amounts of data, choose suitable targets and complementary ligands, plan experiments, and carry them out. Biologicals are frequently used in novel medication therapies, which are also far more expensive than currently available options. As part of the aging process and the ability of emerging nations to pay for medical care rises, there is a parallel increase in the demand for pharmaceuticals. The strongest risk factor for illnesses is age. The potential to use biologics to treat more uncommon ailments directs the business, supports its expansion, and draws more individuals under the pharmaceutical umbrella. The biotech business was where these were primarily created, allowing for the treatment of more rare ailments. Big Pharma acquired the pioneering biotechs to join the biological revolutions somewhat later. Small compounds taken orally are now being used to develop new medicines. The advancements in the treatment of conditions like rheumatoid arthritis and multiple sclerosis show and provide light on the direction of progress. In the last 50 years, the field of drug research and development has made impressive strides. Professor Ross King, a specialist in artificial intelligence and machine learning at Chalmers University of Technology in Sweden, claims that "The entire drug design process, from target selection to screening to QSAR learning to new compound synthesis, will be almost entirely automated in 10–15 years. This will occur because machines are considerably more adept than people at making these kinds of decisions, and the ensuing procedure will be much quicker and more efficient." King is well known for his work on the robot scientists "Adam and Eve," which serves as an example of how high-throughput robots are. They automate the early stages of drug design with Eve. Eve had the ability to automate the assays for synthetic biology for many substances. Eve began testing substances from the library until she found a sufficient number of hits, at which point she halted and began performing more illuminating assays on the hits. Eve chose additional substances from its library that were intended to best enhance the statistical models using "active" machine learning. They further established that this method of drug testing is typically more affordable than screening the entire library (Williams et al. 2015). By 2040, one might anticipate that many diseases will be more successfully identified, avoided, treated, or even managed in some cases using nonpharmacological interventions if we take into consideration new information about disease pathogenesis, the application of cutting-edge technologies, and all-encompassing strategies. One can only hope that the vision is realized. Clinicians and researchers will continue to work very hard and make significant contributions to enhancing

human health and well-being (Villoutreix 2021). Theoretically, discovering effective targets for antivirals is simpler than doing the same for targets for more difficult diseases like Alzheimer's or cancer, where the molecular background is less well understood. Given the tiny genome size of viruses, one may usually focus their efforts either on viral entrance or replication. Because of the considerable sequence homology, especially between the active regions of polymerases from various coronavirus strains, the goal is to precisely block viral reproduction by targeting the polymerase. A strong inhibitor against a strain should have a decent crossover into potential new strains, and we should be able to say with some degree of assurance. The end goal of futuristic drug designing is to be able to develop a unique, safe, efficient, and patient-tailored medication over the course of a few hours. This objective is totally attainable in the near future, although it seems amazing right now.

7.7 Learning Outcomes

After reading this book chapter, readers should be in a position to understand the principles underlying target identification, validation, and strategies used in different target and drug discovery. They shall further be able to explain various drug target classes and demonstrate specific techniques and strategies used for target identification and validation. Understanding the principles underlying target-screening strategies, assays, and approaches used in lead identification and optimization during the preclinical development of drugs should be easy for the reader. They will be able to analyze and appreciate the evolution of target identification from past to present and future strategies. However, the following points are important to remember when it comes to drug screening, designing, and discovery:

(i) The goal of a preclinical drug discovery program is to generate one or more clinical candidate molecules that provide sufficient evidence of biological action at a disease-relevant target, adequate safety, and sufficient drug-like properties to be studied in people.
(ii) Effective drug design depends on determining and preserving the clinical spectrum of the disease and the exact function that a potential therapeutic target plays in the disease. A pharmacological target is a biological object (typically a protein or gene) that reacts with and has its function changed by a particular drug.
(iii) A suitable drug target should be modifiable for therapeutic purposes and pertinent to the illness phenotype. A strong therapeutic window is also required to guarantee that no treatment modality employed to treat the target interferes with the target's physiological role in healthy tissue, which might result in unexpected effects.
(iv) Promising therapeutic targets should have the following characteristics: The target should either be disease-modifying or should have a known function in

the pathophysiology of a particular disease, to evaluate the druggability of the target, the 3D structure of it should be available, high-throughput screening should be made possible by the target's simple "assayability," the toxicity profile of the target should be promising and phenotypic information should be used to anticipate any potential negative consequences, and the proposed target should have an advantageous standing for its intellectual property.

(v) The phenotypic approach to drug discovery, which falls under the category of target deconvolution, exposes cells, separated tissues, or animal models to small molecules in order to determine whether a particular candidate molecule enforces the desired effect, which is indicated by a change in phenotype. For the characterization of tiny compounds and small-scale drug screening methodologies, mammalian cells are typically selected over other animal models given that they are more suitable for high-throughput screening (HTS) and have better physiological relevance. Instead of focusing only on particular proteins or nucleic acids, the phenotypic technique considers the investigation of whole signaling networks. The drug's action is established before the specific biological (drug) targets responsible for the observed phenotypic characteristics are found.

References

Abifadel M, Varret M, Rabès JP, Allard D, Ouguerram K, Devillers M, Cruaud C, Benjannet S, Wickham L, Erlich D, Derré A, Villéger L, Farnier M, Beucler I, Bruckert E, Chambaz J, Chanu B, Lecerf JM, Luc G, Boileau C (2003) Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat Genet 34(2):154–156. https://doi.org/10.1038/ng1161

Allen TE, Goodman JM, Gutsell S, Russell PJ (2014) Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. Chem Res Toxicol 27(12): 2100–2112. https://doi.org/10.1021/tx500345j

Anderson AC (2003) The process of structure-based drug design. Chem Biol 10(9):787–797. https://doi.org/10.1016/j.chembiol.2003.09.002

Anighoro A, Bajorath J, Rastelli G (2014) Polypharmacology: challenges and opportunities in drug discovery. J Med Chem 57(19):7874–7887. https://doi.org/10.1021/jm5006463

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL (2010) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29(3):730–741. https://doi.org/10.1002/etc.34

Antolin AA, Workman P, Mestres J, Al-Lazikani B (2016) Polypharmacology in precision oncology: current applications and future prospects. Curr Pharm Des 22(46):6935–6945. https://doi.org/10.2174/1381612822666160923115828

Armstrong JW (1999) A review of high-throughput screening approaches for drug discovery. Am Biotech Lab 17:26–28

Badalà F, Nouri-mahdavi K, Raoof DA (2008) Optic disk and nerve fiber layer imaging to detect glaucoma. Am J Ophthalmol 144(5):724–732

Berg T (2010) Small-molecule inhibitors of protein–protein interactions. In: Protein-protein complexes: analysis, modeling and drug design. Imperial College Press, London, pp 318–339. https://doi.org/10.1142/9781848163409_0012

Berry M, Fielding B, Gamieldien J (2015) Practical considerations in virtual screening and molecular docking. In: Emerging trends in computational biology, bioinformatics, and systems

biology. Elsevier, Waltham, MA, pp 487–502. https://doi.org/10.1016/B978-0-12-802508-6.00027-2

Bhunia SS, Saxena M, Saxena AK (2021) Ligand- and structure-based virtual screening in drug discovery. Top Med Chem 37:281–339

Burki T (2020) A new paradigm for drug development. Lancet Digit Health 2(5):e226–e227. https://doi.org/10.1016/S2589-7500(20)30088-1

Cavasotto CN, Phatak SS (2009) Homology modeling in drug discovery: current trends and applications. Drug Discov Today 14(13–14):676–683. https://doi.org/10.1016/j.drudis.2009.04.006

Chen H, Zhou X, Gao Y, Chen H, Zhou J (2017) Fragment-based drug design: strategic advances and lessons learned. Drug Discov Technol 2(8):212–232. https://doi.org/10.1016/B978-0-12-409547-2.12319-4

Childers WE, Elokely KM, Abou-Gharbia M (2020) The resurrection of phenotypic drug discovery. ACS Med Chem Lett 11(10):1820–1828. https://doi.org/10.1021/acsmedchemlett.0c00006

Choudhury PR, Saha T, Goel S, Shah JM, Ganjewala D (2022) Cross-species virus transmission and its pandemic potential. Bull Natl Res Cent 46(1):18. https://doi.org/10.1186/s42269-022-00701-7

Clark AM, Dole K, Coulon-Spektor A, McNutt A, Grass G, Freundlich JS, Reynolds RC, Ekins S (2015) Open source Bayesian models. 1. Application to ADME/Tox and drug discovery datasets. J Chem Inf Model 55(6):1231–1245. https://doi.org/10.1021/acs.jcim.5b00143

Clark RL, Johnston BF, Mackay SP, Breslin CJ, Robertson MN, Harvey AL (2010) The drug discovery portal: a resource to enhance drug discovery from academia. Drug Discov Today 15(15–16):679–683. https://doi.org/10.1016/j.drudis.2010.06.003

Croston GE (2017) The utility of target-based discovery. Expert Opin Drug Discov 12(5):427–429. https://doi.org/10.1080/17460441.2017.1308351

Dar AM, Mir S (2017) Molecular docking: approaches, types, applications and basic challenges. J Anal Bioanal Tech 8:356. https://doi.org/10.4172/2155-9872.1000356

Duan Y, Gong K, Xu S, Zhang F, Meng X, Han J (2022) Regulation of cholesterol homeostasis in health and diseases: from mechanisms to targeted therapeutics. Signal Transduct Target Ther 7(1):265. https://doi.org/10.1038/s41392-022-01125-5

Eder J, Sedrani R, Wiesmann C (2014) The discovery of first-in-class drugs: origins and evolution. Nat Rev Drug Discov 13(8):577–587. https://doi.org/10.1038/nrd4336

Feng Y, Mitchison TJ, Bender A, Young DW, Tallarico JA (2009) Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. Nat Rev Drug Discov 8(7):567–578. https://doi.org/10.1038/nrd2876

Fernandes TG, Diogo MM, Clark DS, Dordick JS, Cabral JM (2009) High-throughput cellular microarray platforms: applications in drug discovery, toxicology and stem cell research. Trends Biotechnol 27(6):342–349. https://doi.org/10.1016/j.tibtech.2009.02.009

Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD (2015) Molecular docking and structure-based drug design strategies. Molecules 20(7):13384–13421. https://doi.org/10.3390/molecules200713384

Forsberg EM, Sicard C, Brennan JD (2014) Solid-phase biological assays for drug discovery. Annu Rev Anal Chem 7:337–359. https://doi.org/10.1146/annurev-anchem-071213-020241

França TC (2015) Homology modeling: an important tool for the drug discovery. J Biomol Struct Dyn 33(8):1780–1793. https://doi.org/10.1080/07391102.2014.971429

Frearson J, Wyatt P (2010) Drug discovery in academia: the third way? Expert Opin Drug Discov 5(10):909–919. https://doi.org/10.1517/17460441.2010.506508

Gashaw I, Ellinghaus P, Sommer A, Asadullah K (2011) What makes a good drug target? Drug Discov Today 16(23–24):1037–1043. https://doi.org/10.1016/J.DRUDIS.2011.09.007

Gilbert IH (2013) Drug discovery for neglected diseases: molecular target-based and phenotypic approaches. J Med Chem 56(20):7719–7726. https://doi.org/10.1021/jm400362b

Guo YL, Zhang W, Li JJ (2014) PCSK9 and lipid lowering drugs. Clin Chim Acta 437:66–71. https://doi.org/10.1016/J.CCA.2014.07.008

Gurung AB, Ali MA, Lee J, Farah MA, Al-Anazi KM (2021) An updated review of computer-aided drug design and its application to COVID-19. Biomed Res Int 2021:e8853056. https://doi.org/10.1155/2021/8853056

Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6(3):211–219. https://doi.org/10.1038/nrd2220

Heilker R, Lessel U, Bischoff D (2019) The power of combining phenotypic and target-focused drug discovery. Drug Discov Today 24(2):526–532. https://doi.org/10.1016/J.DRUDIS.2018.10.009

Herrera-Acevedo C, Perdomo-Madrigal C, de Sousa JA, Scotti L, Scotti MT (2022) Drug discovery paradigms: target-based drug discovery. Springer, Cham, pp 1–24. https://doi.org/10.1007/978-3-030-95895-4_1

Hevener KE, Pesavento R, Ren J, Lee H, Ratia K, Johnson ME (2018) Hit-to-lead: hit validation and assessment. Methods Enzymol 610:265–309. https://doi.org/10.1016/bs.mie.2018.09.022

Hollingsworth SA, Dror RO (2018) Molecular dynamics simulation for all. Neuron 99(6):1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011

Hughes JP, Rees S, Kalindjian SB, Philpott KL (2011) Principles of early drug discovery. Br J Pharmacol 162(6):1239–1249. https://doi.org/10.1111/j.1476-5381.2010.01127.x

Isgut M, Rao M, Yang C, Subrahmanyam V, Rida PCG, Aneja R (2018) Application of combination high-throughput phenotypic screening and target identification methods for the discovery of natural product-based combination drugs. Med Res Rev 38(2):504–524. https://doi.org/10.1002/med.21444

Kaminski T, Gunnarsson A, Geschwindner S (2017) Harnessing the versatility of optical biosensors for target-based small-molecule drug discovery. ACS Sens 2(1):10–15. https://doi.org/10.1021/acssensors.6b00735

Kaur H, Ahmad M, Scaria V (2016) Computational analysis and in silico predictive Modeling for inhibitors of PhoP regulon in S. typhi on high-throughput screening Bioassay dataset. Interdiscip Sci 8(1):95–101. https://doi.org/10.1007/s12539-015-0273-x

Kawamura T, Futamura Y, Shang E, Muroi M, Janning P, Ueno M, Wilke J, Takeda S, Kondoh Y, Ziegler S, Watanabe N, Waldmann H, Osada H (2020) Discovery of small-molecule modulator of heterotrimeric Gi-protein by integrated phenotypic profiling and chemical proteomics. Biosci Biotechnol Biochem 84(12):2484–2490. https://doi.org/10.1080/09168451.2020.1812375

Kennedy CR, Goya Grocin A, Kovačič T, Singh R, Ward JA, Shenoy AR, Tate EW (2021) A probe for NLRP3 inflammasome inhibitor MCC950 identifies carbonic anhydrase 2 as a novel target. ACS Chem Biol 16(6):982–990. https://doi.org/10.1021/acschembio.1c00218

Kim S (2016) Getting the most out of PubChem for virtual screening. Expert Opin Drug Discov 11(9):843–855. https://doi.org/10.1080/17460441.2016.1216967

Klebe G (2000) Recent developments in structure-based drug design. J Mol Med 78(5):269–281. https://doi.org/10.1007/s001090000084

Kotz J (2012) Phenotypic screening, take two. SciBX 5(15):380–382. https://doi.org/10.1038/SCIBX.2012.380

Kumar A, Choudhir G, Shukla SK, Sharma M, Tyagi P, Bhushan A, Rathore M (2021) Identification of phytochemical inhibitors against main protease of COVID-19 using molecular modeling approaches. J Biomol Struct Dyn 39(10):3760–3770. https://doi.org/10.1080/07391102.2020.1772112

Lanyon-Hogg T, Ritzefeld M, Zhang L, Andrei SA, Pogranyi B, Mondal M, Sefer L, Johnston CD, Coupland CE, Greenfield JL, Newington J, Fuchter MJ, Magee AI, Siebold C, Tate EW (2021) Photochemical probe identification of a small-molecule inhibitor binding site in hedgehog acyltransferase (HHAT)*. Angew Chem Int Ed Engl 60(24):13542–13547. https://doi.org/10.1002/anie.202014457

Laporte R, Kohan A, Heitzmann J, Wiśniewska H, Toy J, La E, Tariga H, Alagarsamy S, Ly B, Dykert J, Qi S, Wiśniewski K, Galyean R, Croston G, Schteingart CD, Rivière PJM (2011) Pharmacological characterization of FE 202158, a novel, potent, selective, and short-acting

peptidic vasopressin V1a receptor full agonist for the treatment of vasodilatory hypotension. J Pharmacol Exp Ther 337(3):786–796. https://doi.org/10.1124/JPET.111.178848

Lindsay MA (2003) Target discovery. Nat Rev Drug Discov 2(10):831–838. https://doi.org/10.1038/nrd1202

Lu Y, Liu H, Yang D, Zhong L, Xin Y, Zhao S, Wang MW, Zhou Q, Shui W (2021) Affinity mass spectrometry-based fragment screening identified a new negative allosteric modulator of the adenosine A2A receptor targeting the sodium ion pocket. ACS Chem Biol 16(6):991–1002. https://doi.org/10.1021/acschembio.0c00899

Malyutina A, Majumder MM, Wang W, Pessia A, Heckman CA, Tang J (2019) Drug combination sensitivity scoring facilitates the discovery of synergistic and efficacious drug combinations in cancer. PLoS Comput Biol 15(5):e1006752. https://doi.org/10.1371/journal.pcbi.1006752

Mandal S, Moudgil M, Mandal SK (2009) Rational drug design. Eur J Pharmacol 625(1–3):90–100. https://doi.org/10.1016/j.ejphar.2009.06.065

Martis EA (2010) High-throughput screening: the hits and leads of drug discovery-an overview. J Appl Pharm Sci 1(1):2–10

Matlock MK, Zaretzki JM, Swamidass SJ (2013) Scaffold network generator: a tool for mining molecular structures. Bioinformatics 29(20):2655–2656. https://doi.org/10.1093/bioinformatics/btt448

Maybauer MO, Maybauer DM, Enkhbaatar P, Laporte R, Winiewska H, Traber LD, Lin C, Fan J, Hawkins HK, Cox RA, Winiewski K, Schteingart CD, Landry DW, Rivière PJM, Traber DL (2014) The selective vasopressin type 1a receptor agonist selepressin (FE 202158) blocks vascular leak in ovine severe sepsis. Crit Care Med 42(7):e525–e533. https://doi.org/10.1097/CCM.0000000000000300

Meng Zhang HX, Mezei M, Cui MXY (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

Moffat JG, Rudolph J, Bailey D (2014) Phenotypic screening in cancer drug discovery—past, present and future. Nat Rev Drug Discov 13(8):588–602. https://doi.org/10.1038/nrd4366

Moffat JG, Vincent F, Lee JA, Eder J, Prunotto M (2017) Opportunities and challenges in phenotypic drug discovery: an industry perspective. Nat Rev Drug Discov 16(8):531–543. https://doi.org/10.1038/nrd.2017.111

Muhammed MT, Aki-Yalcin E (2019) Homology modeling in drug discovery: overview, current applications, and future perspectives. Chem Biol Drug Des 93(1):12–20. https://doi.org/10.1111/cbdd.13388

O'Reilly LP, Luke CJ, Perlmutter DH, Silverman GA, Pak SC (2014) C. elegans in high-throughput drug discovery. Adv Drug Deliv Rev 69–70:247–253. https://doi.org/10.1016/J.ADDR.2013.12.001

Patrick Walters W, Stahl MT, Murcko MA (1998) Virtual screening—an overview. Drug Discov Today 3(4):160–178. https://doi.org/10.1016/S1359-6446(97)01163-X

Peters JU (2013) Polypharmacology - foe or friend? J Med Chem 56(22):8955–8971. https://doi.org/10.1021/jm400856t

Ramírez D (2016) Computational methods applied to rational drug design. Open Med Chem J 10:7–20. https://doi.org/10.2174/1874104501610010007

Ravikumar B, Aittokallio T (2018) Improving the efficacy-safety balance of polypharmacology in multi-target drug discovery. Expert Opin Drug Discov 13(2):179–192. https://doi.org/10.1080/17460441.2018.1413089

Reddy MR, Parrill AL (1999) Overview of rational drug design. ACS Symp Ser 719:1–11. https://doi.org/10.1021/bk-1999-0719.ch001

Richardson P, Griffin I, Tucker C, Smith D, Oechsle O, Phelan A, Stebbing J (2020) Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. Lancet 395(10223):e30–e31. https://doi.org/10.1016/S0140-6736(20)30304-4

Rizvi SMD, Shakil S, Haneef M (2013) A simple click by click protocol to perform docking: Autodock 4.2 made easy for non-bioinformaticians. EXCLI J 12:830–857. https://doi.org/10.17877/DE290R-11534

Roy K, Kar S, Das RN (2015) Other related techniques. In: Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment. Academic Press, Amsterdam. https://doi.org/10.1016/b978-0-12-801505-6.00010-7

Rubin AE, Tummala S, Both DA, Wang C, Delaney EJ (2006) Emerging technologies supporting chemical process R and D and their increasing impact on productivity in the pharmaceutical industry. Chem Rev 106(7):2794–2810. https://doi.org/10.1021/CR040674I/ASSET/CR040674I.FP.PNG_V03

Sams-Dodd F (2005) Target-based drug discovery: is something wrong? Drug Discov Today 10(2):139–147. https://doi.org/10.1016/S1359-6446(04)03316-1

Satyam R, Bhardwaj T, Goel S, Jha NK, Jha SK, Nand P, Ruokolainen J, Kamal MA, Kesari KK (2020) miRNAs in SARS-CoV 2: a spoke in the wheel of pathogenesis. Curr Pharm Des 27(13):1628–1641. https://doi.org/10.2174/1381612826999201001200529

Sidders B, Karlsson A, Kitching L, Torella R, Karila P, Phelan A (2018) Network-based drug discovery: coupling network pharmacology with phenotypic screening for neuronal excitability. J Mol Biol 430(18):3005–3015. https://doi.org/10.1016/J.JMB.2018.07.016

Singh K, Gupta JK, Kumar S, Singh K, Meenakshi K, Kumar K (2022) PCSK9 inhibitors: pharmacology and therapeutic potential. Preprints:2022050290. https://doi.org/10.20944/PREPRINTS202205.0290.V1

Swamidass SJ, Schillebeeckx CN, Matlock M, Hurle MR, Agarwal P (2014) Combined analysis of phenotypic and target-based screening in assay networks. J Biomol Screen 19(5):782–790. https://doi.org/10.1177/1087057114523068

Swinney DC (2013) Phenotypic vs. target-based drug discovery for first-in-class medicines. Clin Pharmacol Ther 93(4):299–301. https://doi.org/10.1038/CLPT.2012.236

Swinney DC, Anthony J (2011) How were new medicines discovered? Nat Rev Drug Discov 10(7):507–519. https://doi.org/10.1038/nrd3480

Tran-Nguyen VK, Rognan D (2020) Benchmarking data sets from PubChem BioAssay data: current scenario and room for improvement. Int J Mol Sci 21(12):4380. https://doi.org/10.3390/IJMS21124380

Vanover KE, Weiner DM, Makhay M, Veinbergs I, Gardell LR, Lameh J, Del Tredici AL, Piu F, Schiffer HH, Ott TR, Burstein ES, Uldam AK, Thygesen MB, Schlienger N, Andersson CM, Son TY, Harvey SC, Powell SB, Geyer MA, Davis RE (2006) Pharmacological and behavioral profile of N-(4-Fluorophenylmethyl)-N-(1-methylpiperidin-4-yl)-N′-(4-(2-methylpropyloxy) phenylmethyl) carbamide (2R,3R)-dihydroxybutanedioate (2:1) (ACP-103), a novel 5-hydroxytryptamine2a receptor inverse agonist. J Pharmacol Exp Ther 317(2):910–918. https://doi.org/10.1124/JPET.105.097006

Verma S, Prabhakar YS (2015) Target based drug design—a reality in virtual sphere. Curr Med Chem 22(13):1603–1630. https://doi.org/10.2174/0929867322666150209151209

Villoutreix BO (2021) Post-pandemic drug discovery and development: facing present and future challenges. Front Drug Discov 1:728469. https://doi.org/10.3389/FDDSV.2021.728469

Walker FR (2013) A critical review of the mechanism of action for the selective serotonin reuptake inhibitors: do these drugs possess anti-inflammatory properties and how relevant is this in the treatment of depression? Neuropharmacology 67:304–317. https://doi.org/10.1016/j.neuropharm.2012.10.002

Wilke J, Kawamura T, Xu H, Brause A, Friese A, Metz M, Schepmann D, Wünsch B, Artacho-Cordón A, Nieto FR, Watanabe N, Osada H, Ziegler S, Waldmann H (2021) Discovery of a σ1 receptor antagonist by combination of unbiased cell painting and thermal proteome profiling. Cell Chem Biol 28(6):848–854. https://doi.org/10.1016/j.chembiol.2021.01.009

Williams K, Bilsland E, Sparkes A, Aubrey W, Young M, Soldatova LN, De Grave K, Ramon J, De Clare M, Sirawaraporn W, Oliver SG, King RD (2015) Cheaper faster drug development

validated by the repositioning of drugs against neglected tropical diseases. J R Soc Interface 12(104):20141289. https://doi.org/10.1098/rsif.2014.1289

Wilson GL, Lill MA (2011) Integrating structure-based and ligand-based approaches for computational drug design. Future Med Chem 3(6):735–750. https://doi.org/10.4155/fmc.11.18

Ye Z, Chen F, Zeng J, Gao J, Zhang MQ (2021) ScaffComb: a phenotype-based framework for drug combination virtual screening in large-scale chemical datasets. Adv Sci (Weinh) 8(24): e2102092. https://doi.org/10.1002/advs.202102092

Zhang W, Bai Y, Wang Y, Xiao W (2016) Polypharmacology in drug discovery: a review from systems pharmacology perspective. Curr Pharm Des 22(21):3171–3181. https://doi.org/10.2174/1381612822666160224142812

Zheng W, Thorne N, McKew JC (2013) Phenotypic screens as a renewed approach for drug discovery. Drug Discov Today 18(21–22):1067–1073. https://doi.org/10.1016/J.DRUDIS.2013.07.001

Chapter 8
Fragment-Based Drug Design in Lead Discovery

André M. Oliveira and Mithun Rudrapal

8.1 Introduction

The principle that underlies fragment-based drug design (FBDD) is the construction of bioactive structures capable of interacting specifically with a given biological target (like an enzyme), from the junction of several fragment hits that are chosen separately based on their ability to interact with the site (Kirsch et al. 2019). The criteria for choosing the fragments are based on different approaches, both computational and experimental, and represent an excellent starting point for the construction of new chemical entities (NCEs). Rational drug design uses several approaches, which involve, in a different, but not necessarily independent way: (1) the structure and properties of the ligand; (2) the structure of the macromolecular target; and (3) combinations of both. FBDD methods would fall into the latter category, as they enable the construction of ligands from carefully chosen fragments based on the structure of the target. Increasing research on FBDD approaches has been observed worldwide, considering that 3642 publications in this field showed up between 1953 and 2016 (Romasanta et al. 2018).

The overall process of designing a new drug demands huge costs and a long time, requiring the expertise of countless scientific and technical skills and thousands of professionals, associated with the pharmaceutical industry, research institutes and universities. Notwithstanding, the new paradigm of drug discovery based on the

A. M. Oliveira (✉)
Department of Environment Studies, Federal Centre of Technological Education of Minas Gerais, Contagem, Minas Gerais, Brazil
e-mail: andreoliveira@cefetmg.br

M. Rudrapal
Department of Pharmaceutical Sciences, Vignan's Foundation for Science, Technology & Research, Guntur, Andhra Pradesh, India

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_8

search for a large number of targets and ligands, which came out alongside the advent of genomic sciences, rapid DNA sequencing, combinatorial chemistry, cell-based assays and automated high-throughput screening (HTS), does not convince much of the chemists and biologists (Drews 2000). The main drawback of this approach is that a new promising "hit" arising from this search usually gives birth to new leads who will not eventually lead to a safe drug for clinical use. It is a fact, however, that the more promising leads one can achieve throughout the process, the more compounds to be tested in vitro and in vivo will be available.

The foundation of FBDD has been attributed to Jencks (Jencks 1981) and supported by Nakamura and Abeles (Nakamura and Abeles 1985), who treated drug-like molecules as the combination of several binding epitopes or fragments.

According to Feyfant et al. (2011), FBDD presents over HTS screening methods the advantage that fragment libraries can cover more chemical space than HTS screening libraries and the fact that it works with the concept of ligand efficiency, i.e. the average contribution of each atom of the molecule to the binding affinity.

Once the discovery of biological targets underlies the FBDD, diverse computational resources to acquire such targets have been developed, such as (Katsila et al. 2016): structure similarity searching, data mining/machine learning, panel docking and bioactivity spectra-based algorithms. Further improvements are as follows: network-based drug discovery and matching fields like genomics, transcriptomics, proteomics, metabolomics, microbiome and pharmacogenomics (Wang et al. 2012).

An issue that underlies the proposal of suitable fragments is the increasing molecular complexity of the models. It has been proposed that the probability of binding decreases rapidly as the complexity of the ligand increases (Leach and Hann 2011); once as the complexity of the ligand increases, the more probable is to find a match, other than a mismatch. The probability of finding a unique binding mode where the ligand can match the receptor in just one way reaches a maximum alongside complexity increasing, and this probability decreases henceforth. The exploitation of all possible topologies for a limited number of heavy atoms has led to libraries with millions of compounds, as researchers from UCSF have a pursuit by screening approximately AmpC β-lactamase and D4 dopamine receptor inhibitors, finding 30 compounds with sub-micromolar activity (Lyu et al. 2019).

The overall process normally starts from databases of low molecular weight fragments, which are tested onto the X-ray crystallographic structure (or a structure obtained by other techniques, such as nuclear magnetic resonance or surface plasmon resonance). A link database is also used in order to establish larger structures containing the best-selected fragments through a grow–link–optimize protocol. The procedure is oriented by the improvement of affinity constant (K_d or K_i) or interaction energy.

The site of interaction can be predicted from the three-dimensional structure of the enzyme, obtained experimentally. Figure 8.1 shows how complementary techniques such as high-throughput screening (HTS) alongside FBDD work together with crystallographic methods to obtain a suitable drug.

Whereas well-known structure databases such as the CSD (Groom et al. 2016) have thousands of complete structures with diverse molecular weights, fragment

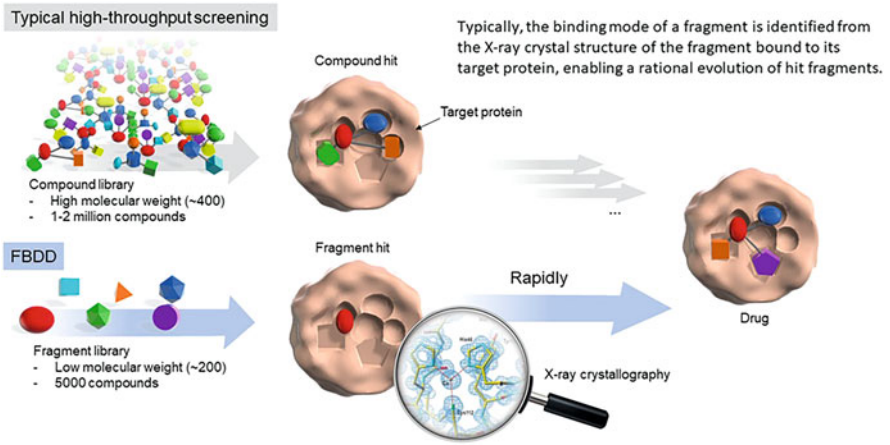


Fig. 8.1 Fundamental steps of HTS/FBDD to new hits proposition. Experimental methods, such as X-ray diffraction, are important to gather the necessary information for the correct dimensioning of the interaction site. Reproduced, by permission (CC BY-NC-ND 4.0), from Atobe (2020)

databases gather small but extremely versatile plenty of small substructures, which combined properly can generate countless candidate ligands (Fig. 8.1). A considerable computational cost may be required for this task, which often requires the use of stochastic statistical techniques to filter the most promising results and avoid redundancies (Zsoldos 2011).

8.2 Fragment Finding

The process of obtaining the fragments that fit the enzyme's interaction sites (called hot spots) involves building the library from which the fragments will be taken, identifying the enzyme's hot spots and calculating the interactions.

8.2.1 Library Building

Fragment libraries should contain substructures with low molecular weight (up to about 300 Da). Structural diversity is important, and there must be guarantees that a minimum of 500–1000 congeners are available. Each fragment must have fewer than 20 heavy atoms and low molecular complexity (Schuffenhauer et al. 2005; Jacquemard and Kellenberger 2019).

The structural diversity of the fragment database can be achieved using a pharmacophoric model as a basis (Xue et al. 2017). It usually employs fragments derived from naturally occurring compounds, as proposed by Liu and Quinn (2019).

Taking into account the huge repertoire of structures that meet the above requirements, other criteria are needed that can ensure the statistical variability of the physicochemical properties, which can be done by applying the so-called rule of three, RO3 (Brown 2016; Kirsch et al. 2019):

- Required features:
  - Molecular weight $\leq$ 300 Da;
  - Number of hydrogen bond donors $\leq$3;
  - Number of hydrogen bond acceptors $\leq$3;
  - log $P \leq 3$.
- Desired features:
  - Number of rotatable bonds (NROT) $\leq$ 3;
  - Polar surface area (PSA) $\leq$ 60.The RO3 keeps a close resemblance with Lipinski's rule of five (RO5) for oral availability of drugs: molecular weight $\leq$ 500 Da, number of hydrogen bond donors $\leq$5, number of hydrogen bond acceptors $\leq$10 and log $P \leq 5$.

Although it is known that the properties of fragments are directly related to their structure, a study focusing on the latter aspect as an independent entity of the former can be done with some success.

Hajduk and co-workers proposed that some "privileged" scaffolds are more frequent in successful fragment screening procedures (Hajduk 2006). Using known drugs as a source for common features and scaffolds is also an option (Bemis and Murcko 1996). Further comprehension can arise from the search of an optimal molecular complexity, as discussed by Hann and co-workers (Hann et al. 2001). In this case, an equilibrium between too few chemical features (which would enlarge prohibitively the fragment dataset) and too many features (that would restrict the dataset to an overwhelming specificity) is desirable.

An important aspect of this study is the synthetic feasibility, which was investigated by Schuffenhauer and co-workers (Schuffenhauer et al. 2005). Structural features that are known to hinder the synthesis of compounds (absence of reactive functional groups, presence of many condensed rings, excess of chiral centres) can be avoided in the design of FBDD fragments. This concern must also be considered taking into account the need for functional groups that can potentially interact with the proposed molecular targets.

The synthetic complexity of a generated compound can be estimated by several functions, like synthetic accessibility (SA) score, which takes into account the occurrence of non-standard structural features (large rings, non-standard ring fusions, number of chiral centres and molecule size) and synthetic complexity (SC) score, based on 12 million reactions from the Reaxys database to impose an inequality constraint to ensure that reaction products are more synthetically complex than the starting reactants (Coley et al. 2018).

8.2.2 Protein Hot Spots Identification

Once the fragment base is ready, the next step is to identify cavities and pockets in the protein structure. These pockets are possible candidates for sites of interaction with the fragments, and their determination is done by different methods, generally of a geometric and spatial nature. An example of a tool that makes it possible to determine these pockets is CASTp (Tian et al. 2018), which uses the mathematical methodologies of Delaunay triangulation, alpha shape and discrete flow (Edelsbrunner 1995; Facello 1995; Edelsbrunner and Shah 1996).

Pockets are empty hollows in the protein structure that give access to solvent molecules (represented by a spherical probe of radius 1.4 Å), but with openings (*mouths*) smaller than the respective concavities. Thus, a shallow depression in the macromolecular structure does not characterize a pocket. A *cavity*, on the other hand, is an empty space inside the structure with no external opening (i.e. no access to the solvent).

The CASTp method is based on a complex triangulation algorithm, which unites the centres of each atom situated on the walls of a concavity, represented by spheres of radius equal to the Van der Waals radii, r_{VDW}, forming polyhedra whose areas, volumes and openings are calculated (Tian et al. 2018). These will correspond to the areas, volumes and openings of the respective concavities and pockets. Thus, the method allows, through a base purely geometric, to determine all the concavities of the molecule. An output file containing the residues that are part of each pocket, as well as how area, volume and crevice size information is then obtained.

There are many further methods for determining possible sites of interaction, as summarized in Table 8.1 (Bartolowits and Davisson 2016).

Experimental methods for determining the structure of macromolecules (such as NMR and X-ray diffraction) do not always allow the identification of possible interaction sites, which makes the aforementioned approaches very useful. However, some experimental information may be useful, such as the use of known ligands, which can be isotopically labelled, allowing the determination of their location in the structure by means of NMR. The underlying principle is that the interaction affects the distribution of charges around the surrounding region, affecting the chemical shifts of neighbouring groups.

Another useful way is using sequence-based computational approaches for predicting protein–ligand binding sites, as described by Ding, Tang and Guo (2017). Computational methods that treat sequences of residues in protein structures generally use matrix algebra in sequence alignment and comparison. This work proposes the use of a mathematical procedure that extracts its feature from the matrix of residues, using the surface accessible to the solvent as the weight to determine the most exposed regions of the protein, which are potential sites of interaction with ligands.

Table 8.1 Methods for determining protein interaction sites

Method	Description	Reference/URL
CatSld	Searches for matches between catalytic sites and proteins	Kirshner et al. (2013)
DoGSiteScorer	Detects protein subpockets and predicts site druggability	https://bio.tools/dogsitescorer
HOMOLOBIND	Identifies residues that are similar to structurally characterized binding sites	http://fredpdavis.com/homolobind/
Med-SuMo	Locates similar regions on protein surfaces that are linked to certain chemical function	http://www.medit.fr
PrISE	Predicts interface residues using local surface structural similarity	Jordan et al. (2012)
SiteComp	Compares binding sites, evaluates residue contribution to binding and identifies subsites with distinct molecular interaction properties	Lin et al. (2012)
DeepPocket	Utilizes 3D convolutional neural networks for the rescoring of pockets identified by an auxiliary software	Aggarwal et al. (2021)
3DLigandSite	Candidate binding sites in proteins are inferred using known binding sites in related protein structures as templates	McGreig et al. (2022)
SURFNET	Generates molecular surfaces and gaps between surfaces from 3D coordinates supplied in a PDB format file	McGreig et al. (2022)
Q-SiteFinder	Uses the interaction energy between the protein and a simple van der Waals probe to locate energetically favourable binding sites	Laurie and Jackson (2005)
SITEHOUND-web	Identify regions of the protein characterized by favourable interactions with a probe molecule.	Hernandez et al. (2009)
P2Rank	Based on the prediction of ligandability of local chemical neighbourhoods that are centred on points placed on the solvent-accessible surface of a protein.	Krivák and Hoksza (2018)

8.2.3 Computational Prediction

Once we have obtained the fragment library and determined the active site, the next step is to build up ligands from the connection between the fragments. This connection process can be done through several strategies: fragment evolution, fragment linking, fragment self-assembly and fragment optimization (Ress et al. 2004).

In *fragment evolution*, an initial fragment is optimized by adding functionality to bind to adjacent regions of the active site (Fig. 8.2a and b).

Fejzo et al. (1999) proposed some p38 kinase inhibitors from fragment evolution, using distribution constants estimated from NMR diffusion measurements (Fig. 8.2c). A 10^6 times-fold increment in affinity is observed throughout the fragment evolution process.

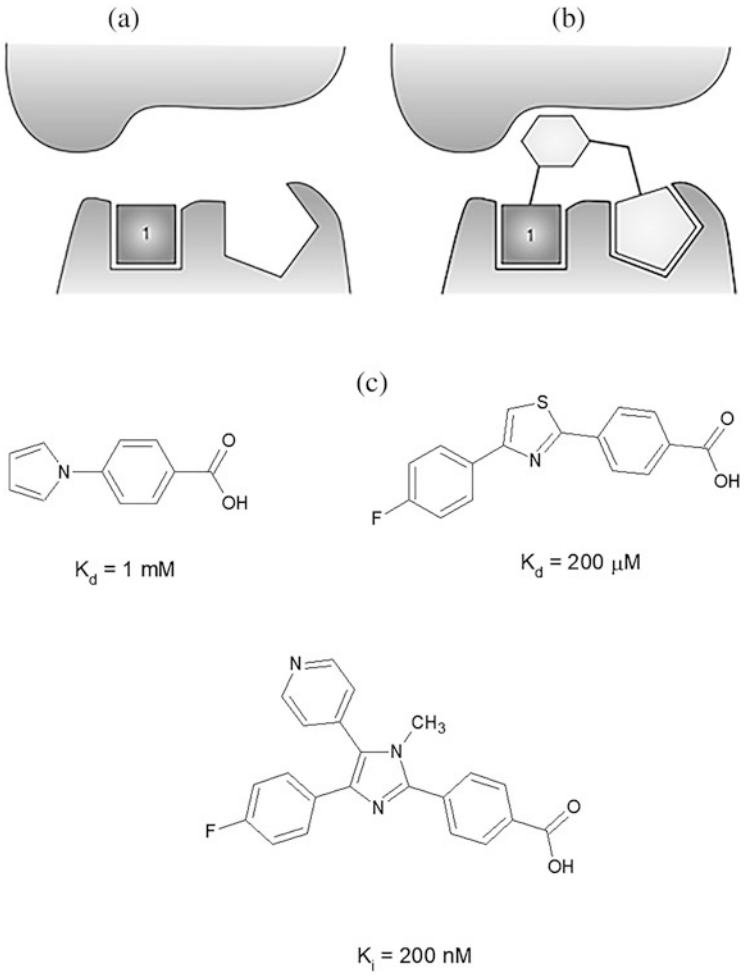


Fig. 8.2 Fragment evolution. (**a**) Fragment 1 binds to the receptor at one site. (**b**) The lead molecule is evolved by building away from the starting fragment and making good contact with the upper surface and then by growing into a second pocket. Reproduced, by permission, from Ress et al. 2004. (**c**) An example of fragment evolution: p38 kinase inhibitor design (Fejzo et al. 1999)

Fragment linking is done when two (or more) fragments, which bind to proximal parts of the active site, are joined together to give a larger, higher affinity binding molecule (Fig. 8.3a–c).

An example of fragment linking that employs an adequate spacer group is described by Pang et al. (1996), illustrated in Fig. 8.3d. The THA ligand binds to different acetylcholinesterase sites, and a new ligand can be designed by connecting both with an aliphatic chain of a size compatible with the site.

In *fragment self-assembly*, fragments with complementary functional groups are allowed to react together in the presence of the protein target and the most potent

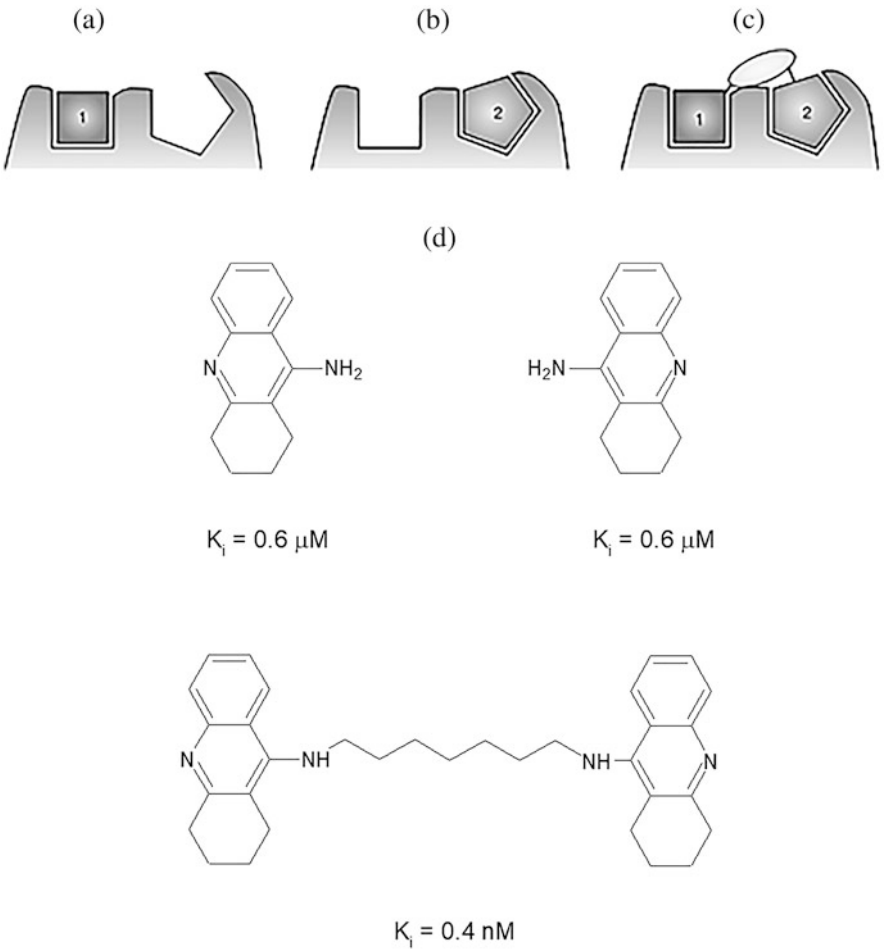


Fig. 8.3 Fragment linking. (**a**) Fragment 1 binds to the receptor at one site. (**b**) Fragment 2 binds to the receptor at an adjacent site. (**c**) Fragments are joined together by a linking group that allows the lead molecule to span both sites. (**d**) An example of fragment linking: design based on the catalytic and peripheral sites of THA in AChE. Reproduced, by permission, from Pang et al. 1996 and Ress et al. 2004

larger molecule is detected (Fig. 8.4a and b). This includes approaches usually called dynamic combinatorial chemistry (Frei et al. 2019).

Figure 8.4c illustrates this approach: the design of a ligand to cyclin-dependent kinase 2 (CDK2) from the chemical junction by means of a Schiff base condensation of two ligands specific to different sites (Congreve et al. 2003).

With *fragment optimization*, fragment approaches are used to optimize drug-like properties of a lead other than just binding affinity (Fig. 8.5a and b).

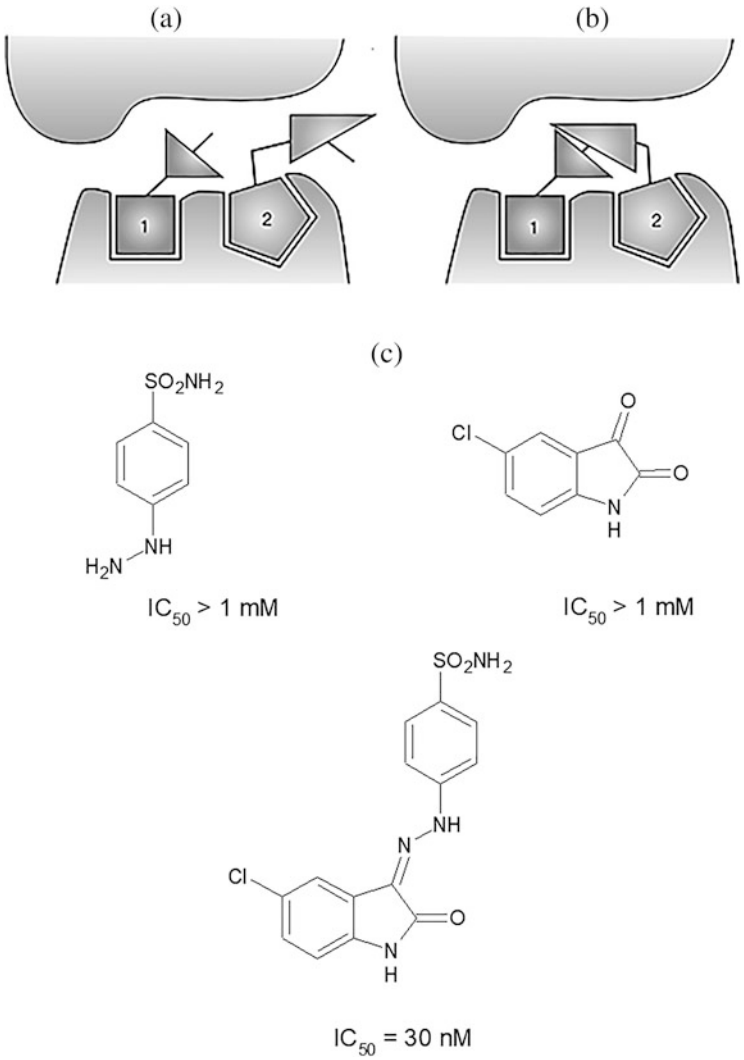


Fig. 8.4 Fragment self-assembly. (**a**) Fragments 1 and 2 bind to receptor sites simultaneously with reacting groups positioned within conformational reach of each other, increasing the effective molarity of reacting groups. (**b**) A lead molecule formed in the active site. Reproduced, by permission, from Ress et al. 2004. (**c**) An example of fragment self-assembly: both ligands can be mounted as one through a Schiff base-type condensation

Urokinase inhibitor shown in Fig. 8.5c can be optimized by introducing a hydroxyl group (Hajduk et al. 2000). That optimization covers both from a pharmacodynamic and pharmacokinetic point of view, once hydroxyl increases the water solubility, diminishing its side effects.

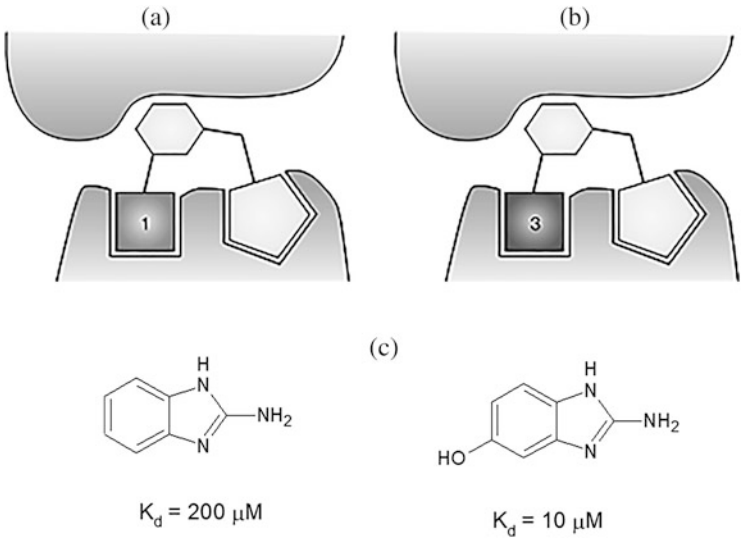


Fig. 8.5 Lead progression via fragment optimization. (**a**) Existing lead molecule discovered by fragment-based approach. (**b**) Lead molecule re-engineered to address optimization of a particular property (e.g. selectivity, cell-based activity, oral activity or efficacy). Reproduced, by permission, from Ress et al. 2004. (**c**) Example of lead progression via fragment optimization: urokinase inhibitor design

8.3 Experimental Identification of Fragments

Computational methods used in drug design, to be validated, must be supported by experimental techniques that are capable of measuring the strength and location of ligand–receptor interactions. Among these techniques, we can mention nuclear magnetic resonance, surface plasmon resonance and X-ray crystallography.

8.3.1 Nuclear Magnetic Resonance

The most common is the use of 1D-NMR, through which changes in the chemical environments of the protein's active site are monitored with the insertion of the ligand. The most used NMR techniques for this purpose are relaxation edited 1D NMR, water ligand observed via gradient spectroscopy (WaterLOGSY) and saturation transfer difference (STD) spectroscopy (Joseph-McCarthy et al. 2014).

In relaxation-edited 1D NMR, ligand binding to a macromolecule is accompanied by shortened relaxation properties of the complexed relative to the free ligand, and $T_{1\rho}$ (longitudinal relaxation) or T_2 (transverse relaxation) filter is applied to the pulse sequence. Fragments that bind to the target exhibit a loss in signal intensity in the NMR spectrum, relative to fragments that do not (Hajduk et al. 1997).

The use of NMR to predict binding ability can be explained by the fact that certain NMR observable parameters (P), such as transverse and longitudinal relaxation rates (R_1 and R_2), the nuclear Overhauser effect (NOE) and the diffusion coefficient, vary according to molecular size and shape and depend on the interactions with larger molecules. Considering the balance between the free ligand (L) and the free protein (E) and in its bound form (E-L),

$$E + \ L \underset{K2}{\overset{k1}{\rightleftarrows}} E - L \tag{8.1}$$

The dissociation constant is given by $K_d = k_2/k_1$; ligands with strong binding present k_2 small values, whereas ligands that bind weakly present higher k_2 values. The magnitude of the binding will depend on the observable parameter (Pobs), which in its turn derives from the free and bound states parameter values (Ni 1994):

$$P_{\text{obs}} = (1 - p_b)P_{\text{free}} + p_b(P_{\text{bound}}) \tag{8.2}$$

In Eq. (8.2), p_b is the fraction of the ligand in the bound state. If the ligand concentration is much smaller than the enzyme's one, there will be slow exchange on the NMR time scale and p_b is nearby 1 and $P_{\text{obs}} \sim P_{\text{bound}}$. On the other hand, ligands with fast exchange shall cause p_b small and $P_{\text{obs}} \sim P_{\text{free}}$. In this last case, parameters will be closer to the free ligand's ones.

NMR gradient spectroscopy (WaterLOGSY) makes use of the interactions between the target and the ligand that are mediated by water molecules, which cause a change in the profile of the relaxation times related to the interactions between ligand and water in the bound state or in the free state (Dalvit et al. 2001). The cross-relaxation rate of dipole–dipole interaction between water and ligand is positive for free ligands and negative for bound ligands nearby water molecules associated with the protein.

The saturation transfer difference NMR (STD-NMR) technique is based on the nuclear Overhauser effect (NOE) and is used for measuring through-space distances. The protein target is irradiated by a radiofrequency field specific for the protein nuclei and removes their magnetic polarization, generating an on-resonance (ISAT) spectrum. The fast exchange between the free and protein-bound form causes a saturation transfer through the protein to the bound ligand and that saturation is carried on to the free ligand where it is detected. A magnetization saturation of the target takes place, which in its turn is transferred to the bound ligand, and not to the free one. A "difference spectrum" is then obtained by subtracting the spectrum obtained with the magnetization saturated target from the spectrum obtained with the non-irradiated target. A deuterated solvent is required, and a shorter selectivity is obtained, when compared to the other techniques (Feyfant et al. 2011; Joseph-McCarthy et al. 2014).

Alongside edited 1D NMR, WaterLOGSY and STD, there are some further methods involving solution NMR that are applicable to the study of protein-binding

interaction and are useful as a means of validating FBDD studies, according to Becker et al. (2018). In *chemical shift mapping*, the ligand bind promotes changes in the local electron density and hydrophobicity at the interaction site, which can be measured through the NMR of the chemical shifts (Williamson 2013). Those changes in chemical shifts are also induced by proximity magnetic susceptibility anisotropic groups, like aromatic rings. This approach is called chemical shift mapping (CSM), chemical shift perturbation (CSP) or complexation-induced changes in chemical shifts (CIS). It is common to use isotopically labelled proteins, as in the ^{15}N-heteronuclear single quantum correlation (HSQC) method, which uses proteins uniformly labelled with ^{15}N, produced from genetically engineered *E. coli*.

Hydrogen exchange is a method that uses residue-specific information about the solvent accessibility in a protein that may derive from exchange rates of backbone amide hydrogen with bulk water. Amide hydrogen protected from bulk water by the ligand exhibit reduced exchange rates, measured by monitoring signal intensity changes of amide protons upon dissolving the protein in deuterated water. Only slowly exchanging amide protons can be measured, as done by Paterson et al. (1990).

Finally, *solvent paramagnetic relaxation enhancement* (sPRE) is a method based upon the addition of an inert and freely soluble paramagnetic agent to a protein solution, which causes an increase in the relaxation of protein nuclei (Pintacuda and Otting 2002; Respondek et al. 2007). The relaxation enhancement shall depend on the distance between the observed nucleus and the paramagnetic probes nearby, proportional to $1/r^6$ (r is the distance between the paramagnetic centre and the observed nucleus). Nuclei closer to the site show lower sPREs than more solvent-exposed ones.

Cross-saturation/transferred cross-saturation (CS) is a technique, related to STD, for mapping the binding area between two proteins, requiring specific isotopic labelling of one protein, namely with ^{2}D and ^{15}N, whereas the other protein is unlabelled and a solvent 10% H_2O and 90% D_2O mixture is used. The complex is irradiated at a frequency, which only affects the unlabelled protein II (Takahashi et al. 2000). Proton–proton NOEs are small and positive for small molecules and large and negative for large molecules, and this behaviour is exploited in *transferred NOE (trNOE)* experiments. A small ligand experiences a rapid exchange between its free and bound forms, with a positive NOE for the first situation and a negative NOE for the second, which can be observed in spectra with an adequate resolution (Clore and Gronenborn 1983; Ni 1994).

Paramagnetic tags are also useful for the sake, when a paramagnetic centre in a protein can be used to probe ligand binding around the interaction site. Unpaired electrons of paramagnetic probes augment relaxation rates of nuclei up to a distance of about 20 Å, in a phenomenon known as paramagnetic relaxation enhancement (PRE), causing a line broadening in the spectra. Organic radicals and paramagnetic lanthanide compounds (such as Dy^{3+}, Tb^{3+} and Tm^{3+}, and more moderately, Er^{3+}, Ho^{3+} and Yb^{3+}) are specially used for this purpose (Reuben and Leigh 1972; Saio et al. 2011).

8.3.2 Surface Plasmon Resonance

Surface plasmon resonance (SPR) is a technique that can be used to measure the binding affinity, specificity and kinetic parameters of the interactions between macromolecules and ligands (Tiwari et al. 2021; Yesudasu et al. 2021). The SPR technique has the advantage of being non-destructive and does not require elaborate sample preparation and can be applied to the study of noncovalent interactions of great interest in biochemistry, pharmacology and other areas (Bakhtiar 2013).

The target protein is immobilized on a gold or silver sensor surface, and a solution of probe flows over the target surface and induces an increase in the refractive index when binding takes place. Studies using SPR involving vaccines (Hearty et al. 2010), drug–DNA interaction (Wolf et al. 2007), antigen–antibody (Ramakrishnan et al. 2009), carbohydrate–nucleic acids (Greenberg et al. 1999), protein–carbohydrate (Smith et al. 2003) and protein–ligand (Sabban 2011), conformational changes in enzymes (Salamon et al. 1994) and in vesicular phospholipids (Salamon et al. 1994) have also been described. From the moment in which the interaction occurs, a change in the refractive index causes the total reflection of the light beam, which is detected, with the angle of reflection being correlated with the interaction parameters (such as inhibition constants).

The immobilization of the sample and preparation of the metallic surface can be done, essentially, in three ways: (1) a direct immobilization in which the ligand or binding molecule is covalently attached to the sensor gold surface using one of several established functional moieties, like amines, thiols, maleimide or aldehyde moieties (Kuroki and Maenaka 2011); (2) indirect noncovalent coupling via a high-affinity capture molecule, which must present sufficient binding to avoid ligand dissociation, such as monoclonal antibodies, avidin-biotin via biotinylation or histidine-tagged recombinant proteins (Jason-Moller et al. 2006); (3) use of membrane protein anchoring agents, such as adsorption of lipids from liposomes or micelles to the sensor chip or assembling a bilayer attachment involving on-surface binding of intact membranes (Vikholm et al. 1996; Jung et al. 2000). A surface-bound vesicle can also be used (a so-called on-surface reconstitution, OSR).

8.3.3 X-Ray Crystallography

Obtaining crystallographic structures of macromolecules represented important progress in the development of new drugs. Banks of X-ray structures have become popular, providing information at an increasing level of accuracy, such as the Brookhaven Protein Database, PDB (Berman et al. 2000). Crystals can be obtained by using the co-crystallization method (the ligand is added to the mixture before crystal formation starts) or the soaking technique (the ligand is added directly to a mixture with pre-existing crystals). An example of such an approach was shown in Fig. 8.1 (in which the target protein is represented by its X-ray diffraction structure).

The crystallographic structures of molecular targets related to a given disease, widely available in databases such as PDB, must be chosen considering some criteria, such as resolution, the organism from which the protein was extracted, the completeness of the structure and the presence of small molecules complexed during the crystallization process, which can serve as a reference for possible sites of interaction (McPherson and Cudney 2014).

X-ray diffraction structures are more abundant in the databases than structures determined by other methods, such as NMR and theoretical study, although those obtained by NMR have the advantage of being composed of different conformations in equilibrium in the aqueous phase, which represents a more realistic system (Puthenveetil and Vinogradova 2019).

An advantage of X-ray diffraction, however, is the possibility of including the water molecules present in the interaction sites in the process of building ligands from fragments with better interaction, which the literature has shown to be important in molecular recognition (Sun et al. 2014; Rudling et al. 2018; Matricon et al. 2021).

A comparison between the determination of structures from X-ray diffraction and structures obtained by nuclear magnetic resonance can be made. X-ray diffraction presents the following advantages:

- It allows direct mathematical reconstruction of the image of the molecule (electron density map).
- It is applicable to large molecules and complexes.
- Data processing is highly automated.
- Presents analysis quality indicators available are well evaluated (R factor, resolution).
- It enables the inclusion of water molecules; in general, it has high accuracy.

Some advantages of X-ray diffraction are as follows:

- There is a need for the formation of stable crystals.
- Crystallization techniques are quite empirical.
- Measurements are performed in a solid state (not physiological).
- There is a need to use radiation ionizing, capable of causing damage to the structure.
- It is difficult to distinguish between static and dynamic disorders.
- Loss of dynamic and trend information occurs to “freeze” flexible regions within the crystal lattice.

NMR, on the other hand, has the following advantages against X-ray diffraction:

- It is applicable in solution, which makes the technique more relevant from a biological point of view).
- It is more informative regarding the dynamics of molecule; it can be applied to isolated domains or large protein “modules”.
- Physiological conditions can be altered (pH, temperature, etc.).

NMR limitations concerning our comparison with X-ray diffraction are as follows:

- In general, the accuracy of the coordinates atomic is smaller; there is a limit to the size of the structures that can be studied (up to 20,000 Da).
- It is difficult to determine the orientation of domains in a structure with more than one domain or modular proteins.
- There is no direct parameter analogous to the resolution of a crystalline system (as in X-ray), which allows the assessment of accuracy; it requires concentrated solutions (not physiological).

Congreve et al. (2003) explain how dynamic combinatorial chemistry (DCC) allows specific members of a combinatorial library to be selected and amplified with the use of a template. The reaction connecting the building blocks is reversible, and an interchange between the different members of the dynamic combinatorial library takes place (Seneci 2000). The authors report an approach in which ligands are observed directly by X-ray crystallography from their electron density maps from crystals exposed to a dynamic combinatorial library mixture.

8.3.4 Thermal Shift Assay

Thermal shift assay (TSA) is a technique that monitors the changes undergone by a protein when it interacts with another molecule by measuring its denaturation temperature. Protein stability depends on a number of factors: pH, ionic strength of the medium, presence of cofactors, mutations, etc. Correlated experimental parameters that are related to structural stability and that are sensitive to a temperature gradient, such as differential scanning fluorimetry (DSF), can be used to measure the ligand–macromolecule interaction (Senisterra et al. 2012). The fluorescence of a protein solution depends on a temperature gradient, and the addition of fluorescence dye exhibits a low fluorescence signal in a polar environment and high signal in a non-polar environment (Lo et al. 2004). The interaction of the protein with the ligand causes exposure of polar regions previously solvated by water molecules that are dispersed with the entry of the ligand, at the same time that it unbalances the distribution of non-polar zones.

This same principle also applies to nucleic acids: the temperature of double-stranded and G-quadruplexes DNA denaturation has been used, for example, to study the interaction of ligands to the minor groove (Record et al. 2003; Senisterra et al. 2012) or as intercalators (Morita et al. 2021). TSA has gained relevance as a validation technique for FBDD studies alongside the years (Romasanta et al. 2018), due to its reproducibility and the possibility of the evaluation of important thermodynamic effects, such as solvation effects (Magsumov et al. 2020).

Fragments containing strong nucleophile ends (like –SH) are expected to lead to compounds capable of establishing covalent bonds with the target, and the thermal shift measured shall be larger than that observed with irreversible ones. Covalent ligands are usual in anticancer chemotherapies (like cisplatin), which in their turn are interesting candidates for drug repositioning (Lotfi Shahreza et al. 2020). The

corresponding temperature curves obtained in those cases must present higher denaturation temperatures, once the covalent bonds require high-energy gains to disrupt the tridimensional protein structure.

8.3.5 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) allows the measuring of thermodynamic parameters such as enthalpy changes (ΔH) and Gibbs free energy (ΔG), which in its turn allows for the calculation of the binding affinity (Leavitt and Freire 2001), as well as stoichiometry n in one single experiment (Damian 2013).

Each ligand injection (i) releases or absorbs a certain amount of heat (qi) proportional to the amount (in moles) of ligand that binds to the protein, which is given by $v \times \Delta L_i$ (v is the volume of ligand, and ΔL_i is the increment in ligand concentration) and the characteristic binding enthalpy (ΔH) for the reaction:

$$q_i = v \times \Delta H \times \Delta L_i \tag{8.3}$$

By repeating the experiment in several temperatures, it is possible to obtain the heat capacity (ΔC_p) associated with binding:

$$\Delta C_p = \frac{\partial \Delta H}{\partial T} \tag{8.4}$$

The ITC instrument has two cells, a reference cell as a buffered control and a sample cell containing the macromolecule and the ligand, which is titrated into the mixture (Wiseman et al. 1989). Modern ITC instruments operate on the heat compensation principle, when the measured signal is expressed as microcalories per second and is the necessary power to maintain constant the temperature difference between the reaction and reference cells (Leavitt and Freire 2001).

The main advantage of this method is the discrimination it allows between enthalpic and entropic contributions to the binding. It is usual to calculate the enthalpic efficacy index (EE), to designate the enthalpic contribution to the binding. Polar interactions are more easily determined, which represents a limitation of the use of this method in FBDD, which employs many non-polar fragments, with weak interactions (Rees 2016). Another disadvantage is a large amount of protein required (Kirsch et al. 2019).

$$\mathrm{EE} = \frac{\Delta H}{Q} \ (Q = \text{number of heavy atoms}) \tag{8.5}$$

A drawback in enthalpic estimations by ITC is the difficulty to determine the energies required to bring protein and ligand from free to bound conformations. The effective binding free energy (ΔG_{bind}) must be

$$\Delta G_{\text{bind}} = \Delta G^{o}_{\text{bind}} + \Delta G_{\text{conf}} \tag{8.6}$$

In Eq. (8.6), $\Delta G^{o}_{\text{bind}}$ is the free energy calculated under the assumption that free and bound conformations are exactly the same, and ΔG_{conf} is the free energy of chance from free to bound conformations (Luque and Freire 2000).An example of the application of ITC in FBDD is given by Drinkwater et al. (2010), with the investigation of phenylethanolamine N-methyltransferase inhibitors, important to the central nervous system essential synthesis of adrenaline. Mashalidis et al. (2013) also describe a screening of a library of low molecular weight compounds (fragments) which combines diverse experimental techniques: a preliminary screening using differential scanning fluorimetry (DSF), followed by NMR spectroscopy validation and the characterization of binding fragments by isothermal titration calorimetry (ITC) and X-ray crystallography.

8.3.6 Mass Spectrometry

Mass spectrometry (MS) is a technique widely used in the structural determination of organic compounds. It is based on the fragmentation of a molecule (e.g. through collision with a beam of high-energy electrons) and the determination of the structure that generated such fragments from the systematic comparison of their masses. Many mass spectrometers are coupled to other instruments, such as gas chromatographs.

MS provides advantages over their counterpart techniques for allowing weak binding detection due to its high sensitivity. Besides, few sample amount is required, no modifications or labelling of the protein target is needed, and direct visualization of all species in solution alongside the binding process is feasible (Chan et al. 2017; Vu et al. 2018). The use of milder ionization techniques, such as non-denaturing electrospray ionization (ESI), is always more interesting because it preserves the structural integrity of the protein:ligand complex. An example of this application is described by Liu and Quinn (2019): 643 natural products from fragment-sized library with low molecular weight were screened against 62 potential protein targets for malaria, which led to 96 low molecular natural products capable of binding and 79 fragments that could inhibit the growth of malaria parasites in vitro. Another worth-citing case is the discovery of a benzimidazole moiety with high affinity to a 29-mer RNA model, identified from an 18,0000 fragment library using MS-based screening methods (Seth et al. 2005).

8.4 FBDD Strategies

Some FBDD strategies can be pointed out, whose choice is done according to the information available or needed.

8.4.1 Chemical Biology Exploration of Biological Targets

This approach is based on the search for possible sites of interaction in the structure of the target molecule and the library fragments. One way to perform this procedure is called MSCS (multiple solvent crystal structures), which uses solvent molecules as bridges between the site and the ligands (Allen et al. 1996). Although water is most commonly used for this sake, organic solvents can also be employed, such as acetonitrile, as described by Fitzpatrick et al. (1993). Another way to use the target structure in the selection of the best fragments is the use of the druggability of the site, using for this purpose properties such as polar and apolar surface area, surface complexity and pocket dimensions, determined from NMR data, as reported by Hajduk et al. (2005).

8.4.2 FBDD as HTS Complimentary

Since the process of developing a new drug depends a lot on the appropriate choice of leads, the methods comprised by FBDD play an important role due to their ability to propose leads of high quality and specificity for the target studied. There is a growing interest in methods that combine FBDD with high-throughput screening (HTS), due to the possibility of covering a wider set of structures. In addition, fragments exhibit the advantage over collections of compounds of presenting better values of ligand efficiency (LE):

$$\mathrm{LE} = -\frac{\Delta G}{N} = -\frac{RT\ln(K_d)}{N} \approx -\frac{RT\ln(\mathrm{IC}_{50})}{N}, \tag{8.7}$$

where ΔG is the binding free energy, N is the number of heavy atoms in the ligand, K_d is the dissociation constant, and IC_{50} is half of the maximum inhibitory concentration, usually expressed by its negative logarithm, pIC_{50} (Hopkins et al. 2004; Ress et al. 2004). Additional criteria may be used, like lipophilic LiPE (Ryckmans et al. 2009),

$$\mathrm{LiPE} = \mathrm{pIC}_{50} - \log P, \tag{8.8}$$

as well as Lipinski's rule of five parameters, as discussed above in this chapter.

Grädler et al. (2019) describe an FBDD/HTS combination protocol to build a chemically diverse hit database of peptidyl-prolyl isomerase cyclophilin D (CypD) ligands, obtaining three different series with either urea, oxalyl or amide linkers with millimolar affinities.

8.4.3 Build-Up Core FBDD for Drug Discovery

Building new hits from "pure" fragments is always an interesting option when there are no likely candidates. Several programs for this purpose are known, including CAVEAT (Lauri and Bartlett 1994), HOOK (Eisen et al. 1994) and SEED (Marchand and Caflisch 2018). CAVEAT is based on an algorithm that searches for the fragment options that best fit the site, using pairs of vectors. HOOK, in its turn, derives functional group sites with defined positions and orientations from known ligand structures or the multicopy simulation search (MCSS) method (Miranker and Karplus 1991), placing molecular "skeletons" from a database into the protein-binding region and connecting them. At last, SEED uses a docking protocol developed for fragment docking and binding energy evaluation by a force field with implicit solvent.

Fragment design using docking methodology must face the common challenges involved in such tasks: (1) the choice of the search algorithms and (2) the scoring functions (Torres et al. 2019). Search algorithms based on systematic search often have to deal with a combinatorial explosion of conformations, which can be remediated by the use of protein and ligand pharmacophore. Stochastic search algorithms (such as Monte Carlo and evolutionary algorithms) perform random changes in the ligand's degrees of freedom, but there is no warranty of convergence to the best solution, which usually requires an iterative process. Finally, deterministic search employs the orientation and conformation of the ligand in each iteration from the previous state, which may demand a large computational cost.

Common scoring functions may include the following: binding energy, free energy or interaction energies. Modern force fields group their scoring functions in force field, empirical and knowledge-based (Guedes et al. 2014).

8.5 Case Studies

8.5.1 Case 1: Pseudo-Natural Products

Liu and Quinn (2019) describe the search of novel anti-parasitic drugs from FBDD strategies, designing "pseudo-natural products" obtained by the modification of native natural products, matching some requirements, such as smaller structures with less reactive sites and more chiral sp^3 centres, molecular weight between 150 and 300 Da, clogP <3. The fragments were derived from well-known drugs

like FK 506 (tacrolima), sanglifehrin, cytochalasin, massarigenin, renieramycin and sparteine.

A synthetic combination of unrelated natural product fragment types was also used in the design of pseudo-natural products. An attempt to combine fragments accessible compounds via classical biosynthetic routes into novel fragments that could be obtained via new artificial biosynthetic pathways, as reported by Klein et al. 2014.

The construction of libraries of fragments that mimic natural products must take into account the set of requirements of Lipinski's rules and the preservation of scaffolds that are essential for biological activity. Several databases of natural products can be used as starting points, such as the Dictionary of Natural Products (Harborne 1995). The authors divided it into 64,650 fragment-sized (MW 100–300 Da) and 145,623 natural products with an MW > 300 and employed Spider software (Reker et al. 2014) to predict the targets of 23,340 (36%) of the low MW natural products. Non-flat (C sp^3-rich) and 2-ring-sized fragments were privileged, accounting for 643 compounds out of the former 23,340 ones.

FBDD-HTS combination comprising pseudo-natural products was used to find novel dengue virus helicase (NS3 DENV Hel) inhibitors and dengue virus methyltransferase (NS5 DENV MTase), with good results, confirmed by thermal shift assays (Coutard et al. 2014; Benmansour et al. 2017). A combination of homology modelling, fragment docking, chemical similarity and structural filters allowed the authors to identify hits against a homology model of DENV NS2B-NS3 protease, generated from five WNV and DENV protease template structures (Knehans et al. 2011). Additional filtering to avoid chemically infeasible molecules and to include compounds containing a terminal cationic or basic moiety to favour interactions with the S2 pocket in DENV NS2B-NS3 protease resulted in 18,803 hits. A further search for compounds capable of interacting with both S1 and S2 pockets and some additional criteria based on docking resulted in 23 promising candidates, submitted to in vitro assays.

8.5.2 Case 2: SARS-CoV-2 Main Protease Inhibitors

The COVID-19 pandemic has posed enormous challenges for the development of specific drugs, and many researches involving rational drug planning methods have been described in the literature. Although greater emphasis has been given to immunological treatments, pharmacotherapy still plays an important role in combating this disease.

Some FDA-approved drugs obtained from in silico approaches are remdesivir, saquinavir and darunavir. Some flavone and coumarin derivatives act as potential inhibitors of human SARS-CoV-2 main protease (Khan et al. 2021). Some drugs have been tested against SARS-CoV-2, like chloroquine, lopinavir, nafamostat, hydroxychloroquine, ritonavir, camostat, corticosteroids and sarilumab (Shaffer 2020).

Andola and co-workers (Andola et al. 2022) describe the search of suitable inhibitors for the main protease in SARS-CoV-2 using FBDD. The authors used a fragment database derived from Auto Core Fragment in silico Screening (ACFIS) 2.0 web server (Hao et al. 2016). The structures obtained were docked against the enzyme using PyRx (Dallakyan and Olson 2015) and AutoDock (Morris et al. 2009), combining the results with a property prediction protocol using SwissADME server (Daina et al. 2017) and a molecular dynamics performed by GROMACS (Hess et al. 2008). A SARS-CoV-2 main protease complexed with 3WL molecule was used, which was defragmented into its chromene and phenyl rings. ADME properties taken into account in the choice of the best compounds were as follows: total polar surface area (Ertl et al. 2000), water solubility, lipophilicity (log P), skin permeation (Potts and Guy 1992) and synthetic accessibility (in a 1–10 scale).

Considering the binding free energies and the ADME properties, four-hit molecules were chosen containing 1H-benzo[d]imidazol-5-yl, 3H-indol-3-yl, isoxazol-5-ylmethyl and 6-fluoro-1,2,3,4-tetrahydronaphthalen-2-yl formate groups as fragments.

8.5.3 Case 3: p38 Mitogen-Activated Protein Kinases

p38 mitogen-activated protein kinases belong to the mitogen-activated protein kinases (MAPKs) group, are responsive to stress stimuli and are involved in cell differentiation, apoptosis and autophagy. Activation of the p38 MAPK pathway in muscle satellite cells due to ageing in a continuous way impairs muscle regeneration (Cosgrove et al. 2014; Segalés et al. 2016).

Fejzo et al. (1999) used an NMR approach to distinguish between binding and non-binding modes of a library of soluble, low molecular weight compounds, against p39 MAPK targets—called "SHAPE". The fragments collection is based on the Comprehensive Medicinal Chemistry (CMC) database. Thirty-two different frameworks, or "shapes", describing ~50% of all known drugs, were yielded from this effort. The inclusion of atom type and bond order in the analysis yielded 2506 complex frameworks describing 5120 entries and 41 frameworks describing 24% of all drugs.

A subsequent search taking into account the synthetic complexity or the absence of sufficiently soluble analogues led to a small, but efficient, dataset of fragments within MW 68–341 Da (average 194 Da), containing 6–22 heavy atoms, and a calculated log P of −2.2 to 5.5.

The one-dimensional (1D NMR) spectra of two small-molecule compounds (nicotinic acid and 2-phenoxybenzoic acid) in the presence and absence of p38 (ligand:protein = 5:1) exhibited line-broadening effects and attenuation of fine structure for 2-phenoxybenzoic acid ($K_d \sim 70$ μM), which indicates binding. A slight relationship between the peak heights and the affinity was also observed.

An additional 2D NMR analysis was a pursuit with this system (two-dimensional transferred NOE experiment—tNOE or tNOESY), allowing to distinguish between a pair of compounds competing for the same targets.

8.6 Conclusion and Future Perspectives

Modern analytical techniques serve as a counterpoint and validation method for the process of generating new bioactive chemical entities, and their steady progress positively affects the quality of use of the fragments.

Recent advances in analytical techniques such as surface plasmon resonance (SPR), discussed in this chapter, allow the addressing of protein interaction sites in a non-destructive and highly specific basis, which can also be applied to nucleic acids, membrane receptors and ion channels (Tiwari et al. 2021; Yesudasu et al. 2021). NMR WaterLOGSY (Bataille et al. 2020) also provides a deeper insight into the nature of the locales and driving forces of the fragment–target interactions.

Dynamic combinatorial chemistry (DCC) is also an interesting approach to generate directed ligand libraries for macromolecular targets, taking into account its ability to reversibly react with building blocks and reach a stable thermodynamic equilibrium. An extension of the technique is called target-directed DCC (tdDCC), which identifies potent ligands for pharmacologically relevant targets (Frei et al. 2019).

A natural evolution of FBDD is its combination with the capabilities of three-dimensional quantitative structure–activity relationships (3D QSAR). While classic QSAR correlates (by means of statistical methods of multivariate regression) the physicochemical properties of a series of compounds with their biological activity (Oliveira 2022), 3D QSAR correlates the values of electronic and steric energy calculated for specific probes, posed around the molecules within a grid of dots, with their biological activity. Such effort yields three-dimensional maps that indicate where positive or negative groups favour interaction with the target and where small or large groups do so.

A method that became popular between the 1990s and 2000s was drug design by the *de novo* approach using the LeapFrog (LF) program (Durdagi et al. 2008), available in the Sybyl package (Sybyl 2001). LF performs molecular evolution or electronic screening through a systematic structural change screening, weighting the results according to the target structure, when available (using the Cavity protocol). If there is no target structure, 3D QSAR maps provide a pharmacophoric model as an adequate portrait of its profile. Three alternative modes can be used for this purpose: OPTIMIZE mode suggests improvement to existing leads; DREAM mode proposes new molecules expected to have good binding, and GUIDE mode supports interactive design by performing and evaluating user modifications (Makhija et al. 2004). Binding energy was calculated encompassing three major components: steric and electrostatic enthalpies; cavity desolvation energy; and ligand desolvation energy.

Holistic methodologies, such as integrated biophysical approach proposed by Silvestre et al. (2013), using combined techniques (thermal shift, 1D NMR, ITC and X-ray diffraction) open up a wide scenario of exploitation of new library portraits with a more robust experimental scaffold, underlying the choice of the best ligands.

Another interesting challenge is to establish a rationale about druggability and molecular shape, as has been studied in the 1990s by Bemis and Murcko (1996) and remains open to investigation. Although this work dealt only with topological features of 2D fragments, a limited number of graph portraits seem to lead to potential drugs. This concept may be useful in the construction of specific fragment libraries with topological constraints.

The possibilities of FBDD depend on the development of higher computational processing capabilities, as well as the availability of specific fragment libraries for specific diseases and conditions. The COVID pandemic issue brought out new demands, mainly concerning drug repurposing, including toxicity studies. Notwithstanding, this field offers a low-cost route to novel chemical entities to be synthesized and submitted to in vitro and in vivo tests, when compared to traditional combinatorial screening as done by pharmaceutical companies.

References

Aggarwal R, Gupta A, Chelur V et al (2021) DeepPocket: ligand binding site detection and segmentation using 3D convolutional neural networks. J Chem Inf Model 62:5069. https://doi.org/10.1021/ACS.JCIM.1C00799/ASSET/IMAGES/LARGE/CI1C00799_0007.JPEG

Allen KN, Bellamacina CR, Ding X et al (1996) An experimental approach to mapping the binding surfaces of crystalline proteins. J Phys Chem 100:2605–2611. https://doi.org/10.1021/jp952516o

Andola P, Pagag J, Laxman D, Guruprasad L (2022) Fragment-based inhibitor design for SARS-CoV2 main protease. Struct Chem 33:1467–1487. https://doi.org/10.1007/S11224-022-01995-Z

Atobe M (2020) Fragment-based design of isoquinoline derivatives as anti-inflammatory drugs. *Research Outreach*. https://researchoutreach.org/articles/fragment-based-design-isoquinoline-derivatives-anti-inflammatory-drugs/. Accessed 21 Sep 2022.

Bakhtiar R (2013) Surface plasmon resonance spectroscopy: a versatile technique in a biochemist's toolbox. J Chem Educ 90:203–209. https://doi.org/10.1021/ED200549G/ASSET/IMAGES/LARGE/ED-2011-00549G_0006.JPEG

Bartolowits M, Davisson VJ (2016) Considerations of protein subpockets in fragment-based drug design. Chem Biol Drug Des 87:5–20. https://doi.org/10.1111/CBDD.12631

Bataille CJR, Rabbitts TH, Claridge TDW (2020) NMR waterLOGSY as an assay in drug development programmes for detecting protein-ligand interactions–NMR waterLOGSY. Bio-protocol 10:e3666. https://doi.org/10.21769/BioProtoc.3666

Becker W, Bhattiprolu KC, Gubensäk N, Zangger K (2018) Investigating protein–ligand interactions by solution nuclear magnetic resonance spectroscopy. ChemPhysChem 19:895–906. https://doi.org/10.1002/CPHC.201701253

Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. J Med Chem 39:2887–2893. https://doi.org/10.1021/JM9602928/ASSET/IMAGES/LARGE/JM9602928FB10A.JPEG

Benmansour F, Trist I, Coutard B et al (2017) Discovery of novel dengue virus NS5 methyltransferase non-nucleoside inhibitors by fragment-based drug design. Eur J Med Chem 125:865–880. https://doi.org/10.1016/J.EJMECH.2016.10.007

Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/NAR/28.1.235

Brown N (2016) In silico medicinal chemistry. The Royal Society of Chemistry, Cambridge

Chan DSH, Whitehouse AJ, Coyne AG, Abell C (2017) Mass spectrometry for fragment screening. Essays Biochem 61:465–473. https://doi.org/10.1042/EBC20170071

Clore GM, Gronenborn AM (1983) Theory of the time dependent transferred nuclear Overhauser effect: applications to structural analysis of ligand-protein complexes in solution. J Magn Reson 53:423–442. https://doi.org/10.1016/0022-2364(83)90215-9

Coley CW, Rogers L, Green WH, Jensen KF (2018) SCScore: synthetic complexity learned from a reaction corpus. J Chem Inf Model 58:252–261

Congreve MS, Davis DJ, Devine L et al (2003) Detection of ligands from a dynamic combinatorial library by X-ray crystallography. Angew Chem Int Ed 42:4479. https://doi.org/10.1002/anie.200351951

Cosgrove BD, Gilbert PM, Porpiglia E et al (2014) Rejuvenation of the muscle stem cell population restores strength to injured aged muscles. Nat Med 20:255. https://doi.org/10.1038/nm.3464

Coutard B, Decroly E, Li C et al (2014) Assessment of Dengue virus helicase and methyltransferase as targets for fragment-based drug discovery. Antiviral Res 106:61–70. https://doi.org/10.1016/J.ANTIVIRAL.2014.03.013

Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 71(7):1–13. https://doi.org/10.1038/srep42717

Dallakyan S, Olson AJ (2015) Small-molecule library screening by docking with PyRx. Methods Mol Biol 1263:243–250. https://doi.org/10.1007/978-1-4939-2269-7_19

Dalvit C, Fogliatto GP, Stewart A et al (2001) WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. J Biomol NMR 21:349. https://doi.org/10.1023/A:1013302231549

Damian L (2013) Isothermal titration calorimetry for studying protein-ligand interactions. Methods Mol Biol 1008:103–118. https://doi.org/10.1007/978-1-62703-398-5_4/COVER

Ding Y, Tang J, Guo F (2017) Identification of protein-ligand binding sites by sequence information and ensemble classifier. J Chem Inf Model 57:3149–3161. https://doi.org/10.1021/ACS.JCIM.7B00307/ASSET/IMAGES/LARGE/CI-2017-003074_0005.JPEG

Drews J (2000) Drug discovery: a historical perspective. Science 287:1960

Drinkwater N, Vu H, Lovell KM et al (2010) Fragment-based screening by X-ray crystallography, MS and isothermal titration calorimetry to identify PNMT (phenylethanolamine N-methyltransferase) inhibitors. Biochem J 431:51–61. https://doi.org/10.1042/BJ20100651

Durdagi S, Mavromoustakos T, Chronakis N, Papadopoulos MG (2008) Computational design of novel fullerene analogues as potential HIV-1 PR inhibitors: Analysis of the binding interactions between fullerene inhibitors and HIV-1 PR residues using 3D QSAR, molecular docking and molecular dynamics simulations. Bioorg Med Chem 16:9957–9974. https://doi.org/10.1016/J.BMC.2008.10.039

Edelsbrunner H (1995) The union of balls and its dual shape. Discret Comput Geom 133(13):415–440. https://doi.org/10.1007/BF02574053

Edelsbrunner H, Shah NR (1996) Incremental topological flipping works for regular triangulations. Algorithmica 153(15):223–241. https://doi.org/10.1007/BF01975867

Eisen MB, Wiley DC, Karplus M, Hubbard RE (1994) HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. Proteins 19:199–221. https://doi.org/10.1002/PROT.340190305

Ertl P, Rohde B, Selzer P (2000) Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. J

Med Chem 43:3714–3717. https://doi.org/10.1021/JM000942E/SUPPL_FILE/JM000942E_S.PDF
Facello MA (1995) Implementation of a randomized algorithm for Delaunay and regular triangulations in three dimensions. Comput Aided Geom Des 12:349–370. https://doi.org/10.1016/0167-8396(94)00018-N
Fejzo J, Lepre CA, Peng JW et al (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. Chem Biol 6:755–769. https://doi.org/10.1016/S1074-5521(00)80022-8
Feyfant E, Cross JB, Paris K, Tsao DHH (2011) Fragment-based drug design. Methods Mol Biol 685:241–252. https://doi.org/10.1007/978-1-60761-931-4_12
Fitzpatrick PA, Steinmetz ACU, Ringe D, Klibanov AM (1993) Enzyme crystal structure in a neat organic solvent. Proc Natl Acad Sci U S A 90:8653. https://doi.org/10.1073/pnas.90.18.8653
Frei P, Hevey R, Ernst B (2019) Dynamic combinatorial chemistry: a new methodology comes of age. Chem A Eur J 25:60
Grädler U, Schwarz D, Blaesse M et al (2019) Discovery of novel cyclophilin D inhibitors starting from three dimensional fragments with millimolar potencies. Bioorg Med Chem Lett 29:126717. https://doi.org/10.1016/J.BMCL.2019.126717
Greenberg WA, Priestley ES, Sears PS et al (1999) Design and synthesis of new aminoglycoside antibiotics containing neamine as an optimal core structure: correlation of antibiotic activity with in vitro inhibition of translation. J Am Chem Soc 121:6527–6541. https://doi.org/10.1021/JA9910356/ASSET/IMAGES/LARGE/JA9910356F00007.JPEG
Groom CR, Bruno IJ, Lightfoot MP, Ward SC (2016) The Cambridge structural database. Acta Crystallogr Sect B Struct Sci Cryst Eng Mater 72:171. https://doi.org/10.1107/S2052520616003954
Guedes IA, de Magalhaes CS, Dadenne LE (2014) Receptor-ligand molecular docking. Biophys Rev 6:75–87
Hajduk PJ (2006) Fragment-based drug design: How big is too big? J Med Chem 49:6972–6976. https://doi.org/10.1021/JM060511H/SUPPL_FILE/JM060511HSI20061002_015718.PDF
Hajduk PJ, Boyd S, Nettesheim D et al (2000) Identification of novel inhibitors of urokinase via NMR-based screening. J Med Chem 43:3862–3866. https://doi.org/10.1021/JM0002228
Hajduk PJ, Huth JR, Fesik SW (2005) Druggability indices for protein targets derived from NMr-based screening data. J Med Chem 48:2518. https://doi.org/10.1021/jm049131r
Hajduk PJ, Olejniczak ET, Fesik SW (1997) One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules. J Am Chem Soc 119:12257. https://doi.org/10.1021/ja9715962
Hann MM, Leach AR, Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. J Chem Inf Comput Sci 41:856–864. https://doi.org/10.1021/CI000403I
Hao GF, Jiang W, Ye YN et al (2016) ACFIS: a web server for fragment-based drug discovery. Nucleic Acids Res 44:W550–W556. https://doi.org/10.1093/NAR/GKW393
Harborne JB (1995) Dictionary of natural products. Phytochemistry 38:279. https://doi.org/10.1016/0031-9422(95)90037-3
Hearty S, Conroy PJ, Ayyar BV et al (2010) Surface plasmon resonance for vaccine design and efficacy studies: recent applications and future trends. Expert Rev Vaccines 9:645–664. https://doi.org/10.1586/ERV.10.52
Hernandez M, Ghersi D, Sanchez R (2009) SITEHOUND-web: a server for ligand binding site identification in protein structures. Nucleic Acids Res 37:W413–W416. https://doi.org/10.1093/NAR/GKP281
Hess B, Kutzner C, Van Der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447. https://doi.org/10.1021/CT700301Q/ASSET/IMAGES/LARGE/CT700301QF00006.JPEG
Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9:430

Jacquemard C, Kellenberger E (2019) A bright future for fragment-based drug discovery: what does it hold? Expert Opin Drug Discov 14:413–416. https://doi.org/10.1080/17460441.2019.1583643

Jason-Moller L, Murphy M, Bruno JA (2006) Overview of Biacore systems and their applications. Curr Protoc protein Sci Chapter 19:Unit 19.13. https://doi.org/10.1002/0471140864.PS1913S45

Jencks WP (1981) On the attribution and additivity of binding energies. Proc Natl Acad Sci U S A 78:4046–4050

Jordan RA, El-Manzalawy Y, Dobbs D, Honavar V (2012) Predicting protein-protein interface residues using local surface structural similarity. BMC Bioinformatics 13:41. https://doi.org/10.1186/1471-2105-13-41

Joseph-McCarthy D, Campbell AJ, Kern G, Moustakas D (2014) Fragment-based lead discovery and design. J Chem Inf Model 54:693–704. https://doi.org/10.1021/CI400731W/ASSET/IMAGES/LARGE/CI-2013-00731W_0005.JPEG

Jung LS, Shumaker-Parry JS, Campbell CT et al (2000) Quantification of tight binding to surface-immobilized phospholipid vesicles using surface plasmon resonance: Binding constant of phospholipase A2. J Am Chem Soc 122:4177–4184. https://doi.org/10.1021/JA993879V/ASSET/IMAGES/MEDIUM/JA993879VE00002.GIF

Katsila T, Spyroulias GA, Patrinos GP, Matsoukas MT (2016) Computational approaches in target identification and drug discovery. Comput Struct Biotechnol J 14:177–184. https://doi.org/10.1016/J.CSBJ.2016.04.004

Khan SA, Zia K, Ashraf S et al (2021) Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. J Biomol Struct Dyn 39:2607–2616. https://doi.org/10.1080/07391102.2020.1751298

Kirsch P, Hartman AM, Hirsch AKH, Empting M (2019) Concepts and core principles of fragment-based drug design. Molecules 24:4309. https://doi.org/10.3390/MOLECULES24234309

Kirshner DA, Nilmeier JP, Lightstone FC (2013) Catalytic site identification—a web server to identify catalytic site structural matches throughout PDB. Nucleic Acids Res 41:W256. https://doi.org/10.1093/NAR/GKT403

Klein J, Heal JR, Hamilton WDO et al (2014) Yeast synthetic biology platform generates novel chemical structures as scaffolds for drug discovery. ACS Synth Biol 3:314. https://doi.org/10.1021/sb400177x

Knehans T, Schüller A, Doan DN et al (2011) Structure-guided fragment-based in silico drug design of dengue protease inhibitors. J Comput Aided Mol Des 25:263–274. https://doi.org/10.1007/S10822-011-9418-0

Krivák R, Hoksza D (2018) P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. J Cheminform 10:1–12. https://doi.org/10.1186/S13321-018-0285-8/TABLES/4

Kuroki K, Maenaka K (2011) Analysis of receptor-ligand interactions by surface plasmon resonance. Methods Mol Biol 748:83–106. https://doi.org/10.1007/978-1-61779-139-0_6/COVER

Lauri G, Bartlett PA (1994) CAVEAT: a program to facilitate the design of organic molecules. J Comput Aided Mol Des 8:51–66. https://doi.org/10.1007/BF00124349

Laurie ATR, Jackson RM (2005) Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. Bioinformatics 21:1908–1916. https://doi.org/10.1093/BIOINFORMATICS/BTI315

Leach AR, Hann MM (2011) Molecular complexity and fragment-based drug discovery: ten years on. Curr Opin Chem Biol 15:489–496. https://doi.org/10.1016/J.CBPA.2011.05.008

Leavitt S, Freire E (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. Curr Opin Struct Biol 11:560–566. https://doi.org/10.1016/S0959-440X(00)00248-7

Lin Y, Yoo S, Sanchez R (2012) SiteComp: a server for ligand binding site analysis in protein structures. Bioinformatics 28:1172. https://doi.org/10.1093/BIOINFORMATICS/BTS095

Liu M, Quinn RJ (2019) Fragment-based screening with natural products for novel anti-parasitic disease drug discovery. Expert Opin Drug Discov 14:1283–1295. https://doi.org/10.1080/17460441.2019.1653849/SUPPL_FILE/IEDC_A_1653849_SM2190.XLSX

Lo MC, Aulabaugh A, Jin G et al (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem 332:153–159. https://doi.org/10.1016/J.AB.2004.04.031

Lotfi Shahreza M, Ghadiri N, Green JR (2020) A computational drug repositioning method applied to rare diseases: adrenocortical carcinoma. Sci Rep 101(10):1–7. https://doi.org/10.1038/s41598-020-65658-x

Luque I, Freire E (2000) The structural stability of binding sites. Consequences for binding affinity and cooperativity. Proteins 4:63–71

Lyu J, Irwin JJ, Roth BL et al (2019) Ultra-large library docking for discovering new chemotypes. Nature 566:224

Magsumov T, Ziying L, Sedov I (2020) Comparative study of the protein denaturing ability of different organic cosolvents. Int J Biol Macromol 160:880–888. https://doi.org/10.1016/J.IJBIOMAC.2020.05.260

Makhija MT, Kasliwal RT, Kulkarni VM, Neamati N (2004) De novo design and synthesis of HIV-1 integrase inhibitors. Bioorg Med Chem 12:2317–2333. https://doi.org/10.1016/J.BMC.2004.02.005

Marchand JR, Caflisch A (2018) In silico fragment-based drug design with SEED. Eur J Med Chem 156:907–917

Mashalidis EH, Śledå P, Lang S, Abell C (2013) A three-stage biophysical screening cascade for fragment-based drug discovery. Nat Protoc 811(8):2309–2324. https://doi.org/10.1038/nprot.2013.130

Matricon P, Suresh RR, Gao ZG et al (2021) Ligand design by targeting a binding site water. Chem Sci 12:960. https://doi.org/10.1039/d0sc04938g

McGreig JE, Uri H, Antczak M et al (2022) 3DLigandSite: structure-based prediction of protein–ligand binding sites. Nucleic Acids Res 50:W13–W20. https://doi.org/10.1093/NAR/GKAC250

McPherson A, Cudney B (2014) Optimization of crystallization conditions for biological macromolecules. Acta Crystallogr Sect F Struct Biol Commun 70:1445. https://doi.org/10.1107/S2053230X14019670

Miranker A, Karplus M (1991) Functionality maps of binding sites: a multiple copy simultaneous search method. Proteins 11:29–34. https://doi.org/10.1002/PROT.340110104

Morita A, Ochi S, Satoh H et al (2021) A novel RNA synthesis inhibitor, STK160830, has negligible DNA-intercalating activity for triggering a p53 response, and can inhibit p53-dependent apoptosis. Life 11:1087. https://doi.org/10.3390/LIFE11101087/S1

Morris GM, Ruth H, Lindstrom W et al (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30:2785. https://doi.org/10.1002/JCC.21256

Nakamura CE, Abeles RH (1985) Mode of interaction of beta-hydroxy-beta-methylglutaryl coenzyme A reductase with strong binding inhibitors: compactin and related compounds. Biochemistry 24:1364–1376

Ni F (1994) Recent developments in transferred NOE methods. Prog Nucl Magn Reson Spectrosc 26:517

Oliveira AM (2022) Quantitative structure-activity relationships (QSARs). In: Rudrapal M, Egbuna C (eds) Computer aided drug design (CADD): from ligand-based methods to structure-based approaches, 1st edn. Elsevier, Cambridge, p 324

Pang YP, Quiram P, Jelacic T et al (1996) Highly potent, selective, and low cost bis-tetrahydroaminacrine inhibitors of acetylcholinesterase. Steps toward novel drugs for treating Alzheimer's disease. J Biol Chem 271:P23646. https://doi.org/10.1074/jbc.271.39.23646

Paterson Y, Englander SW, Roder H (1990) An antibody binding site on cytochrome C defined by hydrogen exchange and two-dimensional NMR. Science 249:755–759. https://doi.org/10.1126/SCIENCE.1697101

Pintacuda G, Otting G (2002) Identification of protein surfaces by NMR measurements with a paramagnetic Gd(III) chelate. J Am Chem Soc 124:372–373. https://doi.org/10.1021/JA016985H/SUPPL_FILE/JA016985H_S1.PDF

Potts RO, Guy RH (1992) Predicting skin permeability. Pharm Res 9:663–669. https://doi.org/10.1023/A:1015810312465

Puthenveetil R, Vinogradova O (2019) Solution NMR: a powerful tool for structural and functional studies of membrane proteins in reconstituted environments. J Biol Chem 294:15914

Ramakrishnan M, Kandimalla KK, Wengenack TM et al (2009) Surface plasmon resonance binding kinetics of Alzheimer's disease amyloid beta peptide-capturing and plaque-binding monoclonal antibodies. Biochemistry 48:10405–10415. https://doi.org/10.1021/BI900523Q

Record MT, Mazur SJ, Melançon P et al (2003) Double helical DNA: conformations, physical properties, and interactions with ligands. Annu Rev Biochem 50:997–1024. https://doi.org/10.1146/ANNUREV.BI.50.070181.005025

Rees D (2016) Fragment-based Drug Discovery: Lessons and Outlook. Edited by Daniel A. Erlanson and Wolfgang Jahnke; Series Editors: Raimund Mannhold, Hugo Kubinyi, and Gerd Folkers. ChemMedChem 11:1667–1667. https://doi.org/10.1002/CMDC.201600256

Reker D, Rodrigues T, Schneider P, Schneider G (2014) Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. Proc Natl Acad Sci U S A 111:4067–4072. https://doi.org/10.1073/PNAS.1320001111/SUPPL_FILE/PNAS.201320001SI.PDF

Respondek M, Madl T, Göbl C et al (2007) Mapping the orientation of helices in micelle-bound peptides by paramagnetic relaxation waves. J Am Chem Soc 129:5228–5234. https://doi.org/10.1021/JA069004F/SUPPL_FILE/JA069004FSI20070207_094619.PDF

Ress DC, Congreve M, Murray CW, Carr R (2004) Fragment-based lead discovery. Nat Rev Drug Discov 3:660–672. https://doi.org/10.1038/nrd1467

Reuben J, Leigh JS (1972) Effects of paramagnetic lanthanide shift reagents on the proton magnetic resonance spectra of quinoline and pyridine. J Am Chem Soc 94:2789–2793. https://doi.org/10.1021/JA00763A042/ASSET/JA00763A042.FP.PNG_V03

Romasanta AKS, van der Sijde P, Hellsten I et al (2018) When fragments link: a bibliometric perspective on the development of fragment-based drug discovery. Drug Discov Today 23:1596–1609. https://doi.org/10.1016/J.DRUDIS.2018.05.004

Rudling A, Orro A, Carlsson J (2018) Prediction of ordered water molecules in protein binding sites from molecular dynamics simulations: the impact of ligand binding on hydration networks. J Chem Inf Model 58:350. https://doi.org/10.1021/acs.jcim.7b00520

Ryckmans T, Edwards MP, Horne VA et al (2009) Rapid assessment of a novel series of selective CB(2) agonists using parallel synthesis protocols: A Lipophilic Efficiency (LipE) analysis. Bioorg Med Chem Lett 19:4406–4409. https://doi.org/10.1016/J.BMCL.2009.05.062

Sabban S (2011) Development of an in vitro model system for studying the interaction of Equus caballus IgE with its high-affinity FcεRI receptor. The University of Sheffield

Saio T, Ogura K, Shimizu K et al (2011) An NMR strategy for fragment-based ligand screening utilizing a paramagnetic lanthanide probe. J Biomol NMR 51:395–408. https://doi.org/10.1007/S10858-011-9566-5/FIGURES/8

Salamon Z, Wang Y, Brown MF et al (1994) Conformational changes in rhodopsin probed by surface plasmon resonance spectroscopy. Biochemistry 33:13706–13711. https://doi.org/10.1021/BI00250A022/ASSET/BI00250A022.FP.PNG_V03

Schuffenhauer A, Ruedisser S, Marzinzik A et al (2005) Library design for fragment based screening. Curr Top Med Chem 5:751–762. https://doi.org/10.2174/1568026054637700

Segalés J, Perdiguero E, Muñoz-Cánoves P (2016) Regulation of muscle stem cell functions: a focus on the p38 MAPK signaling pathway. Front Cell Dev Biol 4:92

Seneci P (2000) Solid-phase synthesis and combinatorial technologies. Wiley-Interscience, New York

Senisterra G, Chau I, Vedadi M (2012) Thermal denaturation assays in chemical biology. Assay Drug Dev Technol 10:128–136. https://doi.org/10.1089/ADT.2011.0390/ASSET/IMAGES/MEDIUM/FIGURE5.GIF

Seth PP, Miyaji A, Jefferson EA et al (2005) SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: Internal ribosome entry site IIA subdomain. J Med Chem 48:7099–7102. https://doi.org/10.1021/JM050815O/SUPPL_FILE/JM050815OSI20050919_010845.PDF

Shaffer L (2020) 15 drugs being tested to treat COVID-19 and how they would work. Nat Med. https://doi.org/10.1038/D41591-020-00019-9

Silvestre HL, Blundell TL, Abell C, Ciulli A (2013) Integrated biophysical approach to fragment screening and validation for fragment-based lead discovery. Proc Natl Acad Sci U S A 110:12984–12989. https://doi.org/10.1073/PNAS.1304045110/SUPPL_FILE/PNAS.201304045SI.PDF

Smith EA, Thomas WD, Kiessling LL, Corn RM (2003) Surface plasmon resonance imaging studies of protein-carbohydrate interactions. J Am Chem Soc 125:6140–6148. https://doi.org/10.1021/JA034165U/SUPPL_FILE/JA034165USI20030318_031755.PDF

Sun H, Zhao L, Peng S, Huang N (2014) Incorporating replacement free energy of binding-site waters in molecular docking. Proteins Struct Funct Bioinforma 82:1765. https://doi.org/10.1002/prot.24530

Sybyl (2001) Sybyl molecular modeling software packages

Takahashi H, Nakanishi T, Kami K et al (2000) A novel NMR method for determining the interfaces of large protein–protein complexes. Nat Struct Biol 73(7):220–223. https://doi.org/10.1038/73331

Tian W, Chen C, Lei X et al (2018) CASTp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Res 46:W363–W367. https://doi.org/10.1093/NAR/GKY473

Tiwari PB, Bencheqroun C, Lemus M et al (2021) SPRD: a surface plasmon resonance database of common factors for better experimental planning. BMC Mol Cell Biol 22:17. https://doi.org/10.1186/s12860-021-00354-w

Torres PHM, Sodero ACR, Jofily P, Silva-Jr FP (2019) Key topics in molecular docking for drug design. Int J Mol Sci 20:4574

Vikholm I, Györvary E, Peltonen J (1996) Incorporation of lipid-tagged single-chain antibodies into lipid monolayers and the interaction with antigen. Langmuir 12:3276–3281. https://doi.org/10.1021/LA950949C/ASSET/IMAGES/LARGE/LA950949CF00006.JPEG

Vu H, Pedro L, Mak T et al (2018) Fragment-based screening of a natural product library against 62 potential malaria drug targets employing native mass spectrometry. ACS Infect Dis 4:431–444. https://doi.org/10.1021/ACSINFECDIS.7B00197/ASSET/IMAGES/LARGE/ID-2017-00197K_0009.JPEG

Wang X, Wei X, Thijssen B, LipkinSM DJ et al (2012) Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nat Biotechnol 30:159–164

Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc 73:1–16. https://doi.org/10.1016/J.PNMRS.2013.02.001

Wiseman T, Williston S, Brandts JF, Lin LN (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. Anal Biochem 179:131–137. https://doi.org/10.1016/0003-2697(89)90213-3

Wolf LK, Gao Y, Georgiadis RM (2007) Kinetic discrimination of sequence-specific DNA-drug binding measured by surface plasmon resonance imaging and comparison to solution-phase measurements. J Am Chem Soc 129:10503–10511. https://doi.org/10.1021/JA072401L/SUPPL_FILE/JA072401LSI20070526_082821.PDF

Xue P, Lu HH, Zhu YY et al (2017) Design and synthesis of hybrids of diarylpyrimidines and diketo acids as HIV-1 inhibitors. Bioorganic Med Chem Lett 27:1640. https://doi.org/10.1016/j.bmcl.2017.03.009

Yesudasu V, Pradhan HS, Pandya RJ (2021) Recent progress in surface plasmon resonance based sensors: a comprehensive review. Heliyon 7:e06321

Zsoldos Z (2011) A fragment-based docking engine: EHiTS: exhaustive fragment pose enumeration and matching using surface contact based statistical score. ACS Symp Ser 1076:91

Chapter 9
Artificial Intelligence and Machine Learning in Drug Discovery

Vivek Yadav, Jurnal Reang, Vinita, and Rajiv Kumar Tonk

9.1 Introduction

An average of $1.3 billion is spent on research and development for individual medicine (Kolluri et al. 2022). For non-oncology drugs, the median period from conception to approval spans from 5.9 to 7.2 years, whereas for oncology drugs, the median time is 13.1 with 13.8% overall probability of success for drug-development (DiMasi et al. 2016). Hence, lowering the success rate and overall costs resulting to lengthy timelines for the modern medication R & D process is a significant issue for both business and academics. Furthermore, the ongoing attrition of drug candidates is the cause of the modern pharmaceutical industry's excessive expenditure. Recent data indicate that animal toxicity (11%), poor pharmacokinetics (39%), and ineffectiveness (30%) account for 80% of the causes of attrition of the drug development process. Unpredictably, the issues raised above are directly connected to the discovery of drugs prior to clinical trials, showing that there is space for improvement (Wong et al. 2019). Since it is practically impossible to synthesis and evaluates all the potential compounds through tests. However, the overall procedure is typically decided by knowledge-based judgments, which might be highly prejudiced.

In the past 10 years, machine learning (ML) and artificial intelligence (AI) techniques have been well-known, thanks to the significant developments in computer technology. Artificial intelligence has the tendency to gather and process massive amounts of data required for research purposes. This helps in finding broad patterns of illness targets using a data-driven method which is a difficult task to recognize due to the complexity of the disease mechanism. In this area, a number of innovative researches have demonstrated the potential use of AI and ML techniques

V. Yadav (✉) · J. Reang · Vinita · R. K. Tonk
Pharmaceutical Chemistry Department, Delhi Pharmaceutical Sciences and Research University, New Delhi, India

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_9

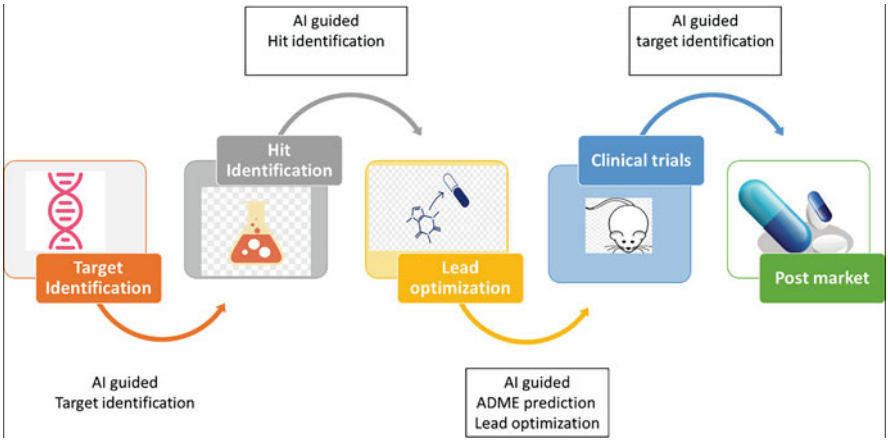


Fig. 9.1 Drug discovery process guided through AI and ML

in drug-target identification as well as their capacity to learn and uncover disease patterns with the corresponding targets without relying on biological proficiency.

The process of finding a medication lead starts with identifying the target of a certain disease, followed by hit identification, and lead optimization (Fig. 9.1). However, the traditional approaches to drug discovery required a lot of human labor, money, and time over an extended period of time. Additionally, research cannot be done with absolute certainty regarding the possibility that a given drug candidate's trial will be successful. The stages of drug discovery where AI is effectively cost-reducible start with the target identification and then identification of the lead or hit molecules through hit identification; furthermore, it helps in lead optimization and even aids in post-marketing surveillance reports.

AI/ML and deep learning systems have the potential to upsurge the probability of accomplishment ratio in drug development process. Moreover, these techniques provide significant progress in a number of R & D fields that includes novel target identification, deep learning and understanding of the target's role in the disease, insights protein structures prediction, and the molecular compound design and optimization. AI further extends their support into the discovery of small molecule by involving in different field deals with new biology, better or distinctive chemistry, and in vivo and in vitro study with higher chances of success and less time making it less expensive as well. In this chapter, artificial intelligence, machine learning, and deep learning in the drug development process with its application will be discussed.

9.2 Artificial Intelligence

9.2.1 Concept of Modernization

The use of computers and computational techniques in research and engineering could be considerably improved with the development of modern artificial intelligence (AI). By assessing clinically pertinent data that directs the discovery of new potential targets, applications of artificial intelligence (AI) in data and chemical synthesis process are directly involved in drug development optimization. The creation and improvement of potential medications' molecular structures can be done using an AI in drug design. Additionally, medication design methodologies comprehend how proteins' specific forms impact their activities in health and malfunction in sickness.

AI is commonly combined with better patient monitoring process performed during clinical trials and medical devices that access specific patient data and advise medical decisions in the organization, optimization, and operation and acquire crucial patient's data for clinical studies (Zhavoronkov et al. 2020). Additionally, it is increasingly feasible to utilize AI approaches to enhance healthcare research and services. However, one such application is risk-based guidance with deep-learning models used to anticipate preventable hospital readmissions (Farghali et al. 2021).

9.2.2 Models

9.2.2.1 AI-Guided Target Identification

A very popular and effective approach to finding new drugs is target-based drug discovery. For the treatment of any particular diseases, one should identify the target responsible for the agonist or antagonist actions. However, because of the choice of targets that are weakly related to the disease or have an unsupported theory, many therapeutic candidates in clinical trials have poor efficacy or elevated toxicity (Kim et al. 2020). Consequently, choosing appropriate targets requires a clearly distinct model for the relationship between the ailment and biological components. To interpret the connections, a variety of omics data types including genomics, proteomics, and metabolomics are required for better results.

The three kinds of conventional target identification techniques include machine learning, network-based models, and statistical analysis (Brown 2007). The most common and traditional methods for target identification have been statistical analyses of omics data for many years. These techniques were developed using the genome-wide study of associations (GWAS) and its emphases on finding genetic differences between samples from healthy and diseased people. By using association tests for the disease's gene expression, such as the Chi-squared test, Fisher's exact test, or t-test, it is possible to pinpoint potential target genes. Numerous study used

different types of data such as for the tumor samples from the Gene Expression Omnibus (GEO) project, miRNA expression data from NCI-60 cancer cell lines, and TNBC and non-TNBC data from the Cancer Cell Line Encyclopedia (CCLE) project to identify three kinase (PKC, CDK6, and MET) targets for triple-negative breast cancer (TNBC) (Chen and Butte 2016). They performed a two-stage bioinformatics investigation that involved a patient-based Kaplan Meier survival test and cell-based gene expression analysis. The disease-related genetic variations can be found using GWAS (Zhu et al. 2016). In order to find the genes linked to a complex human feature, Zhu et al. introduced a technique called Summary data-based Mendelian Randomization (SMR).

9.2.2.2 Network-Based Approaches

Network-based approaches are frequently employed to depict the intricate relationships between the many biological components. Networks are made up of nodes, which stand in biological components, and edges, which show how the nodes interact. Furthermore, this method uses a heterogeneous network to manage the various omics data types. Consequently, a network-based method to target identification is used in numerous investigations.

This network identifies gene sets linked to disease pathways by capturing genes with identical biological process function. Network analysis was utilized by Petyuk et al. to pinpoint a late-onset Alzheimer's target; to determine the gene-protein expression association contours, they built a co-expression network using peptide and transcript data (Petyuk et al. 2018; Mohamed et al. 2020). To add order or path towards network edges, they also created causal predictive networks.

Recently, target identification has also been accomplished using the knowledge graph. Entities, relations, and semantic data are represented in knowledge graphs as a machine-interpretable graph. Based on tensor factorization, a knowledge graph's entity and relationship are encoded into three embedding vectors and efficient through learning by decreasing wrong facts and maximizing accurate facts.

9.2.2.3 Machine Learning-Based Approaches

Finding broad spectrum of illness targets using a data-driven method remains difficult due to the complexity of the disease mechanism. Using the classifiers, we can determine whether a gene is associated with the therapeutic target or not. Through gene-disease association data, Open Targets platform that can be classified into four types such as Random Forest (RF), Support vector machine (SVM), Neural Net, and Gradient Boosting Machine (GBM) (Ferrero et al. 2017). When the four classifiers are performed similarly, the results will be an AUC of 0.75 and an accuracy of about 70%. By combining these regression models with gene expression data from GEO and Array Express, Mamoshina and co-workers created an age prediction system. They used feature importance analysis to determine which

genes were most closely connected with age prediction, and found that five well-known medication targets among the top 20 genes (Mamoshina et al. 2018).

9.2.2.4 AI-Guided Hit Identification

The important milestone in preclinical drug discovery is the identification of drug-target interactions. The molecular interaction among the drug and the chosen target determines the desired effects of the treatment, but unwanted interactions that were not specifically targeted during drug development might also result in side effects and the need to reposition the drug (Keiser et al. 2009). In order to maximize the effectiveness of the initial phases of drug development, numerous computational models are used to detect drug-target interaction and estimate binding affinities, which also has the benefit of delivering unique drug candidates (Ballester and Mitchell 2012; Stepniewska-Dziubinska et al. 2018). There are three basic types of hit identification computational approaches: the first focuses on the structure of the protein, the second on the structure of the ligand, and the third on the chemogenomic methods that describe similarity and feature-based methods (Fig. 9.2).

9.2.2.5 Structure-Based Approaches

The target protein's 3D structures, which are produced by X-ray crystallography (XRC) and proton nuclear magnetic resonance spectroscopy (protein NMR), are utilized by structure-based approaches. A key strategy in structure-based techniques is a molecular docking simulation, which is carried out in two parts (Imrie et al. 2018). The first phase is the search for ligands in conformational space, which thoroughly simulates potential binding poses. Following a conformational search, a scoring function ranks potential ligand poses on the targeted protein structure and calculates binding affinity in the second phase. The evaluation of docking simulations is influenced by the scoring function's quality. Traditionally, binding affinity posture is predicted using empirical or knowledge-based scoring systems.

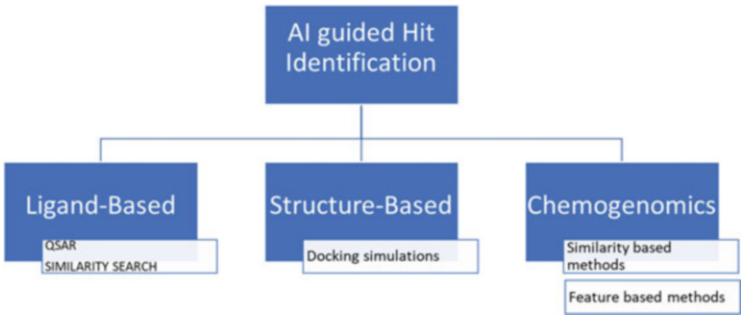


Fig. 9.2 AI guided hit identification methods in drug discovery

Data-driven machine learning scoring functions (MLSF) are created by employing support vector machines (SVM) and a random forest score (RF-Score) towards correcting the bias of classical scoring systems. In order to evaluate binding affinity, numerous deep learning-based scoring functions (DLSF) have recently been created. They have used a variety of deep learning approaches with the given pose, including a 3D convolutional neural network (3D-CNN) and a graph convolutional network (GCN). Each voxel has features that describe internal characteristics including ionization, hydrophobicity, aromaticity, and hydrogen bonds, among others. Deep learning today uses the convolutional neural network (CNN) as a key tool for pattern recognition. 3D-CNN is considered to find a three-dimensional spatial feature, binding pose, and affinity patterns for 3D voxel-based approaches. The potential net performed better than the RF-Score using GCN for the non-covalent (O'Boyle et al. 2011; Ma et al. 2015). Additionally, a number of recent researches recommended an examination of feature weights, which helped to better expand the compound's design.

9.2.2.6 Ligand-Based Approaches

The foundation of ligand-based approaches is the idea that molecules with comparable structural characteristics would interact with the same target. The main strategies in ligand-based methodologies are the quantitative structure-activity relationship (QSAR) models, three-dimensional QSAR, two-dimensional study fingerprint regions for the arrangement of the atoms (2D-QSAR), and estimation of quantitative associations (weights) between structure and its bioactivity (Yadav et al. 2020). A compound's structural and physicochemical characteristics have several connections to its biological activity, for instance, the partition coefficient is closely connected to the hydrophobic effect, which results towards receptor affinity. Many quantitative representations of a chemical can be utilized for prediction, ranging from a straightforward atom count to the Lipinski rule of five (Yadav et al. 2022). In order to develop quantitative molecular descriptors of chemicals, there are numerous tools available. To produce molecular descriptors for bioinformatics and cheminformatics, there are three open-source programs: RDKit, OpenBabel, and chemical development kit (CDK). QSAR creates a model to predict the bioactivity of compounds based on the quantitative descriptors that are generated. On-target bioactivities and ADME attributes were included in the benchmark datasets (Kaggle datasets) for QSAR prediction that is Merck Molecular Activity Challenge (MMAC) issued in 2012. Deep learning-based QSAR calculations have done better than earlier RF QSAR predictions as an advanced deep learning technique methodology.

9.2.2.7 Chemogenomic Approaches

Target proteins and chemicals are both used in chemogenomic techniques. The excellence and variety of chemogenomic techniques are taken advantage of by the exponential growth of data on proteins, compounds, and drug-target interactions (DTI). Chemogenomic techniques are often divided into two groups: similarity methods and feature-based techniques.

(a) *Similarity-Based Approaches*

In order to predict DTIs, similarity-based approaches focus on resemblances between the obtained protein and chemical structures. To develop most suitable similarity index between the proteins and chemicals, a variety of methods can be used such as the topological similarity in graphs and networks, normalized Smith-Waterman scores, Tanimoto coefficient, and pretense distance between protein domains. The bipartite local model (BLM) is a noteworthy study that makes use of the graph-based approach (Bleakley and Yamanishi 2009; Ding et al. 2013). BLM creates a bipartite graph connecting medications and targets, which expects the drug target interactions, and then aggregates both to get a final prediction.

(b) *Feature-Based Approaches*

Target and compound feature vectors, which are fixed-length vectors describing significant physicochemical qualities, are used in feature-based approaches. Drug and target vectors are concatenated, and machine learning models are trained to categorize DTIs using feature vectors of interaction and labels. Moreover, the drug-target characteristics features can be analyzed by protein-protein and drug-drug interaction networks to improve prediction performance (Li et al. 2016; Lee and Nam 2018). Furthermore, applying a deep learning model to feature-based methods has been suggested in numerous papers as a way to improve the results for drug design. However, feature-based approaches have a number of drawbacks, one of which is the information that is lost during feature engineering.

9.2.2.8 AI-Guided ADMET Prediction

Optimizing pharmacokinetic parameters with absorption, distribution, metabolism, excretion, and toxicity is one of the crucial parts in the drug discovery process (ADMET). In order to effectively direct the stages of drug discovery, it is necessary to examine compound's ADMET properties for the detailed understanding of complex biological mechanism (Gola et al. 2006). AI helps in understanding this complex human biological process to determine the results in a faster way with accuracy. The collection of bioactivity and data as well as sophisticated machine learning techniques, the pharmaceutical industries, as well as academic institutions have been drawn to in silico ADMET property predictions.

9.2.2.9 AI-Guided Lead Optimization

The phrase "finding a needle in a haystack" is used to describe the process of identifying a molecule that provides the appropriate pharmacological characteristics or has activity against biological targets. Researchers believe that there are roughly 1030–1060 chemical possibilities in the space of synthesizable compounds, although there are now only about 160 million chemicals listed in Chemical Abstracts Service. Too many resources and computational resources would be needed to fully count this enormous expanse.

These methods make use of deep learning techniques that have shown measurable effectiveness in the fields of machine translation and synthetic image synthesis. Knowledge of the chemical space distribution and performing targeted optimization to get the desired pharmacophore features are key aspects of performing deep generative models in the lead optimization areas (Brown et al. 2019). Although every method has its particular unique advantages, however the AI-guided methods provide number of advantages over other conventional approaches such as it is entirely data-driven and it can decrease human bias. Furthermore, using gradient-based optimization, the chemical space is explicitly modeled as a continuous function (Noorbakhsh-Sabet et al. 2019).

9.3 Applications

Healthcare professionals ought to be prepared for the approaching era of artificial intelligence and welcome the new capabilities that will enable more effective and efficient care. In this article, we examine machine learning's uses, difficulties, ethical issues, and viewpoints in the fields of medicine, translational research, and public health.

9.3.1 Disease Prediction and Diagnosis

Although artificial intelligence is increasingly being used in healthcare, research still mostly focuses on cardiovascular, nervous system, and cancer related as these are the major causes for the ill health and death. Early diagnosis of a variety of diseases can now be accomplished by refining the extraction of clinical understandings and performing these from the well-trained and verified system. For instance, the Food and Drug Administration (FDA) of the United States has approved the use of diagnostic software intended to find wrist fractures in adult patients. More over, 6% of the adult population in the United States suffers from depression. Image heatmap pattern recognition was 74% accurate at predicting severe depressive illness.

Artificial intelligence has the ability to provide prompt and accurate disease diagnosis, according to several researches. For the classification of complicated and multifactorial diseases, supervised approaches are useful tools for capturing nonlinear interactions. Abedi V. et al. discovered in a research involving 260 individuals that the model can identify acute cerebral ischemia more accurately than skilled emergency medical peoples (Abedi et al. 2017). However, the noisy data and experimental constraints diminish the therapeutic value of the models; deep learning methods can solve these constraints by lowering the dimensionality of the data by layered auto-encoding analyses.

9.3.2 Clinical Trials and In Silico-Based Prediction

With the AI method, researchers can partially replace animals or people in a clinical trial and create virtual patients with particular traits to improve the results of such investigations.

The deep learning techniques can be used in pharmacokinetics and pharmacodynamics from the initial preclinical stage to the later post-marketing analysis, and they are especially useful for pediatric or orphan disease trials. In one study, researchers created a sizable in silico randomized, placebo-controlled Phase III clinical trial study in which they treated artificial Crohn's disease patients using virtual therapies. However, with variable drug efficacy results revealed a favorable association between the baseline disease activity score and the decline in disease activity score. The investigational medicine GED-0301 did not receive a high score from the model, and this prediction was confirmed when the business that was conducting the phase III study on GED-0301 halted it after failing to pass an interim futility review. The design and discovery stages of a biomedical product, the identification of biomarkers, the optimization of dose, or the length of the proposed intervention can all benefit significantly from AI-guided in silico clinical trials.

9.3.3 Drug Discovery and Repurposing

Around 25% of altogether medications have been found as a result of unintentional bringing together of various areas. Due to the factors such as high costs of drug research, low success rates in clinical trials, the application of AI and ML is growing significantly and three-dimensional structural data that can aid in the characterization of pharmacological targets, and are used in the drug discovery process. The AI in drug repurposing process not only provides the new targets for the existing drugs but also reduces the expenditure cost.

For example, the DSP-1181 is the first AI-created medication to enter in clinical trials; it is a long-acting, powerful serotonin 5-HT1A receptor agonist. Exscientia is the biotech company which discovered DSP-1181 in collaboration with Sumitomo

Dainippon Pharma of Japan, which noted that the time from screening to the conclusion of preclinical testing was less than 12 months as compared to 4 years utilizing conventional procedures. Researchers at the Massachusetts Institute of Technology (MIT) discovered the medication halicin, which is effective against bacterial type (*Escherichia coli*), using a machine learning algorithm (Stokes et al. 2020).

Moreover, artificial intelligence is used in deep learning of the mechanism of medication toxicity, for example, terbinafine toxicity. The antifungal terbinafine may cause liver damage in some patients, which has very negative health effects. In another example, the machine learning method was performed to determine potential biochemical routes of the terbinafine drug to identify the biotransformation mechanism by the liver. The student discovered that the terbinafine metabolism is a two-step process through the AI/ML algorithm data.

In other examples, sildenafil, a drug first created in 1989 to treat angina, was later discovered to be effective in treating erectile dysfunction and was given the name Viagra. Thalidomide was initially created to treat morning sickness, but it caused serious birth problems, including limb deformity, and was removed off the market. A few years later, scientists learned that thalidomide has an anti-angiogenesis effect and began using it to treat leprosy and multiple myeloma.

To advance the knowledge of understudied biological systems, AlphaFold AI technique revealed the possible predictions of five SARS-CoV-2 targets in 2020, including the SARS-CoV-2's membrane protein, Nsp2, Nsp4, Nsp6, and papain-like proteinase (C terminal domain). The antiviral medications such as atazanavir, remdesivir, efavirenz, ritonavir, and dolutegravir were computationally identified by the MT-DTI technique.

9.4 Machine Learning

Machine learning, a well-known branch of artificial intelligence, used a large number of databases to identify different patterns of variable interactions. The ML can generate novel ideas, uncover previously unknown relationships, and be found to be helpful in obtaining a fruitful path for the drug development and research. Many fields, including data production and analytics, have adopted machine learning (ML). Algorithm-based approaches, like ML, have a strong mathematics and computational theory foundation. Many potential technologies have made use of ML models, including support vector machine-based improved search engines, deep learning (DL) assisted driverless automobiles, and advanced dialogue recognition technology.

Deep learning is a branch of machine learning that creates automated predictions from training datasets by simulating the functioning of the human intellect with numerous layers of artificial neuronal networks (Patel et al. 2020). Deep learning-based models frequently have several parameters and layers; as a result, model overfitting may result in subpar prediction accuracy. Over fitting can be avoided by

enlarging the training sample, reducing the number of hidden layers to obtain the balanced data. The example of the deep neural network application is to reduce the time it took to diagnose new outpatient cerebral hemorrhages by 96% with an accuracy of 84%.

9.4.1 Classifications

The machine learning methods are categorized into two types such as supervised and unsupervised methods (Table 9.1). In supervised learning, labels for fresh samples are determined using training examples with established labels. The regression and classification are useful applications of supervised learning. Examples of applications for supervised learning techniques include the identification of lung nodules from chest X-rays, risk estimation models for anticoagulation therapy, automated defibrillator implantation in cardiomyopathy, categorization of stroke and stroke mimics, identification of arrhythmia in electrocardiograms, and the designing of the in silico clinical trials. In addition to processing labeled input in supervised learning, generative deep neural networks (DNNs) can also be used to analyze unlabeled data. One of the most popular generative network topologies for unsupervised learning is the deep auto-encoder network (DEAN).

Unsupervised learning does not require labeled data and can find unseen patterns in the data that are frequently used for data exploration and the production of innovative ideas. Prior to recognizing patterns in high-dimensional data, the data are typically translated into a lower dimension using unsupervised learning methods. The unsupervised learning utilized to review failed clinical trials with drugs such as spironolactone, enalapril, and sildenafil versus placebo to revisit patients with heterogeneous conditions who had heart failure. The examination was done with three different studies to determine the patient's recovery without any human intervention (Carracedo-Reboredo et al. 2021).

Table 9.1 Components of artificial intelligence

Terms	Description
Supervised	Usage of a previously labeled database to predict outcomes of future events
Unsupervised	Identification of previously uncategorized database to predict peculiar relation between the dataset
Re-enforcement	Interaction of a machine with its environment using sensors, camera, GPS (global positioning system) and robotic interventions
Artificial neural	Computing system that analyses and processes information in a similar way compared to the human brain
Convolutional neural	Performs analyses of visual images
Recurrent neural	Functions by developing connections between nodes from a directed graph along a dynamic temporal sequence

The reinforcement learning method uses trial-and-error to increase accuracy while combining supervised and unsupervised learning. In all stage of the drug discovery process, large amounts of data are essential for the creation, development, and feasibility of efficacious ML algorithms. In precision medicine and therapies within drug discovery, the dependence on large, high-quality datasets and recognized, well-defined training sets is very crucial for the study.

Apart from these classifications, other model classifications frequently used are binary, multiclass, multi-label, and imbalanced. The binary is a two-label classification that employs algorithms like logistic regression, k-nearest neighbors, choices trees, support vector machines, and naive Bayes, while the multiclass involves more than two labels using techniques such choices trees, support vector machines, naive Bayes, random forests, and gradient boosting. In contrast to multiclass, which predicts a single class label for each example, multi-label classifies jobs that have more than two labels. The imbalanced classification model is used to classify the class labels with unevenly distributed jobs.

The deep learning (DL) is a type of machine learning algorithms which is known for using higher level characteristics such as neural networks that are developed from a model of the human brain to enable computers to read, create, and learn complicated hierarchical representations. The input data are transformed into a more compounded output data as a result of this process. There are various kinds of DL architectures, and depending on how the training set is organized, each one may recognize patterns and extract high-level features in a particular way. In this chapter, we briefly discuss on the common architectures, such as the CNN, RNN, and generative networks.

Convolutional neural network (CNN) is one of the most widely used DL designs in various industries, including natural language processing, image and speech identification, and many other natural language processing (NLP). Another sample type of DL architecture is the recurrent neural network (RNN), which was specifically designed to handle sequence data, and has been successfully applied to NLP.

9.4.2 ML Algorithms Used in Drug Discovery

The use of multiple ML algorithms in drug discovery has considerably benefited pharmaceutical businesses. There are different types of ML algorithms models available for forecasting the chemical, biological, and physical properties of molecules in drug advancement method. All phases from the drug identification to the market surveillances of the drug discovery process can benefit from the use of ML algorithms. As an illustration, ML algorithms have been applied to discover novel therapeutic uses, forecast drug-protein interactions, identify medication efficacy, assure the presence of safety biomarkers, and enhance the bioactivity of molecules.

9.4.2.1 Naive Bayes

Machine learning algorithms seek out the most promising theory from a set of relevant data, in particular, for the class of an unknown data sample. According on the description provided by the vector values of each sample's variables, Bayesian classifiers assign each sample to the most likely class. The technique assumes that the variables are independent in its most basic form, making it easier to apply Bayes' Theorem (Madhukar et al. 2019). While the assumption that not all variables are equally significant is impractical, this family of classifiers known as NB (Naive Bayes) that comes from it achieves excellent results, despite the fact that sometimes their set of characteristics exhibits high interdependence. This algorithm provides a straightforward method which is quick and efficient that can handle noisy data. Although it provides better results even though the data volume is very high in terms of the number of samples because of the tiny datasets. It responds each variable as a definite one and employs frequency tables to extract information. However, it is not the best technique for large dimensional issues with many features and requires some kind of transformation when dealing with numerical variables.

9.4.2.2 Naive Bayes in Drug Discovery

The identification of potential drug targets has been done using this approach in drug discovery. They specifically created a Bayesian model that incorporates many data sources, such as data of known side effects or gene expression, and they achieved a model with 90% accuracy on more than 2000 compounds. There are reports that used an experimental approach on machine learning and molecular docking study to identify the potential inhibitors of DNA topoisomerase I enzyme of mycobacterium tuberculosis (MtTOP1) species and evaluated in vitro confirmation of their computational findings (Ekins et al. 2017). The AUC values for these predictions were 74%. In this, the drug prediction models are used in accordance with the ATM (Anatomical Therapeutic Chemical) system using the datasets from STITCH and ChEMB. The different types of molecular descriptors were analyzed for the structural information, and interactions with similar targets are displayed with an accuracy of 65%.

9.4.2.3 Support Vector Machines

Support Vector Machines (SVM) were first presented by Vapnik in the late 1970s. Due to the robustness and capacity to generalize in high-dimensional domains, particularly in bioinformatics, these are among the most extensively utilized approaches (Fernandez-Lozano et al. 2014). Sets of points in a particular space are used in machine learning to figure out how to handle brand new observations. These

points are used by kernel-based approaches to determine how comparable the new observations are and to reach a conclusion.

9.4.2.4 Support Vector Machines in Drug Discovery

The SVM is one of the most often used models in bioinformatics because of its capacity to handle challenging issues that are complicated, nonlinear, high dimensional, and noisy. They have been utilized to classify pharmaceuticals based on their KEGG categorization, with an accuracy score of 83.9%. A brand new method for predicting intricate drug-target interaction networks using interaction matrices with function values of 80% was put forward. Additionally, by calculating several molecular descriptors and chemical indices using ChEMBL datasets with values near 70% in validation, it is able to predict the stability in human liver microsomes. The method used in expression data is an intriguing new method to anticipate a drug's impact on a tumor line by learning more about the genes involved in the drug's response in various tumor types (GEO).

For the prediction of HDAC1 inhibitors, SVMs were also applied to 3D-QSAR descriptors using a feature selection strategy described (Hu et al. 2016). The 2D-QSAR used to predict the compounds that inhibit the P-gp membrane protein target in the cancer study and wrapper feature selection models along with metaheuristic as a genetic algorithm produced promising results that were later confirmed by molecular docking approaches. Multiple Kernel Learning (MKL), which generates various linear combinations of SVMs with various parameters or kernels in an effort solve the problems, is an illustration of a sophisticated application of SVMs. Additionally, this enables the integration of many heterogeneous data sources, although at the expense of raising the computing cost.

9.4.2.5 Tree-Based Models

A decision tree is a hierarchical structure made up of nodes and the connections between them or branches. The method employed for classification issues is distinguished among methods for other sorts of problems, such as regression, survival, or outlier's detection. In this, the root nodes, internal nodes, and terminal nodes were found within a decision tree's hierarchical structure. The root node is found at the top of the tree model with one or more branches emerging from it but no branches reaching towards it. Regarding internal nodes, two or more branches originate from them and reach the next level of the hierarchy. There is no branches originating from the terminal nodes since they are located at the bottom of the hierarchy.

The out of bag error is equivalent to the error that the algorithm would make when the cross-validation is performed. The bagging approach in which the random forest (RF) divides the dataset into one-third part for validation and two-third part for training sets and analyzes to determine generalization error internally from each individual decision tree. Finally, because each decision tree is trained using various

samples and characteristics, it is easy to estimate the importance of each attribute, while ignoring the others and lowering the problem's dimensionality. Because of this, issues with very high dimensionality and noise are particularly well-suited for this technique.

9.4.2.6 Random Forest in Drug Discovery

When it comes to greater performance, speed, and generalizability, the RF model is the greatest among all other models. This model is deemed to be more suited and offers protein interactions with greater than 90% accuracy. They made use of the Open Babel descriptors and the GO and KEGG protein enrichment scores for the validation.

9.4.2.7 Artificial Neural Networks

The artificial neuron is a useful component of the network that accepts input from other components and processes it in some way to provide an output that can be processed by other components before talking about ANN. The artificial neurons may communicate with one another, just like natural neurons, and their connections are represented by weights, which are merely values that attempt to capture the synaptic force of a connection between two neurons. The net value, which sums together all the forces received by an artificial neuron or processing element, is considered first. The output of the processing element is determined by applying a trigger function after the net value calculation. The network of neurons can be created where the outputs of one neuron are used as the input for other neurons. It is important to realize that ANNs require input nodes, or neurons that receive data from the external world; these neurons are referred to as the network's input layer. Additionally, the network involves output nodes, which are located in the hidden layer and transmit ANN results. The network's hidden nodes, which transport data between neurons, are arranged into one or more hidden layers.

9.4.2.8 ANN in Drug Discovery

ACD (Available Chemicals Directory) and CMC (Comprehensive Medicinal Chemistry) data were used to train ANN and tree-based algorithms for drugs and non-drugs, respectively. The 2D descriptors provide the detailed information of the functional groups availability inside the molecules structures. However, 1D descriptors provide the information regarding the molecule's molecular weight and hydrogen bond numbers for each available compound. An ANN with both 1D and 2D descriptors produced the greatest results, with an accuracy of 89%.

To forecast the initial carcinogenesis of substances suggested to be medications includes the calculation of six distinct types of descriptors with a deep learning

model and an accuracy of 86% using 1003 chemicals from the Carcinogenic Potency Database. AUC of 76% can be achieved by beginning an experimental phase in the lab, generating a set of 2130 compounds of potential novel medications of interest for cardiotoxicity, computing each compound's DRAGON 3456 descriptors, and including the analysis in a feature selection procedure.

9.4.2.9 De novo Molecular Design

Recent developments in ML have greatly improved the field of de novo or inverse molecular design. In a very short period of time, many intriguing strategies have been proposed. Recurrent neural networks (RNNs), generative adversarial networks (GANs), and auto-encoders, in particular, have been applied to the optimization of devices and the rational design of organic and inorganic materials. ReLeaSE is a deep reinforcement learning-based technology that produces chemical compounds and focuses on chemical collections with anticipated physical, chemical, and/or bioactivity features (RL). Both generating (G) and predictive (P) neural networks are used in the ReLeaSE method's main workflow. The generative model G serves as an agent in this system by creating new, chemically viable compounds, whereas the predictive model P serves as a critic. P assigns a numerical reward (or penalty) to each created molecule in order to estimate the agent's behavior.

9.4.2.10 Synthesis Planning

Recent advances in research, synthesis planning have made use of ML-based methodologies. Without human support, full syntheses of crucial chemicals for medicine were planned using the computer application Chematica. In order to identify the successful synthetic paths, the reaction guidelines are merged into graphs that connect lots of potential molecules with the chemical reaction knowledge. Retrosynthetic paths can be found using Monte Carlo tree search and symbolic AI without the aid of human expert rules and widely used today in the research organizations. Practically, all organic chemistry-related reported reactions were used to train this neural network. However, the synthetic chemists judged computer-generated pathways to be comparable to approaches described in the literature and with practical results.

9.5 Applications

9.5.1 CNS Disorder

Futuristic CNS drug discovery study will increasingly rely on AI/ML, mostly in the fields such as patient subtyping, identification of crucial disease drivers, estimation

of cell type-specific drug response, sovereign novel drugs design, and with better BBB (blood brain barrier) permeability tests. The role of AI/ML is now being constrained by structural limitations in data and algorithms. However, in the long run, we will be able to create CNS disease treatments that are more potent because of ongoing and new breakthroughs in AI/ML approaches to neuropharmacology (Carpenter and Huang 2018).

9.5.2 Discovering Novel Antimicrobial Agents

Several reported works showed how ML may be used in the context of antibiotic discovery to learn small molecule structural properties from screenings that contain prevailing antimicrobial activity to advance novel antimicrobials. By first creating a genetic library of hypomorph knockdowns for these crucial genes and then screening 50,000 chemical compounds against these hypomorphs, Johnson et al. done a screening for finding biochemical inhibitors of key genes in *M. tuberculosis*.

The supervised ML classification evaluates the novel classes of chemical inhibitors for existing drug targets and recent discovered targets, validated in wild-type cells against standard antibiotics. A deep learning ML model used screening of several molecules with different structural features for antimicrobial activity against *E. coli* in order to predict antimicrobial functions. To predict the inhibition of *Escherichia coli* growth, the scientists used a training set of 2335 molecules for a DNN model. The model was then run on more than 107 million molecules from various chemical libraries.

9.5.3 Epidemic COVID

In order to find effective medications for 65 human proteins (targets) that had shown to interact with SARS-CoV-2 proteins, Kowalewski and Ray created machine learning (ML) models (Kowalewski and Ray 2020). They infer it from inhalation treatments to directly target the injured cells because the virus is known to target the respiratory tract, including nose epithelial cells, upper airway, and lungs. In order to rank the chemicals and identify medications that share the identical chemical space, they gathered 14 million compounds from ZINC databases and used machine learning algorithms to obtain vapor pressure and mammalian toxicity. The objective of the study was to create a short- and long-term pipeline for use in the future. They also developed models that might forecast drug efficacy using SVM and RF.

9.6 Drug Discovery Process

9.6.1 AI and Machine Learning in Precision Drug Discovery

A new approach to disease prevention and treatment called precision medicine considers a person's unique gene, lifestyle, and environmental variations. Based on the genetic profiles of the patients, this technique aids scientists and medical professionals in more precisely preventing and treating disease. Powerful supercomputer infrastructure and innovative algorithms that can autonomously learn in an unheard of fashion from the trained set of data are needed to make the strategy more comprehensive. Medical professionals' cognitive abilities and biomedical data are used by artificial intelligence to achieve results.

With technological advancements, the future of healthcare will change as a result of the creation of large digital datasets obtained through next-generation sequencing (NGS), use of image processing algorithms, patient-related health records, and data resulting from significant clinical trials. Oncology can benefit greatly from machine learning, which is frequently used in precision medicine. Complex neural networks are used to generate diagnostic images and genetic data, which are then used to forecast the likelihood of disease and treatment outcomes (Dlamini et al. 2020). In radiomic field of machines that produces diagnostic images to discover malignant tumors that are undetectable by the human sight, the implementation of AI and ML technologies in healthcare is done to enhance illness management and deliver high-quality medical care (Fig. 9.3).

By highlighting diverse uses of AI in oncology healthcare, such as next-generation sequencing (NGS), advancements in medical imaging, digital pathology, and drug discovery, we present information on AI and precision oncology towards clinical environment for cancer management.

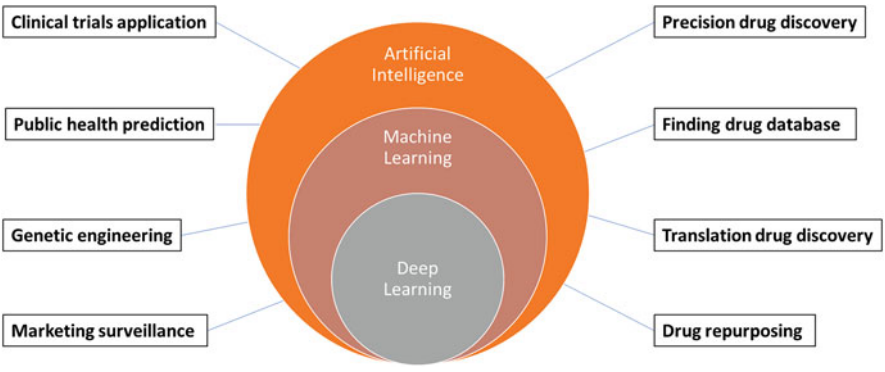


Fig. 9.3 Application of artificial intelligence and machine learning in the drug discovery

9.6.1.1 NGS and Molecular Profiling

The NGS technique utilizes RNA sequencing to discover novel RNA variants and splice sites, or quantify mRNAs for gene expression analysis. Genomic profiling is conceivable and offers promise for the future of precision oncology to the implementation of NGS, which is quickly evolving the field of genomic sequencing for clinical use. Advanced NGS methods can sequence DNA and RNA on a wide scale with high-throughput data and at a lower cost. Numerous sequencing techniques, such as whole-genome, whole-exome, RNA, target, and whole-transcriptome shotgun sequencing, as well as methylation sequencing, are made possible by NGS. DNA or RNA samples from blood samples, tumor samples, cell lines, formalin-fixed paraffin-embedded (FFPE) blocks, and liquid biopsies can all be used for sequencing. As part of the Human Genome Project, the first whole-genome sequencing was carried out at significant expense and over a lengthy period of time. To detect changes in the cellular transcriptome and changed molecular pathways, RNA sequencing is frequently employed in cancer research and diagnosis (Jiang et al. 2017).

The advantages of RNA profiling of cancer models for treatment results have been demonstrated in clinical studies that sequence RNA using precise oncology protocols. RNA profiling is applied to RNA extracted from blood or a tumor sample. According to a study, RNA profiling should be a standard of care for oncology patients because it may have potential clinical benefits, particularly for cancers that are challenging to treat in children and young adults. The study also illustrates the impact of precision oncology. According to the study's findings, about 70% of the gene expression data acquired from RNA sequencing may have clinical applications.

Identification of gene expression signatures to decipher the underlying molecular pathways of cancer and the detection of RNA mutations with implications for alternative splicing are the two most significant and often used applications of RNA sequencing. However, many NGS approaches have drawbacks such as labor-intensiveness, the introduction of sequencing coverage mistakes, and expense. Acquiring pertinent data from NGS datasets is becoming more and more time-effective because of the developments in AI and computational approaches, with some platforms enabling real-time viewing.

9.6.1.2 Biomarkers

Molecular biomarkers are often used in the cancer diagnostics in the early detection of the diseases. Different biomarkers are used, for example, circulating cancer antigen is used to detect ovarian cancer at early stage, carcinoembryonic antigen is used to monitor relapse of colorectal cancer, and estrogen receptor 1 (ESR1) is used for the prognosis prediction and treatment outcomes in breast cancer. Cancer management can be improved by locating biomarkers in the early disease prevention and prognosis prediction for successful treatment. By locating germline DNA

alterations and doing full transcriptome analyses by RNA sequencing, novel molecular biomarkers for various malignancies can be uncovered and utilized to detect the diseases. The potential of RNA sequencing in the development of biomarkers for diagnosis and as a prognostic predictor has been demonstrated in large consortia studies like the Cancer Genome Atlas (TCGA). Aside from pathogenic mutations and changed expression or activity of proteins that regulate significant cellular complexes, these investigations also clarified predicted biomarkers that fuel transformation. Additionally, Shallow full genome sequencing was used to identify copy number variants (CNV) in breast cancer utilizing FFPE samples to diagnoses for breast cancer, lung cancer, and neuroblastoma.

9.6.1.3 Medical Imaging

Applications of AI in radiology are essential for many modalities with enhanced quality, including X-rays, ultrasounds, computed tomography (CT/CAT), magnetic resonance imaging (MRI), positron-emission tomography (PET), and digital pathology. Images are analyzed quickly and accurately using highly specialized algorithms. Accurate diagnosis depends in large part on the ability to distinguish between normal and aberrant medical images. Early cancer detection is extremely important because it will result in a better prognosis and treatments. The future of AI in medical imaging will be focused on increasing speed and lowering costs. AI has already contributed to medical imaging by improving image quality, computer-aided image interpretation, and radiomics (Lewis et al. 2019). The main advancements and breakthroughs of artificial intelligence in healthcare have been widely used for clinical purposes in medical imaging.

9.6.1.4 Radiographic Imaging

In order to accurately diagnose and treat patients, which can take time and be subject to human error and variability, it is necessary to extract pertinent quantitative data from medical images, such as size, symmetry, location, volume, and form. For routine clinical treatment, automated medical imaging analysis is highly necessary. The radiographic imaging includes three stages: the first one is the image segmentation, which detects the image of interest and defines its boundaries; the second one is the image registration, which establishes the spatial three-dimensional relationship between images; and image visualization, which displays pertinent information for precise interpretation, is necessary to analyze the medical images accurately. However, despite of the advancement in the medical imaging, there are still some complications with data complexity, object complexity, and validation issues.

The deep learning-based algorithms for an automated detection system for chest radiography are the recent advancement. However, the chest radiograph analyses for thoracic disease are difficult and error-prone, and the highly skilled radiographers are

required to analyze the images. These AI methods were created to differentiate between common thoracic disorders, including pulmonary malignant tumors.

Imaging in medicine using AI extends beyond radiology. The advent of digital pathology will soon revolutionize pathology laboratories. The gold standard for pathology for many years has been microscopic examination of stained cells and tissues. By reducing labor-intensive microscopic tasks, boosting efficiency, and maintaining the quality for better clinical treatment, technological and AI advancements will transform pathology. Digital pathology that incorporates AI improves workflow, enables doctors to analyze images for precise interpretation, and lowers subjectivity by standardizing processes. Additionally, digital pathology enables reduced fluctuation in color information and larger-scale image viewing. This makes it possible to successfully find distinctive markers linked to disease-specific biomarkers for diagnosis, prognosis, and treatment (Bera et al. 2019).

9.6.2 Repurposed Drug/Drug Discovery by AI/ML Approach

About 25% of all medications have been found as a result of unintentional bringing of various areas. Pharmaceutical companies prefer targeted drug discovery over conventional blind screening because it has a clear mechanism and a better success rate and is less expensive. Due to the following factors such as high costs of drug research, growing accessibility of three-dimensional structural data that can aid in the characterization of pharmacological targets, and shockingly low success rates in clinical trials, machine learning is currently used in the drug discovery process. Cross-domain linkage can be accomplished using machine learning as a bridge. By identifying contextual cues like a discussion of a drug's indication or side effects, it may recognize a newly approved drug.

Despite these innovative methods for drug development, there are still significant obstacles, such as data access and the fact that various datasets are typically kept in a number of separate repositories. Additionally, clinical trial raw data and other preclinical study raw data are often unavailable. The utilization of pharmacological information to gain knowledge into mechanism of action by employing methods like similarity metrics across all diseases to uncover shared pathways is just one example of how artificial intelligence has been successful when applied to available data. Another illustration is the use of NLP to find hidden or unexpected relationships that may be significant in the identification of probable pharmacological side effects based on scholarly articles.

Few organizations have started to make use of these developments to accelerate the release of COVID-19 medications and better understand how the immune system combats the illness. Pharmaceutical companies GlaxoSmithKline (GSK) and Vir Biotechnology teamed together at the beginning of April to accelerate coronavirus treatment development using CRISPR and artificial intelligence. Additionally, in the academic world, the Human Immunomics Initiative, launched recently by the Harvard T. Chan School of Public Health and the Human Vaccines Project, employs

Table 9.2 List of repurposed drugs for COVID-19 through AI

Sl. No.	Drug	Original used	Company
1.	Baricitinib	Rheumatoid joint pain	BenevolentAI
2.	Hydroxychloroquine and Remdesivir	Antimalarial	Innoplexus
3.	Atazanavir	Antiretroviral HIV/AIDS	Deargen
4.	Niclosamide and Nitazoxanide	Viral infections	Gero

artificial intelligence to accelerate the production of antibodies for a variety of illnesses, including COVID-19. A team from Southern Illinois University (SIU) recently developed an information visualization tool that shows users the locations of known COVID-19 instances using GPS data. A contact following application powered by Bluetooth technology has also been developed in cooperation between Google and Apple. These techniques might be successful in collecting a lot of precise data. Businesses that have developed wellness profiles for people based on a fundamental understanding of the infection are conducting research into various medicine delivery methods that have been successfully licensed. The two most well-known examples of this in relation to COVID-19 to date are hydroxychloroquine (recommended for the treatment of malaria) and remdesivir (for the treatment of Ebola). The effectiveness dataset for these drugs may therefore be a decent input for an AI model. The businesses using artificial intelligence (AI) to repurpose currently available drugs for COVID-19 are listed in Table 9.2.

9.7 Limitations of AI/ML Approaches

The use of routine clinical NGS sequencing for cancer diagnosis and management faces significant challenges with data interpretation. Large servers and knowledgeable bioinformaticians are needed for the management and interpretation of big data. The provided datasets for diagnosis contain details on variants that can be classified as benign, likely benign, variant of unknown importance, likely pathogenic, and pathogenic variants. It is crucial to classify all variations into groups and understand their clinical importance. Data acquired can be helpful for cancer management in addition to diagnosis.

However, the drawbacks of whole-genome and exome sequencing include high costs, a heavy computing burden, and challenging data interpretation. In the following 10 years, further development of NGS platforms may result in cost reductions without a reduction in quality.

Despite the advantages of AI, there are still several obstacles to its implementation in the healthcare industry. Big data and costs are on the rise as a result of automated computation. Due to their reliance on specialized computational requirements for rapid data processing, AI systems can be costly. Additional quality procedures are also necessary for these systems. The targeted users must receive training and gain a knowledge of the technology in order to implement AI-based solutions for

everyday clinical practice. Rigby emphasized the moral dilemma presented by AI in healthcare. It is crucial to resolve the ethical problem of using patient data without authorization or justification in light of the big data boom. Additionally, in order to safeguard patient privacy and safety, ethical norms and guidelines are necessary.

Despite appearing to be effective and acceptable in the de novo lead creation approach, the connecting mechanism has some drawbacks. The first restriction is that for proper linking: the linking fragments must be precisely positioned in the cavity. De novo design is additionally assumed to be totally automated, but still requires some arduous manual labor. Furthermore, it is not always simple to manufacture the chemicals created using this method in a lab. Thus, new software that includes de novo compound design and considers synthesis parameters is required.

Although the connecting approach in the de novo lead generation method appears to be effective and acceptable, there are certain restrictions. The first restriction is that for proper linking, the linking fragments must be precisely positioned in the cavity. De novo design is additionally assumed to be totally automated, but still requires some arduous manual labor. Furthermore, it is not always simple to manufacture the chemicals created using this method in a lab. Thus, new software that includes de novo compound design and considers synthesis parameters is required.

Overall, the complexity of small molecule drug discovery will increase. DL ought should be able to manage that complexity since it was made for complicated simulation. Additionally, using DL techniques, we should not limit ourselves to making the conventional predictions about biological activities, ADMET properties, or pharmacokinetic simulations. Instead, it might be possible to systematically integrate all the data and information and reach a new level of AI in drug discovery.

9.8 Conclusions and Future Perspectives

The ultimate goal of machine learning is to create algorithms that can learn continually from fresh information and data in order to find solutions to a wide range of problems. Complex algorithms have appealing prospects for precision medicine, but they also present computing difficulties. To realize this potential, unique solutions are needed for at least three technical problems:

1. The quantity and size of data inputs, outputs, and attributes. This problem can be partially solved by leveraging CPU clusters, data sharing systems, cloud computing, and deep learning techniques.
2. Variety—diverse types of data (picture, video, and text). This problem can be partially solved by integrating data from many sources using novel deep learning techniques.
3. Velocity—the pace of streaming data. To solve this problem, online learning techniques can be developed.

Machine learning techniques used nowadays are very similar to real-world situations. As a result of the quick improvements in technology, algorithms will take on duties that were previously the domain of humans. Radiologists and anatomical pathologists will lose a lot of their jobs as a result of machine learning's capacity to turn data into insight. Clinical medicine, however, has always required physicians to manage enormous amounts of data, from the history and physical examination to the laboratory and imaging examinations, as well as the more recent genetic data. Effective medical professionals have always been able to handle this complexity.

We anticipate that as more scientists become aware of its potential, the usage of ML in VS for drug discovery will continue to expand in the search of new drugs. Drug discovery will undoubtedly become more effective and less expensive, thanks to the combined efforts of computer science and medicinal chemistry.

References

Abedi V, Goyal N, Tsivgoulis G et al (2017) Novel screening tool for stroke using artificial neural network. Stroke 48(6):1678–1681. https://doi.org/10.1161/STROKEAHA.117.017033

Ballester PJ, Mitchell JBO (2012) A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics 26(9):1169–1175. https://doi.org/10.1093/bioinformatics/btq112

Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A (2019) Diagnosis and precision oncology. Nat Rev Clin Oncol 16(11):703–715. https://doi.org/10.1038/s41571-019-0252-y

Bleakley K, Yamanishi Y (2009) Supervised prediction of drug-target interactions using bipartite local models. Bioinformatics 25(18):2397–2403. https://doi.org/10.1093/bioinformatics/btp433

Brown D (2007) Unfinished business: target-based drug discovery. Drug Discov Today 12(23–24): 1007–1012. https://doi.org/10.1016/j.drudis.2007.10.017

Brown N, Fiscato M, Segler MHS, Vaucher AC (2019) GuacaMol: benchmarking models for de novo molecular design. J Chem Inf Model 59(3):1096–1108. https://doi.org/10.1021/acs.jcim.8b00839

Carpenter KA, Huang X (2018) Machine learning-based virtual screening and its applications to Alzheimer's drug discovery: a review. Curr Pharm Des 24(28):3347–3358. https://doi.org/10.2174/1381612824666180607124038

Carracedo-Reboredo P, Liñares-Blanco J, Rodríguez-Fernández N et al (2021) A review on machine learning approaches and trends in drug discovery. Comput Struct Biotechnol J 19: 4538–4558

Chen B, Butte AJ (2016) Leveraging big data to transform target selection and drug discovery. Clin Pharmacol Ther 99(3):285–297. https://doi.org/10.1002/cpt.318

DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: new estimates of R & D costs. J Health Econ 47:20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012

Ding H, Takigawa I, Mamitsuka H, Zhu S (2013) Similarity-basedmachine learning methods for predicting drug-target interactions: a brief review. Brief Bioinform 15(5):734–747. https://doi.org/10.1093/bib/bbt056

Dlamini Z, Francies FZ, Hull R, Marima R (2020) Artificial intelligence (AI) and big data in cancer and precision oncology. Comput Struct Biotechnol J 18:2300–2311

Ekins S, Godbole AA, Kéri G et al (2017) Machine learning and docking models for mycobacterium tuberculosis topoisomerase I. Tuberculosis 103:52–60. https://doi.org/10.1016/j.tube.2017.01.005

Farghali H, Canová NK, Arora M (2021) The potential applications of artificial intelligence in drug discovery and development. Physiol Res 70:715–722. https://doi.org/10.33549/physiolres.934765

Fernandez-Lozano C, Gestal M, González-Díaz H et al (2014) Markov mean properties for cell death-related protein classification. J Theor Biol 349:12–21. https://doi.org/10.1016/j.jtbi.2014.01.033

Ferrero E, Dunham I, Sanseau P (2017) In silico prediction of novel therapeutic targets using gene-disease association data. J Transl Med 15(1):1–16. https://doi.org/10.1186/s12967-017-1285-6

Gola J, Obrezanova O, Champness E, Segall M (2006) ADMET property prediction: the state of the art and current challenges. QSAR Comb Sci 25(12):1172–1180. https://doi.org/10.1002/qsar.200610093

Hu B, Kuang ZK, Feng SY et al (2016) Three-dimensional biologically relevant Spectrum (BRS-3D): shape similarity profile based on PDB ligands as molecular descriptors. Molecules 21(11). https://doi.org/10.3390/molecules21111554

Imrie F, Bradley AR, Van Der Schaar M, Deane CM (2018) Protein family-specific models using deep neural networks and transfer learning improve virtual screening and highlight the need for more data. J Chem Inf Model 58(11):2319–2330. https://doi.org/10.1021/acs.jcim.8b00350

Jiang F, Jiang Y, Zhi H et al (2017) Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2(4):230–243. https://doi.org/10.1136/svn-2017-000101

Keiser MJ, Setola V, Irwin JJ et al (2009) Predicting new molecular targets for known drugs. Nature 462(7270):175–181. https://doi.org/10.1038/nature08506

Kim H, Kim E, Lee I et al (2020) Artificial intelligence in drug discovery: a comprehensive review of data-driven and machine learning approaches. Biotechnol Bioprocess Eng 25:895–930

Kolluri S, Lin J, Liu R, Zhang Y, Zhang W (2022) Machine learning and artificial intelligence in pharmaceutical research and development: a review. AAPS J 24(1):19. https://doi.org/10.1208/s12248-021-00644-3

Kowalewski J, Ray A (2020) Predicting novel drugs for SARS-CoV-2 using machine learning from a >10 million chemical space. Heliyon 6(8):e04639. https://doi.org/10.1016/j.heliyon.2020.e04639

Lee I, Nam H (2018) Identification of drug-target interaction by a random walk with restart method on an interactome network. BMC Bioinformatics 19(Suppl 8):208. https://doi.org/10.1186/s12859-018-2199-x

Lewis SJ, Gandomkar Z, Brennan PC (2019) Artificial Intelligence in medical imaging practice: looking to the future. J Med Radiat Sci 66(4):292–295. https://doi.org/10.1002/jmrs.369

Li ZC, Huang MH, Zhong WQ et al (2016) Identification of drug-target interaction from interactome network with "guilt-by-association" principle and topology features. Bioinformatics 32(7):1057–1064. https://doi.org/10.1093/bioinformatics/btv695

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(2):263–274. https://doi.org/10.1021/ci500747n

Madhukar NS, Khade PK, Huang L et al (2019) A Bayesian machine learning approach for drug target identification using diverse data types. Nat Commun 10(1):1–14. https://doi.org/10.1038/s41467-019-12928-6

Mamoshina P, Volosnikova M, Ozerov IV et al (2018) Machine learning on human muscle transcriptomic data for biomarker discovery and tissue-specific drug target identification. Front Genet 9:1–10. https://doi.org/10.3389/fgene.2018.00242

Mohamed SK, Nováček V, Nounu A (2020) Discovering protein drug targets using knowledge graph embeddings. Bioinformatics 36(2):603–610. https://doi.org/10.1093/bioinformatics/btz600

Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V (2019) Artificial intelligence transforms the future of health care. Am J Med 132:795–801

O'Boyle NM, Banck M, James CA et al (2011) Open babel: an open chemical toolbox. J Cheminform 3(33):1–14

Patel L, Shukla T, Huang X, Ussery DW, Wang S (2020) Machine learning methods in drug discovery. Molecules 25(22). https://doi.org/10.3390/MOLECULES25225277

Petyuk VA, Chang R, Ramirez-Restrepo M et al (2018) The human brainome: network analysis identifies HSPA2 as a novel Alzheimer's disease target. Brain 141(9):2721–2739. https://doi.org/10.1093/brain/awy215

Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P (2018) Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. Bioinformatics 34(21): 3666–3674. https://doi.org/10.1093/bioinformatics/bty374

Stokes JM, Yang K, Swanson K et al (2020) A deep learning approach to antibiotic discovery. Cell 181(2):475–483. https://doi.org/10.1016/j.cell.2020.04.001

Wong CH, Siah KW, Lo AW (2019) Estimation of clinical trial success rates and related parameters. Biostatistics 20(2):273–286. https://doi.org/10.1093/biostatistics/kxx069

Yadav V, Tonk RK, Khatri R (2020) Molecular docking, 3D-QSAR, fingerprint-based 2D-QSAR, analysis of pyrimidine, and analogs of ALK (anaplastic lymphoma kinase) inhibitors as an anticancer agent. Lett Drug Des Discov 18(5):509–521. https://doi.org/10.2174/1570180817999201123163617

Yadav V, Reang J, Vinita TRK (2022) Ligand-based drug design (LBDD). In: Rudrapal M, Egbuna CBT-CADD (CADD): FL-BM to S-BA (eds) . Elsevier, Drug discovery update, pp 57–99

Zhavoronkov A, Vanhaelen Q, Oprea TI (2020) Will artificial intelligence for drug discovery impact clinical pharmacology? Clin Pharmacol Ther 107(4):780–785. https://doi.org/10.1002/cpt.1795

Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR et al (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 48(5):481–487. https://doi.org/10.1038/ng.3538

Chapter 10
Network Pharmacology and Systems Biology in Drug Discovery

Ashish Shah, Vaishali Patel, Manav Jain, and Ghanshyam Parmar

10.1 Introduction

Combating the primary issues the globe has been facing that is related to global health difficulties has become urgently necessary (Noor et al. 2022). Researchers are interested in complex disorders like cancer and diabetes because they typically result from a malfunction of an entire regulatory network rather than a gene mutation or malfunction (Wang et al. 2021; Zuo et al. 2021). In order to combat complex diseases, it is crucial to understand the molecular mechanisms that control disease prognosis (Noor et al. 2021). Currently, natural products make up a sizable component of modern pharmaceutical agents, especially when it comes to the treatment of diseases (Pal and Shukla 2003). Natural products have historically been a vast repository of powerful resources for humanity (Rehman et al. 2022). In order to screen the pharmacological efficacy of herbal remedies in the drug discovery process, high-throughput approaches have been developed (Tan et al. 2020). This chapter provides a comprehensive overview of the methodology, significance, and application of network pharmacology to cure a wide spectrum of complex disease.

A. Shah (✉) · G. Parmar
Department of Pharmacy, Sumandeep Vidyapeeth, Vadodara, Gujarat, India

V. Patel
Laxminarayandev College of Pharmacy, Bharuch, Gujarat, India

M. Jain
Department of Pharmacology, PGIMER, Chandigarh, Punjab, India

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_10

10.1.1 Drug Discovery and Development Process

Drugs are chemicals that aid in disease prevention or provide instructions for regaining normal health. The scientific field of medicinal chemistry is responsible for either the discovery or the creation of these medications. The traditional medications were generally created through chemical alterations of natural compounds or from natural sources. Technologies aid in a more detailed understanding of disease. The deliberate design, manufacturing, and evaluation of therapeutic candidate molecules are becoming more common as science-based tools help us better understand the nature of disease, how cells function, and how medications affect these processes. In the past, finding new medications was done using only randomized search strategies. Such techniques relied heavily on the skill and good fortune of medicinal chemists. The number of chemical compounds increased daily, rendering the randomized approach useless. This strategy required too much time, offered too little assurance of success, and was too expensive. Due to the low likelihood of discovering a novel agent—less than 1 in 10,000—new medication research expenditures have increased. A new rational and scientific approach that shortens the drug development process's time and cost was required to solve this issue. Drug design is a new rational and scientific strategy to finding novel drugs that has been made possible by scientific evolution. When John Langley proposed the receptor theory in 1906, the process of discovering new drugs had just begun. Paul Ehrlich and Sacachiro developed the first logical pharmaceutical. In 1910, they create arsphenamine by using the structure-activity connection technique from atoxyl. Syphilis and the symptoms of sleeping sickness could be treated with atoxyl (Berdigaliyev and Aljofan 2020; Mohs and Greig 2017).

Drug development and discovery are two distinct processes. Finding therapeutically effective chemicals that can be used to treat a variety of ailments is the process of drug discovery. The goal of the drug discovery process is typically to pinpoint a biological target that is essential to the progression of the disease or originates from a molecule with intriguing biological properties. Target validation, lead compound identification, lead compound validation, synthesis, characterization, screening, and therapeutic efficacy assays are all steps in the drug discovery process. Drug development begins once a molecule has demonstrated therapeutic value in tests; this occurs before clinical trials. With an average cost per drug development of between US$897 million and US$1.9 billion, the drug research and development process are both exceedingly expensive and time-consuming. One medicine will take 10 to 15 years to reach the market. Most of the medications have historically been found either through the identification of the active component in conventional treatments or through accidental discovery. In the current technological era, new techniques to drug development have been created to target certain molecules based on an understanding of how disease and infection are controlled at the molecular and physiological level (Hughes et al. 2011).

10.1.2 Role of Computational Methods in Drug Discovery

Numerous improvements in computer-aided drug design (CADD) techniques have been discovered over the past few decades. Structure-based (SBDD) and ligand-based (LBDD) computational approaches are the two primary categories. Both approaches are helpful at different stages of the drug development process. These techniques can be used to design molecules (molecular modeling), predict interactions between proteins or ligands (docking), predict biological activity (QSAR), predict toxicity, and more. In many instances, the compounds that are improved and tested using different CADD approaches have demonstrated good potential in in vitro or in vivo studies. With the development of several in silico technologies, it is now possible to screen millions of chemicals every day using a variety of techniques. The outcomes of computer approaches assist scientists make decisions on which molecules should be rejected or modified or taken into consideration for further research. In fact, the drug discovery and development processes are greatly sped up and cost-effective, thanks to these computational tools. Applications of computational methods have been documented in great numbers, and they are growing daily. Despite these developments and applications, there are still very few practical uses for these techniques (Gupta et al. 2021; Romano and Tatonetti 2019).

10.1.3 Concept and Significance of Network Pharmacology

Network pharmacology (NP), a brand new field, aims to comprehend the actions and interactions of drugs with a variety of targets. The systematic cataloguing of a drug molecule's molecular interactions in a living cell makes use of computing capability. NP emerged as a key tool for comprehending the intricate connections between plant remedies and the entire body. The systematic cataloguing of a drug molecule's molecular interactions within a living cell makes use of computing capability. NP emerged as a key tool for comprehending the intricate connections between plant remedies and the entire body. To choose the appropriate targets and fresh therapeutic molecule scaffolds, these efforts need some direction. Traditional knowledge can be quite useful in the process of formulating new pharmaceuticals and reusing those that already exist. The next generation of promiscuous drugs can be intelligently created by fusing systems biology and NP advancements (Yang et al. 2022; Zhang et al. 2019).

10.1.4 Systems Biology

It is well accepted that rather than working alone, biomolecules interact with one another to complete their varied jobs in the form of the so-called biomolecular networks. For example, a disease rarely arises as a result of an abnormality in a single gene, but rather reflects disruptions or dysfunctions of the complex biological networks that link the systems of tissues and organs. The genes that make up the "disease module" are those that are most likely to interact and show themselves similarly in individuals with similar diseases. Molecular-molecular interactions, cell-cell contact, unipartite networks, bipartite networks (like drug-target interactions), and even tripartite networks can all be used to define the term "network biology" (e.g., drug–disease–protein interaction) (Muzio et al. 2021). Given that biological entities are involved in intricate relationships, learning about biology from network concepts is extremely important. Benefiting from advancements in network science and high-throughput biomedical technology, network biology research has gained a lot of attention recently. Networks, in the form of linkage maps between genes, phenotypes, and the relevant environmental factors, have long been crucial to our understanding of biological systems. A theoretical paradigm called network biology uses a graphical representation of the biological structure to show how functional information flows through it. If we could grasp and model the network structure, we would be better able to accept the dynamical and functional properties of the network, as well as better comprehend the network's evolutionary mechanisms (Fatima Noor et al. 2022a). In this study, we will cover a number of typical features found in the topology of biological networks as well as their metrics.

10.2 Network Pharmacology: Practical Guide

10.2.1 Common Network Pharmacology Databases

Network pharmacology can have an impact on the drug development process' two primary techniques. One is to create a practical network model and anticipate the drug target using data from open databases or previously published studies. The network equilibrium principle should then be examined through the functional drug's mechanism. The alternative strategy involves combining bioinformatics techniques and high-throughput screen (HTS) technology to reconstruct a "drug target disease" network prediction model. In this method, the interaction between the drug and the model was compared in order to investigate the mechanism of pharmaceuticals in the biological network (Zhang et al. 2013). Numerous cases of network pharmacology's use in drug development have been documented in the literature. Li et al. [1] discovered that multilayer networks may underlie the combined processes of herbal formulas by using Liuwei Dihuang pill (a CHM recipe) to anticipate the appropriate network targets in disease treatment (Li et al. 2010). In

addition, nine ingredients in the Fufang Danshen formula were tested based on network pharmacology and found to affect 42 cardiovascular-related genes (Sun and Yang 2019). Additionally, it was shown that salvianolic acid B was appropriate and practical for the treatment of cardiovascular illness (Wang et al. 2013) by combining the abovementioned study approaches.

Multiple databases being utilized for TCM analysis are Super Natural II31, TCM@taiwan32, ChEMBL33, and TCMID34 TCM compounds, whereas STITCH35, STRING36, and OMIM37 are for protein-disease connections. OMIM37 and GAD38 are for compound-protein interactions. However, several features of the current databases are constraints on network pharmacology analysis. For instance, a database like TCMID, whose molecules are far fewer than STITCH, is insufficient. To increase the effectiveness of the database, a sizable amount of redundancy from databases like OMIM and STITCH must also be removed (Zhang et al. 2017) (Table 10.1).

Numerous databases have been proposed in recent years, many of which, despite having distinct data sources, perform many of the same tasks. Input patterns vary from database to database for most databases. For instance, in TCMSP, "chenpi" or "Citrus reticulata" must be used to search for the herb "Dried Tangerine Peel"; however in TCMID, "chen pi" or "Citri reticulatae pericarpium" must be used (Zhang et al. 2019). Additionally, many substances and proteins have different aliases in various databases (e.g., proteins: protein name, gene symbol, node ID, target ID, target drug bank ID; compounds: chemical name, CID number, STITCH ID, CAS number, PubChem CID, EC number, UNII). The current databases would become more succinct because of the uniform input and output formats, and researchers would be able to take.

The data structure and content of diverse databases should be properly comprehended by such a platform. Additionally, it should facilitate the conversion of various herb, chemical, or protein formats to fill the gap between the existing databases for TCM research. Currently, one database can directly use the output of another database as an input. Most databases include related web services; as a result, the converting platform that might transfer data from one database to another can be offered to users as web services that are linked to various databases or to a browser plug-in. The integration of many databases, in addition to database unification, is quite important. Currently, more than 10,000 herbs have been reported as being utilized in more than 100,000 herbal formulations (Zhou et al. 2020).

10.2.2 Research Approaches of Network Pharmacology

Network pharmacology can have an impact on the drug development process' two primary techniques. One is to create a practical network model and anticipate the drug target using data from open databases or previously published studies. The alternative strategy involves combining bioinformatics techniques and high-throughput screen (HTS) technology to reconstruct a "drug target disease" network

Table 10.1 Software and databases related to network pharmacology

Type of databases	Name	Description
TCM-related databases	TCM-mesh	Network pharmacology investigation of TCM formulations using a data-mining technology
	HIT	A thorough and meticulously managed library to supplement existing sources on potential therapeutic targets for proteins as well as the probable precursor substances
	TCM Database@Taiwan	The largest and most complete free tiny molecular database on TCM available anywhere in the globe for virtual screening
	TCMSP	TCMs have a distinctive systems pharmacology platform that captures the connections between treatments, targets, and illnesses
	TCMID	A thorough informational resource that explains everything and connects TCM with contemporary life sciences
Drug-related databases	STITCH	A database of known and anticipated chemical-protein interactions
	ChEMBL	An open data repository with binding, functional, and ADMET details for numerous bioactive compounds with drug-like properties
	PubChem	A system for analyzing small molecule bioactivities that is accessible to the public
Target-related databases	STRING	A database of known and predicted protein-protein interactions
	MINT	A database that emphasis on experimentally verified protein-protein interactions mined from the scientific literature
	ntAct	Open-source database system and analysis tool for molecular interaction data
	Reactome	A free, open-source, curated, and peer-reviewed pathway database
	HAPPI	A comprehensive online library of anticipated and annotated protein interactions for humans
Disease-related databases	OMIM	A complete, reliable reference work on human genes and genetic characteristics
	GAD	A database of genetic association data from complex diseases and disorders
Software	Cytoscape	A software environment for integrated models of biomolecular interaction networks
	Pajek	For complex network analysis

prediction model. This method compared the interactions of the drug and the model to investigate the mechanism of drugs in the biological network. Finding genes associated with substances and diseases, building a protein-protein interaction (PPI) network, and then analyzing and visualizing the network are the main steps in network pharmacology research (Huang et al. 2014). A straightforward beginning

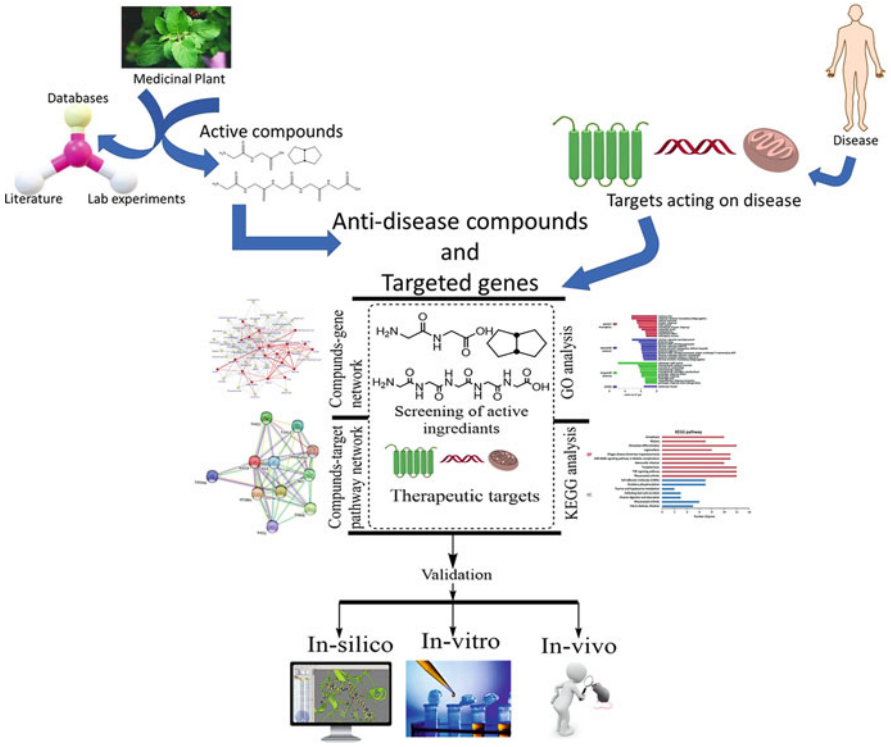


Fig. 10.1 Process for performing network pharmacology

is the creation of molecular networks from big databases. Key nodes are then discovered and important biological pathways are anticipated using network analysis. To successfully validate the interaction between highly active constituents and their potential targets, further network validation is carried out. Modernization of medicinal plants is greatly influenced by advancements in systems biology and bioinformatics, which alter our understanding of the treatment and diagnosis of diseases using medicinal plants from a network pharmacology perspective (Fig. 10.1) (Silverman et al. 2020).

10.2.3 Data Collection and Validation

The efficacy and safety of a proposed medicine and their potent combinations are optimized by network pharmacology. These two steps are also the most crucial in every experimental study. The selection of original data from the trials to construct a biological network is the initial stage in network pharmacology. The experimental verification of the projected network model comes second. Numerous integrated

techniques, such as genomics, proteomics, metabolomics, and HTS/high-content screening (HCS) technologies, can be used to quantify the validated data (Noor et al. 2022b).

HTS/HCS technologies can quickly identify chemicals, detect millions of data samples, influence a specific biochemical pathway, or change the phenotypic of a cell. These have a lot of attractive characteristics, including visualization, real-time dynamic monitoring, and homogenous multidimensional phenotypic detection (Szymański et al. 2012). Additionally, the network data from the trials may be collected using this dual high-throughput technology, which can also confirm the network model. Another method that confirms the network pharmacology approach, discloses the mechanisms behind drug activity, and verifies the drug network or anticipated model is molecular interaction validation technology (Nguyen et al. 2015). High-throughput, high-precision, label-free, and real-time detection are all aspects of these methods (Pe'er and Hacohen 2011).

For locating disease and chemical targets that overlap, Venn diagrams are preferred. This process primarily tries to forecast the genes associated with diseases and then find the genes shared by diseases and substances. The initial measuring sticks for subsequent screening are the common genes. To better understand how medicinal plants heal diseases, network analysis is used (Chen et al. 2015). Due to their high selectivity, adaptability, and versatility, protein-protein interactions (PPI) are of utmost importance. Databases that give information on the functional connections between important targets are used to create the PPI network of key targets (common genes). Later, the hub genes with the highest degree of linkage are predicted using network analysis (Li et al. 2019).

How to extract critical information from networks is the crucial point. By identifying targets, network analysis attempts to reveal significant targets, active substances, and their related pathways. There are many approaches used in network analysis, but network functional analysis is the most used one. It has been found that biological networks have a modular structure, and many helpful medications operate therapeutically by influencing multiple proteins rather than just one. Topological study has revealed several subnetworks with specific roles and topologies in big and complex networks. By examining the related pathways, GO enrichment analysis and KEGG pathway analysis give unique important target properties at the functional level (Mlecnik et al. 2018; Reimand et al. 2019).

10.2.4 Network Analysis and Visualization

Using relevant technology, network analysis concentrates on an existing network and extracts beneficial data that is valuable for further research. There are three different kinds of network analysis. In the first, when specific network data has been extracted and maximally conserved as hidden information within the network, the topological structure and statistical parameters of the network are calculated. Second, by generating suitable modulation, random networks are generated and

compared to existing networks to assess their dependability. Finally, the network is hierarchically clustered, an algorithm is used to predigest the complex network, and potential network information is expected.

Utilizing visualization technologies, network visualization is used to separate the interaction information from inter-association data and switch it into a visual network. There are two steps in this process: (1) enhancing network properties, adding nodes, and boosting network connectivity and (2) defining the network and using a variety of tools to describe the architectural characteristic that accurately and perceptually depicts the network. Currently, the majority of network pharmacology visualization is done using specialized programs like Cytoscape, GUESS, and Pajek. It is possible to validate the effectiveness of projected molecular targets using a variety of techniques. Although in vitro and in vivo procedures are typically seen to be the most effective, they take a long time and cost a lot of money to complete. However, as high-throughput technologies have developed and the genomic era has advanced, several in silico techniques have been developed, offering a useful platform for the validation of results. Finally, it is possible to validate the projected outcomes using both experimental and virtual methods (Wu et al. 2011; Zhang et al. 2013).

To forecast the docking locations of active components and important targets determined from network pharmacology, receptor-ligand molecular docking is used. By effectively bridging the gap between western medicine and herbal medicine, network pharmacology and molecular docking also considerably enhance mechanistic investigations on the synergistic effects of herbal medications. The most useful strategy in the drug development toolbox, molecular docking, has emerged as a lightning rod. The interaction that binds ligands to their corresponding proteins in a bound state can be predicted using molecular docking. For constituent screening, docking score and binding energy are mostly regarded as important factors. Numerous research has shown the value of using molecular docking as a network validation tool.

10.3 Applications of Network Pharmacology in Drug Discovery

The creation and use of network pharmacology has given rise to new ideas for investigating the mechanisms of diverse natural products and formulations. These ideas also assist in the discovery of new medications, components, targets, and disease-treating pathways. The study of the compatibility of traditional medicine with modern medicine, the development of novel indications, the confirmation of the mechanism of pharmacological action, and other areas of traditional medicine research are all made possible by network pharmacology technology (Fig. 10.2).

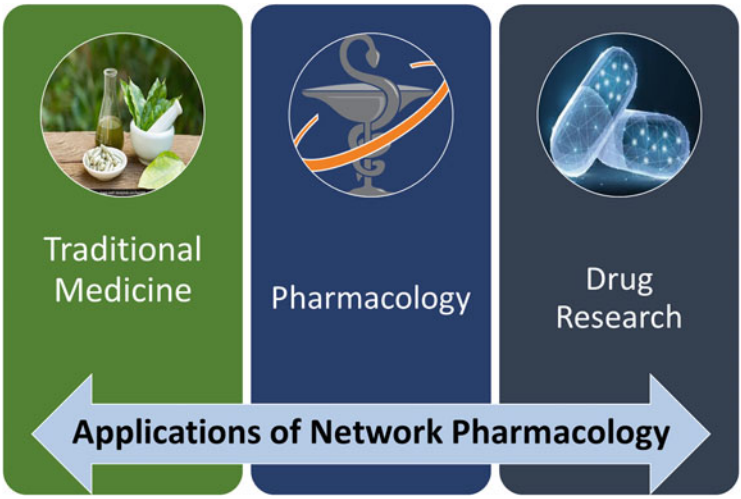


Fig. 10.2 Applications of network pharmacology

10.3.1 Applications of Network Pharmacology for Plant-Based Drug Discovery

10.3.1.1 Case Study I: Network Pharmacology-Based Virtual Screening of Active Constituents of *Prunella vulgaris* L. Against Breast Cancer (Zhang et al. 2020b)

Prunella vulgaris L. is the scientific name for a perennial herbaceous plant in the Prunella genus. It is a traditional Chinese drug that is widely used to treat headaches, eye pain, cancer, and inflammation. Recent pharmacological studies suggest that *Prunella vulgaris* L. may possess antiviral, antibacterial, anti-inflammatory, immunoregulatory, antioxidative, and antitumor effects.

Screening of Phytochemicals

In addition to using the phrase "Prunellae Spica" in the TCMSP search, literature was also searched for on Google Scholar, Pubmed, and the CNKI database. The chemical structures were located using PubChem (https://pubchem.ncbi.nlm.nih.gov/) and ChemSpider (https://www.chemspider.com/). For structures that were not in the database, references from TCMSP and original research articles were used. DL is available on the molsof website (https://www.molsof.com/mprop/) for use in calculating potential components. DL-0.18 was used as the benchmark for detecting act active components. Selected compounds (DL 0.18), which were imported using PreADMET (https://preadmet.bmdrc.kr/), were evaluated by Caco-2, HIA, and PPB. Even though they do not meet this criterion, active substances that unmistakably have biological effects were taken into account. With favonoids, triterpenes, and phenolic acids being the most biologically active compounds, the screening results

showed that 31 constituents were expected to be physiologically active in vivo. These results were consistent with the previously identified anticancer active components of *Prunella vulgaris* L.

Screening of Breast Cancer Targets
The STITCH (https://stitch.embl.de/) and Swiss Target Prediction (https://www.swisstargetprediction.ch/) databases were used to find the gene targets for active compounds. To do this, Homosapiens was selected as the species, the screened components were uploaded to the STITCH database, and the targets were gathered with a combined score of 0.7. The smiles numbers for each component were sent to the Swiss Target Prediction online site. The structural similarity of the target was predicted using reverse pharmacophore matching, and a target with probability of 0.7 was selected.

Network Analysis
According to GO functional analysis, the main targets of *Prunella vulgaris* L. are largely involved in estrogen receptor binding, steroid hormone receptor binding, steroid hormone receptor activity, etc. Using KEGG pathway analysis, relevant signaling pathways linked to *Prunella vulgaris* L.'s anti-breast cancer impact were found. The processes with the highest gene content were proteoglycans in cancer, endocrine resistance, human cytomegalovirus infection, microRNAs in cancer, breast cancer, and Kaposi sarcoma-associated herpesvirus infection. The 20 KEGG signaling pathway results showed that the following genes were significantly enriched: EGF, AKT1, EGFR, ERBB2, SRC, MTOR, MYC, BCL2, JUN, VEGFA, MMP9, and CTNNB1.

PPI Network Analysis
Thirty one target genes associated with antibreast cancer action were incorporated into the STRING database using network design. The interactions that took place when breast cancer first appeared are represented by the nodes in the PPI network. AKT1, ESR1, MYC, JUN, SRC, CASP3, and VEGFA showed higher degrees when the PPI diagram was analyzed using Cytoscape's analysis tool.

10.3.1.2 Case Study II: A Network Pharmacology Approach to Investigate the Anticancer Mechanism and Potential Active Ingredients of *Rheum palmatum* L. Against Lung Cancer (Zhang et al. 2020a)

Rheum palmatum L. (RPL) is a well-known herbal remedy in traditional Chinese medicine with the capabilities of "heat-clearing and damp-drying." Although it has been demonstrated in the past that it can combat lung cancer, little is known about the original methods and substances that enable it to function.

Screening of Phytochemicals
Searches in databases and libraries turned up 1380 compounds in RPL. Based on the screening results for OB and DL, the OB and DL values of 16 compounds were

judged to be "Qualified," indicating that these 16 compounds were potentially active components of RPL in the treatment of lung cancer. In addition, certain important substances with low OB and OL features in RPL and alleged anti-lung cancer tumor properties were also added, such as emodin, resveratrol, chrysophanol, and physcion. Finally, the investigation included 20 substances.

Screening of Targets
A total of 22,418 lung cancer-related targets and 817 potential targets for the 20 RPL medicines were acquired from the OMIM, DisGeNET, TTD, and GeneCards databases. 761 potential anti-lung cancer targets were discovered by the combination of common targets. It is interesting to see that the majority of the 761 targets are shared by the 20 active compounds in RPL.

Network Analysis
The cytoscape program was used to illustrate the protein interaction, employing 761 nodes and 6840 edges, in line with the String's predictions and findings. The top 20 hub genes (INS, AKT1, TP53, ALB, IL6, EGFR, VEGFA, MYC, SRC, TNF, CASP3, HSP90AA1, STAT3, ESR1, MAPK8, CTNNB1, MTOR, CCND1, ERBB2, and APP) were eliminated based on the number of nodes. The genes with the highest node degrees are INS, AKT1, TP53, and ALB. Three main targets—INS, AKT1, TP53, and ALB—have been suggested as the focus of RPL's anticancer action against lung cancer.

KEGG Pathway Enrichment Analysis
The results showed that 46 pathways finally had substantial correlations with the target genes, and a total of 533 genes were implicated in the enrichment (P 0.05). The top 20 pathways are shown, with apoptosis, PI3K-Akt signaling, apoptosis-multiple species, MAPK signaling, and p53 signaling pathways being prominently displayed. These signaling pathways are either intimately linked to the mechanism of RPL's anticancer effects in this illness or directly or indirectly associated to the occurrence and progression of lung cancer.

10.3.1.3 Case Study III: Network Pharmacology-Based Virtual Screening of Active Constituents of *Cnidium monnieri* in Treating Hepatocellular Carcinoma (HCC) (Khan and Lee 2022)

Cnidium monnieri is a Chinese herbal remedy that has been used for over 2000 years. *Cnidium monnieri* has historically been used to treat female vaginal problems, male impotence, and skin ailments. In contemporary TCM therapeutic practice, water decoctions and tinctures of *Cnidium monnieri* are often used alone or in conjunction with other Chinese medicinal herbs to treat persistent skin itch, superficial fungal infections, and atopic dermatitis. *Cnidium monnieri* extracts and components have been found in pharmacological investigations to possess

antibacterial, anticancer, antitumor, and anti-inflammatory properties which may be used to prevent and cure liver infections caused by hepatitis and HCC.

Screening of Phytochemicals
The SwissTargetPrediction database was employed for determining potential protein targets for the active phytochemicals in *Cnidium monnieri*. A total of 1387 potential protein targets were obtained with a probability score > 0. After the removal of redundancies, 532 potential protein targets of active phytochemicals in *Cnidium monnieri* were investigated further.

HCC-Related Genes
A total of 564 HCC-related genes were recovered from OncoDB.HCC (http://oncodb.hcc.ibms.sinica.edu.tw) and Liverome (http://liverome.kobic.re.kr/index.php). Intersecting targets were recognized between the HCC-related genes and potential protein targets of active phytochemicals using the VENNY 2.1.0. online system. A total of 67 intersecting targets were identified between them.

Network Construction and PPI Analysis
The PPI network was constructed by importing the 67 intersecting targets to the STRING database. The PPI network contained 67 nodes and 528 edges. The average PPI enrichment p-value, average local clustering coefficient, and average node degree were $p < 0.00001$, 0.609, and 15.8, respectively. The STRING PPI results were further analyzed by exporting them in a simple textual data format (.tsv) file to Cytoscape software (version 3.9.0). The results showed that the PPI network involved 65 nodes and 1056 edges. The characteristic path length between all node pairs was 2.045. The PPI network radius, diameter, heterogeneity, and density were 3, 5, 0.682, and 0.254, respectively. 25 nodes that achieved the degree centrality (DC) criterion with an average value greater than 32.49 were further extracted and classified as potential anti-HCC core targets. The top six potential anti-HCC core targets, i.e., EGFR, CASP3, ESR1, MAPK3, ERBB2, and CCND1, were chosen.

10.3.2 Applications of Network Pharmacology for Phytoconstituents

10.3.2.1 Case Study I: Network Pharmacology-Based Virtual Screening of Resveratrol Which Can Alleviate (Xiao et al. 2021)

COVID-19-related Hyperinflammation
Resveratrol, an antioxidant phytoalexin with possible chemopreventive qualities, may be isolated from grapes and a variety of other plants (PubChem CID is 445154). By controlling immune cells and preventing the production of pro-inflammatory cytokines, resveratrol has anti-inflammatory properties. Additionally, resveratrol functions as an antiviral drug via a variety of ways. Numerous viruses, including

the human meta pneumonia virus, respiratory syncytial virus, influenza virus, Epstein-Barr virus, enterovirus, and HIV, have been shown to be inhibited by resveratrol.

Screening of Phytochemicals

PubChem (https://pubchem.ncbi.nlm.nih.gov/) was searched for the phrase "Resveratrol," and the PubChem CID (445154) of Resveratrol was found. Target Net (http://targetnet.scbdd.com), an open web server used for netting or predicting the binding of multiple targets for any given molecule, and Comparative Toxicogenomic Database (CTD, http://ctdbase.org/about/), a substantial publicly accessible database that offers manually curated information about chemical-gene/protein interactions and chemical-disease and gene disease relationships, were used to predict potential targets for resveratrol. The names of the targets were provided using the official symbol format from the UniProt Knowledgebase (UniProtKB, http://www.uniprot.org/), because to the non-standard naming.

Screening of SARS-CoV2-Related Gene Targets

The GSE147507 dataset containing the host transcriptional response to SARS-CoV-2 was downloaded from GEO. We selected the NHBE transcriptional data for study. R packages of "impute" and "limma" provided by the Bioconductor project (http://www.bioconductor.org/packages/release/bioc/html/affy.html) were applied to assess the transcriptional results of NHBE. Quantile normalization and log 2-transformation were used to create a robust multiarray average (RMA). Adjusted original p values were obtained via the Benjamini-Hochberg method; the false discovery rate (FDR) procedure was used to calculate fold changes (FC). Gene expression values of $|\log 2\ FC| > 1$ and p value < 0.05 were used as a threshold to filter differentially expressed genes (DEGs).

PPI Network Construction

Intersecting target genes of resveratrol and DEGs associated to SARS-CoV-2 were obtained. The resveratrol-related targets and SARS-CoV-2 DEGs were uploaded to String (https://string-db.org/) with species set as "Homo sapiens," a confidence score $> 0{:}9$ to construct PPI networks, and then, the 2 PPI networks were combined and displayed using Cytoscape 3.7.2 (http://www.cytoscape.org).

PPI Network Analysis

Resveratrol and SARS-CoV-2 DEG targets that overlapped each other were MMP13, PRKCB, PLAT, KCNH2, ICAM1, PDGFRB, TNF, ITGB3, CSF1R, BCL2A1, and MMP9. The systematic visualization and quantification of a given protein's function in a cell was done using a PPI network. The PPI network of SARS-CoV-2 DEGs and resveratrol-related targets was created, and the shared targets were found. Resveratrol's potential therapeutic targets on COVID-19, which include MMP13, PRKCB, PLAT, KCNH2, ICAM1, PDGFRB, TNF, ITGB3, CSF1R, BCL2A1, and MMP9, may be represented by the targets that it shares with the SARS-CoV-2 DEGs. The IL-17 signaling route, the NF-B signaling pathway, and the TNF signaling pathway are the three main ones affected by these. These pathways become more active, which increases the release of cytokines,

which have been proven to be crucial in viral infection. We hypothesize that resveratrol can prevent the activation of these pathways, hence reducing cytokine expression levels and reducing hyperinflammation in COVID-19.

10.3.2.2 Case Study II: Network Pharmacology-Based Virtual Screening of Curcumin against Triple-Negative Breast Cancer (TNBC) (Deng et al. 2022)

An active component of turmeric called curcumin (CUR) has been shown to be able to stop different cancer cells from proliferating, invading, and metastasizing. It was discovered that CUR may be able to stop the development and spread of head and neck cancer cells. CUR can prevent FIC133 cells from migrating and invading (a human thyroid cancer cell line).

Screening of Phytochemicals
Potential CUR-related targets were obtained from the Swiss Target Prediction database (http://www.swisstargetprediction.ch/) and the Encyclopedia of Traditional Chinese Medicine (ETCM) database (http://www.tcmip.cn/ETCM/index.php/Home/). In the Swiss Target Prediction database, potential CUR-related targets were retrieved by the structure of CUR, and species were limited to "Homo sapiens." In the ETCM database, potential CUR-related targets were searched directly using the keyword of "CUR." All targets obtained in both the Swiss Target Prediction database and ETCM were selected as potential CUR-related targets.

Screening of Targets of CUR Against TNBC
Using the keywords "triple negative breast cancer/carcinoma," targets associated to TNBC were found in the Therapeutic Target Database (TTD, http://db.idrblab.net/ttd/), Online Mendelian Inheritance in Man (OMIM, https://omim.org/), and DisGeNET databases. The common targets of CUR and TNBC that were identified as prospective targets of CUR against TNBC were subsequently screened using Venny 2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/index.html).

GO and KEGG Pathway Enrichment Analysis
GO and KEGG pathway enrichment of the common targets of CUR and TNBC were analyzed on the Metascape website (https://metascape.org/). Molecular function (MF), cell component (CC), and biological process (BP) were included in the GO analysis. The analyses were carried out using a p value of less than 0.01. The top 10 items of GO and the top 20 items of the KEGG pathway were selected and visualized on the Bioinformatics website (http://www.bioinformatics.com.cn/).

PPI Network Analysis
A PPI network was acquired by STRING to illustrate the relationships between the 40 common targets of CUR and TNBC. The PPI network has 40 nodes and 223 edges, and the average node degree of the constructed network is 11.2; the PPI enrichment p value is less than 1.0 e−16, and the local clustering coefficient is 0.64. Subsequently, the constructed network was further investigated to screen the

top 10 targets by MCC scores. The higher the MCC score, the more essential role the protein plays. According to the MCC scores, the top 10 targets are STAT3, AKT1, TNF, PTGS2, MMP9, EGFR, PPARG, NFE2L2, EP300, and GSK3B, which may play important roles in the constructed PPI network of potential targets of CUR against TNBC.

Systems Biology

The developments in biological science have produced a vast amount of data in the fields of metabolomics, transcriptomics, and genomics. Studying the organization of biological entities, their interactions, and the dynamic changes in the behavior of these entities under external circumstances is vital for a better understanding of biosystems. To examine the structure and dynamics of cells, tissues, and organisms functioning as a system, the concept of systems biology evolved. Systems biology involves the computational modeling of biological systems (Breitling 2010; Kitano 2002). The main issues with drug development are the rising attrition rate brought on by toxicity, the emergence of drug resistance, and the inconsistent efficacy of medications in various people due to diverse therapeutic responses (Hopkins 2008). Additionally, many prospective medications fail in clinical trials because it is unclear how they work. Most pharmaceutical corporations use reductionist methods in their medical research, which only provide a limited understanding of complicated diseases such as cancer, cardiovascular disease (CVD), and neurogenerative diseases (Zimmermann et al. 2007). Large, integrated signaling networks control various systemic disorders, and many of these signaling pathways contribute to the development of the illness (Breitkreutz et al. 2012; Heineke and Molkentin 2006). Therefore, a more comprehensive knowledge of the illness process and treatment response requires a systemic, integrative approach. Studying physiological and pathological circumstances at the level of regulatory networks, signaling pathways, cells, tissues, organs, and ultimately the entire organism is the goal of systems biology (Butcher et al. 2004). Numerous methods and models used in systems biology help understand the biological complexity of diverse diseases. It offers a platform for combining a lot of reductionist information from genomes, proteomics, and/or metabolomics research to create a network model for analyzing disease and creating new treatments for it.

10.4 Computational Approaches in System Biology

The study of interactions between all system components rather than the features of individual components is a key component of systems biology. Systems biology relies on a combination of experimental research that produces information on the biological components of a system and computational methods that help with the analysis of multiple datasets. Systems biology employs two main computational paradigms: data-driven (top-down approach) and hypothesis-driven (bottom-up approach) (Faratian et al. 2009).

10.4.1 Top-Down Approach

The top-down method involves gathering enormous quantities of omics data and using statistical modeling approaches to assess these datasets (Faratian et al. 2009). One of the most popular data-driven techniques is network modeling (Ma'ayan 2011), which explains how various biological systems' components interact with one another. Analysis of network models reveals the network's topological characteristics, which aid in comprehending the key components and characteristics of the network (Ma'ayan 2011). The fundamental procedures in network modeling and analysis are as follows (Rai et al. 2018):

1. Data mining, which entails extracting interaction information for genes, proteins, metabolites, medicines, etc. from interaction databases
2. Building an interaction network utilizing the retrieved data
3. Verification of the interaction network through comparison to random networks
4. Analysis of the interaction network for topological characteristics, overrepresented pathways, potential disease genes, biomarkers, therapeutic targets, etc.

10.4.2 Bottom-up Approach

On the other hand, smaller systems with comparatively fewer interacting components are studied using the bottom-up method. Since the quantitative aspects of the interactions are unknown, a significant flaw in the bottom-up technique is that it necessitates the formulation of pertinent equations to accurately estimate the values of the parameters connected to the interactions. To quantify relationships between molecular elements and behavior resulting from their interactions, dynamic modeling is the most popular hypothesis-driven method (Faratian et al. 2009). The following steps must be taken to construct a dynamic model (Rai et al. 2018).

1. Create a connectivity diagram that shows each element of a biological process and their connections.
2. Use connectivity diagrams to create mathematical equations.
3. Using a predetermined set of parameters and an initial concentration, the model is calibrated to estimate unknown kinetic parameter values.
4. Model validation through experimental testing of simulation outcomes.

10.4.3 System Biology Application in Drug Discovery

The development of a new therapeutic formulation against any disease requires a thorough understanding of the disease process. As the reductionist approach links a single gene to a single disease, it offers little information on disease mechanisms (Butcher et al. 2004). Systems biology takes into account every element of a system,

their connections with one another, and how various signaling pathways interact to help us better comprehend complex disorders (Rai et al. 2018). Systems biology is therefore frequently utilized in the drug discovery process. The creation of high-throughput datasets of system components (omics data), experimental methods of analysis and data integration, the creation and use of network methodologies, and computationally produced models are all included in systems biology research. Models of cell signaling, pathway, and disease networks are built using omics data sets collected from genomes, transcriptomics, proteomics, and metabolomics, and they are merged to help find new targets and better understand and predict drug action in vivo (Butcher et al. 2004).

10.4.3.1 Target Identification

For identifying critical nodes controlling significant disease pathways based on network topology, recent methodologies aggregate gene expression data together with other information into networks. This is evident from various studies which use system biology for target identification. Kim et al. (2012) constructed a large-scale protein and DNA interaction network using gene expression data, expression quantitative trait loci analyses, and molecular interaction data to discover probable causative genes and dysregulated pathways to identify newer targets in glioblastoma. To integrate many heterogeneous data sources and build a gene network to find possible treatment targets for breast, colon, and lung cancer, an ensemble framework approach based on relevance vector machines (RVM) was used (Wu et al. 2012). A study to better understand lung cancer in female non-smokers combined data from various high-throughput sequencing experiments, with data on gene expression profiling and DNA copy number variation (Kim et al. 2013). Despite the attempts to identify new targets, one of the challenges in this process is that most targets remain undruggable.

10.4.3.2 Mechanism of Action

A lot of decisions about drug development depend on understanding the mechanism of action. Understanding the pathways and biological processes that a pharmacological substance affects is also referred to as the mechanism of action. For example, the network of interactions between genes, miRNA, and proteins offered an effective model for researching cancer's aggressive sensitivity to decitabine (Radpour et al. 2011). Similarly, the capacity to categorize substances according to known mechanisms has been demonstrated with the use of the connection map technique and query signatures using sets of genes that have previously been identified as coregulated or predictive of specific processes (Zhang and Chan 2010). Systems biology approaches are also used to research pharmacological combinations to both understand how they work and discover new combinations (Butcher et al. 2004).

10.4.3.3 Biomarkers Identification

Due to the large number of patients needed to establish and assess biomarker performance, developing biomarkers is difficult. Novel clinical biomarker development has been transformed by omics technology, and systems biology approaches are now having an impact on this field's advancements. Multiple analytes-based biomarker panel Oncotype DX (Paik 2007) for stratification of breast cancer patients with ER-positive subtype and ova1-based classification (Zhang and Chan 2010) of ovarian cancer patients in a high or low-risk group are some of the examples in this field.

10.5 Conclusion and Future Prospects

Systems biology methods could revolutionize drug development and discovery. Results from quantitative and systems pharmacology approach to translational medicine are just now becoming apparent to researchers. Future studies will focus on several areas which include better process and data integration methods, development and sharing of network and newer computational methods to analyze and integrate multiscale information, cell signaling methods combined with computational disease models, and a combination of drug target and drug information approach including drug safety information (Butcher et al. 2004).

In cases where there is no effective treatment available, medicinal plants offer a fresh alternative. Humans have used herbal medicines to treat a variety of illnesses for a very long time. The adverse consequences of synthetic drugs have necessitated advancements in the use of natural remedies for disease management. Future application of developing technologies must be supported by study if astounding benefits are to be made. The majority of commercially available medications come from plants. The network pharmacology technique lays the most recent scientific groundwork for evaluating the effectiveness of multicomponent, multi-target drug formulations and investigating the disease treatment of multiple therapeutic targets. In conclusion, developments in bioinformatics and systems biology will cause an operational change away from reductionism and toward network pharmacology. They will surely result in a paradigm shift in drug research and help modernize and expand the use of natural medicines around the world. Another trend might be the use of various dynamic networks and quantitative networks, and as network pharmacology technology is increasingly employed, costs will drop significantly in the future.

References

Berdigaliyev N, Aljofan M (2020) An overview of drug discovery and development. Future Med Chem 12(10):939–947. https://doi.org/10.4155/fmc-2019-0307

Breitkreutz D, Hlatky L, Rietman E, Tuszynski JA (2012) Molecular signaling network complexity is correlated with cancer patient survivability. Proc Natl Acad Sci U S A 109(23):9209–9212. https://doi.org/10.1073/pnas.1201416109

Breitling R (2010) What is systems biology? Front Physiol 1:9. https://doi.org/10.3389/fphys.2010.00009

Butcher EC, Berg EL, Kunkel EJ (2004) Systems biology in drug discovery. Nat Biotechnol 22(10):1253–1259. https://doi.org/10.1038/nbt1017

Chen Q, Zhou XJ, Sun F (2015) Finding genetic overlaps among diseases based on ranked gene lists. J Comput Biol 22(2):111–123. https://doi.org/10.1089/cmb.2014.0149

Deng Z, Chen G, Shi Y, Lin Y, Ou J et al (2022) Curcumin and its nano-formulations: defining triple-negative breast cancer targets through network pharmacology, molecular docking, and experimental verification. Front Pharmacol 13:920514. https://doi.org/10.3389/fphar.2022.920514

Faratian D, Clyde RG, Crawford JW, Harrison DJ (2009) Systems pathology--taking molecular pathology into a new dimension. Nat Rev Clin Oncol 6(8):455–464. https://doi.org/10.1038/nrclinonc.2009.102

Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK et al (2021) Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Mol Divers 25(3):1315–1360. https://doi.org/10.1007/s11030-021-10217-3

Heineke J, Molkentin JD (2006) Regulation of cardiac hypertrophy by intracellular signalling pathways. Nat Rev Mol Cell Biol 7(8):589–600. https://doi.org/10.1038/nrm1983

Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 4(11):682–690. https://doi.org/10.1038/nchembio.118

Huang C, Zheng C, Li Y, Wang Y, Lu A et al (2014) Systems pharmacology in drug discovery and therapeutic insight for herbal medicines. Brief Bioinform 15(5):710–733

Hughes JP, Rees S, Kalindjian SB, Philpott KL (2011) Principles of early drug discovery. Br J Pharmacol 162(6):1239–1249. https://doi.org/10.1111/j.1476-5381.2010.01127.x

Khan SA, Lee TKW (2022) Network-pharmacology-based study on active phytochemicals and molecular mechanism of Cnidium monnieri in treating hepatocellular carcinoma. Int J Mol Sci 23(10). https://doi.org/10.3390/ijms23105400

Kim J, Gao L, Tan K (2012) Multi-analyte network markers for tumor prognosis. PLoS One 7(12):e52973. https://doi.org/10.1371/journal.pone.0052973

Kim SC, Jung Y, Park J, Cho S, Seo C et al (2013) A high-dimensional, deep-sequencing study of lung adenocarcinoma in female never-smokers. PLoS One 8(2):e55596. https://doi.org/10.1371/journal.pone.0055596

Kitano H (2002) Systems biology: a brief overview. Science 295(5560):1662–1664. https://doi.org/10.1126/science.1069492

Li S, Zhang B, Jiang D, Wei Y, Zhang N (2010) Herb network construction and co-module analysis for uncovering the combination rule of traditional Chinese herbal formulae. BMC Bioinformatics 11(Suppl 11):S6. https://doi.org/10.1186/1471-2105-11-s11-s6

Li H, Wang X, Lu X, Zhu H, Li S et al (2019) Co-expression network analysis identified hub genes critical to triglyceride and free fatty acid metabolism as key regulators of age-related vascular dysfunction in mice. Aging (Albany NY) 11(18):7620–7638. https://doi.org/10.18632/aging.102275

Ma'ayan A (2011) Introduction to network analysis in systems biology. Sci Signal 4(190):tr5. https://doi.org/10.1126/scisignal.2001965

Mlecnik B, Galon J, Bindea G (2018) Comprehensive functional analysis of large lists of genes and proteins. J Proteome 171:2–10. https://doi.org/10.1016/j.jprot.2017.03.016

Mohs RC, Greig NH (2017) Drug discovery and development: role of basic biological research. Alzheimers Dement (N Y) 3(4):651–657. https://doi.org/10.1016/j.trci.2017.10.005

Muzio G, O'Bray L, Borgwardt K (2021) Biological network analysis with deep learning. Brief Bioinform 22(2):1515–1530. https://doi.org/10.1093/bib/bbaa257

Nguyen HH, Park J, Kang S, Kim M (2015) Surface plasmon resonance: a versatile technique for biosensor applications. Sensors (Basel) 15(5):10481–10510. https://doi.org/10.3390/s150510481

Noor F, Saleem MH, Aslam MF, Ahmad A, Aslam S (2021) Construction of miRNA-mRNA network for the identification of key biological markers and their associated pathways in IgA nephropathy by employing the integrated bioinformatics analysis. Saudi J Biol Sci 28(9):4938–4945. https://doi.org/10.1016/j.sjbs.2021.06.079

Noor F, Saleem MH, Javed MR, Chen JT, Ashfaq UA et al (2022) Comprehensive computational analysis reveals H5N1 influenza virus-encoded miRNAs and host-specific targets associated with antiviral immune responses and protein binding. PLoS One 17(5):e0263901. https://doi.org/10.1371/journal.pone.0263901

Noor F, Tahir Ul Qamar M, Ashfaq UA, Albutti A, Alwashmi ASS et al (2022a) Network pharmacology approach for medicinal plants: review and assessment. Pharmaceuticals (Basel, Switzerland) 15(5):572. https://doi.org/10.3390/ph15050572

Noor F, Tahir ul Qamar M, Ashfaq UA, Albutti A, Alwashmi ASS et al (2022b) Network pharmacology approach for medicinal plants: review and assessment. Pharmaceuticals 15(5):572

Paik S (2007) Development and clinical utility of a 21-gene recurrence score prognostic assay in patients with early breast cancer treated with tamoxifen. Oncologist 12(6):631–635. https://doi.org/10.1634/theoncologist.12-6-631

Pal SK, Shukla Y (2003) Herbal medicine: current status and the future. Asian Pac J Cancer Prev 4(4):281–288

Pe'er D, Hacohen N (2011) Principles and strategies for developing network models in cancer. Cell 144(6):864–873. https://doi.org/10.1016/j.cell.2011.03.001

Radpour R, Barekati Z, Kohler C, Schumacher MM, Grussenmeyer T et al (2011) Integrated epigenetics of human breast cancer: synoptic investigation of targeted genes, microRNAs and proteins upon demethylation treatment. PLoS One 6(11):e27355. https://doi.org/10.1371/journal.pone.0027355

Rai S, Raj U, Varadwaj PK (2018) Systems biology: a powerful tool for drug development. Curr Top Med Chem 18(20):1745–1754. https://doi.org/10.2174/1568026618666181025113226

Rehman A, Ashfaq UA, Javed MR, Shahid F, Noor F et al (2022) The screening of phytochemicals against NS5 polymerase to treat Zika virus infection: integrated computational based approach. Comb Chem High Throughput Screen 25(4):738–751. https://doi.org/10.2174/1386207324666210712091920

Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C et al (2019) Pathway enrichment analysis and visualization of omics data using g:profiler, GSEA, cytoscape and enrichment. Map Nat Protoc 14(2):482–517. https://doi.org/10.1038/s41596-018-0103-9

Romano JD, Tatonetti NP (2019) Informatics and computational methods in natural product drug discovery: a review and perspectives. Front Genet 10:368–368. https://doi.org/10.3389/fgene.2019.00368

Silverman EK, Schmidt HH, Anastasiadou E, Altucci L, Angelini M et al (2020) Molecular networks in Network Medicine: Development and applications. Wiley Interdiscip Rev Syst Biol Med 12(6):e1489

Sun Y, Yang J (2019) A bioinformatics investigation into the pharmacological mechanisms of the effect of Fufang Danshen on pain based on methodologies of network pharmacology. Sci Rep 9(1):5913. https://doi.org/10.1038/s41598-019-40694-4

Szymański P, Markowicz M, Mikiciuk-Olasik E (2012) Adaptation of high-throughput screening in drug discovery-toxicological screening tests. Int J Mol Sci 13(1):427–452. https://doi.org/10.3390/ijms13010427

Tan N, Gwee KA, Tack J, Zhang M, Li Y et al (2020) Herbal medicine in the treatment of functional gastrointestinal disorders: a systematic review with meta-analysis. J Gastroenterol Hepatol 35(4):544–556. https://doi.org/10.1111/jgh.14905

Wang J, Xiong X, Feng B (2013) Cardiovascular effects of salvianolic acid B. Evid Based Complement Alternat Med 2013:247948. https://doi.org/10.1155/2013/247948

Wang X, Wang ZY, Zheng JH, Li S (2021) TCM network pharmacology: a new trend towards combining computational, experimental and clinical approaches. Chin J Nat Med 19(1):1–11. https://doi.org/10.1016/s1875-5364(21)60001-8

Wu L, Wang Y, Fan X (2011) Tools for network pharmacology study: network visualization and network analysis. Zhongguo Zhong Yao Za Zhi 36(21):2923–2925

Wu CC, D'Argenio D, Asgharzadeh S, Triche T (2012) TARGETgene: a tool for identification of potential therapeutic targets in cancer. PLoS One 7(8):e43305. https://doi.org/10.1371/journal.pone.0043305

Xiao Z, Ye Q, Duan X, Xiang T (2021) Network pharmacology reveals that resveratrol can alleviate COVID-19-related hyperinflammation. Dis Markers 2021:4129993. https://doi.org/10.1155/2021/4129993

Yang HY, Liu ML, Luo P, Yao XS, Zhou H (2022) Network pharmacology provides a systematic approach to understanding the treatment of ischemic heart diseases with traditional Chinese medicine. Phytomedicine 104:154268. https://doi.org/10.1016/j.phymed.2022.154268

Zhang Z, Chan DW (2010) The road from discovery to clinical diagnostics: lessons learned from the first FDA-cleared in vitro diagnostic multivariate index assay of proteomic biomarkers. Cancer Epidemiol Biomark Prev 19(12):2995–2999. https://doi.org/10.1158/1055-9965.Epi-10-0580

Zhang GB, Li QY, Chen QL, Su SB (2013) Network pharmacology: a new approach for Chinese herbal medicine research. Evid Based Complement Alternat Med 2013:621423. https://doi.org/10.1155/2013/621423

Zhang R-z, Yu S-j, Bai H, Ning K (2017) TCM-mesh: the database and analytical system for network pharmacology analysis for TCM preparations. Sci Rep 7(1):2821. https://doi.org/10.1038/s41598-017-03039-7

Zhang R, Zhu X, Bai H, Ning K (2019) Network pharmacology databases for traditional Chinese medicine: review and assessment. Front Pharmacol 10:123. https://doi.org/10.3389/fphar.2019.00123

Zhang Q, Liu J, Li R, Zhao R, Zhang M et al (2020a) A network pharmacology approach to investigate the anticancer mechanism and potential active ingredients of Rheum palmatum L. against lung cancer via induction of apoptosis. Front Pharmacol 11:528308. https://doi.org/10.3389/fphar.2020.528308

Zhang X, Shen T, Zhou X, Tang X, Gao R et al (2020b) Network pharmacology based virtual screening of active constituents of Prunella vulgaris L. and the molecular mechanism against breast cancer. Sci Rep 10(1):15730. https://doi.org/10.1038/s41598-020-72797-8

Zhou Z, Chen B, Chen S, Lin M, Chen Y et al (2020) Applications of network pharmacology in traditional Chinese medicine research. Evid Based Complement Alternat Med 2020:1646905. https://doi.org/10.1155/2020/1646905

Zimmermann GR, Lehár J, Keith CT (2007) Multi-target therapeutics: when the whole is greater than the sum of the parts. Drug Discov Today 12(1–2):34–42. https://doi.org/10.1016/j.drudis.2006.11.008

Zuo HL, Lin YC, Huang HY, Wang X, Tang Y et al (2021) The challenges and opportunities of traditional Chinese medicines against COVID-19: a way out from a network perspective. Acta Pharmacol Sin 42(6):845–847. https://doi.org/10.1038/s41401-021-00645-0

Chapter 11
In Silico Pharmacology and Drug Repurposing Approaches

Ghanshyam Parmar, Jay Mukesh Chudasama, Ashish Shah, and Ashish Patel

11.1 Introduction

11.1.1 Introduction to Drug Repurposing

Reusing already-approved drugs for new indications is an efficient and creative way to increase the available drug pool. However, it is challenging to find new protein targets for existing drugs. Even though novel strategies for drug repurposing have been developed, there is universal agreement that there is room for development. The COVID-19 pandemic recently swept the globe. So, SARS-CoV-2 (or COVID-19) has spread over the world and become a major health concern for people everywhere. Finding new vaccines and developing new chemicals is a labor- and time-intensive procedure. Drug repurposing refers to the process of selecting potentially medicinal compounds from an existing chemical library.

Drug repositioning (DR) is a term that encompasses a wide range of practices, including drug re-tasking, drug reprofiling, drug rescue, drug recycling, drug redirection, and therapeutic switching. It is the practice of treating disorders for which the original therapeutic use of a medicine was not intended by discovering and developing new pharmacological indications for previously marketed, FDA-approved treatments. It comprises finding new medical uses for well-known pharmaceuticals that have already been approved, shelved, abandoned, or tested (Ashburn and Thor 2004; Dey 2019; Rudrapal et al. 2020).

G. Parmar · J. M. Chudasama · A. Shah (✉)
Department of Pharmacy, Sumandeep Vidyapeeth, Vadodara, Gujarat, India

A. Patel
Department of Medicinal Chemistry, Ramanbhai Patel College of Pharmacy, Charotar University of Science and Technology, Anand, Gujarat, India

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_11

Several reasons have contributed to the recent uptick in focus on drug repositioning. It has been estimated that, on average, it costs wealthy countries $1.24 billion to bring a new medicine to market (Kaitin 2010).

Drug repositioning often referred to as "repurposing" has emerged as a significant source of revenue development in the pharmaceutical sector (Ashburn and Thor 2004). It is easy to see why repurposing medications is appealing, as even unsuccessful drugs have typically undergone extensive preclinical and early human clinical studies and been shown to be safe in a wide range of settings. Repositioning drugs is an attractive strategy to reverse productivity declines (Tobinick 2009). Especially in cases when safety was not the major concern, it may be possible to learn from past mistakes by investigating what went wrong with a compound. This chapter explores some typical instances of drug repositioning.

11.1.2 Conventional/Current Drug Discovery Process Vs. Drug Repurposing, an Old Weapon for New Battle?

There are currently five phases in the medicine development process, which are discovery and preclinical, safety review, clinical research, FDA review, and FDA post-market safety monitoring. It is an inefficient method that will set you back both time and money (Hughes et al. 2011). Drug repositioning, on the other hand, entails only four steps: locating an appropriate compound, acquiring that compound, developing the drug, and finally monitoring its safety after it has been released to the public by the FDA (Rudrapal and Chetia 2016) (Figs. 11.1 and 11.2). Medication repositioning has considerably shortened the time and expense of drug development while lowering the probability of failure, thanks to the advent of bioinformatics/chemoinformatics tools and the availability of enormous biological and structural database. In recent years, the drug purposing process has been sped up even further by the use of in silico methods, structure-based drug design (SBDD), and artificial intelligence (AI) technology (Agrawal 2018; Kalita et al. 2020). However, repositioning, or the practice of applying already-approved treatments to unanticipated therapeutic purposes, has proven effective. Benefits to using this method to find new drugs are clearly illustrated in comparison to the standard drug development process. Sildenafil (Viagra) is a phosphodiesterase-5 (PDE5) inhibitor that was originally developed to treat coronary artery disease (angina). It could have a positive effect on both development time and cost. Metformin (Glucophage), an oral antidiabetic medication commonly used in treating type 2 diabetes mellitus, is currently undergoing phase II/phase III clinical trials as a cancer therapeutic (Ashburn and Thor 2004; Ferreira and Andricopulo 2016). Pharmaceutical repositioning has several benefits over conventional drug discovery strategies. Research and development time is drastically cut as compared to the current medication research program. Current methods indicate that it will take

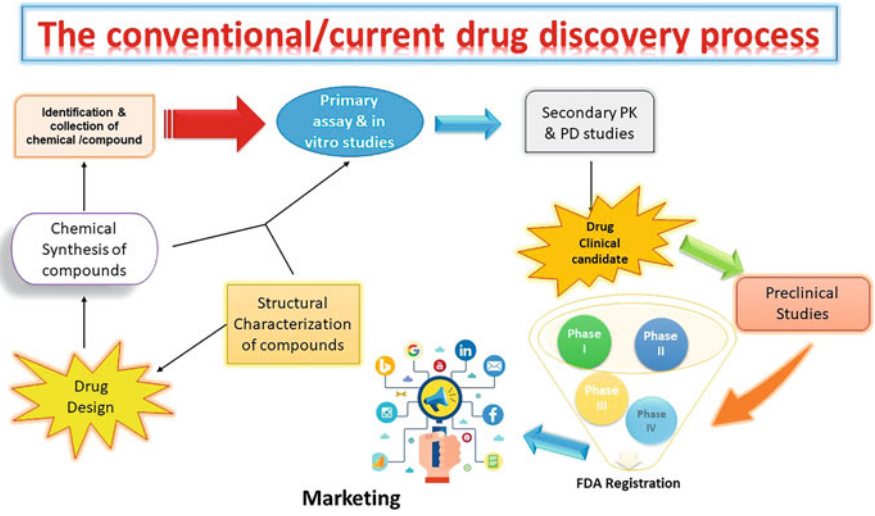


Fig. 11.1 Conventional/current drug discovery process

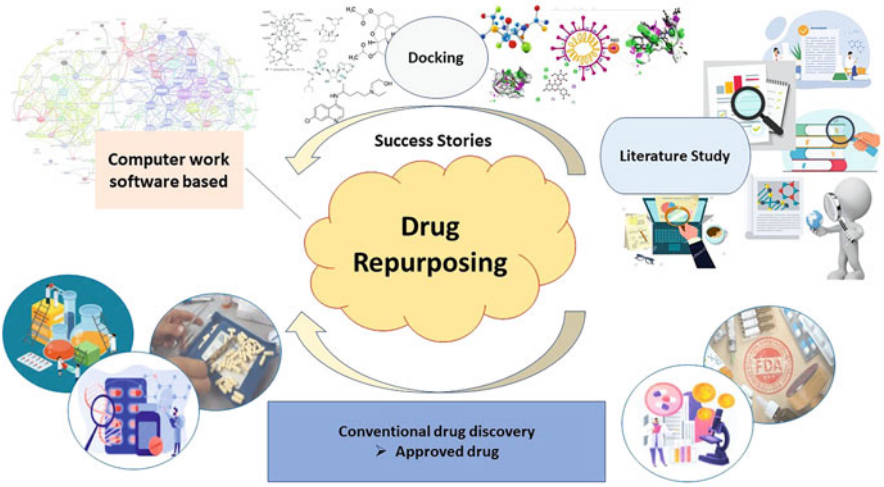


Fig. 11.2 Drug repurposing process

10–16 years to develop a new drug, while DR predicts that it will take only 3–12 years. Drug repositioning allows for new drug development at a cost of $1.6 billion, far less than the current drug development strategy's estimated $12 billion. Additionally, the time it takes to find new pharmacological targets is only 1–2 years, and the time it takes to create a repositioned medicine is around 8 years (Cha et al. 2018; Vickers 2017). Repositioned medicines bypass the first 6–9 years of research typically required for new therapies under the current paradigm and go straight to

preclinical testing and clinical trials, reducing total risk, time, and cost. It is said that the licensing process for repurposed pharmaceuticals can take anywhere from 3 to 12 years at the Food and Drug Administration (FDA) or the European Medicines Agency (EMA). Since the candidate drug has already gone through the structural optimization, preclinical, and/or clinical trial stages of drug development, and since the candidate drug may already be an approved drug with its clinical efficacy and safety profile, a variety of preclinical (pharmacological, toxicological, etc.) and clinical efficacy and safety information is already available at the start of a repositioning project. Consequently, the high risks associated with failures in the early stages of development are mitigated, along with the costs, and the likelihood of greater clinical safety and thus a high success rate is increased (Allarakhia 2013; Jin and Wong 2014).

The benefits of drug repurposing over conventional drug discovery methods include shorter development times, lower development costs, and reduced risks of failure in preclinical studies, thanks to the availability of pharmacokinetic, toxicological, clinical, and safety data at the outset of a repurposing development project (Cha et al. 2018). Developing a relocated medicine is expected to take between 3 and 12 years (as opposed to 10 to 17 years in a normal discovery process), saving the repositioning company a substantial amount of time and money. In comparison to the $1.24 billion required to bring a new medicine to market through the standard drug development approach, the cost of drug repurposing is just about $600 million (Parvathaneni and Gupta 2020). What follows are a few more benefits. Unlike the drug repositioning approach, which focuses on developing drugs for rapidly emerging and re-emerging infectious diseases, difficult-to-treat diseases, and neglected diseases (NTDs), the traditional drug discovery program primarily focuses on discovering drugs to treat chronic and complex diseases. Because of advances in bioinformatics and cheminformatics, as well as the availability of large amounts of data from various "omics" (proteomics, transcriptomics, metabolomics, genomics, etc.), researchers can now use disease-targeted repositioning strategies to learn more about the previously undiscovered mechanisms of action (e.g., drug targets, drug-drug similarities, new biomarkers for diseases, etc.) of currently marketed medications (Allarakhia 2013).

11.1.3 Fundamentals of Drug Repurposing

Drug candidates go through three stages of testing: discovery, where they are screened and identified; preclinical, where they are evaluated in vitro and in animal models; and clinical, when they are examined in humans as part of clinical trials (Fisher Wilson 2006). There are many steps in the discovery process, including finding and validating targets, finding leads using high-throughput screening, and optimizing those leads (development of the most druggable compounds from the lead compounds). Pharmacological effectiveness studies, toxicological assessments, and studies of potential medication interactions are all part of the preclinical

investigation process (Dueñas-González et al. 2008). The utility of even a good chemical in humans is never entirely clear, as preclinical research cannot always account for the physiological differences between people and animals. This means that two of the most common reasons for a drug not making it to market are serious adverse effects and declining efficacy in people throughout clinical trials (Fisher Wilson 2006). Thus, the high-risk/high-reward trade-off in drug research and development is a major problem in the legalization and commercialization of novel medicines.

One alternate approach to drug development is to investigate existing drugs that have either been licensed for use in the treatment of other diseases or have previously had their targets discovered (Dueñas-González et al. 2008). Repositioning, redirection, repurposing, and reprofiling refer to the process of finding new uses for currently available pharmaceuticals (marketed treatments as well as unsuccessful or idle substances) that are not covered by the original indication (Ashburn and Thor 2004). The number of successful repositionings of existing pharmaceuticals is increasing as more pharmaceutical firms look to the pharmacopoeia for repositioning opportunities. The benefits of drug repositioning are compared to those of starting from scratch with a new drug discovery and development process, and the methods used to identify potential repositioning candidates are outlined, as are the common difficulties encountered during the repositioning process, and finally, repositioning initiatives in India are described.

In the event when a temporary or permanent cure for a certain disease condition is discovered early on, drug repurposing is favorable. Medication repurposing is crucial in treating diseases like the COVID-19 pandemic where no vaccines are available.

Today, in vitro or computer modeling studies are used in drug repurposing to prove that the drug may effectively manage disease. Since this reduces the effort and cost needed to find a new therapeutic chemical, it is helpful for researchers. In silico docking is utilized in the repurposing of drugs by comparing the docking score with the outcomes of the interaction between the ligand (drug/chemical) and the receptor (protein molecule).

11.1.4 Advantages of Drug Repurposing Over Typical Drug Development Process

A number of factors have contributed to the recent uptick in focus on drug repositioning. Costs associated with bringing a single new drug to market in high-income nations have been estimated at over USD1.24 billion (Kaitin 2010). Another key issue that worries pharmaceutical firms is their high rates of employee turnover. A study conducted by the Tufts Centre for the Report of Drug Development found that just 16% of medications that began clinical development between 1999 and 2004 really made it to market. Current pharmaceutical R&D productivity is

obviously inadequate, with regulatory bodies in affluent nations only authorizing 18–20 new medications per year despite annual spending in the pharmaceutical business of more than USD50 billion (Ashburn and Thor 2004; Kaitin 2010). It is estimated that just 30% of new goods will bring in enough money to cover their research and development costs. In order to recoup research and development expenditures as quickly as possible, pharmaceutical companies are increasingly focused on drugs to treat chronic and complicated indications like cardiovascular, endocrine, mental, and neurological illnesses and cancers (Kaitin 2010). The current drug discovery paradigm is therefore unprepared to combat rapidly emerging and re-emerging infectious diseases such as mutated influenza strains, drug-resistant microorganisms, and neglected tropical diseases (NTDs), which appear to have a narrower capital system but are crucial to public health (Aronson 2007). Many pharmaceutical firms have had to get creative in their search for new applications for current pharmaceuticals as a result of the productivity problem, price pressures around the world, generic competition, and ever-increasing regulatory obstacles (Ashburn and Thor 2004).

Many pharmaceutical firms have had to get inventive in their pursuit of new uses for existing drugs because of this productivity problem, as well as worldwide price pressures, generic competition, and ever-increasing regulatory restrictions. Since pharmacokinetic, toxicological, and safety data for current medications is already available, the development process for repositioned drugs is quicker and more cost-effective than conventional drug discovery (Tobinick 2009). More stringent regulations have been put in place over the past few years, making it harder and harder for a new drug to access the market under the old standards. There has been a significant increase in both the length of time and total cost associated with creating new pharmaceuticals as a direct result of the stricter regulations (Kaitin 2010). Time and money savings for the repositioning sector are anticipated to be substantial, with the development of a repositioned medicine taking anywhere from 3 to 12 years at significantly reduced costs. Due to the availability of clinical safety data, pharmacokinetics, and a workable dose range at the outset of a repositioned pharmaceutical development project, the risks associated with clinical development are greatly minimized, and fewer failures occur in the later stages of the project (Sleigh and Barton 2012). Despite the challenges of developing new drugs from scratch, this allows pharmaceutical companies, non-profits (such as research institutions, government labs, and organizations representing the interests of patients in both developed and developing countries), and regulators to respond rapidly and effectively to unmet medical needs (Sleigh and Barton 2012).

11.2 Drug Repurposing Strategies

11.2.1 Knowledge-Based Repurposing

Models are created to predict unidentified targets, biomarkers, or disease mechanisms in this repurposing strategy using drug-related information, including drug targets, chemical structures, pathways, adverse effects, etc. (Emig et al. 2013). This tactic entails drug repurposing that is target, pathway, and target mechanism-based.

With the aid of bioinformatics or chemoinformatics techniques, knowledge-based drug repositioning techniques assemble information on medicines, drug-target networks, signaling or metabolic pathways, clinical trial data, and other pertinent drug phenotypic data. The prediction accuracy for the drug repositioning process is significantly increased when using knowledge-based methods because they incorporate a lot of current data.

Understanding the structural resemblance of protein-binding sites can aid in the identification of new therapeutic targets. Numerous proteins have identical binding sites, according to studies. For instance, staurosporine and celecoxib have comparable binding pockets when they bind to carbonic anhydrases and synapsin, respectively (Defranchi et al. 2010; Weber et al. 2004). Therefore, it is most likely that proteins with identical binding sites will bind to the same ligands.

For medication repositioning, certain knowledge-based computational methods create disease-specific pathways from gene expression profiles (Jadamba and Shin 2016; Li and Lu 2013). These techniques are based on the idea that important elements (proteins) in disease pathways could potentially be used as therapeutic targets (Li and Agarwal 2009; Strittmatter 2012). Li and Lu (2013) created a computational technique to relate medications to diseases through target- and gene-involved pathways. The technique found novel uses for existing medications and gave important information for therapeutic repositioning, including the possible re-use of numerous Crohn's disease drugs.

Computational drug-repositioning approaches now use phenotypic data. Clinical phenotypic information mirrors a drug's human phenotypic screen because it is created from patient data. Systematic analyses show phenotypic screening beats target-based techniques for finding new indications (Swinney and Anthony 2011).

11.2.2 Target-Based Drug Repurposing

Drugs have affinity for extra proteins known as off-targets in addition to their therapeutic target proteins with which they interact. These off-target interactions are mostly to blame for pharmacological side effects. These additional interactions, nevertheless, are not always bad for an organism; in certain cases, they may even be advantageous for brand new therapeutic uses. For instance, sildenafil (Viagra) was first created to treat angina but was subsequently modified to address erectile

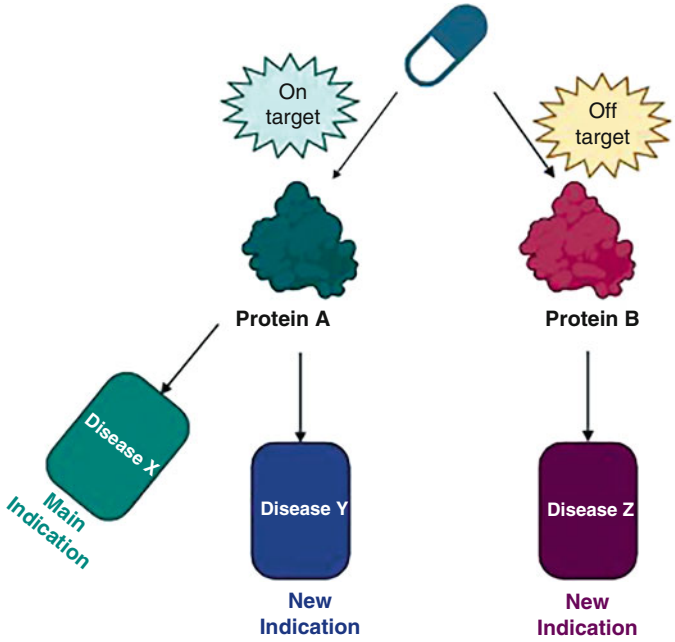


Fig. 11.3 Drug repositioning concept diagram. To cure disease X, a drug molecule interacts with Protein A. Protein A may cause disease Y. The drug may also bind well to off-target Protein B. A drug reported for disease X may treat diseases Y and Z

dysfunction. This rebuilding was carried out when it was discovered that sildenafil interacts with PDE5, the enzyme responsible for the erectile response. Figure 11.3 depicts the conceptual diagram of target-based medication repositioning.

Target-based drug repositioning makes advantage of receptor or ligand structure. These drug-repositioning systems carry out high-throughput in silico screening of chemical libraries using docking and/or pharmacophore models. VS efforts have used computational models successfully, according to the literature (Akhoon et al. 2016; Matter and Sotriffer 2011; Mehra et al. 2016; Xu et al. 2018). These programs rely on small molecule libraries like Pubchem, ZINC, etc. and the target's binding site. Because they can screen nearly all drugs in a matter of days, pharmaceutical corporations choose in silico, target-based drug repositioning methods.

Inhibitors for TGF-1 receptor kinase are one example of in silico target-based methods' potential. Biogen Idec's computational work reproduced Eli Lilly's wet-lab results and found an identical, promising lead compound for the TGF-1 receptor (Shekhar 2008). Such advancements demonstrate the effectiveness of computational methods to screen compounds without synthesis, which is necessary for wet-lab assays, thereby minimizing effort, expense, and time.

Target selection is the first VS step. Proteins are given top priority because of their high specificity, potency, and low toxicity. Targetable biomolecules include polysaccharides, lipids, and nucleic acids. We need the 3D protein structure once one is

chosen. The Protein Data Bank stores 3D structures discovered by X-ray crystallography, NMR spectroscopy, or cryo-electron microscopy (PDB) (Swamidass 2011). In silico affinity prediction is key in target-centered drug repositioning. Molecular docking places a small molecule in a receptor's binding pocket to estimate its affinity. Docking involves sampling and scoring. Sampling generates different drug poses in the target active site; scoring evaluates the target-ligand complex's binding strength. Various scoring functions can predict the Gibbs free energy of binding of ligands and rank them by their binding energies (Doman et al. 2002).

11.2.3 Pathway-Based Drug Repurposing

Predicting the similarity or relationship between disease and drug is the goal of pathway-based drug repurposing, which makes use of information about metabolic pathways, signaling pathways, and protein interaction networks. For instance, disease-specific pathways are recreated utilizing omics data extracted from human patients or animals to serve as new targets for repositioned medications (Jadamba and Shin 2016). These techniques are useful because they can reduce complex signaling networks involving many proteins to a more manageable set using fewer proteins (or target molecules).

Pharmacological manipulation of many physiological and cellular pathways can have a significant impact on the majority of illnesses. Further, many nodes or molecular targets may exist inside each pathway to control the pharmacological effect. As a chemical is put through its paces in an array of in vivo models of different disease states, many different pathways and targets within those pathways are tested all at once. Drug activities and pharmacological responses that are the result of simultaneous modulation of multiple targets and pathways, and activities that are driven by modulation of complex biological networks, can be discovered through screening drugs through an intact in vivo system, which is not the case when testing drugs in vitro or in cell-based systems.

11.2.4 Target Mechanism-Based Drug Repurposing

To identify novel therapeutic mechanisms of action, target mechanism-based repurposing incorporates knowledge of signaling pathways, treatment omics data, and protein interaction networks. Such drug-repurposing strategies are driven by the growing importance of precision medicine. These repurposing strategies have the potential to uncover pathways associated to not only diseases or medications but also drug treatments for individual diseases. Determining disease targets and their related mechanisms of action is crucial to the target mechanism-based drug repurposing method, which in turn leads to the discovery of novel therapeutic applications. To better understand the mode of action and diseases linked with medicinal substances,

this method integrates a computational biological system with the research of biological pathways. Systems pharmacology is another useful tool for identifying potential new uses for existing drugs based on their underlying mechanisms of action. One medication for several targets is the foundation of systems pharmacology, which bridges the gap between systems biology and traditional pharmacology. The chemical–protein, protein–protein, genetic, signaling, and physiological interactions in a biological system can all be predicted with this method (Jin et al. 2012).

The biological system is an example of an emergent property, which is a property of the system as a whole that cannot be explained by looking at the pieces in isolation. Systems biology is an emerging field that provides an all-encompassing setting for investigating the dynamic functional interactions across biological systems throughout time and opening up novel avenues for target mechanism-based medication repurposing. Researchers are now able to undertake multiscale modeling of biological networks, i.e., networks of networks, thanks to the development of systems biology, which has also changed the field of drug discovery. Systems biology uses quantitative mathematical network models in conjunction with network models to determine the dynamic behavior of biological systems. Target mechanism-based medication repurposing approaches are found on the idea that these networks can be mined for information about new therapeutic targets and the unique processes associated with them.

This strategy has considerable applicability for repurposing with a specific mechanism as the target. Pathway analysis is a stepping stone toward accurate systems biology-based target mechanism prediction. Genomic, epigenomic, proteomic, and metabolomics data are all examples of high-throughput "omics" data that can be integrated into pathway analyses to shed light on previously unknown connections across pathways that play a role in both healthy and pathological states. Drug repurposing based on previously unknown mechanisms of action has benefited greatly from the identification of relationships between previously unrelated complicated biological networks. Therapeutic medication repurposing that takes advantage of a medicine's original target mechanism can also be useful in the fight against treatment-resistant patients. Understanding the entire route during drug resistance will be helpful in developing a new medication therapy for diseases like tuberculosis, where resistance to existing drugs is a leading cause of patient mortality (Wu et al. 2016).

Target mechanism-based approaches to drug repurposing are distinct from other predictive methods in that they make full use of millions of compound–protein interactions to investigate the drug candidate's entire target space, rather than just the subset that shares chemical structure similarities or phenotypic effects with known drugs. To further precision medicine, it is useful to incorporate phenotypic similarities during drug repurposing research.

11.2.5 Signature-Based Repurposing

In signature-based repurposing, gene signatures information obtained from disease omics data (Haeberle et al. 2012) is used to discover new off-targets or mechanisms of disease. This approach searches inverse drug-disease relationships by comparing gene expression profiles between drug and disease. In the work by Dudley et al. (2011) potential drug–disease pairs were investigated for inflammatory bowel disease (IBD), where gene expression profiles obtained from the gene expression omnibus database (Barrett et al. 2005) were compared with gene expression profiles comprising 164 drug compounds obtained from the connectivity map (Lamb et al. 2006). As a result, unknown drug–disease pairs were discovered, with one pair validated in preclinical models.

A gene signature, sometimes called a gene expression signature, is a particular pattern of gene expression that can be traced back to a certain cause or situation. Determining the biological and functional significance of genes in the human genome has accelerated the drug-discovery process, according to the Human Genome Project (HGP). One way to get a better grasp on how biological systems work is to examine the cellular gene expression pattern of hundreds of genes simultaneously. High-throughput gene-expression data has allowed for in-depth tracking of transcriptional responses related to a wide range of disease states and therapeutic interventions. Signature-based drug-repositioning approaches build on this information by using it to decipher the gene signatures of different diseases and drugs. Genome-wide expression profiles play a crucial role in drug repurposing because they shed light on how a medication interacts with different cellular processes. Better understanding of drug indications is made possible by signature-based approaches, which reveal hitherto unknown mechanisms of action for many medications. To develop a signature of a disease, it is useful to compare the sets of genes that are either unregulated or downregulated in diseased and healthy states. When performing expression profiling in the presence of a pharmacological molecule, disease signatures are also useful for deciphering the molecular foundation of the drug's mechanism of action.

Microarray and next-generation sequencing (NGS) technology are two examples of cutting-edge gene sequencing methodologies that are widely utilized to obtain high-throughput gene-expression data quickly and cheaply. DNA microarray methods are commonly used for gene signature detection. DNA microarray allows us to examine the expression of hundreds of genes in individual cells. Traditional drug discovery and development methods have been completely upended by the use of microarray technology. Target discovery, validation, and lead optimization are all greatly aided by DNA microarray. By comparing the gene expression profiles of the diseased and healthy states, one can hypothesize the complex disease pathways using the microarray technique. New possibilities for drug repurposing are becoming available to DNA microarray technology. Although off-target indications are not taken into account during the drug design process that uses traditional structure, they can be easily measured using microarray techniques. The expression of microarrays

is a non-biased method for detecting both on-target and off-target impacts of substances.

In in silico molecular docking and gene expression investigations to repurpose various FDA-approved medicines against seven targets involved in Alzheimer's disease (AD). Cell-based microarray experiments from this dataset were deposited in the cMap database, and screening of 1553 FDA-approved medications yielded 74 compounds with low binding energy (less than 10 kcal/mol for all seven targets) (Lamb et al. 2006). Due to this research, four medicines (risperidone, droperidol, glimepiride, and glipizide) were reformulated for use in AD.

The advantage of these approaches is that they identify new mechanisms of action for drugs. Also, unlike knowledge-based methods, more molecular and/or genetic-level mechanisms are involved in these methods.

11.2.6 Phenotype-Based Repurposing

The phenotypic information has become available as a new source of drug repositioning. In recent years, this type of information has been increasingly used by systems approaches to detect genetic traits associated with human diseases (Hebbring 2014). Natural language processing skills applied to electronic health records (EHRs) can reveal additional adverse drug events which were not observed during drug development (Luo et al. 2017). For example, mining EHRs helped in identifying that metformin can be repurposed for cancer treatment (Xu et al. 2015).

For medications that can bring about the desired change in a disease's phenotype, researchers use a screening method known as phenotypic approach, or blinded drug repurposing (Swinney and Anthony 2011). Using a disease-associated trait as a basis, researchers may create a cell-based test in which a large number of chemicals are screened for their ability to reverse the disease phenotype. Several rare illnesses without effective therapy might benefit from this strategy since it is more effective for diseases in which particular therapeutic targets have not been found or validated. However, the phenotypic method is limited in its ability to provide light on the disease's mechanism and the drug's particular targets/actions since it relies only on information about the phenotype (Jin and Wong 2014).

11.3 Methods for Computational Drug Repurposing

11.3.1 Machine Learning

Over from early attempts to evaluate medications, which often relied on a single source of biological or biomedical data, computational drug repurposing has evolved over the past two decades to become a cutting-edge application domain for machine

learning techniques. To expose the underlying correlations between biological and biomedical entities, computational drug repurposing models need to be trained on a vast quantity of data, much like machine learning models in other fields. This data is then used to develop strong decision rules. Massive efforts have been made to develop, study, and apply machine learning methods for discovering novel drug-disease associations and drug repurposing applications, and this has been aided by the enormous increase in the volume of publicly available biological and biomedical data as well as the valuable advancements resulting from machine learning models in other disciplines.

Logistic regression and k-nearest neighbors are two examples of the machine learning (ML) approaches that have been used in medication repositioning (kNN) (Shen et al. 2003), random forest (Susnow and Dixon 2003), support vector machines (SVM) (Meng et al. 2004), deep neural network (NN), and deep learning (DL) for binary/multiclass values prediction (Chen et al. 2018). Among the best logistic regression-based ML methods is PREDICT, which combines drug–drug and disease–disease similarity to reveal integrated similarity values for the prediction of a single drug's effect on a single disease (Gottlieb et al. 2011). Among the best logistic regression-based ML methods is PREDICT, which combines drug–drug and disease–disease similarity to reveal integrated similarity values for the prediction of a single drug's effect on a single disease (Liu et al. 2015). A support vector machine (SVM)-based ML technique, on the other hand, combines molecular target, drug chemical structure, and gene expression similarity in a single similarity matrix to predict therapeutic class (Napolitano et al. 2013). By analyzing the genomic and structural fingerprints of cancer cell lines with random forest regression, the neuronal network [NN]-based ML model is optimal for predicting response to anticancer drugs in cancer cell lines (Menden et al. 2013).

However, the deep learning (DL) method is a helpful drug development tool through the discovery of latent and complex structures in a large database, as well as connected weight adjustment, such as the analysis of complex gene expression data for the prediction of therapeutic categories of drugs with the prediction of toxicity (Aliper et al. 2016; LeCun et al. 2015; Mayr et al. 2016).

11.3.2 Network Models

It is possible to learn a great deal about drug mechanisms of action and indications, drug targets and how they function, therapeutic potential, and drug repurposing applications from analyzing networks. For this reason, it has found widespread application in computational drug repurposing. Biological and medical entities, along with their interactions and relationships, are well-represented in network models.

Using ideas from graph theory, statistical analysis, and computational models, network models can be used to unearth useful connections. The nodes in these network-based models are biological entities like drugs, diseases, gene products,

and so on, while the edges represent relationships between these entities, such as their shared functions or mechanisms of action. Unknown drug–targets, drug–diseases, disease–disease, protein–protein interactions, transcriptional and signaling networks, and so on are also represented by nodes and edges integrated from heterogeneous data using the "guilt-by-association" principle (Azuaje 2013; Iorio et al. 2013). The interpretation of drug–target association and the identification of new drug repurposing molecules, such as novel anticancer drugs, are facilitated by network analysis-based computational drug repurposing approaches (Haeberle et al. 2012).

11.3.3 Text Mining and Semantic Inference

There is a mountain of drug and disease data in the biomedical and pharmaceutical literature, and it is up to us to sift through it all to find new uses for old drugs. It is common practice in computational drug repurposing to use text mining to locate data pertaining to a given gene, disease, or drug and then to classify the relevant entities or knowledge from the retrieved data using either natural language processing or the co-occurrence of the relevant entities (Tari and Patel 2014).

An additional link between drug A and disease C may have been discovered, for instance, if drug A is associated with gene B and gene B is associated with disease C. There are typically four stages to text mining: There are four major types of data analysis: (1) IR, (2) NER, (3) IE, and (4) KBD (knowledge base development) (Zhu et al. 2013).

Semantic technologies have made it simple to aggregate data from various sources in order to forecast therapeutic potentials and new indications for currently available drugs. Therefore, semantic inference technology-based topic modeling aids in the discovery of drug indications by integrating multiple data sources.

For instance, Latent Dirichlet Allocation-based drug repurposing topic model is used to process the phenome information for drug side effects and also identify the relationship of existing approved breast cancer drugs with their associated genes and pathway through the integration of ontology-based knowledge tools (Bisgin et al. 2014; Zhu et al. 2014). Moreover, a semantic linked network-based approach is used to assess drug–target associations, which comprised drugs, protein targets, chemical compounds, diseases, pathways, side effects, and their relations through the identification of the drug and drug-target location in the subgraph (Karthik et al. 2014).

11.4 Validation for Computational Repurposing

When compared to the traditional drug discovery and development process, computational drug repurposing is preferable because it saves both time and money. The problem is that specific validation models may be inaccurate. As a result, the success

of the proposed computational models relies heavily on the understanding and selection of appropriate validation models. Because of factors like high cost, high level of toxicity, and reduced bioavailability, as well as the fact that some drugs have been abandoned or are not preferred by physicians or biologists, selecting the right set of drug repurposing candidates for validation is crucial. Different studies use different validation/evaluation models, and these models may depend, at least in part, on the type of results that are sought. Each model is supported by either (1) in vitro experiments, (2) in vivo experiments, (3) electronic health records, (4) leave-one-out and cross-validation, or (5) benchmarking against previously developed models. Case studies, a literature synthesis, and advice from an expert in the field are all steps 6–8.

Drug repurposing candidates have often been validated using in vitro and in vivo experimental validation models, despite the fact that these approaches have a number of well-known limitations (Lim et al. 2016).

Furthermore, PubMed articles as a model for medical literature cross-referencing are used to verify the efficacy of the computational approach to drug repurposing. Recently, however, there has been an explosion in the use of literature-based validation models as literature mining techniques become more widespread in scientific research (Ozsoy et al. 2018).

11.5 Success Stories of Drug Repurposing

As mentioned above, there are many success stories; here some more recently discovered success stories for drug repurposing are enlisted:

11.5.1 Pimozide (Antipsychotic Drug)

Pimozide is an antipsychotic medication that has been shown in vitro to suppress cell growth in HCC cell lines by inducing apoptosis at the G0/G1 phase. Furthermore, pimozide inhibited HCC stem-like cells, particularly the CD133-positive cell side population. Pimozide was discovered to target STAT3 expression using luciferase assay activity, as well as the transcription levels of downstream oncogenes for STAT3 signaling. Pimozide's anticancer efficacy was confirmed in vivo in hairless mice (Chen et al. 2017).

11.5.2 Valproate (Antiepileptic Drugs)

Valproic acid (VPA) is a widely used antiepileptic medicine that is a potent and selective inhibitor of histone deacetylase (HDAC). As a definable target with a

known molecular signature, HDAC has been recognized for its critical role in the progression of numerous forms of cancer. The mono- and adjuvant therapeutic in vitro actions of VPA and doxorubicin (DOX) against the HepG2 cell line were reported to be selective, efficient, and antiproliferative in a synergistic manner. At the molecular level, activation of caspase-3 and poly (ADP-ribose) polymerase (PARP) facilitated the synergism induced by VPA and DOX to induce apoptosis. The combination of VPA and DOX therapy increased reactive oxygen species levels (ROS) (Saha et al. 2017).

11.5.3 Amiodarone (Antiarrhythmic Drug)

Amiodarone, a class III antiarrhythmic and powerful mTOR inhibitor, was reported to reduce liver tumor development in the rat orthotropic model and the mouse xenograft model by inducing autophagy activity. Furthermore, a big data analysis of 32,625 case-control data from Taiwan's National Health Insurance program demonstrated that long-term regular amiodarone use reduces the incidence of HCC considerably. Amiodarone, as a repurposed medication, has anticancer potential by inducing autophagy activity and thereby suppressing liver tumor development and preventing HCC incidence (Attia et al. 2020).

11.5.4 Sildenafi

Viagra® (sildenafil), a phosphodiesterase 5 (PDE5) inhibitor that was under development for the treatment of angina in the 1990s, is the most commonly quoted example involving drug repurposing. Clinical trials for the medicine were halted because it was discovered that the compound's PK properties were incompatible with the long-term treatment of angina in patients (Ghofrani et al. 2006). However, during these trials, researchers discovered a startling side effect that assisted in the definition of a new disorder: erectile dysfunction (ED). The compound's weak PK qualities, which rendered it unsuitable as an anti-angina therapy, were ideal for a medication given for ED. PDE5 inhibitors have been explored in a range of additional indications after being used for ED and shown to be beneficial in pulmonary arterial hypertension (PAH) (Ghofrani et al. 2006) which sildenafil citrate is now authorized and marketed under the brand name REVATIO®.

11.5.5 Pertuzumab

Pertuzumab, a first-in-class monoclonal antibody that functions as a "HER dimerization inhibitor" and was expected to be the successor to Herceptin®, is another

recent example from Genentech. Pertuzumab Phase II clinical trials in prostate, breast, and ovarian malignancies were unsuccessful in 2005 (Menendez and Lupu 2007). When taken in combination with other chemotherapeutic treatments, pertuzumab caused tumors to vanish in 49% of newly diagnosed early-stage HER-2-positive breast cancer patients, compared to 29% of patients receiving Herceptin® and chemotherapy.

11.5.6 Thalidomide

Thalidomide, introduced by Gr ü nenthal in 1957, was discovered to be an efficient tranquilizer and painkiller (Matthews and McCoy 2003). It was also discovered to be an efficient antiemetic and to reduce morning sickness during pregnancy. Soon after its introduction, hundreds of children were born with severe developmental abnormalities of the limbs and face (phocomelia) as a result of thalidomide use. The medication was taken off the market in 1962. Subsequent research indicated that the substance was an enantiomer and that only one of the two optical isomers was responsible for the teratogenic consequences (Fabro et al. 1967). Unfortunately, because the two isomers interconvert in humans, separating the risk from the benefit in women of reproductive age is impossible. Despite its disastrous effects on the developing fetus, thalidomide has since been used successfully in the treatment of ENL, a painful leprosy complication, and tuberculosis. According to molecular investigations, the efficacy found may be attributable to its potential to block tumor necrosis factor (TNF) alpha signaling. Further research has been conducted to determine the possibility of thalidomide in Kaposi's syndrome (an AIDS consequence) and multiple myeloma (Matthews and McCoy 2003; Ng et al. 2002). In 2008, Celgene earned $550 million from thalidomide sales. As a result, there is increasing interest in thalidomide and its derivatives, and a recent literature search by these authors (Thomson Reuters Integrity database) uncovered research on its usage in more than 30 different indications.

11.5.7 Repurposing in Malaria

11.5.7.1 Dapsone

The development and eventual failure of dapsone as a potential treatment for uncomplicated malaria demonstrate how recognized side effects can derail a medication repurposing endeavor. Clinical studies must be designed to detect rare, albeit significant, known problems of existing medications in order to ensure safety (Guragain et al. 2017). This is especially true in developing countries, where conducting Phase IV post-marketing monitoring studies is almost impossible due to a lack of infrastructure. One prominent consequence that emerged from several of

these trials was the development of anemia in G6PD-deficient patients, which has been related to hemolysis caused by dapsone. Dapsone is known to produce dose-dependent hemolytic anemia in up to 20% of leprosy patients, and this side effect in G6PD patients has been documented since 1966. Dapsone's N-hydroxy metabolite is hypothesized to contribute to lipid peroxidation and the formation of reactive oxygen species. Patients with G6PD deficiency are more vulnerable to oxidative stress because they have lower amounts of glutathione, which is regenerated from glutathione disulfide utilizing nicotinamide adenine dinucleotide phosphate (NADPH), which is created by G6PD action on glucose-6-phosphate. G6PD deficiency is thought to protect against malaria, which has a prevalence of 4% to 28% in Africa. Although some of the studies were insufficiently powered to identify unfavorable outcomes in G6PD-deficient patients, the frequency of this side effect in numerous trials led to GlaxoSmithKline withdrawing hlorproguanil/dapsone in 2008 (Karadsheh et al. 2021).

11.5.7.2 Drug Repurposing in COVID-19

Several studies are being conducted to combat the potentially fatal corona virus infection using antiviral medication therapy. HIV protease inhibitors have been proposed as one possible treatment for COVID-19. Docking tests were carried out on 61 compounds with known antiviral properties in this study. Many HIV protease inhibitors demonstrated outstanding binding affinities with COVID-19 enzymes in this investigation (Shah et al. 2020). Lopinavir, asunaprevir, indinavir, and ritonavir are four protease inhibitors that have been found to be effective. Remdesivir, which acts on viral RNA polymerase, also improves activity in silico. Along with them, novel COVID-19 inhibitors such as methisazone, ABT450 (Paritaprevir), and CGP42112A have been developed (Frediansyah et al. 2021; Sharma et al. 2020) (Table 11.1).

11.6 Opportunities and Limitations of In Silico Drug Repurposing

Drug repurposing has become increasingly important in the pharmaceutical industry in recent years, with roughly one-third of approvals corresponding to repurposed medications, some of which have even attained blockbuster status. The stories of successful medication repurposing have paved the path for new kinds of public–private sector cooperation, a virtuous relationship that has not yet achieved its pinnacle. Governmental organizations and agencies are equipped to provide ways to get over some of the commercial and legal obstacles that medication repurposing programs encounter. Pharmaceutical firms, on the other hand, own the priceless (albeit occasionally underutilized) capital of their proprietary chemical libraries,

Table 11.1 Some repositioned drugs

Drug	Original indication	New indication	References
Allopurinol	Cancer	Gout	Yasuda et al. (2008)
Aspirin	Inflammation, pain	Antiplatelet	Ahmad et al. (2017), Qorri et al. (2022)
Bromocriptine	Parkinson's disease	Diabetes mellitus	Padhy and Gupta (2011)
Bupropion	Depression	Smoking cessation	Ashburn et al. (2004)
Duloxetine	Depression	Stress urinary incontinence	Li and Jones (2012), Maund et al. (2017)
Finasteride	Benign prostatic hyperplasia	Hair loss	McClellan and Markham (1999)
Gabapentin	Epilepsy	Neuropathic pain	Karmarkar et al. (2011)
Gemcitabine	Antiviral	Cancer	Qorri et al. (2022), Toschi et al. (2005)
Methotrexate	Cancer	Rheumatoid arthritis, Parkinson's	Padhy and Gupta (2011)
Propranolol	Hypertension	Migraine headache	Bidabadi and Mashouf (2010)
Raloxifene	Osteoporosis	Breast cancer	Agrawal (2018)
Sildenafil	Angina	Erectile dysfunction	Roundtable on Translating Genomic-Based Research for et al. (2014)
Thalidomide	Sedation, morning sickness	Leprosy, multiple myeloma	Amare et al. (2021), Laffitte and Revuz (2004)
Zidovudine	Cancer	AIDS	Ashburn et al. (2004)
Amantadine	Influenza	Parkinson's disease	Padhy and Gupta (2011)
Atomoxetine	Antidepressant	Attention deficit hyperactivity disorder	Shaughnessy (2011), Upadhyaya et al. (2013)
Colchicine	Gout	Recurrent pericarditis	Shaughnessy (2011)
Retinoic acid	Acne	Acute promyelocytic leukemia	Avvisati and Tallman (2003), Tallman et al. (1997)
Auranofin	Rheumatoid arthritis	Malaria	Roder and Thomson (2015), Sannella et al. (2008)
Bimatoprost	Glaucoma	Promoting eye lashes growth	Monaghan and Murphy (2021)
Clofazimine	Leprosy	Tuberculosis	Zhang et al. (2006)
Dapson	Leprosy	Malaria	Tiono et al. (2009)
Statins	Hyperlipidemia	Inflammatory and autoimmune disease	Smaldone et al. (2009)
Zileuton	Astham	Acne	Zouboulis (2009)
Nortriptine	Depression	Neuronal pain	Derry et al. (2015)

(continued)

Table 11.1 (continued)

Drug	Original indication	New indication	References
Amphotericin B (AMB)	Fungal infections	Leishmaniasis	Machado et al. (2015), Mondal et al. (2010)
Favipiravir	Influenza	COVID-19	Furuta et al. (2013), Singh et al. (2020)
Hydroxychloroquine	Malaria, RA	COVID-19	Infante et al. (2021), Rameshrad et al. (2020)
Ivermectin, anthelmintic	Scabies, river, blindness, helminthiasis	COVID-19	Shirazi et al. (2022)
Lopinavir/ritonavir,	HIV/AIDS	COVID-19	Singh et al. (2020)
Remdesivir	Influenza, Ebola (failed in clinical trial)	COVID-19	Babadaei et al. (2021), Rosa and Santos (2020)
Orlistat	Obesity	Cancer	Schcolnik-Cabrera et al. (2018), Turanli et al. (2018)
Tocilizumab	Rheumatoid arthritis	COVID-19	Perrone et al. (2020), Sultana et al. (2020)
Crizotinib	Lymphoma	NSCLC (non-small cell lung carcinoma), Leukaemia	Boulos et al. (2021), MotieGhader et al. (2022)
Daunorubicin	Antibiotic	Breast cancer	Agrawal (2018), Correia et al. (2021)
Dimethyl fumarate	Psoriasis	Multiple sclerosis	Thomas et al. (2022)
Digoxin	CVDs such as heart failure	Prostate cancer	Bahmad et al. (2022)
Everolimus	Immune suppressant	Pancreatic neuroendocrine tumors	Yao et al. (2011)
Fluorouracil	Cancer	Breast cancer	Aggarwal et al. (2021), Correia et al. (2021)
Fluoxetine	Depression	Premenstrual dysphoria	Romano et al. (1999), Steiner et al. (1995)
Galantamine	Neuromuscular paralysis	Alzheimer's disease	Kumar et al. (2017)
Ibudilast	Asthma	Neuropathic pain	Hama et al. (2012), Sisignano et al. (2022)
Isoniazid	Tuberculosis	Certain types of tumor	Lv et al. (2018)
Milnacipram	Depression	Fibromyalgia	English et al. (2010)
Miltefosine	Cancer	Leishmaniasis, amoeba infection	Latifi (2020), Sunyoto et al. (2018)
Mifepristone	Termination of pregnancy in combination with misoprostol	Cushing's syndrome	Johanssen and Allolio (2007), Morgan and Laufgraben (2013)
Minoxidil	Hypertension	Androgenic alopecia	Badria et al. (2020), Suchonwanit et al. (2019)
Nelfinavir	HIV/AIDS	Breast cancer	Koltai (2015), Subeha and Telleria (2020)

(continued)

Table 11.1 (continued)

Drug	Original indication	New indication	References
Simvastatin	CVDs	Lung cancer	Marciano et al. (2022)
Sunitinib	Imatinib-resistant	Pancreatic neuroendocrine tumors	Delbaldo et al. (2012), Raymond et al. (2011)
Topiramate	Fungal infections	Inflammatory bowel disease	Dudley et al. (2011)
Valsartan	Hypertension, heart attack	Alzheimer's disease	Kim (2015)

which includes unsuccessful and shelved drug ideas that may be saved to meet unmet medical needs (prominently, rare and neglected conditions).

Additionally, pharmaceutical sponsors, governmental organizations, and other stakeholders have access to sensitive information from clinical trials that, if made public, could result in a quality improvement for the drug discovery community. As drug rescue offers excellent, affordable opportunities to commercially exploit abandoned drug projects and maximize the advantages of pharmaceutical companies' portfolios, it is possible that open collaboration models bridging academia with private partners (non-profit organizations and industry) will continue to expand. Drug repurposing has frequently been promoted as an intriguing method to investigate new pharmacological treatments for uncommon and untreated disorders (in reality, many of the accessible therapies for these conditions can be viewed as repurposed drugs). Even though it may not be particularly profitable in terms of pure economics, the pursuit of pharmaceutical treatments for uncommon and undertreated disorders does imply other forms of value, such as corporate social responsibility and the resulting raised social awareness of and perception of pharmaceutical companies. Government organizations also have the ability to support such programs through a variety of financial incentives; thus it is crucial to raise awareness of the fact that the cost of treating neglected and uncommon diseases far outweighs the expenditure needed to create novel therapeutic approaches. It is possible that the two sectors with the best promise for drug repurposing are systems and precision medicine.

11.7 Conclusion

Historically, the discovery of new drug compounds has come about through the process of dug repurposing, with an emphasis on serendipitous findings. This has paved the way, in recent years, for the development of new therapeutics based on licensed medications that are already on the market. Strategic drug repositioning has sparked innovation as pharmacological compounds with unidentified therapeutic indications have been discovered. As a result of their ability to drastically reduce R&D costs, increase success rates, shorten research time, and decrease investment risk, drug repositioning strategies are gaining in popularity. The adoption of novel

repositioning strategy techniques in the drug discovery program for nearly all human diseases is made possible by these benefits, which are beneficial to researchers, consumers, and pharmaceutical corporations. Drug repurposing is an integral part of the drug discovery process, and it can be sped up with the help of in silico methods, SBDD, pharmacophore modeling tools, and AI technologies. In the era of precision medicine, drug repositioning has proven to be incredibly useful through the study of novel disease, metabolic, and signaling pathways, off-targets and target-specific processes, genetic expression profiles, and even genetic diseases. Advances in genomics have enabled us to access massive amounts of genomic and transcriptomic data through the use of methods like next-generation sequencing, microarray data, and transcriptomics, among others. There may be additional benefits from using network biology and systems biology methods to uncover such novel mechanisms of action.

References

Aggarwal S, Verma SS, Aggarwal S, Gupta SC (2021) Drug repurposing for breast cancer therapy: old weapon for new battle. Semin Cancer Biol 68:8–20. https://doi.org/10.1016/j.semcancer.2019.09.012

Agrawal P (2018) Artificial intelligence in drug discovery and development. J Pharm 6(2):1000e1173

Ahmad T, Davies R, Alagona P Jr (2017) The benefits and risks of oral antiplatelet therapy in patients with acute coronary syndrome. J Fam Pract 66(2 Suppl):supp_az_0217

Akhoon BA, Pandey S, Tiwari S, Pandey R (2016) Withanolide A offers neuroprotection, ameliorates stress resistance and prolongs the life expectancy of Caenorhabditis elegans. Exp Gerontol 78:47–56. https://doi.org/10.1016/j.exger.2016.03.004

Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P et al (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. Mol Pharm 13(7):2524–2530. https://doi.org/10.1021/acs.molpharmaceut.6b00248

Allarakhia M (2013) Open-source approaches for the repurposing of existing or failed candidate drugs: learning from and applying the lessons across diseases. Drug Des Devel Ther 7:753–766. https://doi.org/10.2147/DDDT.S46289

Amare GG, Meharie BG, Belayneh YM (2021) A drug repositioning success: the repositioned therapeutic applications and mechanisms of action of thalidomide. J Oncol Pharm Pract 27(3):673–678. https://doi.org/10.1177/1078155220975825

Aronson JK (2007) Old drugs–new uses. Br J Clin Pharmacol 64(5):563–565

Ashburn TT, Thor KB (2004) Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov 3(8):673–683. https://doi.org/10.1038/nrd1468

Attia YM, Ewida H, Ahmed MS (2020) Successful stories of drug repurposing for cancer therapy in hepatocellular carcinoma. In: Drug repurposing in cancer therapy. Elsevier, pp 213–229

Avvisati G, Tallman MS (2003) All-trans retinoic acid in acute promyelocytic leukaemia. Best Pract Res Clin Haematol 16(3):419–432. https://doi.org/10.1016/s1521-6926(03)00057-4

Azuaje F (2013) Drug interaction networks: an introduction to translational and clinical applications. Cardiovasc Res 97(4):631–641. https://doi.org/10.1093/cvr/cvs289

Babadaei MMN, Hasan A, Vahdani Y, Bloukh SH, Sharifi M et al (2021) Development of remdesivir repositioning as a nucleotide analog against COVID-19 RNA dependent RNA polymerase. J Biomol Struct Dyn 39(10):3771–3779. https://doi.org/10.1080/07391102.2020.1767210

Badria FA, Fayed HA, Ibraheem AK, State AF, Mazyed EA (2020) Formulation of sodium valproate nanospanlastics as a promising approach for drug repurposing in the treatment of androgenic alopecia. Pharmaceutics 12(9). https://doi.org/10.3390/pharmaceutics12090866

Bahmad HF, Demus T, Moubarak MM, Daher D, Alvarez Moreno JC et al (2022) Overcoming drug resistance in advanced prostate cancer by drug repurposing. Med Sci (Basel) 10(1). https://doi.org/10.3390/medsci10010015

Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC et al (2005) NCBI GEO: mining millions of expression profiles—database and tools. Nucleic Acids Res 33(Database issue):D562–D566. https://doi.org/10.1093/nar/gki022

Bidabadi E, Mashouf M (2010) A randomized trial of propranolol versus sodium valproate for the prophylaxis of migraine in pediatric patients. Paediatr Drugs 12(4):269–275. https://doi.org/10.2165/11316270-000000000-00000

Bisgin H, Liu Z, Fang H, Kelly R, Xu X et al (2014) A phenome-guided drug repositioning through a latent variable model. BMC Bioinformatics 15(1):1–12

Boulos JC, Saeed MEM, Chatterjee M, Bulbul Y, Crudo F et al (2021) Repurposing of the ALK inhibitor Crizotinib for acute leukemia and multiple myeloma cells. Pharmaceuticals (Basel) 14(11). https://doi.org/10.3390/ph14111126

Cha Y, Erez T, Reynolds IJ, Kumar D, Ross J et al (2018) Drug repurposing from the perspective of pharmaceutical companies. Br J Pharmacol 175(2):168–180. https://doi.org/10.1111/bph.13798

Chen JJ, Cai N, Chen GZ, Jia CC, Qiu DB et al (2017) The neuroleptic drug pimozide inhibits stem-like cell maintenance and tumorigenicity in hepatocellular carcinoma. Oncotarget 8(11):17593–17609. https://doi.org/10.18632/oncotarget.4307

Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T (2018) The rise of deep learning in drug discovery. Drug Discov Today 23(6):1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039

Correia AS, Gartner F, Vale N (2021) Drug combination and repurposing for cancer therapy: the example of breast cancer. Heliyon 7(1):e05948. https://doi.org/10.1016/j.heliyon.2021.e05948

Defranchi E, Schalon C, Messa M, Onofri F, Benfenati F et al (2010) Binding of protein kinase inhibitors to synapsin I inferred from pair-wise binding site similarity measurements. PLoS One 5(8):e12214. https://doi.org/10.1371/journal.pone.0012214

Delbaldo C, Faivre S, Dreyer C, Raymond E (2012) Sunitinib in advanced pancreatic neuroendocrine tumors: latest evidence and clinical potential. Ther Adv Med Oncol 4(1):9–18. https://doi.org/10.1177/1758834011428147

Derry S, Wiffen PJ, Aldington D, Moore RA (2015) Nortriptyline for neuropathic pain in adults. Cochrane Database Syst Rev 1(1):CD011209. https://doi.org/10.1002/14651858.CD011209.pub2

Dey G (2019) An overview of drug repurposing: review article. J Med Sci Clin Res 7(2):3–5

Doman TN, McGovern SL, Witherbee BJ, Kasten TP, Kurumbail R et al (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. J Med Chem 45(11):2213–2221. https://doi.org/10.1021/jm010548w

Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S et al (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci Transl Med 3(96):96ra76. https://doi.org/10.1126/scitranslmed.3002648

Dueñas-González A, García-López P, Herrera LA, Medina-Franco JL, González-Fierro A et al (2008) The prince and the pauper. A tale of anticancer targeted agents. Mol Cancer 7(1):1–33

Emig D, Ivliev A, Pustovalova O, Lancashire L, Bureeva S et al (2013) Drug target prediction and repositioning using an integrated network-based approach. PLoS One 8(4):e60618. https://doi.org/10.1371/journal.pone.0060618

English C, Rey JA, Rufin C (2010) Milnacipran (Savella), a treatment option for Fibromyalgia: P T. 35(5):261–266

Fabro S, Smith RL, Williams RT (1967) Toxicity and teratogenicity of optical isomers of thalidomide. Nature 215(5098):296. https://doi.org/10.1038/215296a0

Ferreira LG, Andricopulo AD (2016) Drug repositioning approaches to parasitic diseases: a medicinal chemistry perspective. Drug Discov Today 21(10):1699–1710. https://doi.org/10.1016/j.drudis.2016.06.021

Fisher Wilson J (2006) Alterations in processes and priorities needed for new drug development: American College of Physicians

Frediansyah A, Tiwari R, Sharun K, Dhama K, Harapan H (2021) Antivirals for COVID-19: a critical review. Clin Epidemiol Glob Health 9:90–98. https://doi.org/10.1016/j.cegh.2020.07.006

Furuta Y, Gowen BB, Takahashi K, Shiraki K, Smee DF et al (2013) Favipiravir (T-705), a novel viral RNA polymerase inhibitor. Antivir Res 100(2):446–454. https://doi.org/10.1016/j.antiviral.2013.09.015

Ghofrani HA, Osterloh IH, Grimminger F (2006) Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond. Nat Rev Drug Discov 5(8):689–702. https://doi.org/10.1038/nrd2030

Gottlieb A, Stein GY, Ruppin E, Sharan R (2011) PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol 7(1):496. https://doi.org/10.1038/msb.2011.26

Guragain S, Upadhayay N, Bhattarai BM (2017) Adverse reactions in leprosy patients who underwent dapsone multidrug therapy: a retrospective study. Clin Pharmacol 9:73–78. https://doi.org/10.2147/CPAA.S135846

Haeberle H, Dudley JT, Liu JT, Butte AJ, Contag CH (2012) Identification of cell surface targets through meta-analysis of microarray data. Neoplasia 14(7):666–669. https://doi.org/10.1593/neo.12634

Hama AT, Broadhead A, Lorrain DS, Sagen J (2012) The antinociceptive effect of the asthma drug ibudilast in rat models of peripheral and central neuropathic pain. J Neurotrauma 29(3):600–610. https://doi.org/10.1089/neu.2011.1863

Hebbring SJ (2014) The challenges, advantages and future of phenome-wide association studies. Immunology 141(2):157–165. https://doi.org/10.1111/imm.12195

Hughes JP, Rees S, Kalindjian SB, Philpott KL (2011) Principles of early drug discovery. Br J Pharmacol 162(6):1239–1249. https://doi.org/10.1111/j.1476-5381.2010.01127.x

Infante M, Ricordi C, Alejandro R, Caprio M, Fabbri A (2021) Hydroxychloroquine in the COVID-19 pandemic era: in pursuit of a rational use for prophylaxis of SARS-CoV-2 infection. Expert Rev Anti-Infect Ther 19(1):5–16. https://doi.org/10.1080/14787210.2020.1799785

Iorio F, Rittman T, Ge H, Menden M, Saez-Rodriguez J (2013) Transcriptional data: a new gateway to drug repositioning? Drug Discov Today 18(7–8):350–357. https://doi.org/10.1016/j.drudis.2012.07.014

Jadamba E, Shin M (2016) A systematic framework for drug repositioning from integrated omics and drug phenotype profiles using pathway-drug network. Biomed Res Int 2016:7147039. https://doi.org/10.1155/2016/7147039

Jin G, Wong STC (2014) Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. Drug Discov Today 19(5):637–644

Jin G, Fu C, Zhao H, Cui K, Chang J et al (2012) A novel method of transcriptional response analysis to facilitate drug repositioning for cancer therapy. Cancer Res 72(1):33–44. https://doi.org/10.1158/0008-5472.CAN-11-2333

Johanssen S, Allolio B (2007) Mifepristone (RU 486) in Cushing's syndrome. Eur J Endocrinol 157(5):561–569. https://doi.org/10.1530/EJE-07-0458

Kaitin KI (2010) Deconstructing the drug development process: the new face of innovation. Clin Pharmacol Ther 87(3):356–361. https://doi.org/10.1038/clpt.2009.293

Kalita J, Chetia D, Rudrapal M (2020) Design, synthesis, antimalarial activity and docking study of 7-Chloro-4- (2-(substituted benzylidene)hydrazineyl)quinolines. Med Chem 16(7):928–937. https://doi.org/10.2174/1573406415666190806154722

Karadsheh NS, Quttaineh NA, Karadsheh SN, El-Khateeb M (2021) Effect of combined G6PD deficiency and diabetes on protein oxidation and lipid peroxidation. BMC Endocr Disord 21(1): 246. https://doi.org/10.1186/s12902-021-00911-6

Karmarkar AB et al (2011) A review of drug discovery, development and clinical research

Karthik L, Kumar G, Keswani T, Bhattacharyya A, Chandar SS et al (2014) Protease inhibitors from marine actinobacteria as a potential source for antimalarial compound. PLoS One 9(3): e90972

Kim TW (2015) Drug repositioning approaches for the discovery of new therapeutics for Alzheimer's disease. Neurotherapeutics 12(1):132–142. https://doi.org/10.1007/s13311-014-0325-7

Koltai T (2015) Nelfinavir and other protease inhibitors in cancer: mechanisms involved in anticancer activity. F1000Res 4:9. https://doi.org/10.12688/f1000research.5827.2

Kumar S, Chowdhury S, Kumar S (2017) In silico repurposing of antipsychotic drugs for Alzheimer's disease. BMC Neurosci 18(1):76. https://doi.org/10.1186/s12868-017-0394-8

Laffitte E, Revuz J (2004) Thalidomide: an old drug with new clinical applications. Expert Opin Drug Saf 3(1):47–56. https://doi.org/10.1517/14740338.3.1.47

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC et al (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313(5795): 1929–1935. https://doi.org/10.1126/science.1132939

Latifi A (2020) Reviewing the effects of Miltefosine and suggesting it for the treatment of coronavirus disease (COVID-19). Infect Dis (Auckl) 13:1178633720977488. https://doi.org/10.1177/1178633720977488

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539

Li Y, Agarwal P (2009) A pathway-based view of human diseases and disease relationships. PLoS One 4(2):e4346. https://doi.org/10.1371/journal.pone.0004346

Li YY, Jones SJ (2012) Drug repositioning for personalized medicine. Genome Med 4(3):27. https://doi.org/10.1186/gm326

Li J, Lu Z (2013) Pathway-based drug repositioning using causal inference. BMC Bioinformatics 14(Suppl 16):S3. https://doi.org/10.1186/1471-2105-14-S16-S3

Lim H, Poleksic A, Yao Y, Tong H, He D et al (2016) Large-scale off-target identification using fast and accurate dual regularized one-class collaborative filtering and its application to drug repurposing. PLoS Comput Biol 12(10):e1005135. https://doi.org/10.1371/journal.pcbi.1005135

Liu Z, Guo F, Gu J, Wang Y, Li Y et al (2015) Similarity-based prediction for anatomical therapeutic chemical classification of drugs by integrating multiple data sources. Bioinformatics 31(11):1788–1795. https://doi.org/10.1093/bioinformatics/btv055

Luo Y, Thompson WK, Herr TM, Zeng Z, Berendsen MA et al (2017) Natural language processing for EHR-based pharmacovigilance: a structured review. Drug Saf 40(11):1075–1089. https://doi.org/10.1007/s40264-017-0558-6

Lv Q, Wang D, Yang Z, Yang J, Zhang R et al (2018) Repurposing antitubercular agent isoniazid for treatment of prostate cancer. Biomater Sci 7(1):296–306. https://doi.org/10.1039/c8bm01189c

Machado PR, Rosa ME, Guimaraes LH, Prates FV, Queiroz A et al (2015) Treatment of disseminated Leishmaniasis with liposomal amphotericin B. Clin Infect Dis 61(6):945–949. https://doi.org/10.1093/cid/civ416

Marciano G, Palleria C, Casarella A, Rania V, Basile E et al (2022) Effect of statins on lung cancer molecular pathways: a possible therapeutic role. Pharmaceuticals (Basel) 15(5). https://doi.org/10.3390/ph15050589

Matter H, Sotriffer C (2011) Applications and success stories in virtual screening, pp 319–358

Matthews SJ, McCoy C (2003) Thalidomide: a review of approved and investigational uses. Clin Ther 25(2):342–395. https://doi.org/10.1016/s0149-2918(03)80085-1

Maund E, Guski LS, Gotzsche PC (2017) Considering benefits and harms of duloxetine for treatment of stress urinary incontinence: a meta-analysis of clinical study reports. CMAJ 189(5):E194–E203. https://doi.org/10.1503/cmaj.151104

Mayr A, Klambauer G, Unterthiner T, Hochreiter S (2016) DeepTox: toxicity prediction using deep learning. Front Environ Sci 3:80

McClellan KJ, Markham A (1999) Finasteride: a review of its use in male pattern hair loss. Drugs 57(1):111–126. https://doi.org/10.2165/00003495-199957010-00014

Mehra R, Rani C, Mahajan P, Vishwakarma RA, Khan IA et al (2016) Computationally guided identification of novel mycobacterium tuberculosis GlmU inhibitory leads, their optimization, and in vitro validation. ACS Comb Sci 18(2):100–116. https://doi.org/10.1021/acscombsci.5b00019

Menden MP, Iorio F, Garnett M, McDermott U, Benes CH et al (2013) Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. PLoS One 8(4):e61318. https://doi.org/10.1371/journal.pone.0061318

Menendez JA, Lupu R (2007) Targeting human epidermal growth factor receptor 2: it is time to kill kinase death human epidermal growth factor receptor 3. J Clin Oncol Off J Am Soc Clin Oncol 25(17):2496–2498

Meng H, Shawe-Taylor J, Szedmak S, Farquhar JDR (2004) Support vector machine to synthesise kernels

Monaghan MG, Murphy CM (2021) Old drugs, new tricks - redefining therapeutic strategies for tissue regeneration. Adv Drug Deliv Rev 173:279–280. https://doi.org/10.1016/j.addr.2021.03.019

Mondal S, Bhattacharya P, Rahaman M, Ali N, Goswami RP (2010) A curative immune profile one week after treatment of Indian kala-azar patients predicts success with a short-course liposomal amphotericin B therapy. PLoS Negl Trop Dis 4(7):e764. https://doi.org/10.1371/journal.pntd.0000764

Morgan FH, Laufgraben MJ (2013) Mifepristone for management of Cushing's syndrome. Pharmacotherapy 33(3):319–329. https://doi.org/10.1002/phar.1202

MotieGhader H, Tabrizi-Nezhadi P, Deldar Abad Paskeh M, Baradaran B, Mokhtarzadeh A et al (2022) Drug repositioning in non-small cell lung cancer (NSCLC) using gene co-expression and drug-gene interaction networks analysis. Sci Rep 12(1):9417. https://doi.org/10.1038/s41598-022-13719-8

Napolitano F, Zhao Y, Moreira VM, Tagliaferri R, Kere J et al (2013) Drug repositioning: a machine-learning approach through data integration. J Chem 5(1):1–9

Ng SS, Brown M, Figg WD (2002) Thalidomide, an antiangiogenic agent with clinical activity in cancer. Biomed Pharmacother 56(4):194–199. https://doi.org/10.1016/s0753-3322(02)00177-4

Ozsoy MG, Özyer T, Polat F, Alhajj R (2018) Realizing drug repositioning by adapting a recommendation system to handle the process. BMC Bioinformatics 19(1):1–14

Padhy BM, Gupta YK (2011) Drug repositioning: re-investigating existing drugs for new therapeutic indications. J Postgrad Med 57(2):153–160. https://doi.org/10.4103/0022-3859.81870

Parvathaneni V, Gupta V (2020) Utilizing drug repurposing against COVID-19 - efficacy, limitations, and challenges. Life Sci 259:118275. https://doi.org/10.1016/j.lfs.2020.118275

Perrone F, Piccirillo MC, Ascierto PA, Salvarani C, Parrella R et al (2020) Tocilizumab for patients with COVID-19 pneumonia. The single-arm TOCIVID-19 prospective trial. J Transl Med 18(1):405. https://doi.org/10.1186/s12967-020-02573-9

Qorri B, Mokhtari RB, Harless WW, Szewczuk MR (2022) Next generation of cancer drug repurposing: therapeutic combination of aspirin and oseltamivir phosphate potentiates gemcitabine to disable key survival pathways critical for pancreatic cancer progression. Cancers (Basel) 14(6). https://doi.org/10.3390/cancers14061374

Rameshrad M, Ghafoori M, Mohammadpour AH, Nayeri MJD, Hosseinzadeh H (2020) A comprehensive review on drug repositioning against coronavirus disease 2019 (COVID19). Naunyn Schmiedeberg's Arch Pharmacol 393(7):1137–1152. https://doi.org/10.1007/s00210-020-01901-6

Raymond E, Dahan L, Raoul JL, Bang YJ, Borbath I et al (2011) Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. N Engl J Med 364(6):501–513. https://doi.org/10.1056/NEJMoa1003825

Roder C, Thomson MJ (2015) Auranofin: repurposing an old drug for a golden new age. Drugs R D 15(1):13–20. https://doi.org/10.1007/s40268-015-0083-y

Romano S, Judge R, Dillon J, Shuler C, Sundell K (1999) The role of fluoxetine in the treatment of premenstrual dysphoric disorder. Clin Ther 21(4):615–633; discussion 613. https://doi.org/10.1016/S0149-2918(00)88315-0

Rosa SGV, Santos WC (2020) Ensayos clinicos de reposicionamiento de medicamentos para el tratamiento de la COVID-19. Rev Panam Salud Publica, 44:NA–NA

Roundtable on Translating Genomic-Based Research for H, Board on Health Sciences P, Institute of M (2014) The National Academies Collection: reports funded by National Institutes of Health drug repurposing and repositioning: workshop summary. National Academies Press (US), Washington, DC

Rudrapal M, Chetia D (2016) Endoperoxide antimalarials: development, structural diversity and pharmacodynamic aspects with reference to 1,2,4-trioxane-based structural scaffold. Drug Des Devel Ther 10:3575–3590. https://doi.org/10.2147/DDDT.S118116

Rudrapal M, Khairnar SJ, Jadhav AG (2020) Drug repurposing (DR): an emerging approach in drug discovery. In: Dekebo A (ed) Molecular aspects, applications. IntechOpen, London

Saha SK, Yin Y, Kim K, Yang G-M, Abdal Dayem A et al (2017) Valproic acid induces endocytosis-mediated doxorubicin internalization and shows synergistic cytotoxic effects in hepatocellular carcinoma cells. Int J Mol Sci 18(5):1048

Sannella AR, Casini A, Gabbiani C, Messori L, Bilia AR et al (2008) New uses for old drugs. Auranofin, a clinically established antiarthritic metallodrug, exhibits potent antimalarial effects in vitro: mechanistic and pharmacological implications. FEBS Lett 582(6):844–847. https://doi.org/10.1016/j.febslet.2008.02.028

Schcolnik-Cabrera A, Chavez-Blanco A, Dominguez-Gomez G, Taja-Chayeb L, Morales-Barcenas R et al (2018) Orlistat as a FASN inhibitor and multitargeted agent for cancer therapy. Expert Opin Investig Drugs 27(5):475–489. https://doi.org/10.1080/13543784.2018.1471132

Shah B, Modi P, Sagar SR (2020) In silico studies on therapeutic agents for COVID-19: drug repurposing approach. Life Sci 252:117652. https://doi.org/10.1016/j.lfs.2020.117652

Sharma A, Tiwari S, Deb MK, Marty JL (2020) Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): a global pandemic and treatment strategies. Int J Antimicrob Agents 56(2):106054. https://doi.org/10.1016/j.ijantimicag.2020.106054

Shaughnessy AF (2011) Old drugs, new tricks. BMJ 342:d741. https://doi.org/10.1136/bmj.d741

Shekhar C (2008) In silico pharmacology: computer-aided methods could transform drug development. Chem Biol 15(5):413–414. https://doi.org/10.1016/j.chembiol.2008.05.001

Shen M, Xiao Y, Golbraikh A, Gombar VK, Tropsha A (2003) Development and validation of k-nearest-neighbor QSPR models of metabolic stability of drug candidates. J Med Chem 46(14):3013–3020. https://doi.org/10.1021/jm020491t

Shirazi FM, Mirzaei R, Nakhaee S, Nejatian A, Ghafari S et al (2022) Repurposing the drug, ivermectin, in COVID-19: toxicological points of view. Eur J Med Res 27(1):21. https://doi.org/10.1186/s40001-022-00645-8

Singh TU, Parida S, Lingaraju MC, Kesavan M, Kumar D et al (2020) Drug repurposing approach to fight COVID-19. Pharmacol Rep 72(6):1479–1508. https://doi.org/10.1007/s43440-020-00155-6

Sisignano M, Gribbon P, Geisslinger G (2022) Drug repurposing to target neuroinflammation and sensory neuron-dependent pain. Drugs 82(4):357–373. https://doi.org/10.1007/s40265-022-01689-0

Sleigh SH, Barton CL (2012) Repurposing strategies for therapeutics. Pharm Med 24:151–159

Smaldone C, Brugaletta S, Pazzano V, Liuzzo G (2009) Immunomodulator activity of 3-hydroxy-3-methilglutaryl-CoA inhibitors. Cardiovasc Hematol Agents Med Chem 7(4):279–294. https://doi.org/10.2174/187152509789541864

Steiner M, Steinberg S, Stewart D, Carter D, Berger C et al (1995) Fluoxetine in the treatment of premenstrual dysphoria. Canadian fluoxetine/premenstrual dysphoria collaborative study group. N Engl J Med 332(23):1529–1534. https://doi.org/10.1056/NEJM199506083322301

Strittmatter WJ (2012) Medicine. Old drug, new hope for Alzheimer's disease. Science 335(6075): 1447–1448. https://doi.org/10.1126/science.1220725

Subeha MR, Telleria CM (2020) The anti-cancer properties of the HIV protease inhibitor nelfinavir. Cancers (Basel) 12(11). https://doi.org/10.3390/cancers12113437

Suchonwanit P, Thammarucha S, Leerunyakul K (2019) Minoxidil and its use in hair disorders: a review. Drug Des Devel Ther 13:2777–2786. https://doi.org/10.2147/DDDT.S214907

Sultana J, Crisafulli S, Gabbay F, Lynn E, Shakir S et al (2020) Challenges for drug repurposing in the COVID-19 pandemic era. Front Pharmacol 11:588654. https://doi.org/10.3389/fphar.2020.588654

Sunyoto T, Potet J, Boelaert M (2018) Why miltefosine-a life-saving drug for leishmaniasis-is unavailable to people who need it the most. BMJ Glob Health 3(3):e000709. https://doi.org/10.1136/bmjgh-2018-000709

Susnow RG, Dixon SL (2003) Use of robust classification techniques for the prediction of human cytochrome P450 2D6 inhibition. J Chem Inf Comput Sci 43(4):1308–1315. https://doi.org/10.1021/ci030283p

Swamidass SJ (2011) Mining small-molecule screens to repurpose drugs. Brief Bioinform 12(4): 327–335. https://doi.org/10.1093/bib/bbr028

Swinney DC, Anthony J (2011) How were new medicines discovered? Nat Rev Drug Discov 10(7): 507–519. https://doi.org/10.1038/nrd3480

Tallman MS, Andersen JW, Schiffer CA, Appelbaum FR, Feusner JH et al (1997) All-trans-retinoic acid in acute promyelocytic leukemia. N Engl J Med 337(15):1021–1028. https://doi.org/10.1056/NEJM199710093371501

Tari LB, Patel JH (2014) Systematic drug repurposing through text mining. Biomed Lit Mining 253–267

Thomas SD, Jha NK, Sadek B, Ojha S (2022) Repurposing dimethyl fumarate for cardiovascular diseases: pharmacological effects, molecular mechanisms, and therapeutic promise. Pharmaceuticals (Basel) 15(5):497. https://doi.org/10.3390/ph15050497

Tiono AB, Dicko A, Ndububa DA, Agbenyega T, Pitmang S et al (2009) Chlorproguanil-dapsone-artesunate versus chlorproguanil-dapsone: a randomized, double-blind, phase III trial in African children, adolescents, and adults with uncomplicated Plasmodium falciparum malaria. Am J Trop Med Hyg 81(6):969–978. https://doi.org/10.4269/ajtmh.2009.09-0351

Tobinick EL (2009) The value of drug repositioning in the current pharmaceutical market. Drug News Perspect 22(2):119–125. https://doi.org/10.1358/dnp.2009.22.2.1303818

Toschi L, Finocchiaro G, Bartolini S, Gioia V, Cappuzzo F (2005) Role of gemcitabine in cancer therapy. Future Oncol 1(1):7–17. https://doi.org/10.1517/14796694.1.1.7

Turanli B, Grotli M, Boren J, Nielsen J, Uhlen M et al (2018) Drug repositioning for effective prostate cancer treatment. Front Physiol 9:500. https://doi.org/10.3389/fphys.2018.00500

Upadhyaya HP, Desaiah D, Schuh KJ, Bymaster FP, Kallman MJ et al (2013) A review of the abuse potential assessment of atomoxetine: a nonstimulant medication for attention-deficit/hyperactivity disorder. Psychopharmacology 226(2):189–200. https://doi.org/10.1007/s00213-013-2986-z

Vickers NJ (2017) Animal communication: when i'm calling you, will you answer too? Curr Biol 27(14):R713–R715

Weber A, Casini A, Heine A, Kuhn D, Supuran CT et al (2004) Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. J Med Chem 47(3):550–557. https://doi.org/10.1021/jm030912m

Wu Z, Lu W, Wu D, Luo A, Bian H et al (2016) In silico prediction of chemical mechanism of action via an improved network-based inference method. Br J Pharmacol 173(23):3372–3385. https://doi.org/10.1111/bph.13629

Xu H, Aldrich MC, Chen Q, Liu H, Peterson NB et al (2015) Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. J Am Med Inform Assoc 22(1):179–191. https://doi.org/10.1136/amiajnl-2014-002649

Xu X, Huang M, Zou X (2018) Docking-based inverse virtual screening: methods, applications, and challenges. Biophys Rep 4(1):1–16. https://doi.org/10.1007/s41048-017-0045-8

Yao JC, Shah MH, Ito T, Bohas CL, Wolin EM et al (2011) Everolimus for advanced pancreatic neuroendocrine tumors. N Engl J Med 364(6):514–523. https://doi.org/10.1056/NEJMoa1009290

Yasuda T, Yoshida T, Goda AE, Horinaka M, Yano K et al (2008) Anti-gout agent allopurinol exerts cytotoxicity to human hormone-refractory prostate cancer cells in combination with tumor necrosis factor-related apoptosis-inducing ligand. Mol Cancer Res 6(12):1852–1860. https://doi.org/10.1158/1541-7786.MCR-08-0012

Zhang Y, Post-Martens K, Denkin S (2006) New drug candidates and therapeutic targets for tuberculosis therapy. Drug Discov Today 11(1–2):21–27. https://doi.org/10.1016/S1359-6446(05)03626-3

Zhu F, Patumcharoenpol P, Zhang C, Yang Y, Chan J et al (2013) Biomedical text mining and its applications in cancer research. J Biomed Inform 46(2):200–211. https://doi.org/10.1016/j.jbi.2012.10.007

Zhu Q, Tao CUI, Shen F, Chute CG (2014) Exploring the pharmacogenomics knowledge base (PharmGKB) for repositioning breast cancer drugs by leveraging web ontology language (OWL) and cheminformatics approaches *Biocomputing 2014*. World Scientific, pp 172–182

Zouboulis CC (2009) Zileuton, a new efficient and safe systemic anti-acne drug. Dermatoendocrinol 1(3):188–192. https://doi.org/10.4161/derm.1.3.8368

Chapter 12
CADD Approaches in Anticancer Drug Discovery

Abanish Biswas and Venkatesan Jayaprakash

12.1 Introduction

In the majority of the world's nations, cancer is currently the main cause of premature death or a close second. Due to the considerable impact of demographic changes, such as population growth and aging, on the disparate patterns in cancer incidence in different locations, it is anticipated that there will be an increase in the number of cancer patients worldwide during the next 50 years. By 2070, the incidence of all malignancies combined is expected to double compared to 2020, if that recent incidence patterns for the major cancer types continue. The expected increases in the national burden decrease with higher levels of national Human Development Index (HDI), with the largest increases predicted in lower-resource settings and in nations that are currently given a low HDI. Therefore, in order for national cancer control programs to produce the anticipated public health and economic benefits in the twenty-first century, countries must launch them quickly (Soerjomataram and Bray 2021).

Recent research on the immune system and cancer has resulted in the creation of novel drugs that boost the immune system to target cancer cells. There are medicines available that make use of certain signals to halt the multiplication of cancer cells, and there are additional medications now under research that target cancer cells specifically and kill them off directly. The process of finding new treatments for complex illnesses may be sped up by using machine learning algorithms to the investigation of these ailments. The analysis of cancer genomes and the identification of the treatment protocols that are the most effective for various subtypes of the

A. Biswas (✉) · V. Jayaprakash (✉)
Department of Pharmaceutical Sciences and Technology, Birla Institute of Technology, Ranchi, Jharkhand, India
e-mail: venkatesanj@bitmesra.ac.in

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_12

disease are two fields of medicine that might stand to gain a great deal from the use of machine learning techniques. Unfortunately, the creation of a whole new medication is an endeavor that is not only expensive but also time-consuming and fraught with risk. When adopting the conventional approach, the process of developing new pharmaceuticals may take up to 15 years and cost more than 1 billion dollars. As a result, computer-aided drug design, also known as CADD, has developed as a successful and promising strategy for the production of safer, more cost-effective, and more rapidly usable therapeutic prototypes. Numerous drug discovery programs currently make use of cutting-edge technology and methodologies, both of which have been developed in recent years to improve the analytical processes involved in drug development and the overall effectiveness of the process. The preclinical screening of candidate compounds is an essential part of the process of discovering novel treatments. The procedures of compound selection have substantially improved, thanks to the introduction of animal testing and in vitro drug screening. The process of finding new drugs requires several steps, one of which is the re-clinical screening of candidate molecules. Through the use of animals as test subjects and in vitro drug screening, the procedures for selecting compounds have been significantly improved and optimized. However, the exploratory tests used for small molecule analysis in cancer treatment discovery are frequently expensive and time-consuming. Therefore, improved methods for creating conventional medications are required (Rosenblum et al. 2018).

Finding new applications for already-approved drugs is significantly less expensive than looking for brand new cancer treatments. Using obtained multi-omics data, anticancer therapy efficacy can be predicted based on drug repositioning. One of the biggest challenges in modern cancer treatment is coping with patient heterogeneity. Since more than 50 years ago, it has been understood that people with different forms of cancer will react to the same treatment in quite diverse ways. Cancer patients frequently receive combination therapy, which combines several different drugs for better clinical outcomes. Over monotherapy, combination therapies and medications have many benefits. We look at how these methods have been used in assisted separation to good effect (Yap et al. 2013). Innovative anticancer therapies may be developed by combining in silico drug design with improvements in cancer research, which have been shown to be effective in the past (Li et al. 2019a).

12.1.1 Structure-Based Computer-Aided Drug Design (SB-CADD) Approach

Structure-based computer-aided drug design, made possible as a result of the sequencing of the human genome, has had a major impact on the procedure of creating novel pharmaceuticals. Possible targets for cancer treatment and new understandings of the illness might be uncovered with its help. SB-CADD should be utilized to identify anticancer drugs with a wide range of shapes, taking use of

cutting-edge technologies such the 3D architectures of cancer-related proteins that are crucial to therapy. The binding site interaction and other variables contributing to specificity may be studied in order to achieve this goal. Protein-centric and ligand-centric structure-based approaches are the two most common types. These two classes should not be overlooked. Studying the structural information of complex ligands may help advance SB-CADD in the quest for novel therapeutics. The main interaction between the target protein and ligand may be extracted from the protein-ligand complex, which may provide information on the protein's activity or the ligand's inhibitory effects. If the protein-ligand approach is not feasible, the protein-based approach might be used to interpret the quality of the relevant protein binding data into pharmacophoric features (Ferreira et al. 2015).

12.1.2 Ligand-Based Computer-Aided Drug Design (LB-CADD) Approach

LB-CADD methods make predictions about new bioactive molecules with comparable biological effects based on prior knowledge about active medications, like their structural and physiochemical characteristics. Drug compound prediction is based on the similarities of features (such as aromaticity, hydrogen-bond acceptors (HBA), hydrogen-bond donors (HBD), surface charge, anion, and cation residues), with the underlying premise that compounds with high physicochemical and structural similarity seem to be more likely to share similar biological activity (Martin et al. 2002). When the target protein's 3D structure is unknown, LB-CADD is generally used. When the data regarding the structure of protein is very limited, Quantitative Structure-Activity Relationship (QSAR) models and Pharmacophore models yield constructive data regarding the interaction between ligand and protein (Prada-Gracia et al. 2016). There are a variety of freely accessible compound libraries for virtualized chemical compound screening (Table 12.1) and for target prediction (Table 12.2). One should remember that SB-CADD and LB-CADD techniques are not contradictory incompatible to each other; sometimes combination techniques are required for screening of large libraries. A comprehensive flowchart for CADD process is shown in Fig. 12.1.

12.2 Computational Approaches for Anticancer Drug Discovery

12.2.1 Anticancer Small Organic Molecules Design

The use of cytotoxic drugs is the foundation of conventional chemotherapy. However, these drugs frequently fail to localize preferentially in the microenvironment of

Table 12.1 List of online available compound database

Database	URL	References
ChEMBL	https://www.ebi.ac.uk/chembl/	Wishart et al. (2008), Kim et al. (2019)
DrugBank	https://go.drugbank.com/	Wishart et al. (2008)
Therapeutic Target Database	http://db.idrblab.net/ttd/	Wang et al. (2019b)
TCM	http://tcm.cmu.edu.tw/	Chen (2011)
ZINC	https://zinc.docking.org/	Irwin et al. (2020)
CTD	http://ctdbase.org/	Davis et al. (2021)
ChemSpider	http://www.chemspider.com/	Pence and Williams (2010)
T3DB	http://www.t3db.ca/	Wishart et al. (2015)
PubChem	https://pubchem.ncbi.nlm.nih.gov/	Kim et al. (2021)

Table 12.2 Available virtual library for target prediction

Sl. No	Name of the server	URL of the server	References
1	DisGenNET	https://www.disgenet.org/	Piñero et al. (2019)
2	Harmonizome	https://maayanlab.cloud/Harmonizome/	Rouillard et al. (2016)
3	MolTarPred	http://moltarpred.marseille.inserm.fr	Peón et al. (2019)
4	Open targets platform	https://platform.opentargets.org/	Ochoa et al. (2021)
5	PPB	http://gdbtools.unibe.ch:8080/PPB/	Awale and Reymond (2017)
6	SuperPred	https://prediction.charite.de/	Dunkel et al. (2008)
7	SwissTargetPrediction	http://www.swisstargetprediction.ch/	Gfeller et al. (2014)

the tumor, which can result in damage to healthy tissue as well as the need to increase the dosage in order to achieve the desired therapeutic effect. Researchers have a tough time increasing selectivity in oncology to lessen off-target damage caused by traditional cancer treatment due to the high degree of similarity that exists between cancer cells and normal cells. Small molecule drug conjugates, also known as SMC, are a promising technique for targeted treatment. These drug conjugates enable small molecules to release a powerful cytotoxic agent particularly in the microenvironment of a tumor, which increases the therapeutic potential of anticancer drugs.

They are preferable than antibody drug conjugates because of their non-immunogenic nature, reduced molecular weight, and regulated synthesis, which make them particularly effective for penetrating cancer cells. The SMCs were developed using a concept that is comparable to that of antibody drug conjugates. Numerous products manufactured by SMC are now participating in clinical research and testing, one of which is 177Lu-DOTATATE, which is already in use in medical facilities. According to this point of view, the numerous SMC design

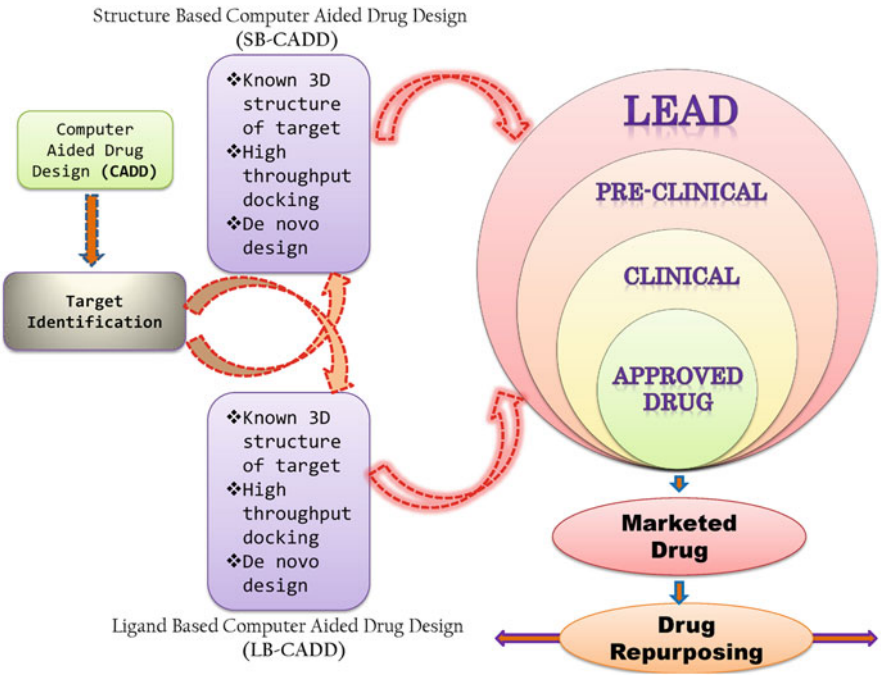


Fig. 12.1 Comprehensive workflow for computer aided drug design process

components, which include therapeutic payloads, targeting ligands, linkers, spacers, and cleavable bridges, have attracted a considerable amount of attention. There is also discussion of the many different types of SMCs, the possible mechanisms of action and therapeutic uses of these SMCs, and a few more SMCs that are now undergoing clinical investigation against a variety of tumours (Patel et al. 2021).

12.2.2 Anticancer Peptide Design

Peptides were long seen as a fringe industry with dim future prospects. This was principally caused by these molecules' inability to traverse the plasma membrane, physiological instability, limited or no oral absorption, and the critical functions that the amino acid chain plays in hormone signaling (Henninot et al. 2018). They can potentially replace natural agonists and can specifically target interactions between protein molecules, among their many other benefits. These drawbacks have been addressed using a variety of strategies, including the use of synthetic amino acids, framework modifications, and innovative formulations, leading to a notable rise in the manufacture of peptide drugs. It is preferable to use a structure-based approach while designing computational peptides.

Structural protein-protein complexes are indeed the major causes of protein sequences for therapeutic peptide engineering. However, this sort of information is rarely accessible, and computational chemistry may thus play a crucial role in this. Initially, it is indeed essential to construct a precise marker of the ligand-protein complex, if feasible, or perhaps more biologically active chains of amino acids from arbitrary repositories or legitimate sources when even a basic model is inaccessible when conducting a computer-aided analysis to develop an amino acid chains the specificity and affinity for a specific target. Consolidated docking approaches are frequently used in drug discovery because peptides probe a broader conformational area than low molecular weight compounds.

These, meanwhile, are unsuitable for this kind of study. Alternatively, where appropriate, contemporary docking methods in conjunction with experimental constraints may help in the identification of the appropriate structure (Salmaso et al. 2017; Ciemny et al. 2018). Once a model of the complex is created, molecular dynamics simulations study predicted that such a technique could be used to help make alterations in peptide structure to increase affinity and specificity (Lammi et al. 2019).

The capacity of the peptide to bind BTLA was supported by ten-nanosecond simulations of MD, which initially backed this notion. The amino acid chain was found to be a highly effective inhibitor of interactions between protein components through experiments. A covalent link between BTLA and the protein may have formed instead of a molecule with the same structure as that shown in the X-ray experiments, however, as this influence is predominantly caused by the appearance of an accessible cysteine residue in the amino acid chain.

12.2.3 QSAR Modeling

One further ligand-based strategy, QSAR (Quantitative Structure Activity Relationship), examines pharmacological biological activities by comparing their unique molecular descriptors (MDs) or fingerprints (FPs). With the evolution of ML algorithms, interaction descriptors can also be used to generate QSAR predictive model using python notebook (Jupyter notebook, google colab), R-Studio. Like its moniker indicates, it uses statistical techniques to establish a quantitative relationship between the experimentally observed biological activity of a molecule and its physiochemical features. The resulting quantitative equation may be used to foretell the bioactivity of a custom-built molecular structure. Several tools are available online (Table 12.3) for building a robust QSAR model.

Several machine learning and deep learning (DL) methods, such as Support Vector Machine (SVM), Random Forest (RF), Polynomial Regression (PR), Multi Linear Regression (MLR), and Artificial Neural Network (ANN), have been used in the process of developing QSAR models (Mendenhall and Meiler 2016). In addition to its use in drug discovery, QSAR has been put to use in a wide variety of other

Table 12.3 List of online both server and single application file used for QSAR modeling

Sl. No	Name of the tool	URL	References
1	QSAR-Co	https://sites.google.com/view/qsar-co	Ambure et al. (2019)
2	Open3DQSAR	http://open3dqsar.sourceforge.net/	Tosco and Balle (2011)
3	SYBYL-X	https://chemweb.ir/downloads/sybyl-x-suite/	Jing et al. (2014)
4	QSAR ToolBox	https://qsartoolbox.org/	Dimitrov et al. (2016)
5	McQSAR	http://users.abo.fi/mivainio/mcqsar/index.php	Vainio and Johnson (2005)

areas of molecular design, including the prediction of novel molecule analogue activity, the optimization of lead, and the prediction of new structural leads.

In the traditional 2D-QSAR methodologies, the bioactivity of molecule is linked to their physiochemical characteristics, such as their steric, electronic, and hydrophobic properties, and the statistical equations that describe the relationships between these factors are used to describe the connections (Hansch and Fujita 1964). Calculations of the force fields are the foundation of more sophisticated 3D-QSAR methods, such as comparative molecular field analysis (Cramer et al. 1988) and molecular similarity indexes in a comparative analysis (Klebe et al. 1994). Both of these approaches were published in 1988 and 1994, respectively. The data on the structures of the compounds is required, and the derived predictive models must be represented as three-dimensional contour maps in order to make visualization and understanding easier.

12.2.4 Pharmacophore Mapping

One of the most crucial methods for the efficient virtual screening of databases containing millions of chemicals is 3D pharmacophore-based methodologies. Virtual screening with such three-dimensional geometric models remains a substantial computational and conceptual problem, despite the fact that their effectiveness is mostly reliant on intuitive interpretation and invention. Most existing systems prioritize rapid screening speed above precision. In this overview, we will compare and contrast the approaches used by several existing pharmacophore mapping (PM) systems to do 3D pharmacophore searches. The price of creating new medicines for medical use is high. To begin, compounds in current chemical libraries are screened to see whether they are active against a target. This calls for a substantial investment of time and energy. Consequently, it is now normal practice to conduct a virtual screening, where computers are used to anticipate the activity of very huge libraries of molecules, in order to determine which ones have the greatest promise for subsequent laboratory studies. The expense of scientific investigations in the fields of medicine and biology may be greatly reduced by using simulation software rather

than actual experiments. In this research, we discuss methods that may rapidly search large databases for compounds that are structurally similar to a given sample compound (Brown et al. 2021).

The field of pharmacophore mapping (PM), which is widely acknowledged to be one of the most valuable tools, has seen a substantial amount of development throughout the course of history. These advancements may be attributed to the efforts of several researchers. This factor has been taken into consideration at a number of different junctures during the process of producing new pharmaceuticals over the course of the last several years. The modeling of pharmacophores has previously been accomplished by using a wide variety of structure-based techniques. It has been shown that the use of PM is advantageous in a number of contexts, including virtual screening, de novo design, and the maximization of lead potential (Yang 2010). The process of finding the protein's active-binding pocket and establishing the critical receptor-ligand interactions may be made a great deal simpler by using a technique that is known as the target-ligand complex approach. Additional examples are "Ligandscout" (Wolber et al. 2007), "pocket v.2" (Chen and Lai 2006), and "GBPM" (Piotrowska et al. 2018). It is very important to keep in mind that ligands cannot be used in contexts in which the identities of the ligands are unknown. The protein that was used in Discovery Studio (Lu et al. 2018), but did not apply a strategy that was based on ligands, gives a real-world example of a method that does not rely on ligands or interactions between receptors and ligands. This method was described as "ligand-independent." This technique does not use any ligand-based approaches in any way. Characterizing the pharmacological properties of the interactions that take place inside the binding site is the job of the piece of software known as LUDI (Yuriev et al. 2015). The produced interaction maps often include a number of unprioritized interactive features, despite the fact that using this rigorous SBP approach to characterize the overall interaction potential of a binding pocket does provide a few advantages.

12.2.5 Simulation of Molecular Docking and Molecular Dynamics of Small Molecules

By analyzing and predicting the sequences and contact interactions between ligands and receptor proteins, molecular docking has become a basic structure-based tool in rational drug design (Ferreira et al. 2015). Rigid molecular docking (RMD) and flexible molecular docking (FMD) are two types of docking studies that are defined by the flexibility or lack thereof of the ligands used in the computational technique (Halperin et al. 2002; Dias and de Azevedo 2008). The rigid docking technique, also known as a critical approach, emphasizes rigidity and the lack of flexibility in the induced-fit theory (Salmaso and Moro 2018) in favor of a focus on fixed geometry and structural and chemical reciprocity between ligands and targeted proteins. Rigid docking is widely used in the process of drug discovery because of its speed and

Table 12.4 List of molecular docking software

Sl. No.	Name if the tool	URL	References
1	GOLD	https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/	Joy et al. (2006)
2	Glide	https://www.schrodinger.com/products/glide	Friesner et al. (2004)
3	FlexX docking	https://www.biosolveit.de/SeeSAR/#FlexX	Kramer et al. (1999)
4	UCSF DOCK	https://dock.compbio.ucsf.edu/	Kuntz et al. (1982)
5	AutoDock Vina	https://vina.scripps.edu/	Trott and Olson (2010)
6	MGL tools	https://ccsb.scripps.edu/mgltools/	Morris et al. (2009)
7	HADDOCK	https://wenmr.science.uu.nl/haddock2.4/	Dominguez et al. (2003)
8	Surflex	https://www.biopharmics.com/	Jain (2007)
9	pyDockWEB	https://life.bsc.es/pid/pydockweb/	Cheng et al. (2007)
10	GEMDOCK	http://gemdock.life.nctu.edu.tw/dock/	Yang and Chen (2004)
11	PATCHDOCK	https://bioinfo3d.cs.tau.ac.il/PatchDock/	Schneidman-Duhovny et al. (2005)
12	ClusPro	https://cluspro.bu.edu/publications.php	Kozakov et al. (2017)

efficiency, and it is implemented by a wide variety of tiny molecular databases. Such data, however, would be more specific and nuanced under a flexible docking strategy. List of various molecular docking tools are shown Table 12.4.

The molecular docking process may be broken down into three distinct steps (Fig. 12.2). Small molecule and target protein structures must be established initially. The second use is predicting ligand binding site conformations, orientations, and positional spaces. Conformational search algorithms achieve this aim by combining the methods of systematic and stochastic searching to anticipate the conformations of binary compounds. Exhaustive search, fragmentation, and conformational ensemble are the three main categories of systematic search techniques. However, a number of other stochastic techniques exist, such as (1) the use of a Monte Carlo (MC) algorithm, (2) the use of a tabu search strategy, (3) the use of an evolutionary algorithm, and (4) the use of a swarm optimization (SO) algorithm. Finally, the scoring function and the potential binding free energy are evaluated by these algorithms to determine which molecules have the greatest binding potential to their targets during molecular docking. Among the many possible scoring functions are the function for consensus scoring, empirical scoring functions, scoring functions based on prior knowledge, and scoring functions based on a theoretical force field (Cui et al. 2020; Kong et al. 2022). Currently, with the evolution of hardware accelerated molecular docking software such as AutoDock-GPU, docking process can be done using Google Colab which provides GPU-enabled local host runtime (Solis-Vasquez et al. 2022).

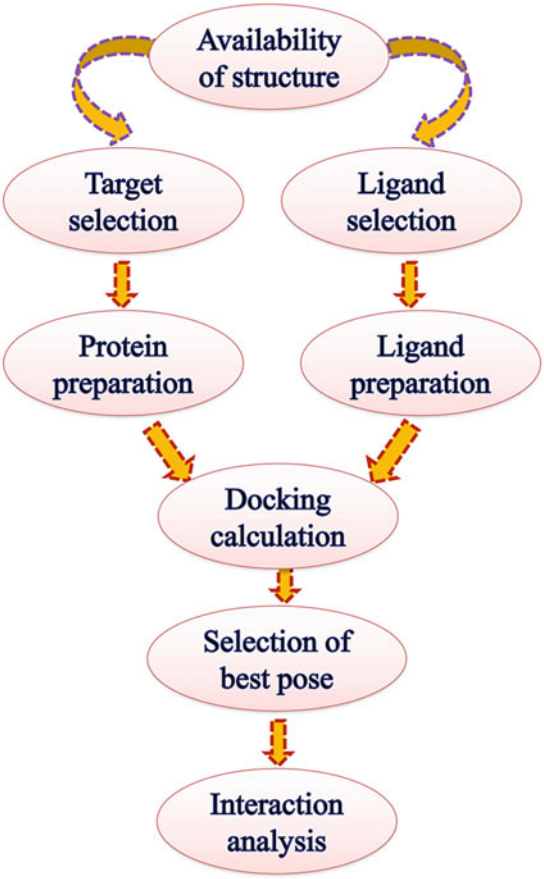


Fig. 12.2 Stepwise workflow for molecular docking process

Molecular Dynamic Simulation

However, because of the complicated atomic interactions between the target protein and ligand molecule, predicting the movements of active binding sites and ligands is computationally expensive. This shortcoming was initially addressed with the advent of molecular dynamics (MD) simulation in the 1970s. Simulation of atomic movements is achieved by solving Newtonian motion, which also helps to simplify the calculations required (McCammon et al. 1977; Hansson et al. 2002). The capacity to virtual screening and get insight into the structural characteristics of proteins and the stability of protein-ligand complexes is a major benefit of MD simulations for the drug development process. The discovery of new binding sites, such as allosteric sites, aids in the development of more potent pharmaceutical drugs (Grant et al. 2011; Nair et al. 2012). In computational drug development, MD simulations are often used to verify the binding stability of the best pose of the complexes. For the sake of brevity, we will assume that AMBER or CharmGUI (Pearlman et al. 1995) is used to construct protein and ligand topologies using

default values. AMBER (Pearlman et al. 1995), CHARMM (Jo et al. 2008), and GROMOS (Christen et al. 2005) are all simulation tools that use force fields to model the dynamics (atomic movement) of the complex. In order to evaluate complex stability, scientists often look at the root mean square fluctuation, the RMSD, the radius of gyration, and the presence of hydrogen bonding structures. Molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) and molecular mechanics/generalized born surface area (MM/GBSA) are two methods for calculating the binding free energy of ligand-protein complexes that are both more accurate than most molecular docking scoring functions and require less computational effort (Rastelli et al. 2009; Wang et al. 2019a, b). These strategies may enhance molecular docking outcomes because they effectively duplicate experimental data. The MMPBSA.py module in the AMBER package may be used to compute the binding free energy, which incorporates numerous electrostatic energies such as the van der Waals energy, the internal energy accumulated from molecular mechanics, and the polar contribution toward solvation energy. A more precise molecular force field is needed to mimic the motion of atoms in target proteins and ligands in order to further enhance MD simulations. The potential increase in processing complexity, however, prevents simulations lasting more than a microsecond (Chodera et al. 2011).

12.2.6 Discovery of New Binding Sites Aided by Molecular Dynamics

Understanding the numerous protein–ligand interrelationships at play in essential biological processes requires extensive knowledge. It is necessary to identify and characterize LBP in order to get an understanding of the mechanism of action of both endogenous ligands and synthesized therapeutic compounds. Targeting G-protein coupled receptor (GPCR) is common practice when it comes to the creation of novel medications (Kinney et al. 2005). A recent research demonstrated that ligands are attached to multiple allosteric sites other than the intended binding sites, in addition to orthosteric points (DeVree et al. 2016). This discovery was made possible by the fact that orthosteric points are involved. In a recent review, primary computational approaches for predicting functional regions such as 3D ligand sites and others were described. These methods include ligand site prediction and protein structure prediction. Nevertheless, these reporting methods usually provide a large number of possible ligand binding sites, which makes it difficult for the user to evaluate which active pocket of the structure is appropriate for the binding of a chemical or medicine. In recent years, methods based on molecular dynamics (MD) have been utilized to get around this constraint. For example, supervised MD is an excellent approach for ligand-binding site identification as well as accurate sampling (Sabbadin and Moro 2014; Cuzzolin et al. 2016). The MD simulations uncovered an additional sodium ion in the vicinity of the orthosteric binding site (Chan et al. 2020). This information might be put to use to discover allosteric sites in protein

kinases, Ras proteins, and *Staphylococcus aureus* sortase, among other enzymes (Tong and Seeliger 2015).

12.3 Recent Advances in Computational Approaches for Anticancer Drug Discovery

12.3.1 Use of Machine Learning (ML) Algorithms

Each and every machine learning method helps with predictive modeling by uncovering previously unseen connections and patterns in data. We may classify methods used in machine learning as either reinforcement, supervised, unsupervised, or semi-supervised. The primary difference between these strategies is in the amount of data used to train the model. In the fight against cancer, chemists have made substantial use of machine learning, particularly supervised learning (Alberi et al. 2019). In order to train the input data and make approximations about the output, supervised machine learning algorithms require target labels. Conventional supervised learning techniques that employ input space transformation to create a new feature space include artificial neural networks (ANNs) and kernel approaches (Sheng et al. 2015; Rupp et al. 2018). For ANNs, feature modification through many input layers is a crucial component. These approaches, on the other hand, are useful for spotting nonlinear connections in the data. A kernel function is used in kernel-based approaches to conduct nonlinear data changes that may then be used with linear algorithms. Both ANNs and DL algorithms are used extensively in the biomedical and pharmaceutical industries. Microarray and gene expression data are utilized to train machine learning algorithms used to discover new cancer treatments and identify biomarkers. Genetic data-driven research was bolstered by the realization that genes have a role in a variety of cancers. Cancers of different sorts are notoriously difficult to diagnose and cure because of their complex microenvironment. When treated with the same medicine, people with the same kind of tumor might have widely variable results (Sheng et al. 2015). There has been an increase in the popularity of deep learning algorithms, despite the fact that standard machine learning methods are also extremely useful for creating biological computational models. The widespread availability of biological and pharmacogenomic datasets (Barretina et al. 2012; Yang et al. 2012) and high computing devices for parallel processing, such as GPUs, is largely responsible for this dramatic shift.

12.3.2 Drug Repurposing (DR) for Anticancer Drug Discovery

Drug repurposing is a method that consists of discovering new indications for previously recognized marketed medications that are used in a variety of therapeutic

contexts or extensively defined compounds, despite the fact that these treatments may have failed in the past. Recently, it has become as an alternative method for quick discovery of novel medicines for a wide variety of uncommon and difficult illnesses that currently do not have any medication therapies that are shown to be successful. In recent years, the success rate of the pharmaceuticals repurposing strategy has been responsible for roughly 30% of the newly authorized medications and vaccines by the FDA. In this review, the state of the repurposing method for different illnesses, such as skin disorders, infectious diseases, inflammatory diseases, cancer, and neurodegenerative diseases, is the primary emphasis. There have been efforts made to offer the structural characteristics of medications as well as their modes of activities.

Here are structures (Fig. 12.3) and summary of some drugs which have already been repurposed for cancer therapy (Pillaiyar et al. 2020):

- Acetylsalicylic acid, better known by its chemical name aspirin (Fig. 12.3(i)), is a kind of nonsteroidal anti-inflammatory medication (NSAID) used for pain relief, fever reduction, and prevention of cardiovascular disease. It was initially reported by Gasic and colleagues that aspirin could be useful in treating cancer. Antiplatelet activity of 20 was shown to be related with a 50% decrease in lung metastasis in tumor-bearing animals (Elwood et al. 2018). A recent research showed that those who took medication (Fig. 12.3(i)) (75 mg) daily saw a strong protective benefit against developing cancers of the digestive tract, esophagus, pancreas, brain, prostate, and lung. Aspirin's mechanism of action is said to include modulating a large number of molecules that are involved in carcinogenesis (Elwood et al. 2009). When activity of cyclooxygenase enzyme is inhibited, it induces cancer via the manufacture of prostaglandins (PGE2) (Simmons et al. 2004) and has been linked to drug 20's anticancer potential in preclinical investigations. As an added bonus, medication (Fig. 12.3(i)) was shown to block the activation of the apoptosis-related transcription factor NF-kB (Takada et al. 2004). Compound (i) was shown to decrease proliferations and accelerate death of cancer cells, as well as delay and overcome acquired resistance to targeted treatment, according to a research published by Li Ling and team. Increased cancer and increased nuclear factor kappa-light-chain-enhancer of activated B cell (NF-kB) signaling in tumors may account for aspirin's ability to inhibit proliferation, apoptosis, and cancer stemness in resistant tumors to a greater extent than in parental, sensitive cells. However, at the same dosages employed on lung and breast cancer cells, aspirin had no impact on the growth of normal lung and mammary epithelial cells. Consequently, aspirin may be considered for use in chemoradiation treatments for lung and breast cancer (Yu et al. 2019). Despite strong evidence linking medication 20's anticancer properties to benefits such as reduced gastrointestinal and renal toxicity, the medicine's overall usefulness remains limited. Accordingly, it is not advised that the general populace all take compound (i). Nonetheless, numerous studies found that those aged 40–85 would benefit more from adopting 20 as a main cancer prevention measure. Therefore, according to the US Preventive Services Task Force (USPSTF),

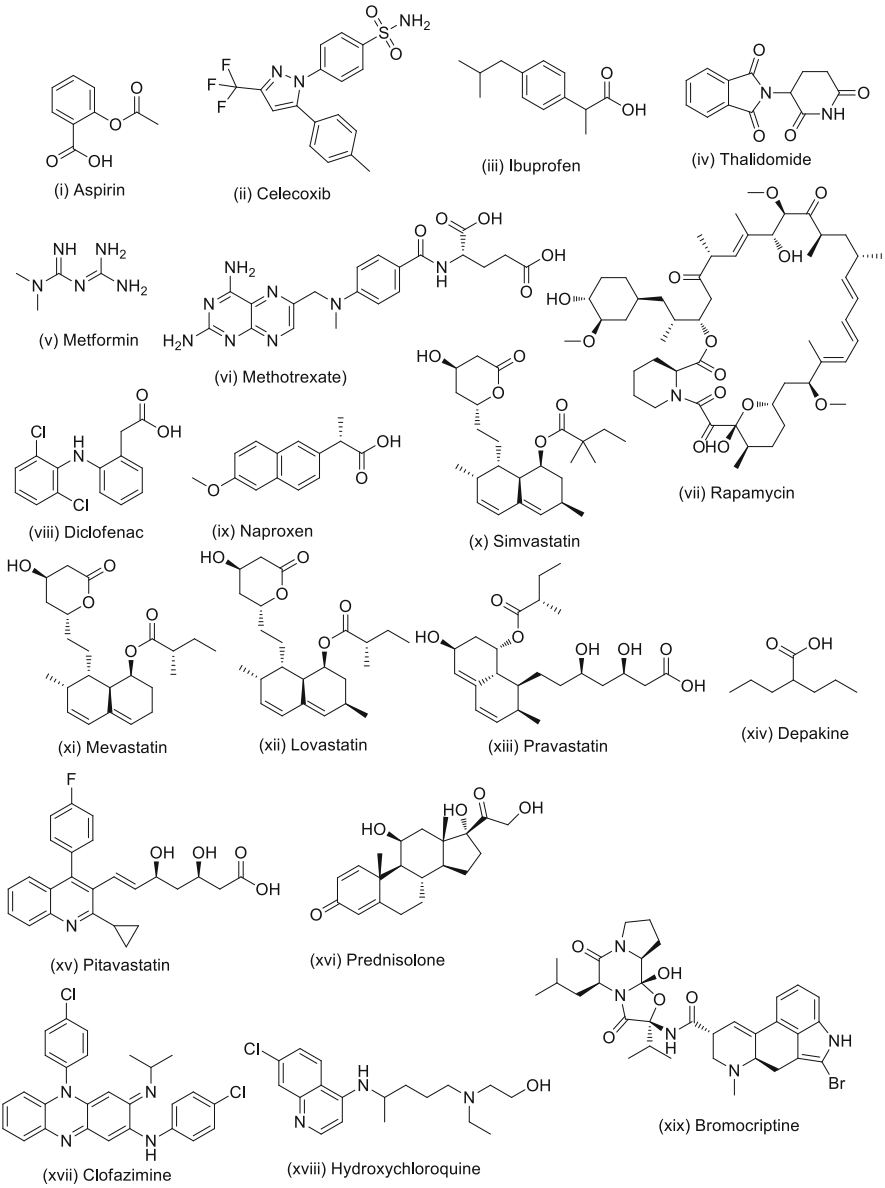


Fig. 12.3 Structures of few drugs that are repurposed as anticancer drugs

people over the age of 40 who have an elevated risk of cardiovascular disease and colorectal cancer should consume 20 mg of folic acid daily.

- The nonsteroidal anti-inflammatory drug family includes celecoxib (Fig. 12.3 (ii)), which has been used to treat rheumatoid arthritis and osteoarthritis pain and inflammation. Celecoxib is a reversible inhibitor of cyclooxygenase-2 (COX-2), a

recognized inflammatory cancer target. It was authorized by the FDA in 1999. Due to its COX-2 inhibitory effect, medication 3(ii)'s anticancer activities have been widely examined, and it has been proven to have chemopreventive properties against several cancer types. In addition to the cyclooxygenase-2 (COX-2) family, celecoxib inhibits the β-catenin, nuclear factor κB (NF-κB), AKT8, and B-cell lymphoma (Bcl-2) families (Jendrossek 2013). Drug 3(ii) (400 mg) was shown to dramatically lower the incidence of colorectal adenomas in patients with familial adenomatous polyposis (FAP). This chemical has been shown to be effective in reducing polyps in the colon and rectal areas in persons with familial adenomatous polyposis (Lynch et al. 2016). On the other hand, it has certain negative side effects, such as on the gut, kidneys, and heart.

- There is a nonsteroidal anti-inflammatory drug (NSAID) called ibuprofen (Fig. 12.3(iii)). Ibuprofen inhibits COX, an enzyme necessary for the production of prostaglandin, at the molecular level. However, it does not discriminate between COX isoforms. Both the United Kingdom (1969) and the United States (1971) promoted the medicine for the treatment of rheumatoid arthritis. Drug's anticancer potential has been studied in a wide range of cancer cell lines. Several studies have shown that ibuprofen may reduce the proliferation of prostate cancer cells (Costea et al. 2019). Drug 3(iii) displayed anticancer effects in gastric adenocarcinoma cells, and these effects were mediated by its ability to inhibit angiogenesis, induce apoptosis, and slow cell proliferation. Drug 3(iii) causes cell death in melanoma cell lines that have spread throughout the body. Ibuprofen has been shown to reduce heat shock protein 70 s (Hsp70s) levels in lung cancer cells, which in turn increases their susceptibility to the chemotherapeutic agent cisplatin. Hsp70s have been linked to apoptosis resistance (Endo et al. 2014) and played a crucial role in the cell's protein folding mechanism. Therefore, because ibuprofen inhibits Hsp70s, it makes cisplatin more effective in inducing apoptosis.
- Immunomodulatory medication thalidomide (Fig. 12.3(iv)) was first created as a sedative-hypnotic to alleviate morning sickness in pregnant women. Since it caused birth defects, however, it was taken off the market. The drug's anti-angiogenic properties were tested to see whether it may be utilized to treat patients with refractory myeloma. The FDA has given the drug 3(iv) the go light to treat multiple myeloma, after positive clinical trials. Acute myeloid leukemia, myelodysplasia, and myelodysplastic syndrome were only a few of the cancers that drug 3(iv) proved effective against (Hiramatsu et al. 2018). The transcription factors Ikaros and Aiolos are rapidly ubiquitinated and degraded by the proteasomal pathway after thalidomide binds to cereblon and forms an E3 ubiquitin ligase complex (Stewart 2014).
- To manage type 2 diabetes, many people turn to metformin (Fig. 12.3(v)), a medicine that may be taken orally. Metformin works by stimulating the enzyme adenosine monophosphate-induced protein kinase, which plays a crucial role in cellular metabolism. Cancer cell survival is affected by the gene rapamycin (mTOR), which is adversely controlled by AMPK. Additionally, metformin may lower mTOR signaling by blocking Rag-mediated mTOR activation

(Kalender et al. 2010). In addition to the higher risk of other cancers, women patients with diabetes have >20% chance of acquiring breast cancer. Drug 3 (v) has been linked in many studies to anticancer effects. Patients with diabetes had a lower chance of developing gastroesophageal cancer when taking 3(v) at a daily dosage of 250–500 mg (Yu et al. 2019). Taking drug 3(v) was linked to a lower risk of death from any cause among diabetic patients, according to a meta-analysis and a number of studies (Noto et al. 2012). According to a recent meta-analysis comparing many diabetes medications, people using drug 3(v) had a 14% lower risk of cancer and a 30% lower death rate. However, there was a correlation between insulin usage and an elevated risk of cancer and death.

- The enzyme dihydrofolate reductase (DHFR) is essential in the production of DNA, RNA, thymidylates, and proteins, and methotrexate (Fig. 12.3(vi)) acts as a competitive inhibitor of DHFR. Drug's anti-leukemia effects may be attributed to its ability to suppress hydrofolate action. In 1988, the Food and Drug Administration (FDA) authorized this chemical for the treatment of osteosarcoma, breast cancer, acute lymphoblastic leukemia, and Hodgkin lymphoma. Methotrexate's anticancer effect may be attributable, at least in part, to the fact that it blocks inflammatory pathways, as has been shown by several studies. For instance, methotrexate has been shown to inhibit NF-kB in cancer cells by causing the release of adenosine (Tabas and Glass 2013).
- The antifungal drug rapamycin (Fig. 12.3(vii)), now more well-known by its generic name sirolimus. However, drug 3(vii) was taken off the market because of its strong anticancer and immunosuppressive properties. Inhibiting mTOR, a protein that is significantly elevated in many tumor cells, is the drug's mechanism of action for suppressing T cells and B cells by making them less sensitive to IL-2. Allograft rejection may be avoided using rapamycin, which the FDA authorized for this purpose in 1999. Since then, this compound's potential cancer-fighting effects have been studied. Patients with acute myeloid leukemia have had less colony formation of leukemia progenitor cells when drug 3(vii) was administered (Alvarado et al. 2011). Moreover, drug 3(vii) demonstrated efficacy in patients with imatinib-resistant chronic myelogenous leukemia by reducing VEGF mRNA levels in leukemia cells while causing minimal adverse effects.
- The acetic acid derivative diclofenac (Fig. 12.3(viii)) belongs to the nonsteroidal anti-inflammatory drug (NSAID) family and has been used to treat pain and inflammatory illnesses including gout. Some researchers hypothesized that by blocking both COX-1 and COX-2, it would reduce prostaglandin synthesis. In 1988, drug 3(viii) was first prescribed in the United States. Several forms of cancer, including hepatoma, colon, fibrosarcoma, pancreatic, and ovarian, have been shown to respond to drug 3(viii)'s anticancer properties. The development rate and degree of vascularization of fibrosarcoma and hepatoma models in rats were dramatically decreased by the drug 3(viii) (Pantziarka et al. 2016). Human colon cancer cell lines were shown to benefit from the antiproliferative effects of the drug 3(viii). Drug 3(viii)'s tumor-inhibiting efficacy was also shown in a model of ovarian cancer (Valle et al. 2013). Diclofenac causes increased reactive oxygen species and so triggers apoptosis by blocking the antioxidant superoxide

dismutase 2 (SOD 2) (Brinkhuizen et al. 2016). Diclofenac was shown to be extremely successful in tumor regression with 64% in a recent Phase II clinical trial research for the treatment of basal cell carcinoma.

- Naproxen (Fig. 12.3(ix)), a nonsteroidal anti-inflammatory drug (NSAID) that belongs to the propionic acid family and is used to treat pain, inflammation, fever, and conditions including rheumatoid arthritis. This medication blocks prostaglandin formation by nonselectively inhibiting cyclooxygenase-1 and -2 enzymes. Drug 3(ix) was first commercialized in the United States for therapeutic use in 1976. Inhibiting cell growth, inducing apoptosis, and suppressing metastasis in several kinds of cancer cells have led to drug 3(ix)'s recent repurposing for its anticancer activities. Drug 3(ix) inhibits PI3K and promotes cell-cycle arrest and death in human urinary bladder cancer cell lines (Kim et al. 2014). When given together, atorvastatin and naproxen dramatically reduced colonic adenocarcinomas in vivo in rats (Suh et al. 2011). For recurrent prostate cancer, it has been studied in a phase II clinical study in conjunction with calcitriol. The findings demonstrated that the combination was safe and well-accepted by the patients. Drugs belonging to the statin-family reduce cholesterol levels by blocking the enzyme 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase. Statins are routinely recommended to people with a high risk of cardiovascular disease in order to decrease cholesterol production. In addition, statins block the production of the compounds mevalonate, farnesyl, and geranyl pyrophosphate, which are all used in cholesterol synthesis. Statins are attractive prospects as cancer therapies because of their ability to inhibit the activity of these molecules, which are essential for cell cycle progression and cell proliferation. Simvastatin (Fig. 12.3(x)), along with other natural statins such as mevastatin (Fig. 12.3(xi)), lovastatin (Fig. 12.3(xii)), and pravastatin (Fig. 12.3(xiii)), showed TNF-induced apoptosis in chronic myeloid leukemia cells through the downregulation of NF-kB-mediated antiapoptotic gene products (Aggarwal et al. 2019). Statins have been studied for their potential anticancer effects, and these effects have been confirmed in animal tests (Li et al. 2019b). Multiple observational studies and a meta-analysis confirm a link between statin use and a decreased risk of cancer in people. Patients using statins had a lower chance of developing gastric cancer (Broughton et al. 2012), esophageal cancer (Ibáñez-Sanz et al. 2019), and hepatocarcinoma (Sehdev et al. 2014), according to meta-analyses. Drug 3(x) substantially decreased the incidence of colorectal cancer in a case-control research when administered at a dose of 40 mg/day for 2–5 years (Broughton et al. 2012).
- Valproic acid, also known as depakine (Fig. 12.3(xiv)), is a short-chain free fatty acid used in first-line treatment of epilepsy, bipolar disorder, and migraine headaches. Blocking voltage-gated sodium channels and elevating gamma aminobutyric acid (GABA) levels are thought to be responsible for its anticonvulsant effect. Drug 3(xiv) was first demonstrated to inhibit histone deacetylase (HDAC) (Rocha et al. 2019), which paved the way for its validation as an anticancer agent in leukemia cells. Researchers have discovered that depakine may inhibit cytokine production and alter the dynamics of the inflammatory

signaling cascade. Human glioma and leukemia cells were treated with drug 3 (xiv), and the results showed a reduction in IL-6 and TNF-a production (Soria-Castro et al. 2019). By downregulating NF-kB activity, drug 3(xiv) blocked IL-6 production in prostate cancer cells (Martin et al. 2015).

- Recently, it was reported (Abdullah et al. 2019) that pitavastatin (Fig. 12.3(xv)) may be useful in the treatment of ovarian cancer, but only if geranylgeraniol intake in the diet is limited. Myopathy is the most prevalent side effect of statins, yet rather large dosages are needed to trigger apoptosis in cancer cells. Since this is problematic, it is important to find alternatives to pitavastatin that might lessen the drug's side effects. An approach based on re-categorizing existing drugs was used to find promising leads. Prednisolone (Fig. 12.3(xvi)) was the most significant result when a library of 100 off-patent were screened (Astual paper) for synergistic action with pitavastatin. Several experiments evaluating proliferation, survival, or apoptosis in different ovarian cancer cell lines found that prednisolone enhanced the efficacy of pitavastatin. Pitavastatin suppressed the development of the examined cell lines with an IC_{50} of 1.1 to 4.8 mM. So, testing the combination of prednisolone and pitavastatin in patients with ovarian cancer may be necessary. Clofazimine (Fig. 12.3(xvii)), an anti-leprosy medication, has recently been shown to be effective against triple-negative breast cancer (TNBC) (Ahmed et al. 2019), thanks to a large body of preclinical data compiled by Ahmed and team. The canonical Wnt signaling pathway is selectively inhibited by clofazimine in an in vitro panel of TNBC cells. In HEK293T and BT-20 cells, for instance, the IC_{50} was 6 mM and 7 mM. Clofazimine inhibits the Wnt pathway in tumors in vivo, and its efficacy in suppressing tumor development has been shown in many mice xenograft models of TNBC. Clofazimine's addition to doxorubicin has been shown to have a synergistic impact on tumor growth reduction without causing any unwanted side effects. Clofazimine's new molecular method of action, together with its good and well-characterized pharmacokinetics profile, makes it an attractive candidate for the repositioning clinical trials. The incurable hematological malignancy known as multiple myeloma (MM, also known as plasma cell myeloma, is a cancer of plasma cells) is caused by a number of mutations and epigenetic changes.
- Hydroxychloroquine (Fig. 12.3(xviii)), a chloroquine derivative and anti-malarial drug also known as an autophagy inhibitor, was recently shown to have antitumor potential, as reported by Raffaella and his research group. New treatment techniques are required to increase survival rates for patients with acute myeloid leukemia (AML) due to the lack of considerable progress in the field over the last few decades (Ferrara and Schiffer 2013).
- Bromocriptine (Fig. 12.3(xix)), a dopamine agonist derived from ergoline, is now being utilized to treat diabetes mellitus (Murteira et al. 2013). Bromocriptine has been found to be a powerful anticancer medication that primarily targets leukemia stem cells according to a repurposing technique conducted by Lara-Castillo and coworkers (Lara-Castillo et al. 2016). Myeloid differentiation was induced by treatment with drug 3(xix), and the apoptotic program was activated, resulting in a decrease in the viability of AML cells. Additionally, the percentage of primitive

AML cells enriched in LSCs was more responsive to bromocriptine. Indeed, bromocriptine inhibited the ability of AML cells to form clones. Interestingly, normal blood cells and hematopoietic stem/progenitor cells show almost little change.

12.3.3 Pseudoreceptor Modeling

The SB-CADD and LB-CADD modeling approaches may be combined in pseudoreceptor models by employing surrogate three-dimensional receptor structures. These architectures change the form and volume of the binding region, as well as the essential interaction parameters between the ligand and the receptor. In order to guarantee correct binding, it is necessary to do research into both aspects of the interaction between the receptor and the ligand. Experimentation, in the form of things like mutation studies, has to be done in order to effectively include the bioactive conformation of these medications into the models. Tanrikulu's review (Tanrikulu and Schneider 2008; Wilson and Lill 2011) discusses a broad variety of methodologies, such as those that are based on partitions, grids, peptides, isosurfaces, atomic levels, fragment levels, and more. A significant number of computer studies relating to cancer have incorporated pseudoreceptor modeling. For illustration purposes, Rodl and coworkers developed a model of a 5-lipoxygenase (5-LO) pseudoreceptor by using reference structures that were obtained from molecules that were discovered by use similarity search algorithms. Experiments using VS showed promising inhibitors that did not have an effect on cell viability, and these compounds have the potential to serve as the basis for additional work to optimize lead compounds. This allowed for the binding interaction pattern to be established (Pei et al. 2001; Peng et al. 2003; Rödl et al. 2011; Lee et al. 2012). It is also important to keep in mind that pseudoreceptors may not always have a binding pocket that is an exact match for the anticipated receptor. This is something that must be taken into consideration. Because this framework is built on a collection of reference molecules, it is probable that it will favor compounds that have the same structure as other molecules (Basith et al. 2018).

12.3.4 Proteochemometric Modeling

Lapinsh and coworkers used the word "proteochemometrics" to describe the methodology they had recently established for analyzing data on interactions between receptors and ligands. Chimeric ligands and receptors have been studied for their binding characteristics (Lapinsh et al. 2001). This capability, in contrast to portraying the key linkages between a single ligand and a single receptor, may be utilized concurrently to depict the interactive relation of a number of ligands with a number of proteins. It is useful for explaining the interplay between receptors and

their ligands. As a result, this method may be used to infer relationships between related QSAR datasets. In addition, the created proteochemometric modeling may be used to other related series (van Westen et al. 2011). For example, Wu and coworkers' used proteochemometric modeling to screen selective histone deacetylase inhibitors as a prototypical example for the use of this approach in the search for cancer medicines. The inhibitory effects of each HDAC inhibitor could be correctly predicted using the fit and robust model used for design. The isoform of the drug might also be assessed, which can get us to the identification of leads with fewer negative side effects (Wu et al. 2012). Recent research has connected COX-2 to colorectal cancer, suggesting its inhibition as a strategy for developing anticancer drugs (Eberhart et al. 1994; Wang and DuBois 2010; Xu et al. 2014).

Case Study

- In the work that Wu and colleagues published in 2012 (Wu et al. 2012), they made use of proteochemometric modeling to test a large number of chemicals to see whether or not they were able to selectively block histone deacetylase (HDAC). During the course of that investigation, multiplication cross-terms were constructed.
- During the course of that study, a total of 18 different proteochemometric models were developed to make predictions about protein-protein interactions using the data from the training set. All of the models exhibited goodness-of-fits (R^2) that were more than 0.9619 and cross-validation coefficients (Q^2_{cv}) that ranged from 0.573 to 0.7162. The model that was created utilizing P1 and GD had the best predictive performance ($Q^2_{cv} = 0.7162$ and $Q^2_{test} = 0.7542$), respectively. The ensuing study was conducted in a manner that was consistent with the P1-GD model.
- The HDACs were characterized in this study by employing three different types of protein descriptors: the sequence similarity descriptor (P0), the structural similarity descriptor (P1), and the geometry descriptor (P2). The sequence identities of HDACs are used as the basis for the sequence similarity descriptor, while the structure similarity descriptor and the geometry descriptor use HDACs' three-dimensional structures as the basis for their definitions. Proteins are characterized by descriptors that are distinct from those used to describe ligands because proteins have more intricate molecular structures.
- Across the board, the predictive performance of the models that were based on the geometry descriptor was the worst. This was determined by using the Q^2_{test}. The General Descriptor (GD) and the Drug-Like Index (DLI) are both examples of typical ligand descriptors that were used here. These are similar to the descriptors used for proteins. According to the findings of our paired t-test, there is not a statistically significant difference between the Q^2 values predicted by models constructed using GD and DLI and shown by a p-value that is more than 0.1. Both the ligand's physical characteristics and its topological indices are taken into consideration while describing it. In our dataset, the predictive ability of these two ligand descriptors did not vary in a way that was statistically significant from one another.

- Their models made use of a multiplied cross-term and revealed the ineffectiveness of this method non-terms of boosting the PCM models' ability to predict outcomes. Models that included cross-terms performed worse on the Q^2_{test} in every category than models that did not include them.
- During the course of that study, a total of 18 different proteochemometric models were developed by making use of the training data in order to make predictions about protein-protein interactions. All of the models exhibited goodness-of-fits (R^2) that were more than 0.9619 and cross-validation coefficients (Q^2_{cv}) that ranged from 0.573 to 0.7162. The model that was created utilizing P1 and GD had the best predictive performance ($Q^2_{\text{cv}} = 0.7162$ and $Q^2_{\text{test}} = 0.7542$), respectively. The inquiry that followed was conducted in a manner consistent with the P1-GD paradigm.

Conclusion

- There have been an increasing number of discoveries of HDAC inhibitors; nonetheless, there is still a scarcity of inhibitors that are selective for a particular class or isoform. In light of this, it was very important to locate particular inhibitors that have the potential to be used as medications for the treatment of tumors that have a low level of toxicity. They used proteochemometric models in order to analyze the inhibitory impact that 1275 different compounds had on 5 different HDAC isoforms.
- The P1-GD model beats the others because of its exceptional predictive strength ($Q^2_{\text{test}} = 0.7542$), as well as its strong ability to distinguish selective HDAC inhibitors from pan inhibitors. Proteochemometric modeling may thus be used to predict how inhibitors will interact with the various isoforms of HDAC. The study group also uncovered evidence that the optimal model that they established might potentially be used to construct anticancer pharmaceutical candidates with the capacity to precisely target a single HDAC or a set of HDAC isoforms. This ability was discovered by the research team.

12.4 Conclusion and Future Perspectives

The spread of cancer is dangerous to people's health. Annually, 9.6 million individuals are afflicted by this illness. Cancer is becoming the leading cause of mortality, surpassing even heart disease. It now takes 12 years and $2.7 billion for modern anticancer medicines. Inadequate knowledge of cancer pathways has made it challenging to discover new, effective therapies for cancer. The creation of a new drug takes a long time and a lot of money. Protein-association network analysis, drug target prediction, restriction site prediction, and virtual screening are only some of the potential uses of computational methods in drug development. These methods have the potential to aid in the development of more efficient cancer treatments. AI has made retro-manufactured routine arrangement, medicine framework age, and

medication controlling fondness expectancies more important. Technology and computational models may help find new cancer therapies more quickly. How to find drugs that can combat cancer was just discussed. Predictions about the efficacy of different medication combinations, optimal drug placement, and targeted cancer therapy were all considered. Multi-omics data give great prospects for precise and inexpensive anticancer medication development. There are still issues in the actual world, despite progress in constructing prediction algorithms. PPI networks and biological pathways are common subjects of the many computerized studies that have already been conducted. Sparse data hampers the reliability and precision of algorithmic forecasts. Context-specific medicine response prediction therapies require greater investigation. Prediction models are built on pan-cancer data studies, which do not take into account the specifics of the disease being studied or the treatment being administered. Predictions of cancer outcomes are heavily influenced by the fact that different tumor types have distinctive molecular profiles of cancer cell lines. Transcriptome profiles, which are molecular profiles, are used in most predictive drug development methods for cancer cell lines. In humans, cancer cell lines cannot recapitulate underlying molecular defects. For therapeutically relevant research, bioinformaticians should be familiar with cell line boundaries. If you choose a sparser data type, like CMap-based models, where data is only accessible for a small number of cell lines across a limited set of tissue types, your model will be less scalable and less useful in clinical settings. There is a lack of appropriate cell lines for modeling response in certain malignancies, and the use of cell lines that do not adequately mimic tumor biology is a major contributor to the failure of computational drug development. The following methods may be used to get around the limitations of computational drug discovery. The clinical usefulness of computer prediction models may be improved by using data formats similar to in vitro patients. Combining diverse information may construct more accurate forecasting models. Clinical data and therapeutically relevant animal models should be used to validate models. Having a doctor involved in the process of using clinical data to predict cancer drugs for therapeutic purposes may improve the chances of success.

References

Abdullah MI, Abed MN, Khanim F, Richardson A (2019) Screening a library of approved drugs reveals that prednisolone synergizes with pitavastatin to induce ovarian cancer cell death. Sci Rep 9:9632. https://doi.org/10.1038/s41598-019-46102-1

Aggarwal V, Tuli H, Varol A et al (2019) Role of reactive oxygen species in cancer progression: molecular mechanisms and recent advancements. Biomol Ther 9:735. https://doi.org/10.3390/biom9110735

Ahmed K, Koval A, Xu J et al (2019) Towards the first targeted therapy for triple-negative breast cancer: repositioning of clofazimine as a chemotherapy-compatible selective Wnt pathway inhibitor. Cancer Lett 449:45–55. https://doi.org/10.1016/j.canlet.2019.02.018

Alberi K, Nardelli MB, Zakutayev A et al (2019) The 2019 materials by design roadmap. J Phys D Appl Phys 52:13001. https://doi.org/10.1088/1361-6463/aad926

Alvarado Y, Mita MM, Vemulapalli S et al (2011) Clinical activity of mammalian target of rapamycin inhibitors in solid tumors. Target Oncol 6:69–94. https://doi.org/10.1007/s11523-011-0178-5

Ambure P, Halder AK, González Díaz H, Cordeiro MNDS (2019) QSAR-co: an open source software for developing robust multitasking or multitarget classification-based QSAR models. J Chem Inf Model 59:2538–2544. https://doi.org/10.1021/acs.jcim.9b00295

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

Barretina J, Caponigro G, Stransky N et al (2012) The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483:603–607. https://doi.org/10.1038/nature11003

Basith S, Cui M, Macalino SJY, Choi S (2018) Expediting the design, discovery and development of anticancer drugs using computational approaches. Curr Med Chem 24:4753–4778. https://doi.org/10.2174/0929867323666160902160535

Brinkhuizen T, Frencken KJA, Nelemans PJ et al (2016) The effect of topical diclofenac 3% and calcitriol 3 μg/g on superficial basal cell carcinoma (sBCC) and nodular basal cell carcinoma (nBCC): a phase II, randomized controlled trial. J Am Acad Dermatol 75:126–134. https://doi.org/10.1016/j.jaad.2016.01.050

Broughton T, Sington J, Beales ILP (2012) Statin use is associated with a reduced incidence of colorectal cancer: a colonoscopy-controlled case–control study. BMC Gastroenterol 12:36. https://doi.org/10.1186/1471-230X-12-36

Brown BP, Mendenhall J, Geanes AR, Meiler J (2021) General purpose structure-based drug discovery neural network score functions with human-interpretable pharmacophore maps. J Chem Inf Model 61:603–620. https://doi.org/10.1021/acs.jcim.0c01001

Chan HCS, Xu Y, Tan L et al (2020) Enhancing the signaling of GPCRs via orthosteric ions. ACS Cent Sci 6:274–282. https://doi.org/10.1021/acscentsci.9b01247

Chen CY-C (2011) TCM database@Taiwan: the World's largest traditional Chinese medicine database for drug screening in silico. PLoS One 6:e15939–e15939. https://doi.org/10.1371/journal.pone.0015939

Chen J, Lai L (2006) Pocket v.2: further developments on receptor-based pharmacophore modeling. J Chem Inf Model 46:2684–2691. https://doi.org/10.1021/ci600246s

Cheng TM-K, Blundell TL, Fernandez-Recio J (2007) pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. Proteins 68:503–515. https://doi.org/10.1002/prot.21419

Chodera JD, Mobley DL, Shirts MR et al (2011) Alchemical free energy methods for drug discovery: progress and challenges. Curr Opin Struct Biol 21:150–160. https://doi.org/10.1016/j.sbi.2011.01.011

Christen M, Hünenberger PH, Bakowies D et al (2005) The GROMOS software for biomolecular simulation: GROMOS05. J Comput Chem 26:1719–1751. https://doi.org/10.1002/jcc.20303

Ciemny M, Kurcinski M, Kamel K et al (2018) Protein–peptide docking: opportunities and challenges. Drug Discov Today 23:1530–1537. https://doi.org/10.1016/j.drudis.2018.05.006

Costea T, Nagy P, Ganea C et al (2019) Molecular mechanisms and bioavailability of polyphenols in prostate cancer. Int J Mol Sci 20:1062. https://doi.org/10.3390/ijms20051062

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

Cui W, Aouidate A, Wang S et al (2020) Discovering anti-cancer drugs via computational methods. Front Pharmacol 11:733. https://doi.org/10.3389/fphar.2020.00733

Cuzzolin A, Sturlese M, Deganutti G et al (2016) Deciphering the complexity of ligand–protein recognition pathways using supervised molecular dynamics (SuMD) simulations. J Chem Inf Model 56:687–705. https://doi.org/10.1021/acs.jcim.5b00702

Davis AP, Grondin CJ, Johnson RJ et al (2021) Comparative toxicogenomics database (CTD): update 2021. Nucleic Acids Res 49:D1138–D1143. https://doi.org/10.1093/nar/gkaa891

DeVree BT, Mahoney JP, Vélez-Ruiz GA et al (2016) Allosteric coupling from G protein to the agonist-binding pocket in GPCRs. Nature 535:182–186. https://doi.org/10.1038/nature18324

Dias R, de Azevedo W (2008) Molecular docking algorithms. Curr Drug Targets 9:1040–1047. https://doi.org/10.2174/138945008786949432

Dimitrov SD, Diderich R, Sobanski T et al (2016) QSAR Toolbox – workflow and major functionalities. SAR QSAR Environ Res 27:203–219. https://doi.org/10.1080/1062936X.2015.1136680

Dominguez C, Boelens R, Bonvin AMJJ (2003) HADDOCK: a protein−protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125:1731–1737. https://doi.org/10.1021/ja026939x

Dunkel M, Gunther S, Ahmed J et al (2008) SuperPred: drug classification and target prediction. Nucleic Acids Res 36:W55–W59. https://doi.org/10.1093/nar/gkn307

Eberhart CE, Coffey RJ, Radhika A et al (1994) Up-regulation of cyclooxygenase 2 gene expression in human colorectal adenomas and adenocarcinomas. Gastroenterology 107:1183–1188. https://doi.org/10.1016/0016-5085(94)90246-1

Elwood PC, Gallagher AM, Duthie GG et al (2009) Aspirin, salicylates, and cancer. Lancet 373:1301–1309. https://doi.org/10.1016/S0140-6736(09)60243-9

Elwood PC, Pickering JE, Morgan G et al (2018) Systematic review update of observational studies further supports aspirin role in cancer treatment: time to share evidence and decision-making with patients? PLoS One 13:e0203957–e0203957. https://doi.org/10.1371/journal.pone.0203957

Endo H, Yano M, Okumura Y, Kido H (2014) Ibuprofen enhances the anticancer activity of cisplatin in lung cancer cells by inhibiting the heat shock protein 70. Cell Death Dis 5:e1027–e1027. https://doi.org/10.1038/cddis.2013.550

Ferrara F, Schiffer CA (2013) Acute myeloid leukaemia in adults. Lancet 381:484–495. https://doi.org/10.1016/S0140-6736(12)61727-9

Ferreira L, dos Santos R, Oliva G, Andricopulo A (2015) Molecular docking and structure-based drug design strategies. Molecules 20:13384–13421. https://doi.org/10.3390/molecules200713384

Friesner RA, Banks JL, Murphy RB 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

Gfeller D, Grosdidier A, Wirth M et al (2014) SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res 42:W32–W38. https://doi.org/10.1093/nar/gku293

Grant BJ, Lukman S, Hocker HJ et al (2011) Novel allosteric sites on ras for lead generation. PLoS One 6:e25711–e25711. https://doi.org/10.1371/journal.pone.0025711

Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins Struct Funct Genet 47:409–443. https://doi.org/10.1002/prot.10115

Hansch C, Fujita T (1964) Additions and corrections - ρ-σ-π analysis. A method for the correlation of biological activity and chemical structure. J Am Chem Soc 86:5710. https://doi.org/10.1021/ja01078a623

Hansson T, Oostenbrink C, van Gunsteren W (2002) Molecular dynamics simulations. Curr Opin Struct Biol 12:190–196. https://doi.org/10.1016/S0959-440X(02)00308-1

Henninot A, Collins JC, Nuss JM (2018) The current state of peptide drug discovery: back to the future? J Med Chem 61:1382–1414. https://doi.org/10.1021/acs.jmedchem.7b00318

Hiramatsu T, Yoshizawa J, Miyaguni K et al (2018) Thalidomide potentiates etoposide-induced apoptosis in murine neuroblastoma through suppression of NF-κB activation. Pediatr Surg Int 34:443–450. https://doi.org/10.1007/s00383-018-4234-4

Ibáñez-Sanz G, Guinó E, Pontes C et al (2019) Statin use and the risk of colorectal cancer in a population-based electronic health records study. Sci Rep 9:13560. https://doi.org/10.1038/s41598-019-49877-5

Irwin JJ, Tang KG, Young J et al (2020) ZINC20—a free ultralarge-scale chemical database for ligand discovery. J Chem Inf Model 60:6065–6073. https://doi.org/10.1021/acs.jcim.0c00675

Jain AN (2007) Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. J Comput Aided Mol Des 21:281–306. https://doi.org/10.1007/s10822-007-9114-2

Jendrossek V (2013) Targeting apoptosis pathways by celecoxib in cancer. Cancer Lett 332:313–324. https://doi.org/10.1016/j.canlet.2011.01.012

Jing P, Zhao S, Ruan S et al (2014) Quantitative studies on structure–ORAC relationships of anthocyanins from eggplant and radish using 3D-QSAR. Food Chem 145:365–371. https://doi.org/10.1016/j.foodchem.2013.08.082

Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 29:1859–1865. https://doi.org/10.1002/jcc.20945

Joy S, Nair PS, Hariharan R, Pillai MR (2006) Detailed comparison of the protein-ligand docking efficiencies of GOLD, a commercial package and ArgusLab, a licensable freeware. In Silico Biol 6:601–605

Kalender A, Selvaraj A, Kim SY et al (2010) Metformin, independent of AMPK, inhibits mTORC1 in a rag GTPase-dependent manner. Cell Metab 11:390–401. https://doi.org/10.1016/j.cmet.2010.03.014

Kim M-S, Kim J-E, Lim DY et al (2014) Naproxen induces cell-cycle arrest and apoptosis in human urinary bladder cancer cell lines and chemically induced cancers by targeting PI3K. Cancer Prev Res 7:236–245. https://doi.org/10.1158/1940-6207.CAPR-13-0288

Kim S, Chen J, Cheng T et al (2019) PubChem 2019 update: improved access to chemical data. Nucleic Acids Res 47:D1102–D1109. https://doi.org/10.1093/nar/gky1033

Kim S, Chen J, Cheng T et al (2021) PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res 49:D1388–D1395. https://doi.org/10.1093/nar/gkaa971

Kinney GG, O'Brien JA, Lemaire W et al (2005) A novel selective positive allosteric modulator of metabotropic glutamate receptor subtype 5 has in vivo activity and antipsychotic-like effects in rat behavioral models. J Pharmacol Exp Ther 313:199–206. https://doi.org/10.1124/jpet.104.079244

Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 37:4130–4146. https://doi.org/10.1021/jm00050a010

Kong W, Midena G, Chen Y et al (2022) Systematic review of computational methods for drug combination prediction. Comput Struct Biotechnol J 20:2807–2814. https://doi.org/10.1016/j.csbj.2022.05.055

Kozakov D, Hall DR, Xia B et al (2017) The ClusPro web server for protein–protein docking. Nat Protoc 12:255–278. https://doi.org/10.1038/nprot.2016.169

Kramer B, Rarey M, Lengauer T (1999) Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins Struct Funct Genet 37:228–241. https://doi.org/10.1002/(SICI)1097-0134(19991101)37:2<228::AID-PROT8>3.0.CO;2-8

Kuntz ID, Blaney JM, Oatley SJ et al (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161:269–288. https://doi.org/10.1016/0022-2836(82)90153-X

Lammi C, Sgrignani J, Arnoldi A, Grazioso G (2019) Biological characterization of computationally designed analogs of peptide TVFTSWEEYLDWV (Pep2-8) with increased PCSK9 antagonistic activity. Sci Rep 9:2343. https://doi.org/10.1038/s41598-018-35819-0

Lapinsh M, Prusis P, Gutcaits A et al (2001) Development of proteo-chemometrics: a novel technology for the analysis of drug-receptor interactions. Biochim Biophys Acta Gen Subj 1525:180–190. https://doi.org/10.1016/S0304-4165(00)00187-2

Lara-Castillo MC, Cornet-Masana JM, Etxabe A et al (2016) Repositioning of bromocriptine for treatment of acute myeloid leukemia. J Transl Med 14:261. https://doi.org/10.1186/s12967-016-1007-5

Lee I-J, Coffman VC, Wu J-Q (2012) Contractile-ring assembly in fission yeast cytokinesis: recent advances and new perspectives. Cytoskeleton 69:751–763. https://doi.org/10.1002/cm.21052

Li K, Du Y, Li L, Wei D-Q (2019a) Bioinformatics approaches for anti-cancer drug discovery. Curr Drug Targets 21:3–17. https://doi.org/10.2174/1389450120666190923162203

Li Y, He X, Ding Y et al (2019b) Statin uses and mortality in colorectal cancer patients: an updated systematic review and meta-analysis. Cancer Med 8:3305–3313. https://doi.org/10.1002/cam4.2151

Lu X, Yang H, Chen Y et al (2018) The development of pharmacophore modeling: generation and recent applications in drug discovery. Curr Pharm Des 24:3424–3439. https://doi.org/10.2174/1381612824666180810162944

Lynch PM, Burke CA, Phillips R et al (2016) An international randomised trial of celecoxib versus celecoxib plus difluoromethylornithine in patients with familial adenomatous polyposis. Gut 65:286–295. https://doi.org/10.1136/gutjnl-2014-307235

Martin YC, Kofron JL, Traphagen LM (2002) Do structurally similar molecules have similar biological activity? J Med Chem 45:4350–4358. https://doi.org/10.1021/jm020155c

Martin F, Ufodiama C, Watt I et al (2015) Therapeutic value of voltage-gated sodium channel inhibitors in breast, colorectal, and prostate cancer: a systematic review. Front Pharmacol 6:273. https://doi.org/10.3389/fphar.2015.00273

McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267:585–590. https://doi.org/10.1038/267585a0

Mendenhall J, Meiler J (2016) Improving quantitative structure–activity relationship models using artificial neural networks trained with dropout. J Comput Aided Mol Des 30:177–189. https://doi.org/10.1007/s10822-016-9895-2

Morris GM, Huey R, Lindstrom W et al (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30:2785–2791. https://doi.org/10.1002/jcc.21256

Murteira S, Ghezaiel Z, Karray S, Lamure M (2013) Drug reformulations and repositioning in pharmaceutical industry and its impact on market access: reassessment of nomenclature. J Mark Access Health Policy 1:21131. https://doi.org/10.3402/jmahp.v1i0.21131

Nair PC, Malde AK, Drinkwater N, Mark AE (2012) Missing fragments: detecting cooperative binding in fragment-based drug design. ACS Med Chem Lett 3:322–326. https://doi.org/10.1021/ml300015u

Noto H, Goto A, Tsujimoto T, Noda M (2012) Cancer risk in diabetic patients treated with metformin: a systematic review and meta-analysis. PLoS One 7:e33411–e33411. https://doi.org/10.1371/journal.pone.0033411

Ochoa D, Hercules A, Carmona M et al (2021) Open targets platform: supporting systematic drug–target identification and prioritisation. Nucleic Acids Res 49:D1302–D1310. https://doi.org/10.1093/nar/gkaa1027

Pantziarka P, Sukhatme V, Bouche G et al (2016) Repurposing drugs in oncology (ReDO)—diclofenac as an anti-cancer agent. Ecancermedicalscience 10:610. https://doi.org/10.3332/ecancer.2016.610

Patel TK, Adhikari N, Amin SA et al (2021) Small molecule drug conjugates (SMDCs): an emerging strategy for anticancer drug design and discovery. New J Chem 45:5291–5321. https://doi.org/10.1039/D0NJ04134C

Pearlman DA, Case DA, Caldwell JW 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

Pei J, Zhou J, Xie G et al (2001) PARM: a practical utility for drug design. J Mol Graph Model 19:448–454. https://doi.org/10.1016/S1093-3263(00)00104-2

Pence HE, Williams A (2010) ChemSpider: an online chemical information resource. J Chem Educ 87:1123–1124. https://doi.org/10.1021/ed100697w

Peng T, Pei J, Zhou J (2003) 3D-QSAR and receptor modeling of tyrosine kinase inhibitors with flexible atom receptor model (FLARM). J Chem Inf Comput Sci 43:298–303. https://doi.org/10.1021/ci0256034

Peón A, Li H, Ghislat G et al (2019) MolTarPred: a web tool for comprehensive target prediction with reliability estimation. Chem Biol Drug Des 94:1390–1401. https://doi.org/10.1111/cbdd.13516

Pillaiyar T, Meenakshisundaram S, Manickam M, Sankaranarayanan M (2020) A medicinal chemistry perspective of drug repositioning: recent advances and challenges in drug discovery. Eur J Med Chem 195:112275. https://doi.org/10.1016/j.ejmech.2020.112275

Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J et al (2019) The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res 48:D845–D855. https://doi.org/10.1093/nar/gkz1021

Piotrowska Z, Costa DB, Oxnard GR et al (2018) Activity of the Hsp90 inhibitor luminespib among non-small-cell lung cancers harboring EGFR exon 20 insertions. Ann Oncol 29:2092–2097. https://doi.org/10.1093/annonc/mdy336

Prada-Gracia D, Huerta-Yépez S, Moreno-Vargas LM (2016) Application of computational methods for anticancer drug discovery, design, and optimization. Bol Med Hosp Infant Mex 73:411–423. https://doi.org/10.1016/j.bmhimx.2016.10.006

Rastelli G, Rio A del, Degliesposti G, Sgobba M (2009) Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. J Comput Chem 31. https://doi.org/10.1002/jcc.21372, 810

Rocha MA, Veronezi GMB, Felisbino MB et al (2019) Sodium valproate and 5-aza-2′--deoxycytidine differentially modulate DNA demethylation in G1 phase-arrested and proliferative HeLa cells. Sci Rep 9:18236. https://doi.org/10.1038/s41598-019-54848-x

Rödl CB, Tanrikulu Y, Wisniewska JM et al (2011) Potent inhibitors of 5-lipoxygenase identified using pseudoreceptors. ChemMedChem 6:1001–1005. https://doi.org/10.1002/cmdc.201100059

Rosenblum D, Joshi N, Tao W et al (2018) Progress and challenges towards targeted delivery of cancer therapeutics. Nat Commun 9:1410. https://doi.org/10.1038/s41467-018-03705-y

Rouillard AD, Gundersen GW, Fernandez NF et al (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database 2016:baw100. https://doi.org/10.1093/database/baw100

Rupp M, von Lilienfeld OA, Burke K (2018) Guest editorial: special topic on data-enabled theoretical chemistry. J Chem Phys 148:241401. https://doi.org/10.1063/1.5043213

Sabbadin D, Moro S (2014) Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR–ligand recognition pathway in a nanosecond time scale. J Chem Inf Model 54:372–376. https://doi.org/10.1021/ci400766b

Salmaso V, Moro S (2018) Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview. Front Pharmacol 9:923. https://doi.org/10.3389/fphar.2018.00923

Salmaso V, Sturlese M, Cuzzolin A, Moro S (2017) Exploring protein-peptide recognition pathways using a supervised molecular dynamics approach. Structure 25:655–662.e2. https://doi.org/10.1016/j.str.2017.02.009

Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res 33:W363–W367. https://doi.org/10.1093/nar/gki481

Sehdev A, Shih Y-CT, Huo D et al (2014) The role of statins for primary prevention in non-elderly colorectal cancer patients. Anticancer Res 34:5043–5050

Sheng J, Li F, Wong STC (2015) Optimal drug prediction from personal genomics profiles. IEEE J Biomed Health Inform 19:1264–1270. https://doi.org/10.1109/JBHI.2015.2412522

Simmons DL, Botting RM, Hla T (2004) Cyclooxygenase isozymes: the biology of prostaglandin synthesis and inhibition. Pharmacol Rev 56:387–437. https://doi.org/10.1124/pr.56.3.3

Soerjomataram I, Bray F (2021) Planning for tomorrow: global cancer incidence and the role of prevention 2020–2070. Nat Rev Clin Oncol 18:663–672. https://doi.org/10.1038/s41571-021-00514-z

Solis-Vasquez L, Tillack AF, Santos-Martins D et al (2022) Benchmarking the performance of irregular computations in AutoDock-GPU molecular docking. Parallel Comput 109:102861. https://doi.org/10.1016/j.parco.2021.102861

Soria-Castro R, Schcolnik-Cabrera A, Rodríguez-López G et al (2019) Exploring the drug repurposing versatility of valproic acid as a multifunctional regulator of innate and adaptive immune cells. J Immunol Res 2019:1–24. https://doi.org/10.1155/2019/9678098

Stewart AK (2014) How thalidomide works against cancer. Science 343:256–257. https://doi.org/10.1126/science.1249543

Suh N, Reddy BS, DeCastro A et al (2011) Combination of atorvastatin with sulindac or naproxen profoundly inhibits colonic adenocarcinomas by suppressing the p65/β-catenin/cyclin D1 signaling pathway in rats. Cancer Prev Res 4:1895–1902. https://doi.org/10.1158/1940-6207.CAPR-11-0222

Tabas I, Glass CK (2013) Anti-inflammatory therapy in chronic disease: challenges and opportunities. Science 339:166–172. https://doi.org/10.1126/science.1230720

Takada Y, Bhardwaj A, Potdar P, Aggarwal BB (2004) Nonsteroidal anti-inflammatory agents differ in their ability to suppress NF-κB activation, inhibition of expression of cyclooxygenase-2 and cyclin D1, and abrogation of tumor cell proliferation. Oncogene 23:9247–9258. https://doi.org/10.1038/sj.onc.1208169

Tanrikulu Y, Schneider G (2008) Pseudoreceptor models in drug design: bridging ligand- and receptor-based virtual screening. Nat Rev Drug Discov 7:667–677. https://doi.org/10.1038/nrd2615

Tong M, Seeliger MA (2015) Targeting conformational plasticity of protein kinases. ACS Chem Biol 10:190–200. https://doi.org/10.1021/cb500870a

Tosco P, Balle T (2011) Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. J Mol Model 17:201–208. https://doi.org/10.1007/s00894-010-0684-x

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

Vainio MJ, Johnson MS (2005) McQSAR: a multiconformational quantitative structure−activity relationship engine driven by genetic algorithms. J Chem Inf Model 45:1953–1961. https://doi.org/10.1021/ci0501847

Valle BL, D'Souza T, Becker KG et al (2013) Non-steroidal anti-inflammatory drugs decrease E2F1 expression and inhibit cell growth in ovarian cancer cells. PLoS One 8:e61836–e61836. https://doi.org/10.1371/journal.pone.0061836

van Westen GJP, Wegner JK, Geluykens P et al (2011) Which compound to select in Lead optimization? Prospectively validated proteochemometric models guide preclinical development. PLoS One 6:e27518–e27518. https://doi.org/10.1371/journal.pone.0027518

Wang D, DuBois RN (2010) The role of COX-2 in intestinal inflammation and colorectal cancer. Oncogene 29:781–788. https://doi.org/10.1038/onc.2009.421

Wang E, Sun H, Wang J et al (2019a) End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. Chem Rev 119:9478–9508. https://doi.org/10.1021/acs.chemrev.9b00055

Wang Y, Zhang S, Li F et al (2019b) Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. Nucleic Acids Res 48:D1031–D1041. https://doi.org/10.1093/nar/gkz981

Wilson GL, Lill MA (2011) Integrating structure-based and ligand-based approaches for computational drug design. Future Med Chem 3:735–750. https://doi.org/10.4155/fmc.11.18

Wishart DS, Knox C, Guo AC 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

Wishart D, Arndt D, Pon A et al (2015) T3DB: the toxic exposome database. Nucleic Acids Res 43: D928–D934. https://doi.org/10.1093/nar/gku1004

Wolber G, Dornhofer AA, Langer T (2007) Efficient overlay of small organic molecules using 3D pharmacophores. J Comput Aided Mol Des 20:773–788. https://doi.org/10.1007/s10822-006-9078-7

Wu D, Huang Q, Zhang Y et al (2012) Screening of selective histone deacetylase inhibitors by proteochemometric modeling. BMC Bioinformatics 13:212. https://doi.org/10.1186/1471-2105-13-212

Xu L, Stevens J, Hilton MB, et al (2014) COX-2 inhibition potentiates antiangiogenic cancer therapy and prevents metastasis in preclinical models. Sci Transl Med 6:242ra84-242ra84. https://doi.org/10.1126/scitranslmed.3008455

Yang S-Y (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

Yang J-M, Chen C-C (2004) GEMDOCK: a generic evolutionary method for molecular docking. Proteins 55:288–304. https://doi.org/10.1002/prot.20035

Yang W, Soares J, Greninger P et al (2012) Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res 41:D955–D961. https://doi.org/10.1093/nar/gks1111

Yap TA, Omlin A, de Bono JS (2013) Development of therapeutic combinations targeting major cancer signaling pathways. J Clin Oncol 31:1592–1605. https://doi.org/10.1200/JCO.2011.37.6418

Yu H, Zhong X, Gao P et al (2019) The potential effect of metformin on cancer: an umbrella review. Front Endocrinol (Lausanne) 10:617. https://doi.org/10.3389/fendo.2019.00617

Yuriev E, Holien J, Ramsland PA (2015) Improvements, trends, and new ideas in molecular docking: 2012-2013 in review. J Mol Recognit 28:581–604. https://doi.org/10.1002/jmr.2471

Chapter 13
CADD Approaches and Antiviral Drug Discovery

Mohammad Yasir, Alok Shiomurti Tripathi, Manish Kumar Tripathi, Prashant Shukla, and Rahul Kumar Maurya

13.1 Introduction

In the beginning, computational techniques were limited as a toolbox for academic researchers interested in the correlation of computed mathematical values denoting the nature of chemical structure to the observed physical properties of the molecules. This situation was due to the lack of computer machines with high calculative powers, and their application to molecular docking (MD) and receptor structure optimization was not practically possible.

The development of CADD techniques and advancements in computer hardware has played an efficient and timely design of drug molecules for novel drug targets (Kapetanovic 2008). The role is further fueled by giving the highest computing power at disposal to computational chemists and biologists, enabling very complex calculations involved in the optimization of novel targets and ligands very rapidly (Sliwoski et al. 2014). It takes around 15 years for a medicine to make it through the existing drug approval process. Between 5000 and 10,000 molecules are created each year, yet only one new medicine makes it to market. The pharmaceutical sector

M. Yasir (✉) · R. K. Maurya (✉)
Amity Institute of Pharmacy Lucknow, Amity University Uttar Pradesh, Noida, India
e-mail: myasir@lko.amity.edu; rkmaurya@lko.amity.edu

A. S. Tripathi
Sree S K Patel Institute of Pharmaceutical Education and Research, Ganpat University, Mehsana, Gujarat, India

M. K. Tripathi
Department of Biophysics, All India Institute of Medical Sciences, New Delhi, India

P. Shukla
Department of Pharmacy School of Health Sciences, University of Petroleum and Energy Studies, Dehradun, India

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_13

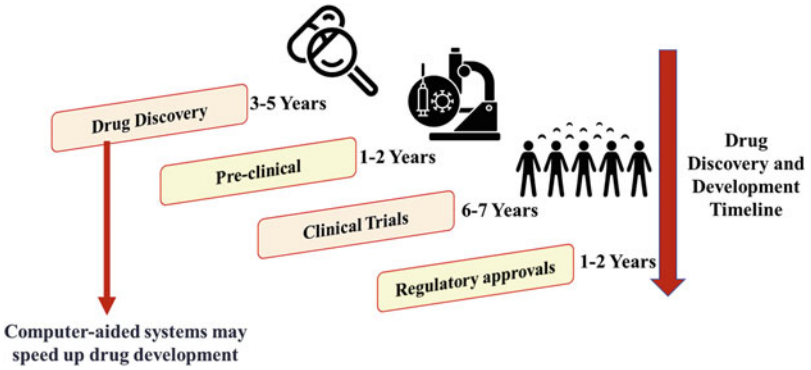


Fig. 13.1 Drug discovery and development timeline of computers in the process of drug discovery

faces difficulties related to lowering research expenses and speeding up the development procedures of new medication discoveries.

CADD and other computational methods have had a major influence on the development of new medications in recent years. DD may be made quicker, cheaper, and more effective with CADD's help, and the field has already yielded useful ideas for treatment (Fig. 13.1).

Viral diseases have haunted the mankind from very early days; the most notorious viral diseases include Spanish flu, HIV, and COVID-19 diseases (Roychoudhury et al. 2020). Although vaccines were able to tame some of the diseases caused by virus infection, diseases like HIV and Hepatitis C infection still possess a significant challenge as they have been proven to be intractable by utilization if vaccine approach. During the early days as viral infection was proved to be dependent upon utilization of cellular machinery for replication, many scientists thought that targeting cellular machinery may lead to certain toxic side effects.

However, as the novel viral proteins involved in replication and interaction with the cells were discovered and their role in the pathogenesis of viral infection was established, the search for agents for targeting those viruses specific molecules led to the fruitful development of some of the most widely used antiviral drugs, e.g., drugs like oselamivir as anti-influenza drugs inhibiting viral neuraminidase and ribavirin as an inhibitor of IMP dehydrogenase (Yin et al. 2021).

13.2 Methodology for CADD

13.2.1 Structure-Based Drug Design (SBDD)

SBDD is related to bioinformatics, concerned with applying mathematics, statistic, and computer science to the study molecular biology. Bioinformatics has been crucial to our knowledge of biochemical and biological processes. Structure-based

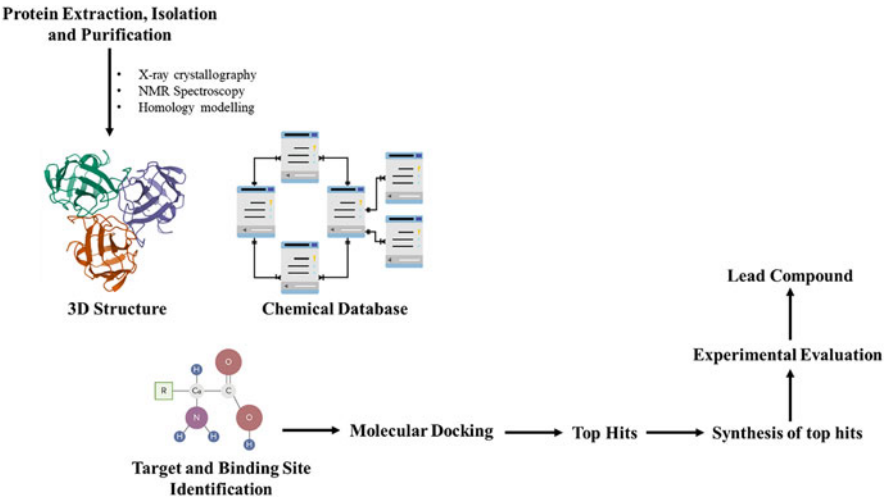


Fig. 13.2 Steps involved in creating a medicine using structure-based design (SBDD)

screening is a new method of molecular composite docking that uses the receptor's three-dimensional structure to automatically pair small molecules from the compound database at the binding site and anticipate their binding mode in order to get a composite energy ranking. The exclusive set of molecular modeling techniques in SBDD includes methods like MD, homology modeling, structure-based virtual screening, and many more (Fig. 13.2). Furthermore, SBDD is an iterative procedure that involves numerous cycles before advancing a refined drug candidate to clinical trials.

A drug discovery process typically consists of four stages: discovery, development, clinical trial, and registry. The first phase involves the identification of a prospective therapeutic target as well as active ligands. To develop effective ligands, current SBDD approaches take into account the main properties of the therapeutic target's binding cavity (Anderson 2003). Moreover, the SBDD is most effective when used as part of a larger drug lead discovery strategy. DD achievable crystal structure of protein commonly ranges from few amino acid to 998 kDa, which provides suitable structural information (Nissen et al. 2000). Diffraction amplitude resolution (also known as resolution), reliability (R) factors, inaccuracy in coordinates, temperature effects, and chemical "correctness" are all important metrics to consider when evaluating a crystal structure. There are two specific reasons like high data to parameter ratio and the obvious position of residues in the electron density map for crystal structures obtained with data (exceeding: 2.5) are generally suitable for design of drug applications. R factor and Rfree values given for a model are indicators of how well it matches up to experimental data (Holton et al. 2014). The Rfree value must be $<28\%$ and $< 25\%$, and the R factor must be much $<25\%$ for the structure to be used in drug design (DD). It is possible to proceed with DD, if structure for a target does not meet the R factor or resolution requirements, but the

outcomes must be assessed with caution (Anderson 2003). The top-ranked compounds with strong affinity for selective regulation of the target protein are evaluated in vitro in biochemical experiments in the second phase.

Hepatitis C (HC) is a liver-damaging infectious illness, caused by the infection of HC virus (HCV), belongs to genus *Hepacivirus* (*Flaviviridae*). One of the feasible strategies for identifying leads for target is structure-based inhibitor creation (Batool et al. 2019). Furthermore, replication of HCV require serine protease, which is an important target to cease the HCV infection. In order to find a potential HCV NS5B inhibitor, the researchers examined the HCV NS5B template-primer complex model proposed for viral thumb domain rearrangement and replication in response to RNA binding (Bressanelli et al. 1999).

Presently, no licensed antiviral available for dengue, as attachment and fusion of dengue virus, occurs with envelope glycoprotein (EG); it is a prospective target for therapeutic research (Anasir et al. 2020). Understanding about the dengue virus EG has prompted researchers to investigate structural-based medication development of antiviral chemicals and peptides against dengue virus infections.

The influenza A virus produces an acute respiratory viral illness that threatens worldwide human health leading to economical loss. Anti-influenza medications remain the most effective therapy due to new virus strains. Viral proteins are required for the growth of influenza A virus, used for the design of its medicine (Crocker et al. 2021).

13.2.2 Ligand-Based Drug Design (LBDD)

LBDD is a method for identifying active compounds against drug targets and forecasting possible active chemical targets. However, this technique has inherent flaws caused by factors such as variable protein conformations with widely varying binding sites or a lack of real target proteins in the database (Yang et al. 2021). LBDD tools provide detailed structure (3D) of target and interaction between ligand and target, which allows for lead discovery and its optimization (Acharya et al. 2011). Furthermore, known compound data for structure activity is used to develop pharmacophore model in the absence of 3D structure of protein (Schuster et al. 2011b). The simplicity of interpretation and customization of pharmacophore models enables to incorporate information about a given binding mode into a straightforward (Murgueitio et al. 2012; Noha et al. 2012). Compound library commercially available was used to develop and screen 3D pharmacophore model of HCV NS5 (NNI site IV). The best chemical among the 18 tested hits has an EC50 value (Murgueitio et al. 2012). 3D space on which chemical feature markers placed carefully to observe molecules that have right interactions, is the ways to use VS. This has led to many successes in explaining ligand affinity in the past and

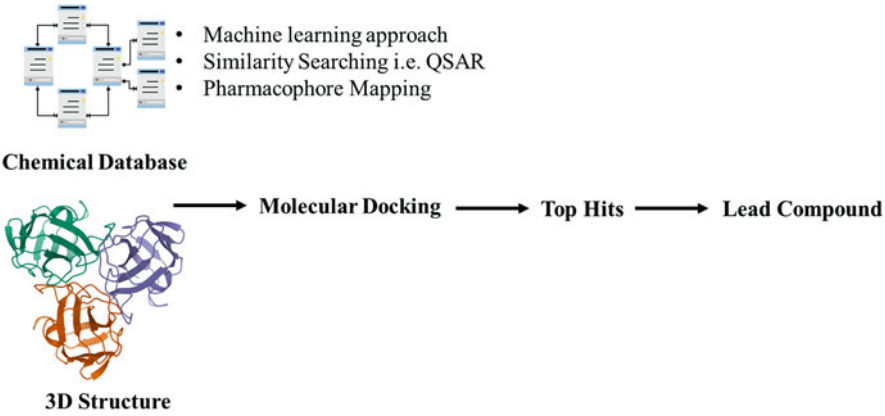


Fig. 13.3 Workflow of LBDD

designing new ligands for the future (Schuster et al. 2011a). LigandScout is a software that extracts accurate hybridization states, analyze ligand geometries, protein-ligand complexes and uses a set of criteria to classify possible protein–ligand interactions (Wolber et al. 2007; Wolber and Langer 2005). Quantitative structure-activity relationship (QSAR) (Achary 2020) is based on molecule shape and other specialized approaches. The best research approach is determined by the amount of information accessible (Bassetto et al. 2017). The more exact the information, the more dependable the results (Fig. 13.3).

The shared pharmacophore of chloroquine and remdesivir predicted that interacting residues would have a significant role in corona therapy (Yang et al. 2021). According to Dushyant et al., it has been revealed that the spike proteins of the new coronavirus COVID-19 interact with angiotensin-converting enzyme (ACE). Chloroquine therapy alleviates the clinical symptoms of corona. Yanqing et al. introduced D3Similarity, a ligand-based technique for predicting active drugs against SARS-CoV-2 and identifying potential bioactive molecules which are probable to target proteins in unavailability of 3D structure of proteins (Yang et al. 2021).

Medication used for the management of AIDS develops by targeting reverse transcriptase (RT) of HIV-1. A study reported on the high-throughput docking of NCI database of 2800 compounds; based on docking score, top 6 hits were tested biologically and 4 of them shown suppression of RT (Bustanji et al. 2009). Another HIV treatment technique is to impede viral DNA integration by targeting HIV-1 Integrase (IN) in the host cell. A database was used to search target-based pharmacophore, out of with biological research was performed on 10 hit molecules. One drug demonstrated high inhibitory activity against IN and modest inhibition of HIV-1 proliferation (Rajamaki et al. 2009).

13.2.3 Machine Learning (ML)

ML has transformed the field of computational drug discovery in recent decades. ML approaches are ideally suited for problem domains characterized by large amounts of data. ML technique emphasizes two major approaches, i.e., Support Vector Machine (SVM) and Artificial Neural Networks (ANN) are the computational techniques used for the process of drug discovery. ANN commonly known as "Neural Network" (NN) is a learning algorithm of functional and structural aspect of biological neural networks (Ge et al. 2020). In other words, ANN is an emulation of a biological neural system. Modern neural networks observe pattern in data, relationship between input and output, and probability of observed variable with linear statistical data modeling tools (Paul et al. 2021). SVM is a regression and classification prediction tool that maximizes predictive accuracy by using theory of ML and avoids over-fit to the data automatically (Shahid et al. 2019).

Data preprocessing, learning model, and evaluation are the steps involved in ML-based prediction model. We prepare the data via discretization and standardization-based methods in data preprocessing steps. In the model learning phase, ML algorithms were implemented (Tripathi et al. 2021a). Finally, the performance of build models was accessed by the sensitivity and accuracy parameters. The ML algorithms have to be evaluated critically for their performance, which is crucial for the algorithm's outcome (Sharma et al. 2023).

Accuracy of ML methods is evaluated using various parameters, which included the following:

Accuracy: It correctly predicts the positive sample percentage.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100$$

Sensitivity: It correctly predicts the negative sample percentage.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100$$

Specificity: It correctly predicts the positive and negative samples.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100$$

Many platforms based on ML methods are available nowadays, which provides simple graphical interface for the development of learning model. WEKA is widely used nowadays and is a java-based ML platform containing algorithms for regression and classification. Some other platforms are also present and widely used for drug discovery purposes, such as Rattle, H2O, SciKit-learn, etc.

ML methods, namely support boosting, vector machine, and random forest are used to predict the interaction between protein and ligand. With increasing data size

and the exponential growth of databases, ML algorithms have become indispensable to handling these extensive databases without compromising speed and accuracy (Abiodun et al. 2019; Chan et al. 2019; Tang et al. 2019). The ML has an essential application in the early stages of computational drug discovery, proven a valuable tool. The ML helps predict new compounds, including not simply the physicochemical properties but also various biological activities such as drug efficacy, ligand binding, and adverse effects of the identified compounds.

13.2.4 Deep Learning (DL)

Drug discovery is a complex process which takes enormous cost and time. Thus, there is a need for the DL method to accelerate the drug discovery process and identify promising drugs against the desired target. In the training step, the ML algorithm analyzes the massive datasets of experimentally determined data to build a predictive model for gaining information from the dataset (Tripathi et al. 2021b). In the ML method, the loss of relevant information during the feature extraction poses a shortcoming of this technique. Thus, DL methods have evolved, which solves the failure of the feature extraction process in ML (Carpenter et al. 2018; Di Gangi et al. 2018). DL methods authorize the automatic generation of higher-level hierarchical abstraction from massive datasets and reduce the demesne for feature generation from ML methods. DL is a branch of ML that uses elaborate neural networks with numerous layers to extract hidden insights from raw experimental data. It has been increasingly used in recent years for computational drug development.

Compared with ML methods, the DL algorithm is efficient and has a process of automatic feature extraction. DL methods such as recursive neural networks (RNN) and convolutional neural networks (CNN) are extensively applied in drug discovery. The DL neural network architecture consists of more than two hidden layers. Successful training of DL neural networks requires a large amount of data, as the parameters are quite large. With the enormously increasing volume of data, DL plays an essential role in mining the huge dataset for identifying drug molecules. The CNN has been widely used in virtual screening (VS) as it implements the feature. DL neural networks are now also commonly used in de de-novo molecule design. The other prominent method of DL includes RNNs (DiPietro and Hager 2020), variational autoencoders (Simidjievski et al. 2019), and generative adversarial networks (Kazeminia et al. 2020). The benefit of DLNN is that it learns new higher-order representations from the data (Di Gangi et al. 2018). Apart from this, vanishing gradient problem is the disadvantage of traditional NNs (Hochreiter 1998), in which there is difficulty in the weight optimization process as the number of layers (depth of the network) increases.

13.2.5 Virtual Screening (VS)

Previously, the pharmaceutical industry identified novel lead molecules by adopting the experimental screening of large chemical libraries. This high-throughput-based screening process faces drawbacks such as high cost and time to process the desired certainty and mechanism of action of the identified lead molecules. Thus, the advancements in computing resources and the growth of targets in public databases led to the use of computational tools as a screening process to identify novel lead molecules against target. Thus, the SBDD with computational methods is used for drug discovery by pharmaceutical industries (Tripathi et al. 2022).

In computer-aided drug discovery, VS is an essential tool for identifying novel molecules. It can be categorized into two types: structure-based virtual screening (SBVS) and ligand-based virtual screening (LBVS) (Tripathi et al. 2019). In the SBVS, the target structure of the target protein is used, while in LBVS, the information of known inhibitors were used for the VS. The features of VS include (1) filtering large compound libraries into the small sets of active compounds, (2) optimizing the identified lead compounds to increase their affinity, and (3) giving a scaffold to develop novel compounds.

In SBVS, the widely used method is docking which predicts the binding mode and orientation of one molecule and helps to understand the mechanism of drug-protein interaction. The docking method has two main components: a search algorithm and a scoring function used to quantify the binding strength of the drug molecule in the target active site. Algorithms such as Monte Carlo, genetic algorithm, systematic search, simulated annealing, etc. are widely used by different softwares for this purpose (Shrivastava et al. 2016; Tripathi et al. 2015, 2016).

The LBVS depends on the dataset of the active and inactive molecules. This method predicts the active ligand with higher activity based on the physiochemical analysis and spatial similarities between the active ligands. This method is used to identify the active ligand when the target structure is unknown, or the structural accuracy is low. Pharmacophore modeling is the most widely used method for the LBVS. ML-based methods also help in identifying the predicted hits by minimizing the rate of false hits prediction.

In silico VS describes macromolecule-ligand interactions using computer models. There are several ways available for this goal, which may be divided into two categories: 2D and 3D approaches. Descriptor-based approaches (2D methods) focus on calculating and comparing scalar molecular qualities to find molecules that are comparable in terms of their derived molecular attributes (Murgueitio et al. 2012). However, most computational efforts for antiviral drug discovery have recently relied on 3D methodologies, which try to characterize the steric and chemical complementarity of the 3D conformations of a macromolecular target and the binding ligand. In MD-based VS procedures, protein-ligand complexes are awarded a score that corresponds with the expected binding affinity, which can be estimated using a physics-based, empirical, or knowledge-based potential function (Sohraby et al. 2019). AutoDock (Morris et al. 2009), AutoDock Vina (Trott and

Olson 2009), MOE (Vilar et al. 2008), GLIDE (Friesner et al. 2004), Discovery Studio (SYSTÈMES 2016), and other VS tools are commonly utilized. In VS, small-molecule databases are docked into the area of interest, and their anticipated interactions with the site are graded, whereas in de novo method, molecules are constructed by placing building blocks at the specific location, scoring them, and linking them virtually. After the connected fragments are generated in silico, the final molecules must be synthesized in a laboratory. The VS and de novo generation labels coincide in several cases.

Furthermore, PAINS-Remover is an Internet service that is intended for screening libraries (Baell and Holloway 2010). Online, the Swiss ADME website forecasts the physicochemical description of chemicals (Daina et al. 2017). These VS and compound optimization filtering technologies are required to increase the success rate of medication development and lessen the problem of lost costs in the latter phases of drug research.

Novel NS5B polymerase inhibitors found using VS were reported, and their inhibitory effects were evaluated in vitro. A VS of compounds that fit this binding pocket from the existing chemical library of 3.5 million compounds was undertaken on the basis of a newly found binding pocket of NS5B, separate from the nucleotide binding site but highly conserved among diverse HCV strains (Ryu et al. 2009). A recent study by Zhang et al. utilized this strategy to screen the inhibitor against the PA endonuclease and its variant. Bilobetin was shown to be capable of competitively inhibiting the PA endonuclease after initial in vitro and in silico screening (Zhang and Wang 2021). Zhang et al. used 3D-QSAR modeling and a docking-based VS technique to identify a raltegravir derivative as a possible new PA endonuclease inhibitor (Zhang et al. 2021a). Ferro et al. also created a three-dimensional pharmacophore model and acquired three "hit drugs" via VS. The MD approach and enzymatic analysis with recombinant PA endonuclease were used to study the binding poses of these hit compounds (Ferro et al. 2018). Pala et al. proved in silico findings through PA enzymatic testing and antiviral activity in the low micromolar range in a cell-based influenza virus assay (Pala et al. 2015).

The influenza hemagglutinin (HA) A virus is a homotrimer that is frequently separated into head and stem regions and is required for viral entrance and uncoating (Harrison 2008). Endocytosis transports influenza virus particles to the endosome once they bind to the host receptor. Proton entrance causes the pH of the endosome to fall (Carr et al. 1997). Bodian et al. discovered efficient HA fusion inhibitors (Bodian et al. 1993). Waldmann et al. use in silico design, chemical synthesis, and binding studies to report a trivalent glycopeptide mimic, a nanomolar multivalent ligand that binds to avian influenza hemagglutinin H5 (Waldmann et al. 2014). Another possible therapeutic target in the Flaviviridae family is the protease/helicase NS3. Takaya's novel induced-fit docking tool (GENIUS) was utilized to uncover 13 new HCV NS3 protease inhibitors in research targeting this protein (Takaya et al. 2011). Shiryaev used the NCI database compounds to dock into the NS2B cofactor binding site of the WNV NS3 protease crystal structure, yielding one physiologically verified inhibitor with nanomolar activity (Murgueitio et al. 2012).

While in silico VS has been developed for several useful applications, it still has a number of significant disadvantages. To provide just one example, the docking score is not a reliable indicator of pharmaceutical effectiveness since most docking methods only consider binding affinity and ignore other properties. It is also important to remember that MD-based VS has a high false-positive rate.

13.3 Host and Viral Proteins as Target

13.3.1 Chemokine Receptors

Chemokines are small proteins, which signal through its receptors, i.e., G-coupled protein receptor (GPCRs). It is chemoattractant cytokines responsible for migration of leukocytes towards the infectious and injured tissue. GPCRs present on the surface of target cells interact with chemokines for it signaling, autoimmunity, and inflammation and occur due to its dysregulation. There are 19 different types of chemokine receptors, and 50 types of chemokine ligands were identified, which can bind to more than a single receptor (Hauser et al. 2017).

Viral infections are pathogenic conditions, altering the response of acquired and innate immunity of infected host. It is evident that chemokine receptors and chemokine ligands are encoded on DNA viruses like HIV, poxviruses, and herpesviruses and secrete chemokine binding proteins (CKBPs). Details of virus-encoded chemokine modulators are given further, which are involved in the development of pathogenesis.

13.3.2 Viral Chemokine Receptors

Viral chemokine receptors like US28 and ORF74 are present on the surface of cytomegalovirus and human herpesvirus 8 (HHV8), respectively (Pontejo and Murphy 2017). Kaposi sarcoma (KS) lesions were reported to be developed due to infection of HHV8, characterized by infiltrates of inflammation, angiogenesis, and spindle-shaped cells. There are 100 open reading frames (ORFs) present on the genome of HHV8. Literature suggests that ORF74-encoded chemokine receptor contributes to the development of angioproliferative lesions, which activates several pathways includes NF-kB to affect biology of lymphatic and vascular endothelium (T. Liu et al. 2017). Moreover, ORF74 activates AP-1, NF-kB, and MEK with the help of phosphatase SHP.

It is also observed that the increased expression of Notch signaling in KS, vGCPR on vascular endothelial cells enhances the expression of several components of Notch and ERK signaling pathway such as the ligand Notch and Jagged1 and the receptor Notch2 downstream targets such as Hey1 (R. Liu et al. 2010). Moreover, promotion of Notch signaling by Notch4 occurs due to DLL4 (Notch ligand)

interaction, which upregulates viral GPCR on lymphatic endothelial cells. HHV8 upregulates DLL4-stimulated signaling and alters the cellular components of uninfected neighboring cells which affect cellular plasticity and quiescence (Emuss et al. 2009).

Cancer and HCMV have a common molecular connection, that is, US28; study reveals that NIH-3 T3 cells transfect US28 that originates in tumor of nude mice (Maussang et al. 2006). COX-2 expression is highly upregulated in several forms of cancer. Celecoxib is reported to delay the formation of US28-transfected tumor cells in nude mice (Baryawno et al. 2011).

13.3.3 Viral Chemokine Ligand

There are number of chemokine ligands encoded on several viruses which contribute in the development of pathogenic conditions. HHV-8 is encoded with three chemokine ligands such as vCCL1, vCCL22, and vCCL3: CCR8 interacts with its agonist vCCL1; CCR1, CCR2, CCR5, CX3CR1, and CXCR4 interact with its antagonist vCCL2; and XCR1 receptor interacts with its agonist vCCL3 (Alcami and Lira 2010). Virus spread is reported to be facilitated by vCCL1 and vCCL3 as it regulates the influx. Moreover, replication of virus contributes in the development of pathogenesis of KS, which is facilitated by vCCL1 and vCCL2 (Greene et al. 2007).

There are two CXC ligands (UL146 and UL147) encoded on HCMV, vCXCL1 protein encodes gene UL146, which act as agonist of vCXCR1 and CXCR2; however it has low affinity and potency (Lüttichau 2010).

13.3.4 Viral CKBPs

Proteins secreted by viruses which regulate the chemokine activity during infection are commonly called chemokine binding proteins. There are several CKBPs identified, which are encoded by ticks.

The first CKBP was identified on gamma herpesvirus 68 (MHV-68) known as M3 protein. There are two binding sites for chemokines on M3 protein, which binds to broad range of chemokines. Intracellular signaling and cellular receptor interaction with chemokine are prevented by viral CKBPs. Moreover, alpha herpesviruses contain two CKBP (MHV-68 gG and M3), which alters the chemokine interaction with GAGs and chemokine receptors. M3 chemokine complex was responsible for inhibition of chemokine-GAG interaction (Parry et al. 2000).

13.3.5 Glycoproteins

Viruses on its outermost surface consist of glycoprotein, and thus pathogenic viruses commonly interact with glycoproteins, i.e., glycan receptor. Glycoprotein plays an important role on virulence capacity, entry, and infectivity of virus.

13.3.6 Glycoprotein of Virus

Glycoprotein present on the enveloped viruses contributes to entry of virus in the cell by interacting with the receptors present on the cell. Human immunodeficiency virus 1 (HIV-1) on its surface contains gp41 transmembrane protein and gp120 surface protein. Chemokine receptors (CCR5) and CD4 bind HIV-1 through gp120 to facilitate the entry of virus in target cells. However, Ebola virus (EBOV) entry was enhanced by removing N-glycan from envelope glycoprotein.

Influenza A virus consists of two different surface glycoproteins such as neuraminidase (NA) and hemagglutinin (HA). HA protein of influenza virus interacts with glycoprotein (terminal sialic acid [SA]) of host cell and SA residues of glycoproteins cleaved by NA of enveloped virus, which promotes the infectivity of virus (Kosik and Yewdell 2019).

Glycoproteins present on the surface of virus alter the virulence capacity of virus by antigenic property and stimulation of immune response and modulate the binding of host receptor with virus. Rabies viruses (RABVs) contain several proteins like large transcriptase protein, GP, matrix protein, phosphoprotein, and nucleoprotein. There are two N-glycosylation sites (Asn37 and Asn319) of GP present on RABVs (Mebatsion et al. 1999). Propagation of virus enhances due to Asn37 and pathogenicity of it reduces. Moreover, higher number of HA glycosylation sites contributes to reduction in virulence capacity of H1N1.

13.3.7 Glycoprotein of Host

Viral infection occurs due to binding of host receptor with virus; viral glycoprotein has major responsibility for the viral infection. However, viral glycoprotein binds to the host receptor, and thus host receptor is also important for the cellular entry of virus (Maginnis 2018). Host receptors contribute in entry of virus and promotion of viral infection. Viral receptor is separated into two different categories according to their functions such as attachment factor and entry receptor. Virus can bind to entry receptors present on the host cell surface which concentrate virus on the surface of cell-like HIV can bind to CD4 (Checkley et al. 2011).

13.3.8 Kinases

Tyrosine kinase receptors are present on the cellular membrane, which contribute in the cellular response against virus by alteration of survival of cells, proliferation, differentiation, and migration. Tyrosine kinase receptors having single helix protein with ligand binding site at the extracellular region. Kinases were found to be 500 types present in human proteome. Multiple virus life cycles interpretation and propagation of virus are regulated by protein kinase during viral infection (Pillaiyar and Laufer 2022).

Viral infection results in physiological regulation of majority of host protein kinases; multiple viruses used different kinases which targets host signaling cascades. Moreover, cellular protein kinase inhibitors are used to assess the antiviral property as it blocks the viral replication in the cellular culture. It alters the several steps involved in the life cycle of viruses (Pillaiyar and Laufer 2022). There are several kinases targeted for the development of antiviral drugs, which are given further.

13.3.9 Lipid Kinase

Lipid kinases are responsible for the phosphorylation of cellular lipids, contributing in maintenance of cellular functions and lipid homeostasis. Lipid kinases are present in several substrate forms such as phosphatidylinositol kinases, sphingosine kinases, and diacylglycerol kinases. Lipid kinase inhibitors targeted to the host cell are used for antitumor, antidiabetic, anti-inflammatory, etc. activity. Kinases are reported to be activated during the viral infection; thus kinase is a potential therapeutic target for the management of viral infection. Kinase inhibitors are used for the antiviral property (Merida et al. 2019).

13.3.10 Numb-Associated Kinases (NAKs)

NAKs belong to Ser/Thr kinase family; serine/threonine kinase 16 (STK16), BMP-2 inducible kinase (BIKE/BMP2K), cyclin G-associated kinase (GAK), and adaptor-associated kinase 1 (AAK1) are the four different members that come under NAK. All these kinases are involved in the development of neurodegenerative disorders and cancer. Cellular process endocytosis involves GAK and AAK1, which also occur in viral infection (Sorrell et al. 2016).

Endocytosis is required for the cellular penetration of virus responsible for infection. There are several enveloped RNA viruses that use the adaptor protein for the process of infection. Moreover, clathrin-mediated endocytosis process contributes in the other process of viral growth (entry, replication, assembly, reverse

transcription, and DNA synthesis) and infection. Thus, kinase (GAK and AAK1) involved in the process of infection is used for the screening of broad-spectrum antiviral drugs (Yángüez et al. 2018).

13.3.11 Receptor Tyrosine Kinases (RTKs)

Cell cycle and its metabolism are regulated with a receptor present on the cell surface commonly known as RTK. Approximately 58 types of RTKs have been identified; dysregulation of these receptors relates to diseases including inflammation, diabetes, and cancer. Human papilloma viruses contain viral E5 gene product with epidermal growth factor receptor which stimulates viral replication by activating EGFR pathway. Inhibitors of EGFR tyrosine kinase are reported to possess antiviral property (Lemmon and Schlessinger 2010).

13.3.12 Mitogen-Activated Protein Kinases (MAPKs)

Cellular response to the external stimuli was also regulated with the help of signaling cascade MAPK, as it is responsible for the growth receptor's interaction. Three different kinases such as p38, JNKs, and ERKs of MAPK family get activated which produces key molecule responsible for cellular activities like differentiation, apoptosis, survival, motility, metabolism, and mitosis (Cargnello and Roux 2011).

MAPK cascade is reported to be induced in infected cells by a few RNA and DNA viruses. Moreover, influenza A virus activity is observed to be decreased by reducing its proliferation with the inhibition of ERK signaling. Japanese encephalitis occurs due to infection of Japanese encephalitis virus (JEV) and JNK-1 inhibitor found to be effective against JEV by reducing inflammatory cytokines (Ashraf et al. 2021).

13.3.13 Src Kinases

Src's kinases play a role in several cellular processes like survival, progression of cell, differentiation, and motility. There are 11 different types of Srcs kinase identified, which are non-receptor tyrosine kinase. Srcs inhibitors are used for the management of cancer, and several US FDA-approved medicines are available for it (Sen and Johnson 2011).

Src inhibitors are also observed to possess antiviral activity; a drug dasatinib reduces the viral infection in a dose-dependent manner against DENV infection. Moreover, dasatinib is also used against HIV-1, as it reduces the T-cell activation

which reduces AP-1 and NF-kB leads to decline in viral replication shows potential benefit of it against HIV-1 infection (Chu and Yang 2007).

13.3.14 Cyclin-Dependent Kinases (CDKs)

CDKs control the cell cycle process by phosphorylation of Ser/Thr residues, as it binds with the ATP (Ding et al. 2020). CDK inhibitors are used for the treatment of cancer. Moreover, CDK inhibitors were targeted against SARS-CoV-2 in an in vitro study, and 16 molecules were observed to effect against it. CMV infection is also ameliorated by treating with CDK inhibitors (Zhang et al. 2021b). However, its molecular effect is yet to be understood clearly, which need to be focused by researchers. There are several other kinases available which can be targeted for the development of antiviral drugs (García-Cárceles et al. 2022).

13.3.15 Other Proteins

There are some proteins apart from already explained one, which are involved in the interaction with viral genome or protein and inhibit viral life cycle like virus replication, assembly, and egress. It involves in the development or control of viral infection. These proteins are explained in detail as follows:

13.3.16 Cytoskeleton Protein (Actin)

Cytoskeleton network present in the host cell involves transportation of viral components, entry, and exit in the cells. Expression of viral gene of RNA viruses requires actin- and tubulin-like cytoskeleton proteins. Cytoskeleton protein like actin is required for the entry of influenza virus inside the cells and facilitates the budding of filamentous virus particles (Ploubidou and Way 2001). There are several viruses like SV40 and herpes which produce disassembly of microfilament containing actin. Viral infection disturbs the cytoskeleton for growth to produce infection, but still exact mechanism involved in it not proved (Wu et al. 2019). However, targeting microfilament actin could be used for the development of antiviral drugs.

13.3.17 Annexins

Annexin is a cellular protein present on the surface; and in the cytoplasm, it is also found in the influenza virus. There are several members of annexins available,

annexin 2 is involved in the interaction with several virus to promote the infection. Annexin 2 promotes the binding and fusion of HCMV and also the assembly of HIV-1 (Ryzhova et al. 2006).

There are several other cellular proteins such as tetraspanins, Cyclophilin A, CD59, and glycolytic enzymes which could be used for the development of antiviral drugs.

13.4 Conclusion and Future Perspectives

Antiviral medications tend to be successful in limiting seasonal pandemics, particularly in the early phases of fast transmission. The viral proteins play important roles in virus's life cycle and might be used as therapeutic targets for treating the disease. The creation of anti-flu medications has been greatly aided by CADD, the most important technology in modern drug research and development.

Different CADD approaches to creating small-molecule virus inhibitors are shown. VS, 3D-QSAR, molecular dynamics, pharmacokinetic calculations, etc. are all examples that pertain to SBDD or LBDD in small-molecule inhibitor development.

The Food and Drug Administration has authorized the sale of numerous very effective medications, but the spread of drug-resistant viruses has rendered these treatments ineffective. This means that new mechanisms and strategies to attack the viruses, which are constantly adapting, are desperately needed.

In silico VS is another important technology that has been developed for many useful applications, although it still has several obvious drawbacks. For instance, the docking score is not a helpful indication of therapeutic effectiveness since most docking approaches focus only on binding affinity and disregard other factors. There has been an urgent need for intriguing development in structural informatics that promises to hasten our progress toward a deeper understanding of how protein structure affects human health and medicine.

CADD technology is a vital resource for locating promising new lead compounds, which speeds up the discovery of effective new antiviral medicines. We anticipate a revolution in the discovery of novel drugs in the near future as CADD tools powered by AI become more sophisticated and comprehensive in their coverage of the whole process.

References

Abiodun OI, Jantan A, Omolara AE, Dada KV, Umar AM, Linus OU, Arshad H, Kazaure AA, Gana U, Kiru MU (2019) Comprehensive review of artificial neural network applications to pattern recognition. IEEE Access 7:158820–158846. https://doi.org/10.1109/ACCESS.2019.2945545

Achary PGR (2020) Applications of quantitative structure-activity relationships (QSAR) based virtual screening in drug design: a review. Mini Rev Med Chem 20(14):1375–1388. https://doi.org/10.2174/1389557520666200429102334

Acharya C, Coop A, Polli E, MacKerell D (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

Alcami A, Lira SA (2010) Modulation of chemokine activity by viruses. Curr Opin Immunol 22(4): 482–487. https://doi.org/10.1016/j.coi.2010.06.004

Anasir MI, Ramanathan B, Poh CL (2020) Structure-based design of antivirals against envelope glycoprotein of dengue virus. Viruses 12(4):367. https://doi.org/10.3390/v12040367

Anderson AC (2003) The process of structure-based drug design. Chem Biol 10(9):787–797. https://doi.org/10.1016/j.chembiol.2003.09.002

Ashraf U, Ding Z, Deng S, Ye J, Cao S, Chen Z (2021) Pathogenicity and virulence of Japanese encephalitis virus: neuroinflammation and neuronal cell damage. Virulence 12(1):968–980. https://doi.org/10.1080/21505594.2021.1899674

Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem 53(7):2719–2740. https://doi.org/10.1021/jm901137j

Baryawno N, Rahbar A, Wolmer-Solberg N, Taher C, Odeberg J, Darabi A, Khan Z, Sveinbjörnsson B, FuskevÅg O-M, Segerström L, Nordenskjöld M, Siesjö P, Kogner P, Johnsen JI, Söderberg-Nauclér C (2011) Detection of human cytomegalovirus in medulloblastomas reveals a potential therapeutic target. J Clin Invest 121(10):4043–4055. https://doi.org/10.1172/JCI57147

Bassetto M, Leyssen P, Neyts J, Yerukhimovich MM, Frick DN, Brancale A (2017) Shape-based virtual screening, synthesis and evaluation of novel pyrrolone derivatives as antiviral agents against HCV. Bioorg Med Chem Lett 27(4):936–940. https://doi.org/10.1016/j.bmcl.2016.12.087

Batool M, Ahmad B, Choi S (2019) A structure-based drug discovery paradigm. Int J Mol Sci 20(11):2783. https://doi.org/10.3390/ijms20112783

Bodian DL, Yamasaki RB, Buswell RL, Stearns JF, White JM, Kuntz ID (1993) Inhibition of the fusion-inducing conformational change of influenza hemagglutinin by benzoquinones and hydroquinones. Biochemistry 32(12):2967–2978. https://doi.org/10.1021/bi00063a007

Bressanelli S, Tomei L, Roussel A, Incitti I, Vitale RL, Mathieu M, De Francesco R, Rey FA (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Proc Natl Acad Sci U S A 96(23):13034–13039. https://doi.org/10.1073/pnas.96.23.13034

Bustanji Y, Al-Masri IM, Qasem A, Al-Bakri AG, Taha MO (2009) In silico screening for non-nucleoside HIV-1 reverse transcriptase inhibitors using physicochemical filters and high-throughput docking followed by in vitro evaluation. Chem Biol Drug Des 74(3):258–265. https://doi.org/10.1111/j.1747-0285.2009.00852.x

Cargnello M, Roux PP (2011) Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. Microbiol Mol Biol Rev 75(1):50–83. https://doi.org/10.1128/MMBR.00031-10

Carpenter KA, Cohen DS, Jarrell JT, Huang X (2018) Deep learning and virtual drug screening. Future Med Chem 10(21):2557–2567. https://doi.org/10.4155/fmc-2018-0314

Carr CM, Chaudhry C, Kim PS (1997) Influenza hemagglutinin is spring-loaded by a metastable native conformation. Proc Natl Acad Sci U S A 94(26):14306–14313. https://doi.org/10.1073/pnas.94.26.14306

Chan HCS, Shan H, Dahoun T, Vogel H, Yuan S (2019) Advancing drug discovery via Artificial Intelligence. Trends Pharmacol Sci 40(10):801. https://doi.org/10.1016/j.tips.2019.07.013

Checkley MA, Luttge BG, Freed EO (2011) HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. J Mol Biol 410(4):582–608. https://doi.org/10.1016/j.jmb.2011.04.042

Chu JJH, Yang PL (2007) c-Src protein kinase inhibitors block assembly and maturation of dengue virus. Proc Natl Acad Sci U S A 104(9):3520–3525. https://doi.org/10.1073/pnas.0611681104

Crocker TF, Brown L, Lam N, Wray F, Knapp P, Forster A (2021) Information provision for stroke survivors and their carers. Cochrane Database Syst Rev 2021(11):CD001919. https://doi.org/10.1002/14651858.CD001919.pub4

Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7:42717. https://doi.org/10.1038/srep42717

Di Gangi M, Lo Bosco G, Rizzo R (2018) Deep learning architectures for prediction of nucleosome positioning from sequences data. BMC Bioinformatics 19(14):418. https://doi.org/10.1186/s12859-018-2386-9

Ding L, Cao J, Lin W, Chen H, Xiong X, Ao H, Yu M, Lin J, Cui Q (2020) The roles of cyclin-dependent kinases in cell-cycle progression and therapeutic strategies in human breast cancer. Int J Mol Sci 21(6):1960. https://doi.org/10.3390/ijms21061960

DiPietro R, Hager GD (2020) Deep learning: RNNs and LSTM. In: Zhou SK, Rueckert D, Fichtinger G (eds) Handbook of medical image computing and computer assisted intervention, pp 503–519. Academic Press. https://doi.org/10.1016/B978-0-12-816176-0.00026-0

Emuss V, Lagos D, Pizzey A, Gratrix F, Henderson SR, Boshoff C (2009) KSHV manipulates Notch signaling by DLL4 and JAG1 to alter cell cycle genes in lymphatic endothelia. PLoS Pathogens 5(10):e1000616. https://doi.org/10.1371/journal.ppat.1000616

Ferro S, Gitto R, Buemi MR, Karamanou S, Stevaert A, Naesens L, De Luca L (2018) Identification of influenza PA-Nter endonuclease inhibitors using pharmacophore- and docking-based virtual screening. Bioorg Med Chem 26(15):4544–4550. https://doi.org/10.1016/j.bmc.2018.07.046

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47(7):1739–1749. https://doi.org/10.1021/jm0306430

García-Cárceles J, Caballero E, Gil C, Martínez A (2022) Kinase inhibitors as underexplored antiviral agents. J Med Chem 65(2):935–954. https://doi.org/10.1021/acs.jmedchem.1c00302

Ge Q, Huang X, Fang S, Guo S, Liu Y, Lin W, Xiong M (2020) Conditional generative adversarial networks for individualized treatment effect estimation and treatment selection. Front Genet 11:1578. https://doi.org/10.3389/fgene.2020.585804

Greene W, Kuhne K, Ye F, Chen J, Zhou F, Lei X, Gao S-J (2007) Molecular biology of KSHV in relation to AIDS-associated oncogenesis. Cancer Treat Res 133:69–127. https://doi.org/10.1007/978-0-387-46816-7_3

Harrison SC (2008) Viral membrane fusion. Nat Struct Mol Biol 15(7):690–698. https://doi.org/10.1038/nsmb.1456

Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE (2017) Trends in GPCR drug discovery: new agents, targets and indications. Nat Rev Drug Discov 16(12):829–842. https://doi.org/10.1038/nrd.2017.178

Hochreiter S (1998) The vanishing gradient problem during learning recurrent neural nets and problem solutions. Int J Uncertain Fuzz 6:107–116. https://doi.org/10.1142/S0218488598000094

Holton JM, Classen S, Frankel KA, Tainer JA (2014) The R-factor gap in macromolecular crystallography: an untapped potential for insights on accurate structures. FEBS J 281(18):4046–4060. https://doi.org/10.1111/febs.12922

Kapetanovic IM (2008) Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. Chem Biol Interact 171(2):165–176. https://doi.org/10.1016/j.cbi.2006.12.006

Kazeminia S, Baur C, Kuijper A, van Ginneken B, Navab N, Albarqouni S, Mukhopadhyay A (2020) GANs for medical image analysis. Artif Intell Med 101938. https://doi.org/10.1016/j.artmed.2020.101938

Kosik I, Yewdell JW (2019) Influenza hemagglutinin and neuraminidase: Yin–Yang proteins coevolving to Thwart immunity. Viruses 11(4):346. https://doi.org/10.3390/v11040346

Lemmon MA, Schlessinger J (2010) Cell signaling by receptor tyrosine kinases. Cell 141(7): 1117–1134. https://doi.org/10.1016/j.cell.2010.06.011

Liu R, Li X, Tulpule A, Zhou Y, Scehnet JS, Zhang S, Lee J-S, Chaudhary PM, Jung J, Gill PS (2010) KSHV-induced notch components render endothelial and mural cell characteristics and cell survival. Blood 115(4):887–895. https://doi.org/10.1182/blood-2009-08-236745

Liu T, Zhang L, Joo D, Sun S-C (2017) NF-κB signaling in inflammation. Signal Transduct Target Ther 2(1):17023. https://doi.org/10.1038/sigtrans.2017.23

Lüttichau HR (2010) The cytomegalovirus UL146 gene product vCXCL1 targets both CXCR1 and CXCR2 as an agonist. J Biol Chem 285(12):9137–9146. https://doi.org/10.1074/jbc.M109.002774

Maginnis MS (2018) Virus-receptor interactions: the key to cellular invasion. J Mol Biol 430(17): 2590–2611. https://doi.org/10.1016/j.jmb.2018.06.024

Maussang D, Verzijl D, van Walsum M, Leurs R, Holl J, Pleskoff O, Michel D, van Dongen GAMS, Smit MJ (2006) Human cytomegalovirus-encoded chemokine receptor US28 promotes tumorigenesis. Proc Natl Acad Sci U S A 103(35):13068–13073. https://doi.org/10.1073/pnas.0604433103

Mebatsion T, Weiland F, Conzelmann KK (1999) Matrix protein of rabies virus is responsible for the assembly and budding of bullet-shaped particles and interacts with the transmembrane spike glycoprotein G. J Virol 73(1):242–250. https://doi.org/10.1128/JVI.73.1.242-250.1999

Merida I, Arranz-Nicolás J, Torres-Ayuso P, Ávila-Flores A (2019) Diacylglycerol kinase malfunction in human disease and the search for specific inhibitors. In: Handbook of experimental pharmacology, vol 259, pp 133–162. https://doi.org/10.1007/164_2019_221

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30(16):2785–2791. https://doi.org/10.1002/jcc.21256

Murgueitio MS, Bermudez M, Mortier J, Wolber G (2012) In silico virtual screening approaches for anti-viral drug discovery. Drug Discov Today Technol 9(3):e219–e225. https://doi.org/10.1016/j.ddtec.2012.07.009

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. Science (New York, N.Y.) 289(5481):920–930. https://doi.org/10.1126/science.289.5481.920

Noha SM, Jazzar B, Kuehnl S, Rollinger JM, Stuppner H, Schaible AM, Werz O, Wolber G, Schuster D (2012) Pharmacophore-based discovery of a novel cytosolic phospholipase A2α inhibitor. Bioorg Med Chem Lett 22(2):1202–1207. https://doi.org/10.1016/j.bmcl.2011.11.093

Pala N, Stevaert A, Dallocchio R, Dessì A, Rogolino D, Carcelli M, Sanna V, Sechi M, Naesens L (2015) Virtual screening and biological validation of novel influenza virus PA endonuclease inhibitors. ACS Med Chem Lett 6(8):866–871. https://doi.org/10.1021/acsmedchemlett.5b00109

Parry CM, Simas JP, Smith VP, Stewart CA, Minson AC, Efstathiou S, Alcami A (2000) A broad spectrum secreted chemokine binding protein encoded by a herpesvirus. J Exp Med 191(3): 573–578. https://doi.org/10.1084/jem.191.3.573

Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK (2021) Artificial intelligence in drug discovery and development. Drug Discov Today 26(1):80–93. https://doi.org/10.1016/j.drudis.2020.10.010

Pillaiyar T, Laufer S (2022) Kinases as potential therapeutic targets for anti-coronaviral therapy. J Med Chem. 65(2):955–982. https://doi.org/10.1021/acs.jmedchem.1c00335

Ploubidou A, Way M (2001) Viral transport and the cytoskeleton. Curr Opin Cell Biol 13(1): 97–105. https://doi.org/10.1016/s0955-0674(00)00180-0

Pontejo SM, Murphy PM (2017) Chemokines encoded by herpesviruses. J Leukoc Biol. 102(5): 1199–1217. https://doi.org/10.1189/jlb.4RU0417-145RR

Rajamaki S, Innitzer A, Falciani C, Tintori C, Christ F, Witvrouw M, Debyser Z, Massa S, Botta M (2009) Exploration of novel thiobarbituric acid-, rhodanine- and thiohydantoin-based HIV-1

integrase inhibitors. Bioorg Med Chem Lett 19(13):3615–3618. https://doi.org/10.1016/j.bmcl.2009.04.132

Roychoudhury S, Das A, Sengupta P, Dutta S, Roychoudhury S, Choudhury AP, Ahmed ABF, Bhattacharjee S, Slama P (2020) Viral pandemics of the last four decades: pathophysiology, health impacts and perspectives. Int J Environ Res Public Health 17(24):9411. https://doi.org/10.3390/ijerph17249411

Ryu K, Kim ND, Il CS, Han CK, Yoon JH, No KT, Kim K-H, Seong BL (2009) Identification of novel inhibitors of HCV RNA-dependent RNA polymerase by pharmacophore-based virtual screening and in vitro evaluation. Bioorg Med Chem 17(8):2975–2982. https://doi.org/10.1016/j.bmc.2009.03.024

Ryzhova EV, Vos RM, Albright AV, Harrist AV, Harvey T, González-Scarano F (2006) Annexin 2: a novel human immunodeficiency virus type 1 Gag binding protein involved in replication in monocyte-derived macrophages. J Virol 80(6):2694–2704. https://doi.org/10.1128/JVI.80.6.2694-2704.2006

Schuster D, Kowalik D, Kirchmair J, Laggner C, Markt P, Aebischer-Gumy C, Ströhle F, Möller G, Wolber G, Wilckens T, Langer T, Odermatt A, Adamski J (2011a) Identification of chemically diverse, novel inhibitors of 17β-hydroxysteroid dehydrogenase type 3 and 5 by pharmacophore-based virtual screening. J Steroid Biochem Mol Biol 125(1–2):148–161. https://doi.org/10.1016/j.jsbmb.2011.01.016

Schuster D, Markt P, Grienke U, Mihaly-Bison J, Binder M, Noha SM, Rollinger JM, Stuppner H, Bochkov VN, Wolber G (2011b) Pharmacophore-based discovery of FXR agonists. Part I: Model development and experimental validation. Bioorg Med Chem 19(23):7168–7180. https://doi.org/10.1016/j.bmc.2011.09.056

Sen B, Johnson FM (2011) Regulation of SRC family kinases in human cancers. J Signal Transduct 2011:865819. https://doi.org/10.1155/2011/865819

Shahid N, Rappon T, Berta W (2019) Applications of artificial neural networks in health care organizational decision-making: a scoping review. PloS One 14(2):e0212356–e0212356. https://doi.org/10.1371/journal.pone.0212356

Sharma P, Sethi G, Tripathi MK, Rana S, Singh H, Kaur P (2023) Role of nature-inspired intelligence in genomic diagnosis of antimicrobial resistance BT. In: Raza K (ed) Nature-inspired intelligent computing techniques in bioinformatics. Springer Nature, Singapore, pp 223–245. https://doi.org/10.1007/978-981-19-6379-7_12

Shrivastava R, Yasir M, Tripathi M, Singh P (2016) In silico interaction of methyl isocyanate with immune protein responsible for Mycobacterium tuberculosis infection using molecular docking. Toxicol Ind Health 32(1):162–167. https://doi.org/10.1177/0748233713498447

Simidjievski N, Bodnar C, Tariq I, Scherer P, Andres Terre H, Shams Z, Jamnik M, Liò P (2019) Variational autoencoders for cancer data integration: design principles and computational practice. Front Genet 10(1205). https://doi.org/10.3389/fgene.2019.01205

Sliwoski G, Kothiwale S, Meiler J, Lowe EW (2014) Computational methods in drug discovery. Pharmacol Rev 66(1):334–395. https://doi.org/10.1124/pr.112.007336

Sohraby F, Bagheri M, Aryapour H (2019) Performing an in silico repurposing of existing drugs by combining virtual screening and molecular dynamics simulation. In: Clifton NJ (ed) Methods in molecular biology, vol 1903, pp 23–43. https://doi.org/10.1007/978-1-4939-8955-3_2

Sorrell FJ, Szklarz M, Abdul Azeez KR, Elkins JM, Knapp S (2016) Family-wide structural analysis of human numb-associated protein kinases. Structure 24(3):401–411. https://doi.org/10.1016/j.str.2015.12.015

SYSTÈMES D (2016) BIOVIA discovery studio. In Dassault Syst mes BIOVIA, Discovery Studio Modeling Environment, Release 2017 Dassault Syst mes. http://accelrys.com/products/collaborative-science/biovia-discovery-studio/

Takaya D, Yamashita A, Kamijo K, Gomi J, Ito M, Maekawa S, Enomoto N, Sakamoto N, Watanabe Y, Arai R, Umeyama H, Honma T, Matsumoto T, Yokoyama S (2011) A new method for induced fit docking (GENIUS) and its application to virtual screening of novel HCV

NS3-4A protease inhibitors. Bioorg Med Chem 19(22):6892–6905. https://doi.org/10.1016/j.bmc.2011.09.023

Tang B, Pan Z, Yin K, Khateeb A (2019) Recent advances of deep learning in bioinformatics and computational biology. Front Genet 10(214). https://doi.org/10.3389/fgene.2019.00214

Tripathi MK, Yasir M, Gurjar VS, Bose P, Dubey A, Shrivastava R (2015) Insights from the molecular docking of hydrolytic products of methyl isocyanate (MIC) to inhibition of human immune proteins. Interdiscip Sci 7(3):287–294. https://doi.org/10.1007/s12539-015-0012-3

Tripathi MK, Yasir M, Singh P, Tayubi IA, Gupta R, Shrivastava R (2016) Toxic effect of chemicals dumped in premises of UCIL, Bhopal leading to environmental pollution: an in silico approach. Asian Pac J Trop Dis 6(4):284–290. https://doi.org/10.1016/S2222-1808(15)61032-5

Tripathi MK, Sinha J, Srivastava SK, Kumar D (2019) Bioinformatics in skin cancer: a system biology approach to understanding the molecular mechanisms and it's regulations BT - skin aging & cancer: ambient UV-R exposure (A Dwivedi, N Agarwal, L Ray, AK Tripathi, eds). Springer, Singapore, pp 101–111. https://doi.org/10.1007/978-981-13-2541-0_10

Tripathi MK, Nath A, Singh TP, Ethayathulla AS, Kaur P (2021a) Evolving scenario of big data and Artificial Intelligence (AI) in drug discovery. Mol Divers 25(3):1439–1460. https://doi.org/10.1007/s11030-021-10256-w

Tripathi MK, Sharma S, Singh TP, Ethayathulla AS, Kaur P (2021b) Computational intelligence in drug repurposing for COVID-19 BT - computational intelligence methods in COVID-19: surveillance, prevention, prediction and diagnosis (K Raza, ed). Springer, Singapore, pp 273–294. https://doi.org/10.1007/978-981-15-8534-0_14

Tripathi MK, Ahmad S, Tyagi R, Dahiya V, Yadav MK (2022) Fundamentals of molecular modeling in drug design. In: Rudrapal M, CBT-CADD (CADD): FL-BM to S-BA Egbuna (eds) Drug discovery update. Elsevier, pp 125–155. https://doi.org/10.1016/B978-0-323-90608-1.00001-0

Trott O, Olson AJ (2009) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2). https://doi.org/10.1002/jcc.21334

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(18):1555–1572. https://doi.org/10.2174/156802608786786624

Waldmann M, Jirmann R, Hoelscher K, Wienke M, Niemeyer FC, Rehders D, Meyer B (2014) A nanomolar multivalent ligand as entry inhibitor of the hemagglutinin of avian influenza. J Am Chem Soc 136(2):783–788. https://doi.org/10.1021/ja410918a

Wolber G, Langer T (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model 45(1):160–169. https://doi.org/10.1021/ci049885e

Wolber G, Dornhofer AA, Langer T (2007) Efficient overlay of small organic molecules using 3D pharmacophores. J Comput Aided Mol Des 20(12):773–788. https://doi.org/10.1007/s10822-006-9078-7

Wu Y, Wei F, Tang L, Liao Q, Wang H, Shi L, Gong Z, Zhang W, Zhou M, Xiang B, Wu X, Li X, Li Y, Li G, Xiong W, Zeng Z, Xiong F, Guo C (2019) Herpesvirus acts with the cytoskeleton and promotes cancer progression. J Cancer 10(10):2185–2193. https://doi.org/10.7150/jca.30222

Yang Y, Zhu Z, Wang X, Zhang X, Mu K, Shi Y, Peng C, Xu Z, Zhu W (2021) Ligand-based approach for predicting drug targets and for virtual screening against COVID-19. Brief Bioinform 22(2):1053–1064. https://doi.org/10.1093/bib/bbaa422

Yángüez E, Hunziker A, Dobay MP, Yildiz S, Schading S, Elshina E, Karakus U, Gehrig P, Grossmann J, Dijkman R, Schmolke M, Stertz S (2018) Phosphoproteomic-based kinase profiling early in influenza virus infection identifies GRK2 as antiviral drug target. Nat Commun 9(1):3679. https://doi.org/10.1038/s41467-018-06119-y

Yin H, Jiang N, Shi W, Chi X, Liu S, Chen J-L, Wang S (2021) Development and effects of influenza antiviral drugs. Molecules 26(4):810. https://doi.org/10.3390/molecules26040810

Zhang J, Wang Y (2021) Bilobetin, a novel small molecule inhibitor targeting influenza virus polymerase acidic (PA) endonuclease was screened from plant extracts. Nat Prod Res 35(24): 5968–5971. https://doi.org/10.1080/14786419.2020.1808636

Zhang C, Xiang J, Xie Q, Zhao J, Zhang H, Huang E, Shaw P, Liu X, Hu C (2021a) Identification of influenza PAN endonuclease inhibitors via 3D-QSAR modeling and docking-based virtual screening. Molecules 26(23):7129. https://doi.org/10.3390/molecules26237129

Zhang M, Zhang L, Hei R, Li X, Cai H, Wu X, Zheng Q, Cai C (2021b) CDK inhibitors in cancer therapy, an overview of recent development. Am J Cancer Res 11(5):1913–1935

Chapter 14
CADD Approaches in Anti-inflammatory Drug Discovery

Nigam Jyoti Maiti and Nisha Kumari Singh

14.1 Introduction

Biology depends on chemistry because it offers a technique for examining and comprehending the chemicals that makeup cells (Kore et al. 2012). When it comes to the density, acidity, size, and shapes of atoms, scientists investigate the different matter, whereas biologists are interested in how living creatures react in its surroundings. To comprehend the hidden language of processes taking place in living creatures, a variety of essential sciences, including math, statistics, biology, chemistry, and physics must collaborate with the most recent advancements in research and innovation. Comprehensive overview of the procedures involved in the creation and finding of drugs, sizeable molecular information libraries about illnesses, DNA, RNA, proteins, and tiny molecules are generated using this technology (Ajith and Nair 2017). Computational design technologies are employed by all of the major pharmaceutical and biotechnology sectors (Kore et al. 2012). The contributions, at the simplest basic level, reflect the substitution of crude technique with structural representations that seem to be a far more realistic depiction of chemical truth and also can demonstrate motion and solvent effects (Kore et al. 2012). Moreover, estimates based on theoretical frameworks make it possible to estimate binding free energies and other relevant biological features. Empirical quantum mechanics, statistical mechanics, and molecular mechanics are some of the theoretical tools available. This most recent innovation has made it possible to incorporate explicit solvent effects. The availability of top-notch computer graphics, which are principally supported by workstations, is the foundation for this entire effort.

N. J. Maiti (✉) · N. K. Singh (✉)
Department of Pharmaceutical Sciences and Technology, Birla Institute of Technology, Ranchi, Jharkhand, India

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_14

14.2 Drug Development Strategy and CADD Ideas

Inside the present era of medication research, a molecular approach is used, in which the process underlying sickness or illness is described depending on the incorrect. Based on recent knowledge, computational models of molecular targets linked to particular pathogenic infection candidates are projected. Their interaction is explored using molecular mechanics and dynamics principles to determine the amount of energy linked to the molecules. To determine the optimum forecasting models, the actual mechanisms involved are significantly complicated and include a variety of mathematical, statistical, and data-analyzing techniques. Additionally, as the conclusions are simply descriptive, appropriate methodologies must be used to confirm and validate them (Ajith and Nair 2017). The newest technique employed in the medical field to solve the challenge of managing health and illness using computational simulation is called computer-aided drug discovery (CADD). According to the amount of molecular data known about the target (enzyme/receptor) and the ligands, several CADD strategies are used. The two main modeling techniques utilized for the drug design process are "direct" and "indirect" design. Creation in an indirect manner is based upon a comparison between characteristics of established compounds. Enzyme's characteristics are immediately considered indirect design.

Scientists might encounter little or no structure-related data in the initial stages of a drug discovery procedure. The high-throughput screening (HTS) group must now start developing and evaluating assays right away (Oldenburg 1998), and chemists must focus on filtered results or other early pieces of data right now. The molecules filtered might be naturally occurring substances, commonly available molecules, groups of internally generated molecules, or molecules originating from combinatorial libraries. However, researchers have to assist in selecting the molecules for HTS. Finding leads can be accomplished by choosing a group of molecules that exhibit variation in their physicochemical qualities instead of undertaking additional screening. Such techniques seek to examine and choose fewer compounds while learning from the data set (Moos et al. 1993). Any decrease in the count of molecules intended to be examined could have a targeted impact on the effectiveness of the investigation and the associated expenses (Ooma 2000), provided that it merely reduces redundancy within a database and does not introduce unwanted spaces. Two-dimensional fingerprints were employed in the experiments as molecular data with the effectiveness of rational selection approaches vs. a randomized technique was examined (Halliday et al. 1992; Moos et al. 1993; Martin 1998; Oldenburg 1998; Reigner and Blesch 2002). The initial step in generating a fresh lead, also known as a secondary lead, was generated based on the stereo-electronic characteristic of primary leads. The main leads ought to be chosen from a group of substances with a broad range of chemical compositions interacting with the same target through the same binding interaction. A pharmacophore is determined through analysis of the stereoelectronic characteristics of principal leads. A pharmacophore model is a modeling approach of elements or groups of elements that are assumed to be responsible for pharmacological action (Kubinyi 1993). The remaining molecular

components act as a skeleton in this method to keep the groups in their proper positions. The produced pharmacophores typically include 3–5 components spaced at regular intervals.

The Biomedical Information Science and Technology Initiative (BISTI) was established by the National Institutes of Health (NIH) a few years ago to assess the condition of bioinformatics in the country at the time. The bioinformatics procedure explained by BISTI included its application to scientific research, particularly for initiatives aimed at medication creation. Bioinformatics was viewed as an emerging method for locating medications, advancing those through clinical studies, and ultimately releasing those to the public (Bernard et al. 2005). A specialized field called computer-aided drug design (CADD) uses computational tools to investigate drug-receptor interactions. IT databases, software applications, and computing resources construct the framework for bioinformatics. CADD approaches are highly dependent on bioinformatics tools and applications. Bioinformatics techniques are heavily used in biological sciences, genomics, proteomics, other developing fields (including metabolomics and transcriptomics), and CADD research on the scientific side of the hub. In a number of crucial areas, bioinformatics aids CADD research.

Most CADD work focuses on proteins as targets with tiny compounds acting as ligands or leads, primarily because proteins have a significant structural impact.

CADD can be classified into two categories based on the strategy adopted:

1. *Structure-based method*
2. *Ligand-based methods*

The most widely used approach, *structure-based* CADD, depends on the target protein's structure. This method selects the lead compounds from a broad chemical space that best matches a certain target. The use of computer-aided methodologies is crucial in several of the group's efforts involving rational drug development. Studies into the atomic-level molecular pathways of ligand-target identification can be conducted using NMR spectroscopy in addition to molecular modeling and other spectral methodology techniques (Scherer et al. 2000). The development of innovative treatments and the foretelling of drug interactions with targets require this information. The group has also researched the specifics of ligand binding to DNA's minor groove using compounds like Hoechst 33,258 or tRNA (Irwin et al. 2002). The team also employs NMR techniques to investigate how proteins react with their ligands. The team has made this known to 500 MHz high-field instruments installed at the Chemistry Department as well as 300 MHz instrumentation available within the institution. The team works closely with Professor Gareth Morris, the creator and forerunner of numerous contemporary NMR methodologies, applying cutting-edge methods to complex biological issues (Taft 2008).

The QSAR models used in the *ligand-based* technique are based on the chemical search for structural analogs (Ajith and Nair 2017). This strategy is used when the target structure is known, and lead compounds are selected based on how closely they resemble existing ligands that are effective against particular therapeutic actions.

The structure may be created using CADD tools and protein sequence data. Comparative protein modeling, also known as template modeling, is the process of determining an unknown protein structure by comparing sequence data to a known protein structure known as the "template." This includes

(i) *Homology* modeling, where the template is chosen based on the highest degree of sequence similarity
(ii) *Threading*, wherein more than one template is chosen for the entire range of protein sequences

Now a days search for a therapeutic molecule that binds to the body's protein of interest was carried out with the advanced approaches of CADD (Balaban et al. 1994). CADD and in silico drug design are nearly synonymous. Almost all stages of the drug development process, from target selection to lead identification, from best optimization to preclinical or clinical trials, are now covered by computer-aided drug design (CADD) platforms. This is especially true in the post-genomic era.

The Three-Step In Silico Drug Discovery Process

(a) Stage 1: Entails selecting a pharmacological target and creating a diverse collection of small molecules to be evaluated against it. The creation of a high-throughput screening technique comes next, which is initially started by either dock small molecules from the data source or creating these structures in the active site using de novo design techniques.
(b) Stage 2: These selected picks are docked at known drug target binding sites to test their selectivity.
(c) Stage 3: The molecules that pass these detailed in silico ADMET assessment investigations on these chosen hits are called leads.

Docking is the in silico simulation of a molecular interaction achieved by interacting the target protein with a ligand (Müller 1994). This interaction provides several conformations from which the protein-ligand complex with the lowest energy is considered. Rigid body docking occurs when the docking is done without the molecules changing their conformation, whereas flexible docking allows the molecules to vary their shape to find the optimal match (Williams and Spector 2009).

The protein and the ligand molecules are cleaned and prepared before the docking procedure to remove unwanted atoms, ligands, and charges, finally reducing the energy of the ligands. When a large dataset of ligands is involved, virtual screening is processed for filtering or screening ligand molecules to locate target compounds with the highest affinity for a therapeutic target. The best druggable compounds will be the leads chosen (NAIR et al. 2017) (Fig. 14.1).

Its first crucial step in the pipeline for discovering new drugs is the selection and confirmation of targets. Nevertheless, selecting and validating drug-able targets from among the tens of thousands of potential macromolecules remains a difficult challenge. Currently, many techniques have already been created to address the aims. Target detection systems that use proteomic and genomic methodologies are among the most common. A proteomic technique, for instance, compares the protein

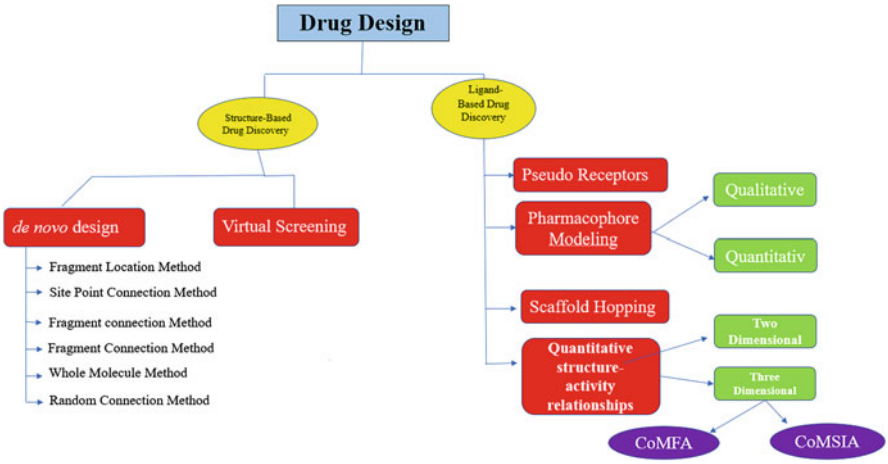


Fig. 14.1 Classification of computer-aided drug discovery (Aparoy et al. 2012)

expression profiles of a specific cell or tissue in the presence or absence of a given small chemical to find binding proteins for that cell or tissue.

The use of computational methods to forecast drug safety during drug creation and research is increasingly popular. Late-stage clinical trials or post-marketing significant unfavorable toxicological results for medicine might result in significant financial losses and put patients in danger. It is preferable if such compounds are discovered and the drug discovery process is stopped as soon as possible.

Metabolic drug-drug interactions (MDDI) have already brought up certain high-profile issues in the creation of new drugs, leading to their limited use, removal, or denial by regulatory agencies. In vitro methods are now frequently used in the medication planning process to assess the possibility of MDDI. However, in the absence of a comprehensive strategy, their interpretation and worth remain up for debate, and the critical distinction between a helpful "simulation" and an accurate "prediction" is not frequently understood. There is already a variety of in silico software that can simulate MDDI. SIMCYP is one of these programs (Ajith and Nair 2017).

Chosen lead compounds are docked against the biological target during virtual screening. The grading pattern comes next. For this, a variety of software is offered. A few are freely usable, while others are offered for sale.

14.3 Inflammation and Its Mechanisms

Inflammation is a protective response involving immune cells, blood vessels, and chemical mediators that play a critical role in the intricate biological response of human tissues to potentially harmful stimuli, such as pathogens, damaged cells, or

irritants. WBC of your body releases substances during inflammation to defend the human system against foreign intruders. As a result, the area of injury or illness receives increased blood flow. Burning and swelling may result from it. Some chemicals lead to edema in human cells by causing fluid flow into them. This defense mechanism may irritate the neurons and hurt them. Increased levels of white blood cells along with substances produced within the joints throughout time irritate the joint lining, create inflammation, and lead to joint damage (the cushioning at the ends of bones). The systems may become inflamed as a result of an inflammatory illness. The afflicted organs will determine the symptoms. For instance:

Myocarditis, an inflammation of the heart, can result in edema or breathing difficulties.

The tiny tubes that carry air to your lungs may become inflamed and result in breathing difficulties.

Nephritis, an inflammation of the kidneys, can result in kidney failure or elevated blood pressure.

Since many organs lack pain-sensitive receptors, people may not even experience discomfort when they have an inflammatory condition (Ahmed 2011). It is characterized by flushed skin at the site of the injury, pain or tenderness, joint pain, abdominal pain, fatigue, fever, swelling, chest pain, etc. There are two types of inflammation (Ansar and Ghosh 2016):

(i) *Acute inflammation*: The reaction to an immediate physical injury, like cutting your finger, and the body sends inflammatory cells to the wound to speed up healing. Finally, the healing process is proceeded by these cells (Bensman 2020).
(ii) *Chronic inflammation*: Even when there is no threat from the outside, the body keeps releasing inflammatory cells. For instance, in rheumatoid arthritis, inflammatory cells and chemicals assault the joint tissues, causing an intermittent inflammation that can seriously harm joints and result in pain and deformity (Dagvadorj et al. 2008).

One of the most frequent causes of chronic inflammation is autoimmune diseases, where the system assaults healthy tissue, including lupus.

Certain lifestyle variables also influence inflammation in the body. Chronic inflammation may be more likely to occur when someone abuse alcohol, except if muscular; have a high body mass index (BMI) that falls within the boundaries for overweight, if one do not workout sufficiently or regularly at utmost effort; and have persistent tension, smoke (Kore et al. 2012).

Numerous medications can reduce pain, edema, and inflammation. Additionally, they might lessen or stop the inflammatory condition. Physicians frequently recommend more than one. The drugs consist of medications that are non-steroidal anti-inflammatory (NSAIDs, such as naproxen, aspirin, or ibuprofen), corticosteroids (such as prednisone) (such as prednisone), antibiotics for malaria (such as hydroxychloroquine), and various other medications, including sulfasalazine, cyclophosphamide, leflunomide, methotrexate, and azathioprine.

Tocilizumab, adalimumab, certolizumab, etanercept, infliximab, golimumab, rituximab, and abatacept are examples of biological medications (Ahmed 2011).

By adopting healthy lifestyles, people can lower their risk of chronic inflammation. These practices include getting to and maintaining a healthy weight, limiting or stopping smoking, physical workout at a minimum of three to four times a week (daily exercise is best), reducing your alcohol intake (maximum 2 ounces per day), and reducing stress using beneficial techniques like journaling or yoga (Kore et al. 2012).

14.3.1 Mechanism of Inflammation (Muzamil et al. 2021)

The mechanism of inflammation is expressed in Fig. 14.2.

14.3.2 Molecular Process of Inflammation

Pattern recognition receptors (PRRs) are explained by cells of both the adaptive and innate immune systems, with unique intracellular receptors that the host cells use to first recognize inflammatory stimuli. PRRs are germline-encoded receptors that detect the occurrence of cell injury in addition to the existence of infectious microbes. Researchers achieve it by recognizing endogenous chemicals caused by internal damage, known as danger-associated molecular patterns (DAMPs), with

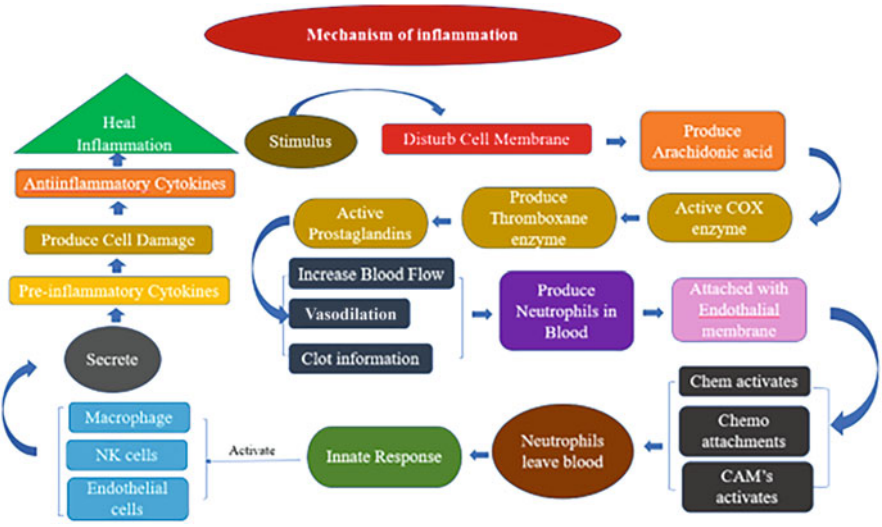


Fig. 14.2 Mechanism of inflammation

pathogen-associated molecular patterns (PAMPs), which are traits retained in microbes (DAMPs). So far, several PRRs have been discovered, including C-type lectin receptors (CLRs), NOD-like receptors, RIG-1-like receptors (RLRs), and Toll-like receptors (TLRs). Such PRRs have the specific ability to detect DAMPs, PAMPs, or both (NLRs). When such receptors connect with specific stimuli, signals are sent to the nucleus, where transcriptional and posttranscriptional mechanisms use them to activate a specific group of genes. The results of these genes specifically regulate the expression of proinflammatory cytokines like IL-1, IL-6, and TNF in response to bacterial effect. The production of IL-1's mRNA, or pro-IL-1 in the first stage, is started by the TLR-dependent expression of the IL-1 zymogen. The second phase is the generation of IL-1 by the caspase-1-mediated cleavage of pro-IL-1. This procedure necessitates a high molecular weight complex known as the inflammasome that "caspase-1-activates" caspase-1. NLRs are one of the scaffold proteins used in the oligomerization process to construct inflammasomes. When a virus is present, type-1 interferons trigger the phosphorylation and nuclear translocation of a complex known as interferon-stimulated gene factor 3 (ISGF3), which is made up of STAT1 and STAT2 as well as an interferon-regulatory factor (IRF) 3. The antiviral gene's protein kinase R (PKR) and 2′,5′-ISGF3 activate oligoadenylate synthase in turn (OAS). PKR inhibits the growth of virus-infected cells, whereas OAS prevents viral replication by cleaving viral nucleotides. The synthesis of proinflammatory cytokines and chemokines is regulated by a similar set of signaling pathways that are activated by signal transductions from PRRs. Identifying transcription factors and DNA motifs on their target genes occupied a significant portion of earlier investigations. The selected transcription factor with sequence-specific DNA binding activity that the right stimulus can only activate is NKB. One of the most researched transcriptional regulators, NK-B, has shed light on the complex control system that allows for the targeted activation of a particular subset of gene expressions. The mammalian NKB family comprises some specific proteins. The Rel homology region (RHR), which promotes the development of permanent homodimers and heterodimers, is a structural homology region that now the members of the NK-κB family share with the retroviral oncoprotein v-Rel in their N terminus. I-B proteins with ankyrin repeats are responsible for keeping the majority of NK-κB proteins in the cytoplasm of unstimulated cells. The precursor proteins p105 and p100, which have an I-B-like ankyrin repeat domain at their C termini, are the building blocks for p50 and p52. To trigger specific gene expressions in response to stimulation, NK-κB dimers in the cytoplasm are liberated from I-B and translocated to the nucleus. Both the phosphorylation of I- κB, which results in its ubiquitylation and proteosome-mediated destruction and the inducible proteolytic cleavage of the p100:RelB heterodimer's ankyrin-repeat domain can be used to separate NK-κB dimers from I-κB. In addition to NK-κB, different transcription factors are essential for the targeted stimulation of inflammatory genes. To name a few, these are the activator protein-1 (AP-1), a heterodimer of the basic leucine zipper proteins c-Jun and c-Fos, cyclic AMP (cAMP) response element binding protein (CREB), a cAMP-induced factor E2F, a transcription factor activated by the adenovirus E1A protein in adenovirus-infected cells serum responses factor (SRF),

and the associated term. Numerous posttranslational mechanisms, including the phosphorylation or dephosphorylation of these transcription factors or their inhibitors, are required to stimulate these signaling pathways in reaction to inflammatory stimuli. A rising body of research indicates that a variety of epigenetic processes tightly regulate the regulation of inflammation. Since the acetylation of histones is linked to chromatin configurations that are relaxed and allow transcription, it appears to be essential for the stimulation of numerous inflammatory genes. For instance, it has been shown that acetylation of histone H3 at the promoter region of numerous inflammatory genes causes an increase in the recruitment of NF-B to these promoters during inflammation. It is widely acknowledged that numerous inflammatory genes are activated by histone acetylation, whereas histone deacetylase (HDAC) activity represses these genes. By retaining chromatin in a flexible or compacted condition, however, histone methylation, on the other hand, can either activate or repress gene transcription (Ahmed 2011).

14.4 Anti-inflammatory Drugs and Their Classification and Mechanism of Action (MOA)

COX-1 and COX-2 are the two cyclooxygenase isoenzymes. In the body, COX-1 is constitutively produced and is involved in maintaining the lining of the gastrointestinal tract, renal function, and platelet aggregation. The body does not express COX-2 constitutively; rather, it is induced to do so during an inflammatory reaction. The major NSAIDs are nonselective COX-1 and COX-2 inhibitors. The adverse effect profile of COX-2-selective NSAIDs, such as celecoxib, is distinct since they solely target COX-2. Importantly, COX-2-selective NSAIDs should offer the anti-inflammatory treatment without harming the gastric mucosa because COX-2 is primarily involved in inflammation, and COX-1 is the primary mediator for maintaining gastric mucosal integrity.

Patients with rheumatoid arthritis have a concentration of a PGE2-like chemical in their synovial fluid of about 20 ng/ml. Aspirin users saw a drop in this value to zero, confirming the drug's clinically proven impact on PG synthesis (Higgs et al. 1974). Rats were given s.c. implants of polyester sponges filled with carrageenan to cause artificial inflammation (Simmons et al. 1983). The concentration of PGE2 grew over the course of the 24-hour trial, according to the periodic assessment of the inflammatory exudate present inside the sponges. TXA2 and LTB4 output also increased, peaked around 4–6 h, and then decreased throughout the duration of the experiment. Vasodilatation and hyperalgesia are brought on by PGE2, and polymorphonuclear leukocytes are likely drawn to the area by LTB4's chemotactic properties (Ford-Hutchinson et al. 1984). TXA2's function in the inflammatory response is not fully understood, though. Carrageenan was used to cause inflammation in the rat paw, providing proof that PGs play a part in the inflammatory response. Aspirin stopped the release of endogenous PGs, and when low dosages of exogenous PGE2

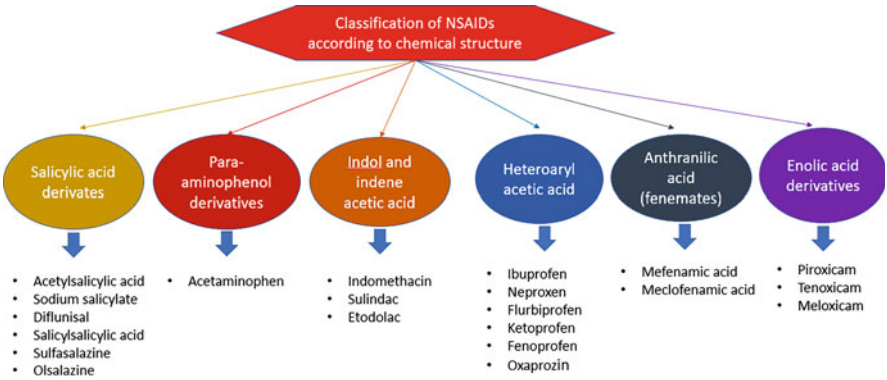


Fig. 14.3 Classification of NSAIDs

(1.0 ng) or prostacyclin (10 ng) were administered, the edema increased (Moncada et al. 1973).

Experimentally, it was determined that aspirin-like medications had no effect on the release of histamine or bradykinin, and more research was conducted to demonstrate a correlation between the anti-enzyme action of aspirin-like medications and their anti-inflammatory properties. Tomlinson et al. (1972) compared the effects of naproxen's two optical isomers and found that the one with anti-inflammatory effects (in adjuvant arthritis and carrageenan edema) was also a strong inhibitor of PGE2 generation. In every test, the other isomer was significantly less active. Even after accounting for protein binding, peak plasma concentrations of a wide variety of NSAIDs at therapeutic doses were more than enough to prevent PG production in a solitary enzyme preparation (Flower 1974). It was important to find out whether an analogous mechanism underpins the side effect profile of aspirin once it was established that the anti-inflammatory actions of NSAIDs are mediated by the inhibition of PG production. Aspirin's potential to cause ulcers has been discussed, and it is now known that prostacyclin is a crucial cytoprotective substance produced by the gastric mucosa. Experimental stomach ulcers can be reversed or prevented by administering various PGs, and some recently produced PG compounds are now accessible for clinical usage. On the other hand, the ability of certain NSAIDs to erode the stomach mucosa is correlated with their anti-enzyme activity. NSAIDs inhibit the development of mucosal PG in the clinic. Salicylate has a very low erosion index, but it also reduces PG concentration in inflammatory exudate without compromising stomach mucosa formation (Whittle et al. 1980). Salicylate differs from other aspirin-like medications in this way for unknown reasons (Fig. 14.3).

MOA of Steroidal Drugs in Inflammation

The action of phospholipase A2, which is required for the releasing of AA, is inhibited by corticosteroids. Corticosteroids finally prevent the development of PGs, TX, and LTs. By causing the production of an antagonistic protein, anti-inflammatory steroids inversely block phospholipase A2. There have been reports

of molecules as large as 15, 30, and 40 kDa for this substance, which has also been referred to as macrocortin, lipomodulin, or renocortin. A pure, copied form of lipocortin, known by the name agreed upon (Flower 1986) and recently made accessible, is said to be a strong anti-inflammatory drug (Wallner et al. 1986). Since lipocortins and calpactins seem to have the same mechanism of action, there is now some debate regarding this mechanism. It has been proposed that the ability of calpactins to link calcium and phospholipid, instead of directly inhibiting phospholipase A2, is what causes the decrease in eicosanoid production (Davidson et al. 1987).

14.5 Anti-inflammatory Drug Discovery Using CADD

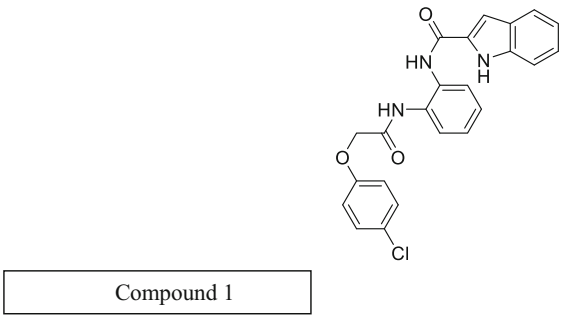
Biologically potent substances including prostaglandins (PGs), leukotrienes (LTs), and epoxyeicosatrienoic acids (EETs), usually considered as eicosanoids, are produced through the cyclooxygenase, lipoxygenase, and epoxygenase pathways. The production of prostaglandins (PGs) from the substrate AA is catalyzed by cyclooxygenase (COX), known as prostaglandin H2 synthase (PGHS). Prostaglandin (PG), thromboxane, and prostacyclin biosynthesis's first two steps are catalyzed by the membrane-bound enzyme COX. The first step involves oxygenating polyunsaturated fatty acids to produce hydroperoxy endoperoxide-PGG2 (cyclooxygenase activity), whereas the next step involves reducing PGG2 to hydroxy endoperoxide-PGH2 (cyclooxygenase activity) (peroxidase activity).

Whereas the cyclooxygenase (COX) enzyme is present until now. It is thought to be in control of catalyzing the conversion of AA to PGG2. It is recently known that this enzyme exhibits with two different isoforms as COX-1 and COX-2. Specifically, COX-2 was found to be an inducible variant that regulates inflammation and other progressive illnesses, including cancer, while COX-1 is a constitutive cytoprotective enzyme. Specialized prostaglandins made by the COX-1 enzyme in the stomach protect the lining's natural saliva production. Stomach cavity classic NSAIDs include medications like aspirin and ibuprofen. Naproxen inhibits both COX-1 and COX-2 isoenzymes. Consistent use, therefore, results in ulcerogenic side effects. On the other hand, the new generation of COX-2-selective NSAIDs has no such gastrointestinal adverse effects (Ahmed 2011).

14.6 Anti-inflammatory Drugs Discovered Using CADD Approaches: An Update

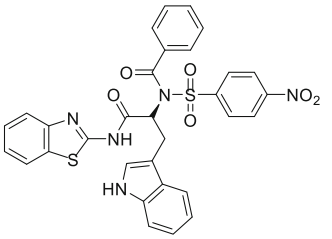
Fares et al. (2019) reported N-(2-(2-(4-chlorophenoxy)acetamido)phenyl (compound 1), a novel indole acetamide. The chemicals N-(2-aminophenyl)-2-(4-chlorophenoxy)acetamide and 1H-indole-2-carboxylic acid were stirred in dry

dichloromethane (DCM), followed by the addition of lutidine, and N,N,N′,N′-O-(benzotriazole-1-yl)-tetramethyluronium tetrafluoroborate (TBTU) in spectroscopic analyses was used to describe the resulting chemical (MS, FT-IR, 1H NMR, 13C NMR, UV-visible, and elemental). The in silico modeling study, which targets the cyclooxygenase COX-1 and COX-2 domains, verified the anti-inflammatory activity. With the use of single-crystal X-ray diffraction analyses, the three-dimensional structure was identified. Density functional theory computations using the B3LYP hybrid functional basis set were used to optimize the compound's geometry. It was discovered through a vibrational examination of the substance that the optimal structure is not in an excited state. To comprehend the electrical charge transfer within the molecule, the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) were examined. Hirshfeld surface analysis was used to examine the crystal's intermolecular interactions. Energy frameworks were built to look at the compound's stability. The various intramolecular interactions were verified using atom-in-molecule (AIM) computations (Al-Ostoot et al. 2020).



Compound 1

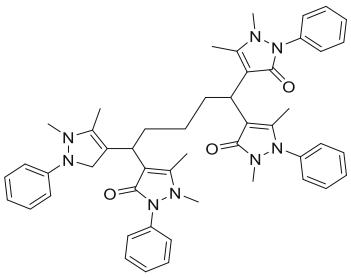
Mardia et al. (2018) reported novel heterocyclic compounds with antipyrine and pyrazolone moieties included have been devised and synthesized as part of a focused program for the development of new active drugs. Novel Mannich base derivatives have been developed beginning with the 4-arylidene-3-methyl-1-phenyl-5-pyrazolone derivative compound 2. For their ability to reduce inflammation, they were produced and biologically tested. Additionally, these substances' effectiveness as COX-1 and COX-2 inhibitors has been curiously investigated. Elemental analysis, IR, 1H NMR, 13C NMR, and mass spectrometry techniques were used to determine the structure of the produced compounds. For both COX-1 and COX-2, the high inhibition values were reported. The results acquired in vitro and in vivo have been validated using a molecular modeling technique. The findings can be applied to the development of more active agents in the future (el Sayed et al. 2018).



Compound 2

Compound 2

David et al. (2018) reported the in vivo anti-inflammatory, analgesic, and ulcerogenic effects of 12 novel benzothiazole compounds carrying benzenesulphonamide and carboxamide were studied. The synthesized compounds displayed excellent binding interactions with the receptors, with compound 3 displaying the highest binding energy (−12.50 kcal/mol). At 1 h, 2 h, and 3 h, respectively, compound 3 reduced carrageenan-induced rat paw edema at 72, 76, and 80%. In the analgesic activity trial, compound 3 had ED50 (mM/kg) values of 96, after 0.5 h; 102, 1 h; and 89 mM/kg after 2 h, respectively. These values were comparable to celecoxib's values of 156, 72, and 70 mM/kg. The ulcerogenic index for the most potent derivative, compound 3, was 0.82 compared to celecoxib's 0.92. The novel derivatives' physicochemical investigations demonstrated that oral bioavailability issues would not be an issue (Ugwu et al. 2018).

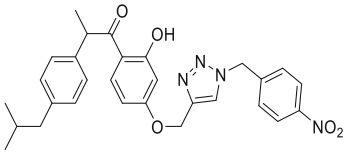


Compound 3

Compound 3

Kishore et al. (2016) reported new 1,4-disubstituted 1,2,3-triazole-containing compounds based on ibuprofen have been made using click chemistry, which has proven to be an effective method. When these substances were tested for their in vivo anti-inflammatory (AI) activity, compound 4 was found to have a more potent impact than the reference AI medication ibuprofen at the same dose (10 mg/kg body weight). There was noticeable AI activity in the compounds 4. The bactericidal

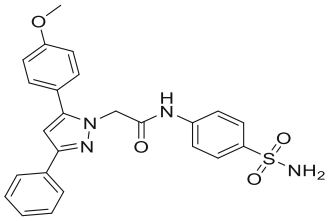
profile of these triazole compounds was also examined. Gram-positive and gram-negative organisms were both significantly resistant to the bactericidal effects of the compound. Additionally, in order to forecast the affinity and orientation of these novel drugs, molecular docking studies were also conducted on the cyclooxygenase-2 active site (Angajala et al. 2016).



Compound 4

Compound 4

Mohammed et al. (2014) reported the vital intermediary in the creation of novel hydrazones and pyrazole derivatives was 2-hydrazinyl-N-(4-sulfamoylphenyl)acetamide. All substances underwent tests to see how well they inhibited PGE2 synthesis in rat serum samples and had anti-inflammatory effects in vivo. The most effective drugs' in vitro IC50 values for inhibiting COX-1 and COX-2 enzymes were established, and the drugs were also examined for their ulcerogenic potential. To determine their manner of binding to the amino acids, molecular docking was used on the COX-2 active site. The majority of the produced compounds had strong anti-inflammatory action, particularly compound 5 which outperformed diclofenac as the reference medication. Compound 5 was less ulcerogenic than the reference medication indomethacin. In a molecular docking analysis, the majority of the produced compounds interacted with Tyr 385 and Ser 530, with compound 5 forming an extra hydrogen bond. In vitro, compound 5 had a high selectivity index value of 11.1 for COX1/COX-2 inhibition (Mohammed and Nissan 2014).



Compound 5

Compound 5

14.7 Case Study (Omar et al. 2018)

Conventional nonsteroidal anti-inflammatory drugs (NSAIDs) such as diclofenac, indomethacin, and aspirin are both COX inhibitors, i.e., COX-1 and COX-2 drugs reduce inflammation but increase the risk of kidney damage and stomach ulcers when used for a prolonged period of time. The removal of rofecoxib occurred in 2004 due to recent research linked with cardiovascular complications (Sibbald 2004).

In light of this, we think that selective inhibition of COX-2 and 15-LOX may be beneficial for lowering inflammatory diseases while limiting side effects with prospective application. The developed compound IV (Geronikaki et al. 2008) containing thiazole and 4-thiazolidinedione rings was demonstrated to be a strong COX-1 and 15-LOX inhibitor using CADD, while COX-2 was not inhibited until 200 M. A similar molecule, V, with a methoxyphenyl group on 4-thiazolidinedione was independently described and had superior 15-LOX inhibitory action (IC50 = 17.11 M) compared to quercetin. So focusing our target on COX-2 and 15-LOX, we designed a novel class of molecules comprising both 4-thiazolidinedione and 1,3,4-thiadiazole.

14.7.1 Design

Based on the examples of novel pharmacophores 1,3,4-thiadiazole and 4-thiazolidinone, we hypothesized the design of novel pharmacophores 1,3,4-thiadiazole and 4-thiazolidinone within the molecular framework may exhibit potent anti-inflammatory activity with fewer side effects. This hypothesis is consistent with our goal of developing thiazolidinone hybrids (Mahdavi et al. 2017) and anti-inflammatory drugs with COX-2 and LOX inhibitory effects (El-Nagar et al. 2018). It was anticipated that the two scaffolds combined would inhibit COX-2 (1,3,4-thiadiazole) and LOX (4-thiazolidinone) and give better level of selectivity for COX-2 over COX-1 enzyme. Because of their enormous bulk, they cannot fit in the small COX-1 pocket for binding (Blobaum and Marnett 2007). Compounds 1, 3, and 6a have been synthesized to test these theories. The molecular volume, potency, and selectivity of the new hybrid drug, which contains both 1,3,4-thiadiazole 6a and 4-thiazolidinone, were compared to those of its component molecules, 1,3,4-thiadiazole (3) and 4-thiazolidinone (1), against the COX-2 and 15-LOX enzymes.

14.7.2 Chemistry

Compound 1 (5Z), the compound 5-benzylidene-2-(4-hydroxyanilino)-1,3-thiazol-4 (5H)-one, is synthesized by utilizing a method that has already been described. With

a 50% total yield, the new intermediate compound 5 was created in four phases as per the designed scheme. In order to synthesize the compound, 2, 4-acetoxybenzoyl chloride was first treated with thiosemicarbazide in dry THF. When this product was then refluxed with phosphorus oxychloride, it was deacetylated and cyclized to create compound 3. This was verified by the IR spectra upon cyclization to produce compound 3. By carefully adding chloroacetyl chloride over the course of 5 h, compound 3 was chloroacetylated, resulting in compound 4. By refluxing compound 4 with ammonium thiocyanate in ethanol, heterocyclization to give compound 5 was accomplished. The effectively synthesizing intermediate compound 5 facilitated our study of the variety of aldehydes and the active methylene of the 4-thiazolidinone ring undergoing Knoevenagel condensation reactions as per the scheme. Addition of piperidine and the corresponding aldehyde to an ethanolic solution of compound 5 catalyzed this reaction, resulting in the formation of ylidene derivatives 6–9 in reasonably good yields of 50–83%. By analyzing with several spectral data, all products were identified.

14.7.3 Biology

In Vitro Studies of Cyclooxygenase Lipoxygenase Inhibitory Activity
Each compound showed that inhibiting capacity against both human recombinant COX-2 and ovine COX-1 was examined. Selection index (COX-1 IC50/COX-2 IC50) and the concentration of the investigated drugs that caused 50% inhibition (IC50, M) were computed. In comparison to analogs containing each scaffold separately, the hybrid 6a had superior selectivity and efficacy. It targeted both COX-2 and 15-LOX (compounds 1 and 3). Selection index of 54 for the COX-2 enzyme (SI of 3 = 27) and the COX-2 IC50 of 6a are both 0.085 M (IC50 of 3 = 0.33 M). IC50 for 15-LOX in 6a is 5.74 M (IC50 for 1 is 8.24 M). We were encouraged by these minute variations and hypothesized that further modifying the arylidene moiety of 6a might allow us to fine-tune the activity of hybrids towards the COX-2 and 15-LOX enzymes.

In Vitro Lipoxygenase Inhibitory Activity
All substances underwent testing for the soybean 15-LOX enzyme. The newly created compounds are more effective against the 15-LOX enzyme than the reference medication zileuton. Substance 5, which does not have an arylidene moiety, was the most active. The IC50 for compound 5 is 3.11 M. The most effective derivative of arylidene-containing compounds was 6a, which had an unsubstituted phenyl ring (IC50 of 5.74 M).

In Vivo Anti-inflammatory Activity
Rats were utilized to choose compounds (6a, 6f, 6i, 6 l, 6 m, and 9) for in vivo research using the carrageenan-induced paw edema method at a dose of 28 mol/kg. Diclofenac sodium and celecoxib were employed as benchmark medications. The data as edema inhibition percentage at 1-, 2-, 3-, and 4-h intervals was carried out.

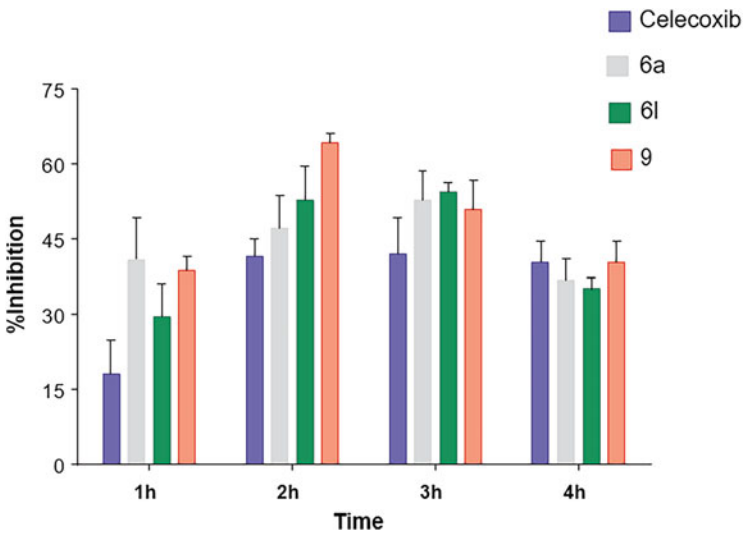


Fig. 14.4 Percentage of studied drugs' ability to reduce edema compared to celecoxib (6a, 6 l, and 9) (Omar et al. 2018)

All substances that were put to the test showed a progressive rise in anti-inflammatory activity that peaked after 3 h. The most potent derivatives were compounds (6a, 6 l, and 9), which also showed similar inhibitory action to the reference medicines (Fig. 14.4).

14.7.4 Molecular Modeling

Using the human COX-2 enzyme (5KIR), a docking study was conducted on COX-2 C, and MOE 2016.09 software was used to calculate volume, logP, and docking scores. The docking scores were between 7.018 and 7.948, and they roughly correlated with the produced drugs' IC50 values. The three compounds (6 l, 9, 6i), in terms of binding and interaction, explore good binding scores with the COX-2 enzyme. Compound 6 l, for example, forms three hydrogen bonds with Arg 513 (2.62), Phe 518 (2.05), and Ile 517 (2.42), as well as one hydrophobic interaction with Ala 527. Additionally, the hydrogen atoms Arg513 and he518 create two hydrogen connections with atom 9. Finally, 6i interacts hydrophobically with Ser353 and generates a pair of hydrogen bonds with Arg513 (2.79) and Leu352 (1.92).

Study of docking on 15-LOX. For the docking study, the human 15-LOX enzyme (4NRE) was employed. The docking scores, which ranged from 3.672 to 5.542, roughly correlated with the produced drugs' IC50 values. The molecular volume and IC50 of the produced compounds show a strong association, with smaller molecules

having higher potency than bigger ones. The three most effective chemicals (5, 6a, and 8) bind to and interact with Glu 613 by forming a hydrogen bond at a distance of 2.26, 2.25, and 2.28, respectively. Lys 612 and compound 8 create an extra hydrogen bond (2.36).

14.8 Conclusion and Future Scope

Based on merging the pharmacophores of COX inhibitor 2-amino-1,3,4-thiadiazole and a 15-LOX inhibitor 4-thiazolodine into a single component, the recent research established a unique, rational design as dual COX-2 and 15-LOX inhibitors. The experimental data clearly proved that the hybrids are significantly more potent and specific than the components of COX-2 and 15-LOX individually. Additionally, compounds 6a, 6f, 6i, 6 l, 6 m, and 9 demonstrated anti-inflammatory action that was on par with or even superior to celecoxib (Abdelall and Kamel 2016). Cyclooxygenase inhibitors are appealing and promising candidates for different inflammatory and cancerous diseases in humans. The COX-2 crystal structure could serve as a model for comparing human COX-1, 2, and 3 models to improve the understanding of the active sites involved in the protein-inhibitor binding mechanism. Currently, using the structure to be complex, multiple crystal structures of COX-2 complexes with inhibitors are readily available, allowing for employing methods for developing drugs that are based on structural knowledge of the targeted and powerful inhibitors. The technique focused on the structure to using docking studies; new COX-2 inhibitors can be created in addition to displaying active site-ligand interactions. Several methods and CADD approaches have been developed and are now being improved to exploit structural information for the design of analogs, which is becoming increasingly helpful. Comparing the predicted and experimental relative binding affinities for COX2 inhibitors with similar structural make-ups revealed qualitative, semi-quantitative, and quantitative agreement with experimental results for the molecular mechanic's techniques, QSAR, and FEP. These results unmistakably demonstrate that one can predict relative binding affinities using these CADD techniques prior to the production and biochemical testing of new analogs, accelerating, and lowering the cost of the drug development process.

References

Abdelall EKA, Kamel GM (2016) Synthesis of new thiazolo-celecoxib analogues as dual cyclooxygenase-2/15-lipoxygenase inhibitors: determination of regio-specific different pyrazole cyclization by 2D NMR. Eur J Med Chem 118:250–258

Ahmed AU (2011) An overview of inflammation: mechanism and consequences. Front Biol (Beijing) 6:274–281

Ajith V, Nair A (2017) Computer aided drug discovery – an overview

Al-Ostoot FH, Geetha D, Mohammed YHE et al (2020) Design-based synthesis, molecular docking analysis of an anti-inflammatory drug, and geometrical optimization and interaction energy studies of an indole acetamide derivative. J Mol Struct 1202:127244

Angajala KK, Vianala S, Macha R et al (2016) Synthesis, anti-inflammatory, bactericidal activities and docking studies of novel 1, 2, 3-triazoles derived from ibuprofen using click chemistry. Springerplus 5:1–15

Ansar W, Ghosh S (2016) Inflammation and inflammatory diseases, markers, and mediators: role of CRP in some inflammatory diseases. In: Biology of C reactive protein in health and disease. Springer, pp 67–107

Aparoy P, Kumar Reddy K, Reddanna P (2012) Structure and ligand based drug design strategies in the development of novel 5-LOX inhibitors. Curr Med Chem 19:3763–3778

Balaban AT, Basak SC, Colburn T, Grunwald GD (1994) Correlation between structure and normal boiling points of haloalkanes C1-C4 using neural networks. J Chem Inf Comput Sci 34:1118–1121

Bensman A (2020) Non-steroidal anti-inflammatory drugs (NSAIDs) systemic use: the risk of renal failure. Front Pediatr 7:517

Bernard D, Coop A, MacKerell AD Jr (2005) Computer-aided drug design: structure-activity relationships of delta opioid ligands. Drug Des Rev Online (Discontinued) 2:277–291

Blobaum AL, Marnett LJ (2007) Structural and functional basis of cyclooxygenase inhibition. J Med Chem 50:1425–1441

Dagvadorj J, Naiki Y, Tumurkhuu G et al (2008) Interleukin-10 inhibits tumor necrosis factor-α production in lipopolysaccharide-stimulated RAW 264.7 cells through reduced MyD88 expression. Innate Immun 14:109–115

Davidson FF, Dennis EA, Powell M, Glenney JR Jr (1987) Inhibition of phospholipase A2 by "lipocortins" and calpactins. An effect of binding to substrate phospholipids. J Biol Chem 262:1698–1705

el Sayed MT, El-Sharief MAMS, Zarie ES et al (2018) Design, synthesis, anti-inflammatory activity and molecular docking of potential novel antipyrine and pyrazolone analogs as cyclooxygenase enzyme (COX) inhibitors. Bioorg Med Chem Lett 28:952–957

El-Nagar MKS, Abdu-Allah HHM, Salem OIA et al (2018) Novel N-substituted 5-aminosalicylamides as dual inhibitors of cyclooxygenase and 5-lipoxygenase enzymes: synthesis, biological evaluation and docking study. Bioorg Chem 78:80–93

Flower RJ (1974) Drugs which inhibit prostaglandin biosynthesis. Pharmacol Rev 26:33–67

Flower RJ (1986) Background and discovery of lipocortins. Agents Actions 17:255–262

Ford-Hutchinson AW, Brunet G, Savard P, Charleson S (1984) Leukotriene B4, polymorphonuclear leukocytes and inflammatory exudates in the rat. Prostaglandins 28:13–27

Geronikaki AA, Lagunin AA, Hadjipavlou-Litina DI et al (2008) Computer-aided discovery of anti-inflammatory thiazolidinones with dual cyclooxygenase/lipoxygenase inhibition. J Med Chem 51:1601–1609

Halliday RG, Walker SR, Lumley CE (1992) R & D philosophy and management in the world's leading pharmaceutical companies. J Pharmaceut Med 2:139–154

Higgs GA, Vane JR, Hart FD, Wojtulewski JA (1974) Effects of anti-inflammatory drugs on prostaglandins in rheumatoid arthritis. Prostaglandin Synthetase Inhibitors 165–173

Irwin J, Lorber DM, McGovern SL et al (2002) Docking and drug discovery. Comput Nanosci Nanotechnol 2:50–51

Kore PP, Mutha MM, Antre R et al (2012) Computer-aided drug design: an innovative tool for modeling

Kubinyi H (1993) 3D QSAR in drug design: volume 1: theory methods and applications. Springer Science & Business Media

Mahdavi M, Saeedi M, Nadri H et al (2017) Synthesis of thiazolone derivatives as novel soybean 15-LOX inhibitors. Lett Org Chem 14:186–191

Martin YC (1998) Perspective in drug discovery and design, vol 12. Springer, New York, pp 3–23

Mohammed KO, Nissan YM (2014) Synthesis, molecular docking, and biological evaluation of some novel hydrazones and pyrazole derivatives as anti-inflammatory agents. Chem Biol Drug Des 84:473–488

Moncada S, Ferreira SH, Vane JR (1973) Prostaglandins, aspirin-like drugs and the oedema of inflammation. Nature 246:217–219

Moos WH, Green GD, Pavia MR (1993) Recent advances in the generation of molecular diversity. Annu Rep Med Chem 28:315–324

Müller K (1994) 5-lipoxygenase and 12-lipoxygenase: attractive targets for the development of novel antipsoriatic drugs. 5-lipoxygenase und 12-lipoxygenase: Attraktive target-enzyme für die Entwicklung neuer Antipsoriatika. Arch Pharm (Weinheim) 327:1–19

Muzamil A, Tahir HM, Ali S et al (2021) Inflammatory process and role of cytokines in inflammation: an overview. Punjab Univ J Zool 36:237–252

Nair V, Das KP, Banerjee S, Bagchi S (2017) Achieving oral health the natural way: pomegranate. J Med Health Res 2:110–118

Oldenburg KR (1998) Annual report in medicinal chemistry, vol 33. Academic Press, London, pp 301–307

Omar YM, Abdu-Allah HHM, Abdel-Moty SG (2018) Synthesis, biological evaluation and docking study of 1, 3, 4-thiadiazole-thiazolidinone hybrids as anti-inflammatory agents with dual inhibition of COX-2 and 15-LOX. Bioorg Chem 80:461–471

Ooma F (2000) Molecular modeling and computer aided drug design. Examples of their application in medicinal chemistry. Curr Med Chem 7:141–158

Reigner BG, Blesch K (2002) Estimating the starting dose for entry into humans: principles and practice. Eur J Clin Pharmacol 57:835–845

Scherer D, Dubois P, Sherwood B (2000) VPython: 3D interactive scientific graphics for students. Comput Sci Eng 2:56–62

Sibbald B (2004) Rofecoxib (Vioxx) voluntarily withdrawn from market. CMAJ 171:1027–1028

Simmons PM, Salmon JA, Moncada S (1983) The release of leukotriene B4 during experimental inflammation. Biochem Pharmacol 32:1353–1359

Taft CA (2008) Silva VB da, Silva CHT de P da. Current topics in computer-aided drug design. J Pharm Sci 97:1089–1098

Tomlinson R, Ringold HJ, Qureshi MC, Forchielli E (1972) Relationship between inhibition of prostaglandin synthesis and drug efficacy: support for the current theory on mode of action of aspirin-like drugs. Biochem Biophys Res Commun 46:552–559

Ugwu DI, Okoro UC, Ukoha PO et al (2018) Novel anti-inflammatory and analgesic agents: synthesis, molecular docking and in vivo studies. J Enzyme Inhib Med Chem 33:405–415

Wallner BP, Mattaliano RJ, Hession C et al (1986) Cloning and expression of human lipocortin, a phospholipase A2 inhibitor with potential anti-inflammatory activity. Nature 320:77–81

Whittle BJR, Higgs GA, Eakins KE et al (1980) Selective inhibition of prostaglandin production in inflammatory exudates and gastric mucosa. Nature 284:271–273

Williams FMK, Spector TD (2009) A new 5-lipoxygenase inhibitor seems to be safe and effective for the treatment of osteoarthritis. Nat Rev Rheumatol 5:132–133

Chapter 15
Drug Repurposing and Computational Drug Discovery for Viral Infections and COVID-19

M. V. Manohar, Amogha G. Paladhi, Bhoomika Inamdar, Kotthapalli Prashanth, Sugumari Vallinayagam, and Mithun Rudrapal

15.1 Introduction

In 2019, the world saw an outbreak of severe acute respiratory disorder causing virus called coronavirus that is pathogenic in humans. It is found that the virus also has an outbreak of pandemic in 2003. In 2012, the outbreak of the same virus termed as middle east respiratory syndrome-related coronavirus (MERS-CoV) had a higher mortality rate compared to that of SARS-CoV-2 of 2003 (Mohamed et al. 2020). The novel coronavirus disease (COVID-19) was emerged from Wuhan where it caused infections in Hubei Province, China in March 2020. WHO (World Health Organization) considered it as a pandemic and termed the virus as SARS-CoV-2 which had higher potential of spreading the infection (Hanaei and Rezaei 2020).

M. V. Manohar
JSS Medical College (Deemed to be University), Mysuru, Karnataka, India

A. G. Paladhi
CHIRST (Deemed to be University), Bengaluru, Karnataka, India

B. Inamdar
JSS Medical College (Deemed to be University), Mysuru, Karnataka, India

K. Prashanth
University of Mysore, Mysuru, Karnataka, India

S. Vallinayagam
Department of Biotechnology, Mepco Schlenk Engineering College, Sivakasi, India

M. Rudrapal (✉)
Department of Pharmaceutical Sciences, Vignan's Foundation for Science, Technology & Research, Guntur, Andhra Pradesh, India

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_15

Afterwards, variants of this virus were created through mutations, resulting in the manipulation of the transmission path of coronavirus and to more potential in spreading and duplication. Humans play an important role in the mutation of coronavirus as their major host to complete and replicate especially in the cells of respiratory zone of humans. Thus, governments of respective countries and WHO worked up few regulations for the citizens to prevent the infection and the transmission of the virus, the so-called COVID appropriate rules. It was the duty of civilians to bind to such rules to avoid the transmission. The regulations included social distancing, self-quarantine, use of sanitizers and frequent hand washing, and maintaining a hygienic environment because the transmission of virus took place due to droplets that come out of sneezing by infected persons (Rezaei 2020a, b). In spite of having preventive actions, the negligence and lack of education among people result in the widespread of the virus, failing the efforts of regulation. The origin, diagnosis, treatment, and management of transmission of coronavirus are challenges that still exist in healthcare system due to failure in abiding the regulations (Basiri et al. 2021; Moazzami et al. 2020).

15.2 Viral Infections and COVID-19

Viral infections in human are common all the time, but, in 2019, a new strain of pathogenic virus that is harmful to humans was found with the potential of causing a pandemic creating an emergency. The breakout of CoV-2 was found to be initiated in Wuhan Virology lab. Later, the citizens of the region who visited the sea food market noticed the abrupt deaths leading to interest on the cause of death. It was found that a virus that infects bats and other civets to complete its lifecycle had entered humans making them an important host. Mutations happened such that it gave rise to the potential of human to human transmission through droplets or aerosols caused by the spray of sneeze or cough. Coronavirus is a pneumonia causing class of virus. It contains spike proteins that are responsible for the pathogenicity of these viruses in host.

Classification of the virus responsible for current pandemic (2019–till date) is found to be Order Nidovirales, Family Coronaviridae, Subfamily Orthocoronavirinae, Genus *Betacoronavirus*, and subgenus *sarbecovirus*. The spike proteins attached to ACE2 receptor in humans are present in the respiratory tract. The structure of coronavirus typically contains genetic material, a protein coat with spike proteins that are present on the membrane of protein coat. These spike proteins (S protein) are found to adhere to ACE2 receptors that lead to the infection of virus in the host body, where the generic material of the virus is pushed inside the host cell and this material attaches itself to the DNA or genetic material of the host. Basically raw materials of the genome are prepared and produced and later the same codes are used for membrane proteins. Thus, all the components needed for each virus are prepared, and later it combines together in an appropriate manner.

15.3 Drug Repurposing

Repurposing or repositioning of drugs is defined as the therapy research of existential drugs. The aim of this method is to identify and connect the trinity of drugs, targets, and diseases. Over 400 human targets are used for current drugs, and these targets have been chosen from the genes within the human genome (Oprea and Overington 2015). Repurposing takes already approved therapy and screens for disease. This implies that novel therapies for diseases are created after an understanding of the underlying disease process. The advantage is that rectifying drugs is faster and more cost-effective option for diseases than conventional drug development methods (Scherman and Fetro 2020; Mahdian et al. 2020).

15.3.1 Drug Repurposing for COVID-19

The Food and Drug Administration (FDA) approves drugs for global release, and these drugs are then repurposed to develop disease therapies. Remdesivir, favipiravir, and ribavirin are drugs which are nucleotide analogue inhibitors used to treat COVID-19 viral infection (Beigel et al. 2020). These drugs work by interfering with the activity of RdRp (RNA dependent RNA polymerase), which prevents viral replication by inducing mutations. Fluorouracil and acyclovir are two additional drugs that are used to treat the disease (Dong et al. 2020).

15.3.1.1 Favipiravir

Favipiravir is an antiviral drug that was originally developed for influenza research and is now being used to treat COVID-19 infections. This is a pro-drug used in producing intracellularly phosphoribosylated active metabolite. This metabolite inhibits a broad array of RNA viruses, including *Arenavirus*, *Bunyavirus*, *Flavivirus*, and *Filoviruses* that cause hemorrhagic fever (Du and Chen 2020). This drug is taken orally and acts by suppressing RdRp of RNA viruses. By inducing mutations in nascent viral RNA, this drug contributes in the chain termination process. The metabolite acts as a mutagen after this mutation, blocking the coronavirus repair mechanism. This blockage may aid in decline of viral replication and infectious viral RNA particles (Joshi et al. 2021; Furuta et al. 2017). One research study looked at two groups, one with standard therapy and the other with favipiravir, and found that the group receiving favipiravir had a powerful impact against the infection and showed progressive clinical recovery (Sheahan et al. 2020). Another report looks at two groups of people—experimental and control—and found that the drug had a stronger antiviral effect than other drugs like lopinavir (Chu et al. 2004).

15.3.1.2 Remdesivir

Remdesivir, an RdRp inhibitor and antiviral drug, induces RNA mutation in SARS-CoV-2, making it a potential therapy option for COVID-19 disease. Remdesivir is a pro-drug, also known as a pro-tide drug that functions as a drug inhibiting viral replication in the host by converting to an active form (Saqrane et al. 2021). Remdesivir's metabolite (remdesivir triphosphate) interacts with adenosine triphosphate in terminating the RNA chain. Before terminating the developing RNA chain, Remdesivir inserts three additional nucleotides inhibiting exonuclease activities resulting in lack of resistance (Byléhn et al. 2021). Remdesivir is a drug that treats Ebola virus disease (EVD) and coronavirus infections including Middle-East respiratory syndrome (MERS) and severe acute respiratory syndrome (SARS) (Dyall et al. 2014). After failing to eradicate EVD, this drug was repurposed to treat SARS-CoV-2 infection (Tchesnokov et al. 2019). Some animal studies support the use of repurposed remdesivir in the treatment of the disease, as well as the drug's efficacy against coronaviruses (Sheahan et al. 2017; de Wit et al. 2020). In a new analysis, it was discovered that compassionate use of remdesivir improved clinical outcomes in 70% of severe COVID-19 patients in a cohort trial (Grein et al. 2020). Another trial found that remdesivir was better to placebo in reducing the time to recuperation in adult COVID-19 patients hospitalized (Wang et al. 2020a). Remdesivir is only administered intravenously in severe cases of infection, not in moderate cases, because it has side effects that the patient cannot tolerate, such as nausea and abrupt respiratory failure (Singh et al. 2020; Goldman et al. 2020).

15.3.1.3 Ribavirin

Ribavirin is an antiviral medication used to treat severe respiratory failure. Ribavirin is an RdRp inhibitor that suppresses viral replication and is useful in treating coronavirus infection. Ribavirin is a guanosine analog that interacts with DNA and RNA replication and interferes with RNA capping, limiting RNA degradation (Khalili et al. 2020). Several therapeutic interventions are implemented during a coronavirus disease epidemic, and repurposed drugs have also been found to have in vitro action against related subjects of SARS-CoV-2, a beta-coronavirus. Ribavirin and interferon-α are administered intravenously to individuals at high risk of disease development, and it is indicated in the treatment protocol for immediate use (Ali et al. 2020). This drug has the adverse effect of developing anemia and did not result in any recovery when administered during the SARS and MERS outbreaks, but it was progressively modified and repurposed for use during the COVID-19 outbreak. Ribavirin has a beneficial impact in treating coronavirus infection when administered prior to the onset of symptoms of pneumonia and organ failure. This drug was originally developed to treat Hepatitis C virus infection; however, it has now been repurposed to treat COVID-19 infection (Li et al. 2021). In a cohort trial, ribavirin was provided along with a combination of intravenous corticosteroids and

oral prednisolone, which had a favorable impact on the infection while omitting the antibiotic medication that had previously been used to diagnose the virus, and this study has not come across any adverse effects (Wang et al. 2020b). Another research was undertaken with a higher dose of the drug, which revealed certain adverse effects such as hemolysis in more than 70% of the trial participants, as well as liver toxicities based on increased transaminases. By monitoring the consequences, this therapy was promptly halted (Tong et al. 2020).

15.3.1.4 Darunavir

Although this medication is an HIV protease inhibitor and is not currently used to treat respiratory syndrome, its repurposing might be beneficial in the treatment of the condition. Despite the fact that it is ineffective in the therapy, it is coupled with a lower dose of ritonavir and lopinavir, a pharmacoenhancer. Despite being combined with ritonavir, lopinavir, or cobicistat, there is no conclusive proof or study that Darunavir is utilized in viral diseases other than HIV.

15.3.1.5 Ritonavir

In the treatment of respiratory syndrome, this antiviral medication ritonavir should be used in conjunction with lopinavir. Basically, lopinavir is an HIV-1 protease inhibitor that is also a SARS-COV protease inhibitor when used in combination. This medication administration resulted in a lower chance of infection, as well as a lower risk of adverse clinical outcomes and viral load.

15.3.1.6 Arbidol

Arbidol is now being used to treat influenza A and B viruses, as well as hepatitis C virus (HCV). ARB may inhibit virus interaction and penetration into host cells by preventing the lipid coat of virus from fusing with the cell membrane. ARB has been found to suppress COVID-19 infection.

15.3.1.7 Chloroquine and Hydroxychloroquine

For COVID infection, chloroquine was first suggested, followed by hydroxychloroquine. The combination of chloroquine and hydroxychloroquine promotes glycosylation of the ACE-2 receptor, to which the SARS-COV-2 binds, rendering the cells resistant to infection. Both drugs have immunomodulatory effects, and HCQ is currently widely used in the treatment of autoimmune disorders. They may be useful in treating the COVID infection and decreasing its severity by suppressing the immunological response to SARS-COV-2.

15.3.1.8 Tocilizumab

Tocilizumab is an immunosuppressive medication that is used to treat COVID-19 patients who are hospitalized with symptoms of pneumonia. Clinical studies for this drug, however, have had mixed results in COVID-19 disease. Fever had decreased substantially after treatment of this drug and symptoms such as cough and pulmonary inflammation had decreased with or without favipiravir combination.

15.3.1.9 Oseltamivir

Oseltamivir is a first-line antiviral medication, particularly in primary care settings. However, with continuing coronavirus disease 2019 (COVID-19), oseltamivir has been utilized by the majority of symptomatic COVID-19 patients. Oseltamivir must be evaluated in the treatment of COVID-19 due to its widespread use and critical role as an antiviral drug. This drug is neuraminidase inhibitor approved by FDA.

15.3.1.10 REGN-COV2

This is an antibody cocktail used in the treatment of COVID infection that has demonstrated to decrease viral load. REGN-COV2 is composed of two human immunoglobulins that target the receptor-binding region of the SARS-CoV-2 spike protein, preventing viral entry into human cells through the use of the angiotensin-converting enzyme 2 (ACE-2) receptor (Table 15.1).

15.4 Computational Methods

Considering the fact that discovery of the new drug is time consuming and expensive, various approaches have been made to treat the current outbreak of coronavirus in which drugs have been reused to treat the symptoms due to its positive results, analysis were made regarding the computational approach where different present day drugs that are used to treat other viral infections were repurposed by formulating them in such a manner that can possibly control or reduce the infection of coronavirus (Wang and Guan 2021).

15.4.1 Molecular Docking Methods

Molecular docking methods are the techniques used as desired approaches in drug repurposing due to its property of binding of ligands to proteins that have multidimensional structures. These are very helpful in approaching higher proteins (Masoudi-Sobhanzadeh et al. 2019).

Table 15.1 Drugs repurposed by computational approaches for antiviral therapies

Target proteins	Drug administered	References
SARS-CoV-2 main protease (Mpro or 3CLpro)	Darunavir, mitoxantrone, nelfinavir, moexipril, daunorubicin, rosuvastatin, saquinavir, metamizole, bepotastine, benzonatate, atovaquone	Mittal et al. (2021)
SARS-CoV-2 main protease (Mpro or 3CLpro)	Leupeptin, hemisulfate, pepstatin A, nelfinavir, birinapant, lypressin, octreotide	Mittal et al. (2021)
RdRp	Ribavirin, remdesivir, sofosbuvir, galidesivir, tenofovir, hydroxychloroquine, cefuroxime, favipiravir, setrobuvir, YAK, IDX-184	Elfiky (2020)
SARS-CoV-2 main protease (Mpro or 3CLpro)	Ritonavir, emetine, lopinavir, indinavir (only listed part of the results)	Das et al. (2021)
SARS-CoV-2 main protease (Mpro or 3CLpro), human transmembrane protease serine 2 (TMPRSS2)	Talampicillin, lurasidone, rubitecan, loprazolam (only listed part of the results)	Elmezayen et al. (2020)
SARS-CoV-2 main protease (Mpro or 3CLpro)	Perampanel, carprofen, celecoxib, alprazolam, trovafloxacin, sarafloxacin, ethyl biscoumacetate	Gimeno et al. (2020)
SARS-CoV-2 main protease (Mpro or 3CLpro), RdRp, Helicase, 3′–5′ exonuclease, endoRNAse, 2′-O-ribose methyltransferase	Atazanavir, ganciclovir, lopinavir, ritonavir, darunavir, and so forth (only listed part of the results)	Beck et al. (2020)
SARS-CoV-2 main protease (Mpro or 3CLpro), Spike (S) protein	Cangrelor, NADH, flavin adenine dinucleotide (FAD) adeflavin, comeprol, Coenzyme A, tiludronate, zanamivir, bortezomib, saquinavir, cangrelor, carfilzomib, indinavir, remdesivir	Hall and Ji (2020)
SARS-CoV-2 envelope (E) protein	Belachinal, macaflavanone E, vibsanol B	Gupta et al. (2021)
Spike (S) protein	Suramin sodium, 5-hydroxytrytophan, dihydroergocristine mesylate, quinupristin, nilotinib, dexamethasone-21-sulfobenzoate, tirilazad, selamectin, acetyldigitoxin, doramectin	de Oliveira et al. (2021)
Spike (S) protein or Spike (S) protein–ACE2 interface complex	Pemirolast, sulfamethoxazole, valaciclovir, sulfamerazine, tazobactam, nitrofurantoin	Batra et al. (2020)
Spike (S) protein	CR3022 human antibody, F26G19 mouse antibody, D12 mouse antibody	Park et al. (2020)

15.4.2 Network-Based Techniques

The network-based techniques are designed in such a way that they are usually not used to treat unique or rare diseases for which the metabolic pathway of the disease is known where the approach of molecules using the particular drug that is repurposed can be known by investigating the metabolic pathways. Because the metabolism of rare or unknown disease is not known, it is difficult for the designing of drugs to repurpose and also it risks the success rate of treatment (Masoudi-Sobhanzadeh et al. 2019).

15.4.3 Connectivity-MAP (CMAP) Methods

The technique includes relationship between the diseases and genes where it requires a large amount of genomic data. Connectivity-MAP method cannot be used in conditions where the subject are different cell lines or platforms where the data is not similar. Repurposing of the drugs in such cases does not concern the treatment and lacks success (Masoudi-Sobhanzadeh et al. 2019).

15.4.4 Data for Specific Goal

Obtaining data from various sources by different procedures such as machine learning and data mining is an important way to obtain the knowledge of novel drug uses which helps in repurposing specific drugs and also helps compute the drugs to a particular rare, unknown, or new disease based on the available data about those drugs. The repositioning or repurposing increases the chances of success in the use of computational drug therapies (Masoudi-Sobhanzadeh et al. 2019).

15.5 Summary and Conclusion

Techniques like repurposing drugs and computing them to treat a disease with less known or unknown data hinder the success rate of the treatment due to the failure of blocking molecules that causes pathogenicity in the host. Methods like network-based techniques and metabolic pathway-based methods are designed in such a way that the treatment works only if the metabolic pathway of the particular disease is well known. Also, C-MAP and data-specific treatment also depend on genetic information and the available data of the disease; enhancing the efficiency of disease approach can be done by monitoring the structure of the drugs used. There might be limitations due to the lack of data which can be corrected by the use of technology

and software engineering techniques and skills. The investigator or analyst that designed the repurposing of drugs should be aware of the cost effectiveness and time consumption.

From the above, we can conclude that repurposing of the drugs for computational drug therapy is one of the best ways of approach to treat diseases like COVID-19 and other viral issues. The method of synthesis of desired drug takes a longer time and also demands more economical support. Repurposing of drugs is done as per the available data due to which the efficiency of the prepared computational drug is pre-known and the data present helps in mapping medicine for viral infections and corona. Researchers must use the known data to engineer drugs in a sequence after repurposing.

References

Ali MJ, Hanif M, Haider MA, Ahmed MU, Sundas FNU, Hirani A, Khan IA, Anis K, Karim AH (2020) Treatment options for COVID-19: a review. Front Med 7(July):1–10. https://doi.org/10.3389/fmed.2020.00480

Basiri A, Heidari A, Nadi MF, Fallahy MTP, Nezamabadi SS, Sedighi M, Saghazadeh A, Rezaei N (2021) Microfluidic devices for detection of RNA viruses. Rev Med Virol 31(1):1–11. https://doi.org/10.1002/rmv.2154

Batra R, Chan H, Kamath G, Ramprasad R, Cherukara MJ, Sankaranarayanan SKRS (2020) Screening of therapeutic agents for COVID-19 using machine learning and ensemble docking studies. J Phys Chem Lett 11(17):7058–7065. https://doi.org/10.1021/acs.jpclett.0c02278

Beck BR, Shin B, Choi Y, Park S, Kang K (2020) Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput Struct Biotechnol J 18:784–790. https://doi.org/10.1016/j.csbj.2020.03.025

Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, Hohmann E, Chu HY, Luetkemeyer A, Kline S, Lopez de Castilla D, Finberg RW, Dierberg K, Tapson V, Hsieh L, Patterson TF, Paredes R, Sweeney DA, Short WR, Lane HC et al (2020) Remdesivir for the treatment of Covid-19 - final report. N Engl J Med 383(19):1813–1826. https://doi.org/10.1056/NEJMoa2007764

Byléhn F, Menéndez CA, Perez-Lemus GR, Alvarado W, De Pablo JJ (2021) Modeling the binding mechanism of Remdesivir, Favilavir, and ribavirin to SARS-CoV-2 RNA-dependent RNA polymerase. ACS Central Sci 7(1):164–174. https://doi.org/10.1021/acscentsci.0c01242

Chu CM, Cheng VCC, Hung IFN, Wong MML, Chan KH, Chan KS, Kao RYT, Poon LLM, Wong CLP, Guan Y, Peiris JSM, Yuen KY (2004) Role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. Thorax 59(3):252–256. https://doi.org/10.1136/thorax.2003.012658

Das S, Sarmah S, Lyndem S, Singha Roy A (2021) An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study. J Biomol Struct Dyn 39(9):3347–3357. https://doi.org/10.1080/07391102.2020.1763201

de Oliveira OV, Rocha GB, Paluch AS, Costa LT (2021) Repurposing approved drugs as inhibitors of SARS-CoV-2 S-protein from molecular modeling and virtual screening. J Biomol Struct Dyn 39(11):3924–3933. https://doi.org/10.1080/07391102.2020.1772885

de Wit E, Feldmann F, Cronin J, Jordan R, Okumura A, Thomas T, Scott D, Cihlar T, Feldmann H (2020) Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection. Proc Natl Acad Sci 117(12):6771 LP–6776776. https://doi.org/10.1073/pnas.1922083117

Dong L, Hu S, Gao J (2020) Discovering drugs to treat coronavirus disease 2019 (COVID-19). Drug Discov Therapeut 14(1):58–60. https://doi.org/10.5582/ddt.2020.01012

Du Y-X, Chen X-P (2020) Favipiravir: pharmacokinetics and concerns about clinical trials for 2019-nCoV infection. Clin Pharmacol Therapeut 108(2):242–247. https://doi.org/10.1002/cpt.1844

Dyall J, Coleman CM, Hart BJ, Venkataraman T, Holbrook MR, Kindrachuk J, Johnson RF, Olinger GG, Jahrling PB, Laidlaw M, Johansen LM, Lear-Rooney CM, Glass PJ, Hensley LE, Frieman MB (2014) Repurposing of clinically developed drugs for treatment of Middle East respiratory syndrome coronavirus infection. Antimicrob Agents Chemother 58(8):4885–4893. https://doi.org/10.1128/AAC.03036-14

Elfiky AA (2020) Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): a molecular docking study. Life Sci 253:117592. https://doi.org/10.1016/j.lfs.2020.117592

Elmezayen AD, Al-Obaidi A, Şahin AT, Yelekçi K (2020) Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. J Biomol Struct Dyn 0:1–13. https://doi.org/10.1080/07391102.2020.1758791

Furuta Y, Komeno T, Nakamura T (2017) Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase. Proc Jpn Acad Ser B Phys Biol Sci 93(7):449–463. https://doi.org/10.2183/pjab.93.027

Gimeno A, Mestres-Truyol J, Ojeda-Montes MJ, Macip G, Saldivar-Espinoza B, Cereto-Massagué A, Pujadas G, Garcia-Vallvé S (2020) Prediction of novel inhibitors of the main protease (M-pro) of SARS-CoV-2 through consensus docking and drug reposition. Int J Mol Sci 21(11). https://doi.org/10.3390/ijms21113793

Goldman JD, Lye DCB, Hui DS, Marks KM, Bruno R, Montejano R, Spinner CD, Galli M, Ahn M-Y, Nahass RG, Chen Y-S, SenGupta D, Hyland RH, Osinusi AO, Cao H, Blair C, Wei X, Gaggar A, Brainard DM, Subramanian A et al (2020) Remdesivir for 5 or 10 days in patients with severe Covid-19. N Engl J Med 383(19):1827–1837. https://doi.org/10.1056/nejmoa2015301

Grein J, Ohmagari N, Shin D, Diaz G, Asperges E, Castagna A, Feldt T, Green G, Green ML, Lescure F-X, Nicastri E, Oda R, Yo K, Quiros-Roldan E, Studemeister A, Redinski J, Ahmed S, Bernett J, Chelliah D, Flanigan T et al (2020) Compassionate use of remdesivir for patients with severe Covid-19. N Engl J Med 382(24):2327–2336. https://doi.org/10.1056/nejmoa2007016

Gupta MK, Vemula S, Donde R, Gouda G, Behera L, Vadde R (2021) In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel. J Biomol Struct Dyn 39(7):2617–2627. https://doi.org/10.1080/07391102.2020.1751300

Hall DC, Ji HF (2020) A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease. Travel Med Infect Dis 35:101646. https://doi.org/10.1016/j.tmaid.2020.101646

Hanaei S, Rezaei N (2020) COVID-19: developing from an outbreak to a pandemic. Arch Med Res 51(6):582–584. https://doi.org/10.1016/j.arcmed.2020.04.021

Joshi S, Parkar J, Ansari A, Vora A, Talwar D, Tiwaskar M, Patil S, Barkate H (2021) Role of favipiravir in the treatment of COVID-19. Int J Infect Dis 102. https://doi.org/10.1016/j.ijid.2020.10.069

Khalili JS, Zhu H, Mak NSA, Yan Y, Zhu Y (2020) Novel coronavirus treatment with ribavirin: groundwork for an evaluation concerning COVID-19. J Med Virol 92(7):740–746. https://doi.org/10.1002/jmv.25798

Li H, Xiong N, Li C, Gong Y, Liu L, Yang H, Tan X, Jiang N, Zong Q, Wang J, Lu Z, Yin X (2021) Efficacy of ribavirin and interferon-α therapy for hospitalized patients with COVID-19: a multicenter, retrospective cohort study. Int J Infect Dis 104:641–648. https://doi.org/10.1016/j.ijid.2021.01.055

Mahdian S, Ebrahim-Habibi A, Zarrabi M (2020) Drug repurposing using computational methods to identify therapeutic options for COVID-19. J Diabetes Metab Disord 19(2):691–699. https://doi.org/10.1007/s40200-020-00546-9

Masoudi-Sobhanzadeh Y, Omidi Y, Amanlou M, Masoudi-Nejad A (2019) Trader as a new optimization algorithm predicts drug-target interactions efficiently. Sci Rep 9(1):1–14. https://doi.org/10.1038/s41598-019-45814-8

Mittal L, Kumari A, Srivastava M, Singh M, Asthana S (2021) Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach. J Biomol Struct Dyn 39(10):3662–3680. https://doi.org/10.1080/07391102.2020.1768151

Moazzami B, Razavi-Khorasani N, Dooghaie Moghadam A, 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

Mohamed K, Yazdanpanah N, Saghazadeh A, Rezaei N (2020) Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website. Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active. Bioorganic Chemistry Computational drug discovery and repurposing for the treatment of COVID-19: A systematic review. January

Oprea TI, Overington JP (2015) Computational and practical aspects of drug repositioning. Assay Drug Dev Technol 13(6):299–306. https://doi.org/10.1089/adt.2015.29011.tiodrrr

Park T, Lee S-Y, Kim S, Kim MJ, Kim HG, Jun S, Kim S, Kim BT, Park EC, Park D (2020) Spike protein binding prediction with neutralizing antibodies of SARS-CoV-2. 1–22. https://doi.org/10.1101/2020.02.22.951178

Rezaei N (2020a) COVID-19 affects healthy pediatricians more than pediatric patients. Infect Control Hosp Epidemiol 41(9):1106–1107. https://doi.org/10.1017/ice.2020.139

Rezaei N (2020b) COVID-19 affects healthy pediatricians more than pediatric patients an easy protective measure in ophthalmology against medical supply shortage. May 2021, 2020–2021. https://doi.org/10.1093/cid/ciaa149.6

Saqrane S, El Mhammedi MA, Lahrich S, Laghrib F, El Bouabi Y, Farahi A, Bakasse M (2021) Recent knowledge in favor of remdesivir (GS-5734) as a therapeutic option for the COVID-19 infections. J Infect Public Health 14(5):655–660. https://doi.org/10.1016/j.jiph.2021.02.006

Scherman D, Fetro C (2020) Drug repositioning for rare diseases: knowledge-based success stories. Therapie 75(2):161–167. https://doi.org/10.1016/j.therap.2020.02.007

Sheahan TP, Sims AC, Graham RL, Menachery VD, Gralinski LE, Case JB, Leist SR, Pyrc K, Feng JY, Trantcheva I, Bannister R, Park Y, Babusis D, Clarke MO, Mackman RL, Spahn JE, Palmiotti CA, Siegel D, Ray AS, Baric RS et al (2017) Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. Sci Transl Med 9(396):eaal3653. https://doi.org/10.1126/scitranslmed.aal3653

Sheahan TP, Sims AC, Leist SR, Schäfer A, Won J, Brown AJ, Montgomery SA, Hogg A, Babusis D, Clarke MO, Spahn JE, Bauer L, Sellers S, Porter D, Feng JY, Cihlar T, Jordan R, Denison MR, Baric RS (2020) Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. Nat Commun 11(1). https://doi.org/10.1038/s41467-019-13940-6

Singh TU, Parida S, Lingaraju MC, Kesavan M, Kumar D, Singh RK (2020) Drug repurposing approach to fight COVID-19. Pharmacol Rep 72(6):1479–1508. https://doi.org/10.1007/s43440-020-00155-6

Tchesnokov EP, Feng JY, Porter DP, Götte M (2019) Mechanism of inhibition of ebola virus RNA-dependent RNA polymerase by remdesivir. Viruses 11(4):1–16. https://doi.org/10.3390/v11040326

Tong S, Su Y, Yu Y, Wu C, Chen J, Wang S, Jiang J (2020) Ribavirin therapy for severe COVID-19: a retrospective cohort study. Int J Antimicrob Agents 56(3):1–5. https://doi.org/10.1016/j.ijantimicag.2020.106114

Wang X, Guan Y (2021) COVID-19 drug repurposing: a review of computational screening methods, clinical trials, and protein interaction assays. Med Res Rev 41(1):5–28. https://doi.org/10.1002/med.21728

Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, Fu S, Gao L, Cheng Z, Lu Q, Hu Y, Luo G, Wang K, Lu Y, Li H, Wang S, Ruan S, Yang C, Mei C, Wang C et al (2020a) Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. Lancet 395(10236):1569–1578. https://doi.org/10.1016/S0140-6736(20)31022-9

Wang Y, Li W, Jiang Z, Xi X, Zhu Y (2020b) Assessment of the efficacy and safety of ribavirin in treatment of coronavirus-related pneumonia (SARS, MERS and COVID-19): a protocol for systematic review and meta-analysis. Medicine 99(38):e23751