Software for Drug Discovery and Protein Engineering: A Comparison Between the Alternatives and Recent Advancements in Computational Biology

Tathagata Adhikary and Piyali Basak

Abstract "Omic" technologies (such as genomics, transcriptomics, proteomics, and metabolomics) generate huge databases that demand computational approaches to state novel conclusions. With the advent of machine learning and artificial intelligence algorithms, the analysis of biological data and protein engineering has taken a step forward. Different virtual screening servers and standalone software paved their importance in the initial phase of drug discovery, aiding in drug repurposing and high-throughput screening. Besides, interaction networks, often encountered in polypharmacology and network pharmacology, guide a researcher in target fishing and developing drug combinations. Visualization and prediction of molecular structures, modeling antibodies, and peptides including homology modeling are crucial to bioinformaticians and clinical biologists. Biological network analysis, pharmacophore modeling, molecular docking, and dynamics simulation are broadly exploited in the domain of computational biology and elucidate the mechanisms underlying biomolecular interactions, consequently revealing the orchestra of biological pathways. Considering the intended purposes, advantages, and limitations of the existing software, this chapter highlights only a fraction of popular platforms and encourages the readers to explore other alternatives in various domains of drug discovery and protein engineering.

Keywords Big data · Virtual screening · Biological networks · Pharmacophore · Homology modeling

1 Introduction: The Need for Computational Biology

Discovering novel drug molecules strictly demands huge investments in terms of time, infrastructure, and labor to identify, optimize, and validate the drug-likeliness of such molecules by conducting in vitro, in vivo, and preclinical experiments

T. Adhikary · P. Basak (✉)
School of Bioscience and Engineering, Jadavpur University, Kolkata, India

241

(Lei et al. 2016; Rifaioglu et al. 2019). To ease the process, a shift toward the application of computational tools is witnessed in the early stages of identifying drug-like molecules. Constant advancements in software and its algorithms aid to bridge the "innovation gap" that exists due to higher investments and lower approval rates. The process of drug discovery sequentially includes the identification and validation of disease targets, lead compound identification and its optimization, and finally success in clinical trials. Accordingly, establishing a drug can take around 10 to 13 years with huge capital expenditure (Malathi et al. 2018). Challenges arising due to the pleiotropic nature of biomolecules and the interaction of chemical compounds with multiple pharmacological targets (often encountered in combinatorial/multi-targeted approaches) can be addressed by chemo- and bioinformatics tools that make use of databases on physicochemical characteristics and therapeutic use of compounds (Lagunin et al.). The fact that the primary healthcare of 80% of the population in developing countries counts on the conventional herbal remedies and the steep rise of 380% in plant-based supplements' sales in the United States from 1990 to 2000 encouraged the development of numerous databases on ethnomedicine (Dunkel et al. 2006; Mosihuzzaman and Choudhary 2008). This expands the prospects of utilizing traditional knowledge on medicinal plants in modern-day drug discovery and drug repurposing.

In 1971, the database Protein Data Bank (PDB) came into existence, being the first open-access digital repository in the field of biology. It is a collection of 3D structures (resolved by laboratory experimentations namely X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy) of biological macromolecules and receptor–ligand complexes (https://www.rcsb.org/) (Burley et al. 2017). ChemCom (Chemical Comparator) is an application based on Java Web Start (JavaWS) technology and includes UnionBit Tree Algorithm to search and compare large chemical libraries (Saeedipour et al. 2015). The list of such repositories can be long enough (Lagunin et al.). A few open-source databases on medicinal plants, phytochemicals and other chemical compounds can be listed as follows: Plants For A Future (PFAF), Indian Medicinal Plants Phytochemistry And Therapeutics 2.0 (IMPPAT 2.0), Native American Ethnobotany database, SuperNatural 3.0, The Natural Compound (NC) collection, NCBI PubChem, ChEMBL, Collection of open natural products (COCONUT), Traditional Chinese Medicine Information Database (TCMID), Dr. Duke's Phytochemical and Ethnobotanical Databases, Aromatic and Medicinal Plants Index (by Purdue University), Agricultural Science and Technology (or AGRIS supported by the Food and Agriculture Organization (FAO) of the United Nations), Compendium of Ayurveda Medicinal Plants of Sri Lanka, Botanical.com, Chinese Herbal Medicine Dictionary (by Complementary and Alternative Healing University), Clinicaltrials.gov database, Medicinal Plant Database (by Botanical Survey of India), EcoPort, TIPdb (a database of indigenous and endemic plant species in Taiwan), Traded Medicinal Plants Database, Herbal Medicines Compendium Medicinal Herbs and Plant Database, Drugs Herbs and Supplements by MedlinePlus, ZINC database, Marowina database medicinal support, Natural Medicines, Herbs at a Glance, Prelude Medicinal Plants Database, Raintree Tropical Plant Database, The World Flora Online, and TRAMIL database

(Xie et al.; Duke 2020). Commercial databases with paid access include Chemical Abstracts Service (CAS), HerbalThink-TCM, Dictionary of Natural Products (DNP), and HerbMed.

The increasing data relating to the bioactivities of a chemical compound, the composition of phytoconstituents in an extract, and the target receptors responsible for specific bioactivity need to be stored and be able to be retrieved systematically. "Omic" technologies have led to the development of diverse databases, and interpreting/interconnecting them or data mining from them is a major challenge to human capabilities. Hence, computational approaches including artificial intelligence and machine learning algorithms (e.g., artificial neural networks (ANN), Naive Bayes, K-Means, support vector machine (SVM), random decision forest, etc.) are widely adopted to provide solutions to complex biological questions (Gupta et al. 2021; Muzammil et al. 2023). Continuous development of in silico tools for chemoinformatics and bioinformatics provides insight to the vast multiomics data and adds different perspectives to the scientists in the domain of drug discovery. Chemoinformatics particularly aims to model a statistical correlation between the observed bioactivity and structural parameters. These approaches relating to computer-aided drug design have gained noteworthy momentum in the drug discovery process. Genome-wide functional genetic screening (e.g., using deep learning algorithms) is a cutting-edge technique that has led to the discovery of genotype–phenotype interconnections and established new phenotypes (Zhang et al. 2011). Genomics and proteomics analyses in high-throughput screening have shown promising results to rationalize the drug discovery process; however, the cost inflation incurred due to these technologies does not meet the expected growth of the drug's approval rate. Freely available software that are frequently employed in machine learning and statistical analysis of data are R, PSPP by the GNU Project, and WEKA while commercial ones include MATLAB, SAS/STAT, SIMCA, SPSS Statistics by IBM, and TIBCO Data Science/Statistica (Dzemyda et al. 2019).

The first thing that needs to be checked while selecting a software for computer-aided drug designing is its vendor and license—whether it is under academia, commercial, open-source, or in-house software. Open-source software are popular among academic personnel as, unlike commercial software, no license fee is required, their source code is made available freely and can be modified by a user. Based on the intended use, license fee, and characteristic features of the software/platforms, attempts are made to categorize and list the in silico tools employed in the various domains of drug discovery (Singh et al. 2021). Molecular docking, pharmacophore modeling, methods relating (Q)SAR, molecular dynamics simulation, network pharmacology and machine learning algorithms accelerate the drug discovery process and complement the traditional bioactivity-guided fractionation, high-throughput screening and systems biology approaches. In this chapter, the tables summarizing the in silico tools only provide a fraction of popular platforms and encourage the readers to explore other alternatives in various domains of drug discovery and protein engineering.

2   Visualization of Molecular Structures

Molecular graphics enhances the experience of representing, modeling and analyzing multifaceted biochemical systems. Besides modeling the 3D architecture of molecular structures, 2D illustrations of molecules have gained interest among chemical scientists and biologists in the field of theoretical chemistry and discovery because of their clear representation of structural characteristics and interactions between atoms (Zhou and Shang 2009). Visualizing molecular structures in any virtual reality environment demands rapid high-quality rendering of geometries to build molecular models with intuitive and informative interactions. Different visualization techniques (such as the space-filling model, the ball-and-stick model, and the reconstruction of the surface of the secondary structures alpha helixes and beta sheets) are used while representing a molecular model, and some available platforms to design, analyze and visualize molecular structures are listed in Table 1.

3   Prediction of Pharmacokinetic/Pharmacodynamic Profile

Evaluating the ADMET (absorption, distribution, metabolism, excretion and toxicity) properties of a molecule is a major step in discovering novel drug compounds. In general, compounds having natural origin tend to have desirable ADMET properties compared to synthetic compounds. Early prediction of ADMET properties of a chemical compound can be of utmost importance since most drug failures occur in the later phases due to undesirable pharmacokinetics and toxicological characteristics. Lipinski's rule of five is often checked to predict the drug-likeliness (in humans) of an oral-administered compound (Lipinski 2004; Rego et al. 2022). According to it, a drug molecule can have at most one violation among these five rules: (a) ligand's molecular weight should be less than or equal to 500 Daltons, (b) the number of H-bond donors should be less than 5, (c) the number of H-bond acceptor should be less than 10, (d) value of octanol partition coefficient (miLogP) should be less than 5 and (e) the number of rotatable bonds should be less than 10.

Most of the software packages that predict the ADMET of compounds (e.g., their affinity toward transporter proteins, blood proteins and drug-metabolizing enzymes P450 cytochromes isoforms, etc.) consider their structural/physicochemical characteristics to develop (Q)SAR models. Derek Nexus (Lhasa Ltd.), TOPKAT (Accelrys), OSIRIS Property Explorer, MCASE (Multicase) and PASS can be opted to predict various toxicities and report the teratogenic, mutagenic, cardiotoxic, hepatotoxic, carcinogenic, and renal-toxic nature of the compounds (Kar et al. 2018). The online server of GUSAR (www.way2drug.com) predicts the LD_{50} values of query compounds on rodents when administered via four different routes. Some other software/web-servers to study the ADMET properties are listed in Table 2.

Table 1 Software/servers for designing and visualizing molecular structures

Software/platform	Description	URL
3DNA	It is used to analyze, rebuild and visualize 3D structural architecture of nucleic acids and ligand/protein–nucleic acid complexes. The parameters relating to the spatial relationship of the base pairs present in the nucleic acids are evaluated by using a simple matrix-based scheme (Lu and Olson 2003)	https://x3dna.org/
iMol	A free tool exclusively designed for the Mac OS X operating system to visualize molecular structures and molecular dynamics trajectories	https://www.pirx.com/iMol/index.shtml
VMD (Visual Molecular Dynamics)	It is an open-source software that utilizes graphics processing units (GPUs) to model, visualize and analyze complex biomolecular systems comprising proteins, DNAs, RNAs, etc. Animation and analysis of molecular dynamics simulation trajectories can also be performed using VMD, thus acting as a graphical interface for an external simulation program. It can read Protein Data Bank (PDB) file format (Humphrey et al.)	https://www.ks.uiuc.edu/Research/vmd/
Swiss-PdbViewer or DeepView	It allows simultaneous analysis of several proteins and their superimposition that helps compare the active/binding sites, amino acid mutations, H-bonds, atomic angles, and distances. It is a 32 bits application and is not supported by OSX Catalina or newer versions of OSX (Johansson et al. 2012)	https://spdbv.unil.ch/
RasTop	Adapted from RasMol, it is an open-source platform under the General Public License (GPL) for molecular visualization (specifically for academic purposes). It is supported in Windows and Linux operating systems	https://www.geneinfinity.org/rastop/
The Ramachandran Plot Explorer	It analyses the conformation of a polypeptide along with the effect of conformational changes in the energy parameters of the system (Hayward et al. 2014)	https://boscoh.com/ramaplot/
QuteMol	It exploits the GPU of a computer system (through OpenGL shaders) for high-quality molecular visualization (in 3D) of biological systems (Tarini and Cignoni 2006)	https://qutemol.sourceforge.net/
PyMOL	It is a user-sponsored commercial molecular visualization product and is compatible with Windows, Linux and macOS (Yuan et al. 2017)	https://pymol.org/2/

(continued)

Table 1 (continued)

Software/platform	Description	URL
ProteinShader	It is a free, open-source (under GNU GPL) molecular visualization program. It requires Java 1.5 (or newer versions) and a graphics card that supports at least OpenGL 2.0. It allows cartoon-type (artistic) illustrations of the protein structures (Weber 2009)	https://proteinshader.sourceforge.net/
MOLMOL	It is a molecular graphics project to display, analyze, and manipulate the 3D structure of biological macromolecules and study their structures obtained from NMR spectroscopy	https://sourceforge.net/projects/molmol/
Jmol/JSmol	It is a molecular viewer for 3D structures. It can take several many file types as its input. This includes PDB, CIF, SDF, MOL, PyMOL PSE files, and Spartan files, as well as output from Gaussian, GAMESS, MOPAC, VASP, CRYSTAL, CASTEP, QuantumEspresso, VMD, and other quantum chemistry programs	https://sourceforge.net/projects/jmol/
ICM-Browser	It is compatible with Windows, Mac and Linux platforms. It can build fully annotated and interactive 3D molecular structure files. ActiveICM plugin allows those files to display directly in PowerPoint and on the web. It can load and display sequence file types and generates publication-quality images. It allows the superimposition of protein structures and calculates atomic distances and angles (both planar and torsion angles) (Magnotti et al. 2019)	https://www.molsoft.com/icm_browser.html#ibFeatures
The BIOVIA Discovery Studio Visualizer	It is a free molecular visualization software for molecular modeling and analyzing the structures (of both proteins and small molecules) with advanced features such as depth cueing, blur and shading	https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/visualization/
NOC	It is a free molecular structure visualizer. Its source code is open for research purposes and allows nonprofit commercial redistribution	https://noch.sourceforge.net/#Downloading
Cn3D	It can simultaneously display structure, sequence, and alignment with advanced options for annotation and alignment editing (Stasinakis et al. 2017)	https://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml

Name	Description	URL
SmilesDrawer	It is used to parse and draw SMILES strings using client-side JavaScript (Probst and Reymond 2018)	https://github.com/reymond-group/smilesDrawer
PlexView	It generates the 2D image of protein-ligand interactions that includes hydrogen bonding and pi-pi stacking. If the protein is not protonated, PlexView will protonate it at pH 7; however, the ligand needs to be protonated previously to correctly highlight the H-bonds	https://playmolecule.com/PlexView/
Ketcher	It is developed using JavaScript and serves as an open-source web-based editor to modify/draw structures of chemical compounds and reactions	https://lifescience.opensource.epam.com/ketcher/
NCIPLOT	It provides a visualization index for noncovalent interactions. It plots reduced density gradient (RDG) as a function of the density across a molecule (Laplaza et al. 2021)	https://www.lct.jussieu.fr/pagesperso/contrera/index-nci.html
PoseView	It is used in the 2D illustrations of protein-ligand interactions. It identifies hydrogen bonds, hydrophobic contacts and the interacting amino acids residues are selected using the FlexX-library (Stierand and Rarey 2010)	https://www.zbh.uni-hamburg.de/en/forschung/amd/server/poseview.html
bioRENDER	It helps to create professional scientific images using premade icons and templates. Its basic version is free for educational purposes but other features require a paid subscription	https://biorender.com/
Servier Medical Art	It allows the free use of over 3000 medical images to illustrate in publications/presentations	https://smart.servier.com/
LeView (Ligand Environment Viewer)	It is a java program (requiring Java Runtime Environment (JRE) to be preinstalled and enabled) that generates 2D representations of protein-ligand interaction from their complex (in PDB file)	http://www.pegase-biosciences.com/leview-ligand-environment-viewer/
Smi2Depict	It is used in the generation of 2D figures of molecular structure from SMILES	http://cdb.ics.uci.edu/cgibin/Smi2DepictWeb.py
IBS 2.0	It is a web-based open resource used for visualizing biological sequences (including both proteins and nucleotides) and highlights the functional elements to assist experimentalists in generating publication-quality images (Xie et al. 2022)	https://ibs.renlab.org/#/home

Table 1 (continued)

Software/platform	Description	URL
CIME (ChemInformatics Model Explorer)	It is an interactive web-based tool for inspecting data sets of chemical compounds and their subgroups and visualizing models (Humer et al. 2022)	https://github.com/jku-vds-lab/cime
Img2Mol	It aims to automatically recognize the molecular content and predict the SMILES strings from molecular graphics (Clevert et al. 2021)	https://chemrxiv.org/engage/chemrxiv/article-details/60c756c6f9a00bff5288b68
DECIMER (Deep lEarning for Chemical ImagE Recognition)	It is an Optical Chemical Structure Recognition (OCSR) tool used to generate the SMILES strings by taking the input of the chemical structures as image files or PDF documents (Rajan et al. 2021)	https://decimer.ai/
ChemPix	It uses deep learning algorithms and neural networks to map and recognize hand-drawn hydrocarbon molecules and present the corresponding machine-readable SMILES (Weir et al. 2021)	https://github.com/mtzgroup/ChemPixCH
Chemtool	It is a LINUX-based program to draw chemical structures	http://ruby.chemie.uni-freiburg.de/~martin/chemtool/
Molsketch	It is an editor for 2D molecular structures that generates high-quality vector images	https://sourceforge.net/projects/molsketch/
JChemPaint (JCP)	It is a java application that can edit and view 2D molecular structures	http://jchempaint.github.io/
SketchEl	It is an interactive sketching tool for chemical structures and exports molecule diagrams as Scalable Vector Graphics (SVG)	https://sketchel.sourceforge.net/
BKChem	It is a python based free software to draw chemical structures	https://bkchem.zirael.org/
DataWarrior	It is an open-source platform with embedded chemical intelligence to visualize and analyze data using scatter plots, box plots, bar charts and pie charts. Considering various chemical descriptors, it can predict physicochemical properties and structure–activity relationships and display activity cliffs (Sander et al. 2015)	https://openmolecules.org/datawarrior/index.html

Name	Description	URL
Skylign	It creates logos to represent sequence alignments and profile hidden Markov model	http://skylign.org/
TMRPres2D (TransMembrane protein Re-Presentation in 2 Dimensions)	It is a tool that provides a graphical user interface to automatically create uniform 2D graphical models of alpha-helical or beta-barrel transmembrane proteins from the corresponding sequences (Spyropoulos et al. 2004)	http://bioinformatics.biol.uoa.gr/TMRPres2D/
ChemTreeMap	It is an open-source interactive tool used in the visualization of chemical similarity/diversity in a library of chemical compounds (Lu et al.)	http://ajing.github.io/ChemTreeMap/
LigPlot+	It automatically generates a high-quality 2D image of ligand-protein interaction	https://www.ebi.ac.uk/thornton-srv/software/LigPlus/
Protein Contacts Atlas	It provides insight into the protein structures by visualizing and analyzing noncovalent contacts	http://pca.mbgroup.bio/

Table 2 In silico tools used in the prediction of ADMET properties of small molecules

Software/platform	Description	URL
DMFGAM	It uses a deep learning algorithm on molecular fingerprints to predict hERG channel blockers and cardiotoxicity (Wang et al.)	https://github.com/zhaoqi106/DMFGAM
TOXRIC	It is an online platform/database consisting of toxicity-related data of 113,372 compounds. It uses molecular fingerprints, transcriptome profiles, metabolic reactions and other chemical descriptors to predict 13 toxicity categories (such as acute toxicity, cardiotoxicity, endocrine disruption, etc.) with 1474 toxicity endpoints (Wu et al.)	https://toxric.bioinforai.tech/home
DIADpredictor	Using machine learning it employs support vector machine (SVM) and uses SMILES strings to predict drug-induced autoimmune diseases	http://diad.sapredictor.cn/
DDInter	It is an open-access database to find drug-drug interactions and performs risk assessment in medications for improved clinical decisions	http://ddinter.scbdd.com/
cardioToxCSM	It is a web-server that takes SMILES strings of small molecules (a maximum of 1000 molecules) as the input for predicting six types of cardiotoxicities (namely arrhythmia, cardiac failure, heart block, hERG toxicity, hypertension, and myocardial infarction)	https://biosig.lab.uq.edu.au/cardiotoxcsm/
ToxSTAR	It combines in vitro, in vivo, and in silico data to build the prediction models that include the prediction of drug-induced cholestasis, cirrhosis, hepatitis, steatosis, and drug-induced liver injury (DILI)	https://toxstar.kitox.re.kr/about_toxstar
Deep-B³	It is a deep learning-based model that uses molecular descriptors and fingerprints, images, and SMILES strings to predict the blood-brain barrier permeability of molecules (Tang et al.)	http://cbcb.cdutcm.edu.cn/deepb3/
toxCSM	It is used in the prediction of toxicity profiles of chemical compounds including environmental toxicity	https://biosig.lab.uq.edu.au/toxcsm/
NR-Profiler	It is used in the prediction of nuclear receptor profiling and identifies the potential nuclear receptor modulators	https://github.com/jywangECUST/NR-Profiler
NURA (NUclear receptor activity dataset)	It integrates several toxicological and pharmacological databases and provides information on the bioactivity of 15,206 molecules and 11 nuclear receptors (Valsecchi et al.)	https://michem.unimib.it/download/data/nura/

(continued)

Name	Description	URL
Super-PRED	It is based on the Anatomical Therapeutic Chemical (ATC) classification system published by the World Health Organization (WHO) that is used for the target prediction of small molecules using Pubchem name or SMILES string	https://prediction.charite.de/index.php
PredMS	One can use the SMILES string or draw the ligand structure in a molecule editor for the prediction of its metabolic stability in human liver microsomes	https://predms.netlify.app/
ICDrug	Apart from ADMET prediction, it can be used as a ligand designer, protein designer, and to analyze protein–protein or protein–ligand interactions	http://www.icdrug.com/
PK-DB	It is a pharmacokinetics open database built using the results of numerous clinical and preclinical research (Grzegorzewski et al. 2021)	https://pk-db.com/
ToxicoDB	It is a curated database (four datasets consisting of 234 compounds and 32,000 genes) of toxicogenomics data and allows mining multiple databases	https://www.toxicodb.ca/
vNN-ADMET	Based on variable nearest neighbor (vNN) methodology, it is a publicly available web server that uses 15 prediction models to predict the ADMET properties (Schyman et al. 2017)	https://vnnadmet.bhsai.org/vnnadmet/login.xhtml
SwissADME	It inputs multiple SMILES strings to study the pharmacokinetics and drug-likeness of chemical compounds (Daina et al.)	http://www.swissadme.ch/
admetSAR	It is a free tool that can predict multiple (24) endpoints such as human intestinal absorption (HIA), renal clearance, organ/genomic/environmental toxicity, etc.	http://lmmd.ecust.edu.cn/admetsar2/
OCHEM	It is a platform for ADMET prediction and can generate quantitative structure–activity and structure–activity models	http://www.eadmet.com/en/ochem.php
SApredictor	It identifies and collects structural alerts (i.e., structural fragments) that have the potential to influence chemical toxicity	http://www.sapredictor.cn/
VenomPred	It uses various machine learning algorithms to predict the potential mutagenicity, hepatotoxicity, carcinogenicity and estrogenic effect of chemical compounds using their SMILES strings (Galati et al. 2022)	http://www.mmvsl.it/wp/venompred/

4 Prediction of Structures Including Homology Modeling

Molecular modeling based on structure-based drug designing requires 3D structures of the receptor and ligand molecules (experimentally determined by X-ray crystallography and NMR spectroscopy). In cases where experimental data are unavailable, the existing data and sequences can be used to predict the structures by homology-based modeling, sometimes referred to as comparative modeling of protein. The amino acid sequence of a protein (acquired from NCBI or UniPort) is used to generate the structure using computational tools. Evolutionarily related proteins share a similarity in sequences and homologous proteins exhibit similarity in their protein structure (substitution matrices such as Blosum 60 describe such homology). The 3-D protein structure is found to be evolutionarily more conserved compared to the sequence conservation alone (Kaczanowski and Zielenkiewicz 2010). Homology modeling starts with recognizing a template that shows similarity in sequence (searching is accomplished by employing BLAST (Basic Local Alignment Search Tool) or PSI-BLAST (Position-Specific Iterated BLAST) or fold recognition methods) and subsequent alignment of the known structures (resolved by experiments) in the database. A similarity of less than 30% is generally not preferred in homology modeling. BLAST compares a query sequence with the existing database and identifies the most suitable sequence with significant similarity, i.e., it identifies the homologous sequences. Alignments with an expectation value (E-value) closer to zero indicate a higher similarity index. A higher E-value makes the alignment of two sequences strenuous, thus considering sequences from other homologous proteins can help in this scenario (Pearson 2013; Alves et al. 2023). Multiple Sequence Alignment programs, e.g., CLUSTALW, can align sequences by performing insertions and deletions. Alignment correction, if done not properly, will generate defective structures. Some of the methods that are used to build models are spatial restraint, rigid-body assembly, segment matching and artificial evolution. Modeling tools, namely Modeller or CASP, can be used to build the backbone from the aligned sequences. Most often, aligning the model sequence with the template sequence creates gaps that can be resolved by considering conformational changes, insertions/deletions/substitutions of amino acid residues. Thus, refining the model includes loop modeling and side-chain modeling following the principles of molecular dynamics, Monte Carlo, and genetic algorithms. After modeling, structures are energetically minimized by employing force fields (for instance OPLS, AMBER, MM3, and CHARMM22 force fields) (Lewis-Atwell et al.). Loop modeling can be knowledge-based or energy-based. Knowledge-based loop modeling, sometimes referred to as template-based or homology-based, searches existing databases to identify known loop conformations that match the input sequence and geometric descriptors about the anchoring points (Karami et al.). It does not require complex simulations and high computation power; however, it relies on the availability of appropriate loop conformations present in the existing repositories of protein structures to consider the entire conformational space. Energy-based loop modeling corresponds to nontemplate-based or de novo methods that use an energy function and

minimizes it by Monte Carlo methods or molecular dynamics to optimize the loop conformation. Proteins that share structural similarity also exhibit similarity in torsion angle about Ca-Cb bond (psi angle) when side-chain conformations are considered. The entire conserved residues can be taken from the template and copied to the model to yield highly accurate results when compared to the methods that copy the backbone or predict the side chains. Modeling of side chains includes knowledge-based methods to extract a library of rotamers from known crystallographic structures and substitutes the side chains on the backbone structure. After side chain modeling, the analysis is done by using their root mean square deviation (RMSD) values. The errors found in the final model are dependent on the extent of similarity between the template and the target. If it is >90%, then the crystallographic structure is fairly predicted, whereas for a value <90%, the r.m.s.d errors will be significant. The estimation of errors can be done by using a force field to calculate the model's energy and checking if the bond lengths and angles are exhibiting a value in the normal range (Dolan et al. 2012; Wink et al. 2019; Lima et al. 2022). However, this method does not evaluate the folding nature of the model and the misfolding in proteins is taken care of by 3D distribution functions. Model validation is necessary to establish the prediction accuracy.

The stereochemical aspects of the protein can be explored by WHATCHECK, WHAT IF, VADAR, and PROCHECK. Ramachandran plot, obtained by plotting the torsional angles of amino acids φ (phi) and ψ (psi) in a protein sequence two-dimensionally, is used to analyze the stereochemical and geometrical nature of the structure and verifies the presence of geometries in the electrostatically unfavored regions of the plot. A higher proportion of residues in the favored region indicates the structural feasibility of the model (Agnihotry et al. 2022). Popularly used tools for homology-based modeling are MODELLER, SWISS PDB VIEWER, SWISS MODEL and COMPOSER (Malathi et al. 2018). MODELLER is also used for sequence searching, comparing and clustering protein structures or sequences. In brief, steps in homology modeling take into account template identification, sequence alignment, structural modification, energy minimization and model validation to predict the 3D structure.

5   Interaction Networks

Hopkins in 2007 brought the concept of network pharmacology that makes use of network analysis algorithms (on the existing knowledge of biological networks consisting of structural/physicochemical properties of proteins/ligands, the interaction of a protein/gene with another protein/gene/ligand, signaling and metabolic pathways) to predict the therapeutic action of small molecules, elucidate their mechanism of action, and understand the drug-disease relationships at the system-level (Csermely et al. 2013). Visualization of biological networks (such as pie-nodes and edge-pie matrix visualization) and network comparison (by employing network alignment and computing pair-wise similarity between selected networks) is

essential for network analysis, identification of key components/nodes/interactions in a concerned biological system, and highlighting the union/intersection/complement regions in a set of biological networks. Networks have the capability to highlight the interacting elements within a complex biochemical system, thus aiding in the visualization and exploration of big data. However, the challenges relating the large size and high complexity of biological networks generate the so-called "hairballs" in the networks. Hence, there is a need for an efficient and interactive graphical user interface for network comparison and visualization (Pirch et al. 2021; Almeida et al. 2022). One needs to consider several types of relationships (namely "target–effect," "target–pathway," "pathway–effect," and "target–pathway–effect" relationships) to investigate the pleiotropic and synergistic effects of a drug compound or a combination of drug compounds. The benefit of conceptualizing such "cause–effect" relationships unfold gradually—if the bioactivity of a drug relates to certain molecular targets and their corresponding pathways are established, then other mode-of-actions of influencing those pathways can yield similar effects.

Analysis of biological pathways (such as signaling pathways, regulatory pathways, metabolic pathways, signal transduction pathways, etc.) makes use of various pathway databases (Lagunin et al.). To name a few, WikiPathways (https://www.wikipathways.org), HumanCyc (https://humancyc.org/), NetPath (http://www.net-path.org/), Reactome (https://reactome.org/), KEGG (https://www.genome.jp/kegg/pathway.html), SignaLink (http://signalink.org/), and Small Molecule Pathway Database (https://www.smpdb.ca/). QIAGEN Ingenuity Pathway Analysis (IPA) is an online platform that is used to analyze, integrate, model and interpret the nexus of data from "omic" technologies including RNAseq experiments and Single-Nucleotide Polymorphism (SNP) microarrays. It aids in the identification of genes and pathways that functionally interact with the drug molecules and compares the gene regulatory circuits involved in the phenotypic responses. Connectivity Map (CMap) connects the genes and the drugs (currently in use) underlying various diseases and enables us to perform data-driven analysis of repurposing/reprofiling/repositioning of drugs (it does so by analyzing the disease-specific and drug-specific gene signatures). A user provides the "gene hit lists" (aka "signatures") to the CMap for its comparison with a gene differential expression (DE) database (obtained by perturbation of cell lines with numerous drug-like molecules) to output a rank of compounds that exhibit similarity in expression patterns considering the query hit list. The CMap resource hosts over 1.5 million gene expression profiles from around 5000 chemical compounds and 3000 genetic reagents that are tested in various cell lines (Lim and Pavlidis 2021). The similarity in the gene expression profiles based on drug–drug, drug–disease, and disease–disease relationships is used to create the disease–drug networks for studying the potential side effects, targets and pathways associated with the drug compound. Aside CMap, Gene Expression Omnibus (GEO), and the Comparative Toxicogenomics Database (CTD) can be opted to create such disease-specific gene expression signatures. DIGEP-Pred is a free web-based platform that considers the structural characteristics of compounds to predict drug-induced variations in gene expression profiles (Lagunin et al. 2013). Natural Product-based Drug Combination and Its Disease-specific Molecular Regulation

(NPCDR) is an interactive database that shares knowledge on drug combinations (of natural products) with clinical or experimental validations. It also provides information on disease-specific molecular recognition and pathways and allows integration of available databases, easing the research on network pharmacology and medicinal chemistry (Sun et al. 2022).

The platforms that are free for academic use in bioinformatics and systems biology researches to analyze complex data from "omic" technologies are OmicsNet (https://www.omicsnet.ca/) (Zhou and Xia 2018), Cell Illustrator (http://www.cellillustrator.com/home), Cytoscape (https://cytoscape.org/), ConsensusPathDB (http://cpdb.molgen.mpg.de/), Gene Set Enrichment Analysis or GSEA (https://www.gsea-msigdb.org/gsea/index.jsp), The Database for Annotation, Visualization and Integrated Discovery or DAVID (https://david.ncifcrf.gov/), VANESA (https://cbrinkrolf.github.io/VANESA/). Other software with paid licenses include the geneXplain platform (https://genexplain.com/), QIAGEN Ingenuity Pathway Analysis (https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/), and Elsevier's Pathway Studio (https://www.elsevier.com/en-in/solutions/pathway-studio-biological-research). Other alternatives that can be explored in this domain of research are presented in Table 3.

6   Pharmacophore Modeling and Molecular Docking

From a large library of chemical compounds, virtual screening identifies the lead compounds having a specific bioactivity. There exists structure-based and ligand-based virtual screening. The former approach utilizes the 3D structure of the target protein and performs molecular docking to report the potential active compounds that exhibit good binding affinity/score with the target receptor structure. Molecular docking is a structure-based approach and is used in the prediction of the 3D orientation of the ligand molecule with respect to a particular conformation of the receptor molecule when both are interacting and forming a stable complex (Sahoo et al.). It is one of the first-line tools used in discovering/designing novel drug molecules that predict the binding affinity of a chemical compound with the target receptor and ranks the ligands based on their respective docking scores. Molecular docking predicts the atomistic model of the receptor–ligand interactions and their binding orientations. In site-specific or targeted docking, the active sites of the target protein are reviewed or predicted by using programs such as CASTp, Q-SiteFinder, LigA Site, and MetaPocket, while blind docking considers the entire protein structure as the probable region of ligand interaction (Wong and Kwan 2015). Searching algorithms that fish out favorable conformations from infinite possibilities include matching algorithms, incremental construction methods, multiple copy simultaneous searching, Monte Carlo and genetic algorithms. Scoring functions (either empirical, force field, or knowledge-based) of a docking software estimate the binding affinity of the ligand with the target receptor and rank the ligands based on docking scores.

Table 3 Some software/servers to generate, visualize, and analyze biological networks

Software/platform	Description	URL
Cytoscape	It is an open-source platform used in the integration and graphical interpretation of complex networks and analysis of human-curated pathway datasets such as WikiPathways, Reactome, and KEGG (Shannon et al.)	https://cytoscape.org/
PathVisio	Developed in Java, it is a free open-source software for pathway analysis that allows a user to draw, edit, and analyze biological pathways. It is the pathway editor for WikiPathways	https://pathvisio.org/
ToxPi (Toxicological Prioritization Index)	It is a free interface distributed under the GNU GPL. It allows the integration of multiple data sources to generate visual profiles and clustering of data (Marvel et al. 2018)	https://toxpi.org/
ReactomeFIViz	It is used to build subnetworks and search pathways and network patterns related to several diseases including cancer. It has access to the Reactome Functional Interaction (FI) network. It needs Cytoscape to be preinstalled where it can be found as "ReactomeFIPlugIn"	https://reactome.org/tools/reactome-fiviz#Overview
PyPathway	It is a python package for analyzing and visualizing biological networks	https://pypi.org/project/pypathway/
Conan	It is a C++/Python library developed for generating, inferencing and analyzing complex networks (Honorato-Zimmer et al. 2010)	https://github.com/rhz/conan/
CyTargetLinker	This tool allows the extension of biological networks and comes with Cytoscape automation feature	https://apps.cytoscape.org/apps/cytargetlinker
multiSLIDE (Multi-omics Systems-Level Interactive Data Exploration)	It is a web server to analyze multiomics data and explore the interconnection of components of biological pathways (the connected molecular features are visualized in heatmaps) (Ghosh et al. 2021)	https://github.com/soumitag/multiSLIDE
VRNetzer	It allows efficient visualization, interactive exploration, customization, integration of external databases and extension of highly complex networks (Pirch et al. 2021)	https://menchelab.com/vrnetzer
MONGKIE (Modular Network Generation and Visualization Platform with Knowledge Integration Environments)	It is a single platform that analyzes (such as performing network clustering and over-representation analysis) and visualizes integrated networks generated from multiomics data	http://yjjang.github.io/mongkie/
CellDesigner	This modeling tool is a kind of editor that provides an intuitive GUI to draw gene-regulatory and biological networks (Funahashi et al.)	https://www.celldesigner.org/index.html

(continued)

Table 3 (continued)

Software/platform	Description	URL
RING (Residue Interaction Network Generator)	It is used in the identification of noncovalent interactions in a given protein structure. It can create probabilistic networks and conformational-dependent contact maps	https://ring.biocomputingup.it/submit
MinePath	It analyses gene expression data and identifies differentially expressed functional paths or subpaths within a gene regulatory network (Koumakis et al. 2016)	http://www.minepath.org/
Pathview	It is used for integrating and mapping diverse biological data on pathways and produces top-notch hyperlinked graphs (Luo et al. 2017)	https://pathview.uncc.edu/
NAPS (Network Analysis of Protein Structures)	It is a web-based platform that models proteins as a network of noncovalent interactions between amino acid residues (rather than the traditional way of analyzing the secondary structure and fold arrangement) to describe the topological characteristics and structure–function relationship (Chakrabarty and Parekh 2016)	https://bioinf.iiit.ac.in/NAPS/
ProSNEx (Protein Structure Network Explorer)	The web service is used to construct Protein Structure Networks (PSNs) and allows sequence conservation, annotation, and analysis of amino acid flexibility (Aydlnkal et al. 2019)	http://prosnex-tool.com/demos/technical/

Qualitative "structure–activity relationships" (i.e., SAR) and quantitative structure–activity relationships (i.e., QSAR) are used in virtual screening (and target fishing) if the structures of the chemical compounds are available or predicted/designed. These approaches assume the bioactivity of a ligand as a function of its structural or physicochemical characteristics. Analysis and comparison of the structures are achieved with the help of some descriptors (such as structural fragments, fingerprints, constitutional, topological, electro-topological, quantum-chemical and physicochemical descriptors) (Lagunin et al.). Pharmacophore modeling considers a group of atoms in the structure whose presence directs the pharmacological effect of the ligand. Ligand-based virtual screening employs QSAR approaches that aim to develop mathematical models to study the correlation between the observed bioactivities and structural/physicochemical characteristics. Software such as Sybyl-X 2.0 and E-Dragon perform QSAR studies (Browne et al.; Fedyushkina et al. 1990).

Two techniques, namely comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), are encountered in 3D QSAR for ligand-based drug designing (Chavda and Bhatt 2019). In CoMFA, a library of ligands comprising their physicochemical characteristics and biological activities is created. These bioactive compounds vary between themselves by some substitutions. Seventy percent of data in the database is fed as the input to the training set (regression models are generated using it following the Partial Least Squares (PLS) regression and correlating the models with the pIC50 value), whereas the rest

of the 30% data is kept as the test set (used to establish the prediction accuracy of the QSAR regression models). Finally, the models undergo leave-one-out (LOO) cross-validation. The descriptors in CoMFA are determined by the sp^3 probe. Columbic potential energy calculates the electrostatic field and Lennard Jones potential energy describes the bond energy curves for Van der Waals bonding. The 3D steric and electrostatic contour plots depict the variation in bioactivity with the alteration of molecular fields. The SEAL similarity method in CoMSIA takes into account the electrostatic, steric, hydrogen bonding and hydrophobic descriptors to predict the similarity between molecules using Gaussian functions. The contour plots produced by the CoMSIA portray the favorable and unfavorable regions for the interaction of ligands (Bordás et al. 2003).

Approaches to build pharmacophore-based models identify the molecular characteristics that direct the macromolecular recognition of ligands, thus triggering the biological response. The aromaticity, hydrophobicity, presence of hydrogen bond acceptors/donors and anion/cation residues are considered to model pharmacophores that act as a query to search the potential bioactives from a database of compounds in virtual screening. Developing pharmacophore models can follow either structure-based or ligand-based approaches. The former approach relies on the availability of X-ray crystallographic or NMR spectroscopic 3D structure of the receptor molecule/target protein. The active sites and the spatial interactions are described by certain physicochemical properties that complement the interacting ligands and selectively identify the compounds with high binding affinity. A good model must incorporate protein flexibility to consider the structural changes that occur during the formation of the receptor–ligand complex. Ligand-based modeling is useful in cases where the 3D molecular structure of the receptor molecule is not available and the pharmacophores are generated by studying the common features (e.g., hydrophobic and electrostatic interaction, hydrogen bonding, etc.) that exist at the same position in the ligand structures. In ligand-based pharmacophore modeling, chemical compounds in the training set create a conformational space that takes care of ligand flexibility (Braga et al.). HipHop, DISCO, HypoGen, and PHASE are some software for generating pharmacophore models.

The structural data generated by NMR, X-ray crystallography, and homology modeling are static in nature that fails to describe the dynamic nature of the biorecognition process during receptor–ligand binding. These experimental data highlight the binding sites for some endogenous agonists; however, other active sites (including the allosteric and cryptic binding sites) are often not identified. Neither the receptor nor the ligand is a frozen/rigid entity; instead, the structures are interacting under constant motion in a solution (any biological fluid). Moreover, an approaching ligand can cause a series of conformational changes in the receptor structure to improve its binding affinity. In order to consider the flexibility of the macromolecular structures, the relaxed complex scheme (RCS) has been developed that extracts several conformations of the receptor sampled using simulation and then performs molecular docking of the ligand with each of the conformations. The scoring functions often consider conformational entropy and solvation energy as

negligible parameters while calculating the binding affinity to make the process computationally less expensive (but compromising with the model's accuracy).

Often researchers employ both QSAR modeling and molecular docking to predict the bioactivity and investigate the mechanism of action of compounds. In a study, the immunomodulatory effect of the ligands is evaluated by employing forward stepwise multiple linear regression to develop a QSAR model with 52 physical-chemical descriptors (important ones are namely dipole moment, steric energy, amide group count, λmax (UV-visible) and molar refractivity) using the SCIGRESS platform. Finally, molecular docking is performed to predict their binding affinity with immunomodulatory targets namely TLR-4, iNOS, COX-2, CD14, IKK b, CD86, and COX-1 (Yadav et al. 2010). A similar QSAR model with 50 descriptors from SYBYL-X 1.3 is used to study the cytotoxicity of ursolic acid analogs against human glioblastoma and lung cancer cell lines. The model exhibited a good regression coefficient (r^2) and the cross-validation regression coefficient (r_{cv}^2) (values ranging from 0.8 to 0.96). The relevant parameters for cytotoxicity are found to be LUMO energy, ring count, dipole vector and solvent-accessible surface area (Kalani et al. 2012).

Some freely available software and webservers to generate descriptors (that include arithmetical, topological, constitutional, geometrical, electrostatic, thermodynamic, quantum-chemical descriptors and other molecular fingerprints) are AFGen (http://glaros.dtc.umn.edu/gkhome/afgen/overview), ISIDA-fragmentor (https://complex-matter.unistra.fr/equipes-de-recherche/laboratoire-de-chemoinformatique/software-development/#c89382), E-DRAGON (http://www.vcclab.org/lab/edragon/), Open3DQSAR (https://open3dqsar.sourceforge.net/), ToMoCoMD-CARDD (http://tomocomd.com/), MOLGEN (http://molgen.de/?src=documents/molgenqspr.html), Mold2 (https://www.fda.gov/science-research/bioinformatics-tools/mold2), Toxicity Estimation Software Tool or TEST by United States Environmental Protection Agency (https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test) and Open Babel (http://openbabel.org/wiki/Main_Page) while commercial alternatives are The CODESSA PRO project (http://www.codessa-pro.com/). Along with the model's high internal accuracy (i.e., $R^2 > 0.9$ and $Rcv^2 > 0.8$ calculated using the training set only), external validation of the (Q)SAR model with experimental data is desirable as per the OECD guidelines (www.oecd.org/env/ehs/risk-assessment/37849783.pdf). In order to better correlate the structural characteristics with the bioactivities, one must use molar units (such as mol/kg, or mmol/kg) instead of mass units (i.e., mg/kg) in the models (Dearden et al. 2009). In inverse docking or target fishing, identification of the possible targets/receptors for the query ligand is performed by software such as GOLD, FlexX, TarFisDock, TarSearch-X, and TarSearch-M.

The evaluation of the bioactivities of a novel compound (i.e., the potential drug targets) can be accomplished by using pair similarity with known compounds (e.g., ChEMBL database calculates the Tanimoto coefficient based on fingerprints), molecular docking, pharmacophore modeling, Bayesian statistics and designing substructural descriptors or fingerprints. However, one must take to avoid the

"activity-cliff" problem in the model that arises when the compounds share analogous structural characteristics but exhibit dissimilar bioactivity spectra. Despite being a rapid and efficient technique in virtual screening, pharmacophore modeling essentially relies on the knowledge of reported active ligands, necessitates sampling conformers using a search algorithm, and is based on a rigid framework for searching hit compounds from the database (Horvath 2010; Kaushik et al. 2018; Lans et al. 2020).

Some platforms to perform protein-protein or protein-DNA docking include SPServer (http://aleph.upf.edu/spserver/), pyDockDNA (https://model3dbio.csic.es/pydockdna), CoDockPP (http://codockpp.schanglab.org.cn/), DOCKSCORE (http://caps.ncbs.res.in/dockscore/), PIIMS Server (http://chemyang.ccnu.edu.cn/ccb/server/PIIMS/index.php), GalaxyDomDock (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=DOMDOCK_INTRO), P3DOCK server, HDOCK server, and GRAMM (Global RAnge Molecular Matching) (https://gramm.compbio.ku.edu/). ezCADD is a fast 2D/3D molecular visualization software that allows small-molecule docking, protein-protein docking, prediction of binding sites, identification of drug targets, homology modeling and structure quality assessment (Tao et al. 2019). FragVLib is an open-source software (distributed under the GNU General Public License) that generates a virtual library of ligand fragments (used for structure-based drug designing) by searching the binding pocket similarity considering a database of ligand-receptor complexes (Khashan 2012). eMolFrag is used for the virtual fragmentation of molecules and extracts the molecular fragments to build a library for virtual screening (Liu et al. 2017). Other software/servers used in virtual screening (structure-based and/or ligand-based) can be listed as follows, although other popular platforms do exist: DENVIS (https://github.com/deeplab-ai/denvis), ReMODE (Receptor-based MOlecular Design for de novo drug designing available at http://cadd.zju.edu.cn/relation/remode/), Pocket2Drug (https://github.com/shiwentao00/Pocket2Drug), DrugRep (http://cao.labshare.cn/drugrep/), DockingPie (a docking plugin for PyMOL), CB-Dock2 (https://cadd.labshare.cn/cb-dock2/php/index.php), PharmRF (https://github.com/Prasanth-Kumar87/PharmRF), DeepDock (https://github.com/OptiMaL-PSE-Lab/DeepDock), Knime workflow (https://hub.knime.com/), RNALigands (http://rnaligands.ccbr.utoronto.ca/php/downloads.php), AutoDock Vina (https://vina.scripps.edu/), eSPC (https://spc.embl-hamburg.de/), RASPDplus (https://github.com/HITS-MCM/RASPDplus), LigRMSD (https://ligrmsd.appsbio.utalca.cl/), LeDock (http://www.lephar.com/index.htm), VSpipe (https://github.com/sabifo4/VSpipe), PyRx (https://pyrx.sourceforge.io/), LiSiCA ((Ligand Similarity using Clique Algorithm available at http://insilab.org/lisica/), ALIDE (http://chemyang.ccnu.edu.cn/ccb/server/AILDE/), Open3DALIGN (https://open3dalign.sourceforge.net/), PrepFlow (https://ifm.chimie.unistra.fr/prep-flow), QSAR-Co-X (https://github.com/ncordeirfcup/QSAR-Co-X), PyRMD (https://github.com/cosconatilab/PyRMD), SwissSimilarity (http://www.swisssimilarity.ch/), PharmMapper (http://lilab-ecust.cn/pharmmapper/), and ZINCPharmer (http://zinc-pharmer.csb.pitt.edu/).

7 Molecular Dynamics Simulation

The deterministic approach of the quantum-mechanical model of motion in the macroscopic world contrasts the use of probability functions that describe the motion in the microscopic world. This is because the electron clouds (that interact while bonding) exhibit wave-particle duality and not simple mechanical bonding. Simulating the system of proteins and other receptor molecules interacting with ligands at the atomistic level has paved its importance to the drug discovery process. The breakthroughs in hardware-based computational power and the development of new algorithms ease the calculation of molecular forces that exist in the system. The limitations of the conventional "lock and key" model of receptor–ligand interaction (where the receptor is held rigid and conformational sampling of the ligand is done, restricting the atomistic motions to keep the model simple) are overcome by such simulations. This considers the dynamic nature of the proteins, thus sampling numerous conformational states and selectively stabilizing them when an agonist or antagonist interacts. Any simulation starts with the modeling of the receptor–ligand system (using the data obtained from NMR, crystallography, or homology modeling), subsequently, the forces experienced by each atom (present in the system) are estimated and the positional changes of atoms are done following Newton's laws of motion. These forces are the results of bonded interactions (i.e., charged/electrostatic interactions that use Coulomb's law to generate the model) and nonbonded interactions (i.e., van der Waals interactions that use the Lennard-Jones 6–12 potential for modeling). Virtual springs and sinusoidal functions are used in the estimation of the difference in potential energy between eclipsed and staggered conformations. The parameters used in the functions identify the stiffness and lengths of the springs, estimate the atomic angles (and dihedral angles), calculate the partial atomic charges (responsible for electrostatic interactions), and predict the van der Waals atomic radii. These parameterizations form the basis of a "force-field" that depicts the nature of molecular dynamics under the influence of several atomic forces. Finally, the simulation time is advanced (by 1–2 femtoseconds, i.e., 10^{-15} s), and the process is iterated (in the order to 10^6) (Durrant and McCammon 2011). Different force fields exist depending on how they are parameterized, although they mostly generate similar outputs. AMBER, CHARMM, and GROMOS force fields are generally encountered in simulation modeling. Molecular dynamics simulation demands performing a huge number of calculations; hence, computer clusters or supercomputers with numerous processors need to operate parallelly. Message Passing Interface (MPI) compatible simulation software like NAMD, CHARMM, and AMBER help in connecting multiple processors so that they can be simultaneously used to execute a complex assignment. Such simulations can estimate the values of NMR-related parameters (e.g., spin relaxation), thus allowing comparison between the theoretical prediction and experimental value.

Simulating molecular systems follows Newton's laws of motion. Such simulations output trajectory graphs for evaluating the stability of the target protein or its docked complexes. In order to perform molecular dynamics simulation, the protein

topology is generated by applying force fields such as Amber and Gromos (using GROMACS or LEaP program), while the PRODRG server can be used for obtaining ligand topology (Strasser and Wittmann 2013). The structures are placed inside a cube and solvation is done using the flexible simple point-charge (SPC) water model. Followed by system neutralization, the steepest descent algorithm minimizes the energy of the system. At a particular temperature (let's say 300 K), position-restraining simulations are performed for a certain period of time under constant volume and temperature dynamics (NVT) and pressure and temperature dynamics (NPT). LINear Constraint Solver (LINCS) algorithm is frequently reviewed for molecular simulations with bond constraints (Hess et al. 1997). The Particle Mesh Ewald algorithm estimates the electrostatic energy (Madelung energy) of the complex/crystal. After performing the molecular dynamics simulation, the trajectories (w.r.t. time) are generated by the XMGrace tool and the parameters namely the root mean square deviation (RMSD), root mean square fluctuation (RMSF), the radius of gyration (Rg), and intermolecular hydrogen bond formations are considered to analyze the stability of the protein-ligand complex (Van Der Spoel et al. 2005). The advantages of molecular dynamics simulation come with a cost—the process becomes computationally expensive. Lower simulation time will reflect the inadequacy (of the model) in the conformation sampling step. Force fields are used in the approximation of the quantum-mechanical model of motion at the atomistic level; hence, molecular dynamics simulations fail largely for the systems having dominant quantum effects such as bonds involving transition metal atoms (Durrant and McCammon 2011). The tools/platforms that can be employed to perform molecular dynamics simulations and analyze the output files post simulation are reviewed in Table 4.

8   Conclusion

To ease the process of drug discovery, a shift toward the application of computational tools is witnessed in the current era of research. Challenges arising due to the pleiotropic nature of biomolecules and the interaction of chemical compounds with multiple pharmacological targets (often encountered in combinatorial/multitargeted approaches) can be addressed by chemo- and bioinformatics tools that make use of databases on physicochemical characteristics and therapeutic use of compounds. Early prediction of ADMET properties of a chemical compound can be of utmost importance since most drug failures occur in the later phases due to undesirable pharmacokinetics and toxicological characteristics. Simulating the system of proteins and other receptor molecules interacting with ligands at the atomistic level has paved its importance to identifying novel drug-like compounds. This considers the dynamic nature of the proteins, thus sampling numerous conformational states and selectively stabilizing them when an agonist or antagonist interacts. The breakthroughs in hardware-based computational power and the development of new algorithms ease the calculation of molecular forces that exist in the system. Biological

Table 4 Platforms to perform molecular dynamics simulations and analysis of output files

Software/platform	Description	URL
YASARA (Yet Another Scientific Artificial Reality Application)	It is compatible with Windows, Linux, macOS and Android and requires a license fee to design photorealistic molecular graphics, and models and perform simulations	http://www.yasara. org/
Abalone	It is a package of molecular graphics used for molecular modeling, geometry optimization, and simulations of proteins, nucleic acids, and ligands. It includes features such as semiautomated parameterization of force fields, GPU acceleration, interfacing with quantum chemical programs, and scripting	http://www. biomolecular- modeling.com/ Abalone/index.html
NAMD	It is based on Charm++ parallel objects and exhibits high-performance molecular dynamics simulation of complex systems (having a large number of atoms). It uses VMD (a tool for molecular graphics) for analyzing trajectories	https://www.ks. uiuc.edu/Research/ namd/
LARMD Server (Ligand and Receptor Molecular Dynamics)	It is used to study the molecular dynamics of protein-ligand interactions. It helps in structure preparation, building force field libraries for molecules, energy calculation (MM/PBSA), and analysis of hydrogen bonds, trajectories, Root-Mean-Square Deviation (RMSD), Radius of Gyration (Rg), Fraction of Native Contacts (Q), Root Mean Square Fluctuation (RMSF), B-factor, Principal Component Analysis (PCA) being some features among others	http://chemyang. ccnu.edu.cn/ccb/ server/LARMD/
GROMACS	A free and open-source platform for performing molecular dynamics simulations and their output analysis	https://www. gromacs.org/
GROMITA	Written in Perl language, it is a GUI (providing both a window-based environment and a terminal mode) front-end for GROMACS (Sellis et al. 2009)	http://gromita.bio. demokritos.gr/
YAMACS	It is a collection of plugins for performing GROMACS simulations through the YASARA platform	https://github.com/ YAMACS-SML/ YAMACS
3dRS (3-dimensional structure Representation Sharing)	It is used to visualize 3D biological structures and molecular dynamics trajectories (Bayarri et al. 2021)	https://mmb. irbbarcelona. org/3dRS/
LAMMPS	It stands for Large-scale Atomic/Molecular Massively Parallel Simulator. It runs on a single processor and also supports Message Passing Interface for parallel operation	https://www. lammps.org/
BitClust	It is based on Python and follows Daura's algorithm and is used for efficient clustering of relatively long molecular dynamics trajectories	https://pypi.org/ project/bitclust/
MDAnalysis	It is a Python library exclusively used in the analysis of molecular dynamics trajectories	https://www. mdanalysis.org/

(continued)

Table 4 (continued)

Software/platform	Description	URL
Tinker-HP	It utilizes multiple CPUs and GPUs and supports MPI to ease long molecular dynamics simulations	https://tinker-hp.org/?Download-instructions
QwikMD	It provides an easy connection between VMD and NAMD for performing hassle-free simulations following a few steps	http://www.ks.uiuc.edu/Research/qwikmd/
PREFMD	It stands for Protein structure REFinement via Molecular Dynamics. The input file must be in PDB format	http://feig.bch.msu.edu/prefmd
Amber	It is a collection of molecular dynamics simulation programs (Salomon-Ferrer et al. 2012)	http://ambermd.org/
DynOmics ENM server	It is used to study the dynamics of biological systems. It integrates two elastic network models (ENMs)—the Gaussian Network Model (GNM) and the Anisotropic Network Model (ANM) (Li et al.)	https://dyn.life.nthu.edu.tw/oENM/

networks have the capability to highlight the interacting elements within a complex biochemical system, thus aiding in the visualization and exploration of big data. In brief, molecular docking, pharmacophore modeling, methods relating (Q)SAR, molecular dynamics simulation, network pharmacology, and machine learning algorithms accelerate the drug discovery process and complement the traditional bioactivity-guided fractionation, high-throughput screening, and systems biology approaches. The examples that are listed/tabularized in this chapter highlight only a fraction of popular software/platforms and encourage the readers to explore other alternatives in various domains of drug discovery and protein engineering.

References

Agnihotry S, Pathak RK, Singh DB, Tiwari A, Hussain I (2022) Protein structure prediction. In: Bioinformatics: methods and applications. Academic Press, pp 177–188. https://doi.org/10.1016/B978-0-323-89775-4.00023-7

Almeida VM, Dias ÊR, Souza BC, Cruz JN, Santos CBR, Leite FHA, Queiroz RF, Branco A (2022) Methoxylated flavonols from Vellozia dasypus Seub ethyl acetate active myeloperoxidase extract: in vitro and in silico assays. J Biomol Struct Dyn 40:7574–7583. https://doi.org/10.1080/07391102.2021.1900916

Alves FS, Cruz JN, de Farias Ramos IN, do Nascimento Brandão DL, Queiroz RN, da Silva GV, da Silva GV, Dolabela MF, da Costa ML, Khayat AS, de Arimatéia Rodrigues do Rego J, do Socorro Barros Brasil D (2023) Evaluation of antimicrobial activity and cytotoxicity effects of extracts of Piper nigrum L and piperine. Separations 10. https://doi.org/10.3390/separations10010021

Aydlnkal RM, Serçinoğlu O, Ozbek P (2019) ProSNEx: a web-based application for exploration and analysis of protein structures using network formalism. Nucleic Acids Res 47:W471–W476. https://doi.org/10.1093/NAR/GKZ390

Bayarri G, Hospital A, Orozco M (2021) 3dRS, a web-based tool to share interactive representations of 3D biomolecular structures and molecular dynamics trajectories. Front Mol Biosci 8. https://doi.org/10.3389/FMOLB.2021.726232/FULL

Bordás B, Kömíves T, Lopata A (2003) Ligand-based computer-aided pesticide design. A review of applications of the CoMFA and CoMSIA methodologies. Pest Manag Sci 59:393–400. https://doi.org/10.1002/PS.614

Braga R, chemistry CA-C topics in medicinal, 2013 undefined Assessing the performance of 3D pharmacophore models in virtual screening: how good are they? ingentaconnect.com

Browne R, Thomas S, in JR-C, 2021 undefined Bioinfomatics as a tool in drug designing. Wiley Online Library

Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S (2017) Protein Data Bank (PDB): the single global macromolecular structure archive. Methods Mol Biol 1607:627–641. https://doi.org/10.1007/978-1-4939-7000-1_26/COVER

Chakrabarty B, Parekh N (2016) NAPS: network analysis of protein structures. Nucleic Acids Res 44:W375–W382. https://doi.org/10.1093/NAR/GKW383

Chavda J, Bhatt H (2019) 3D-QSAR (CoMFA, CoMSIA, HQSAR and topomer CoMFA), MD simulations and molecular docking studies on purinylpyridine derivatives as B-Raf inhibitors for the treatment of melanoma cancer. Struct Chem 30:2093–2107. https://doi.org/10.1007/S11224-019-01334-9

Clevert DA, Le T, Winter R, Montanari F (2021) Img2Mol – accurate SMILES recognition from molecular graphical depictions. Chem Sci 12:14174–14181. https://doi.org/10.1039/D1SC01839F

Csermely P, Korcsmáros T, Kiss HJM, London G, Nussinov R (2013) Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. Pharmacol Ther 138:333. https://doi.org/10.1016/J.PHARMTHERA.2013.01.016

Daina A, Michielin O, reports VZ-S, 2017 undefined SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. nature.com

de Lima AM, Siqueira AS, Möller MLS, de Souza RC, Cruz JN, ARJ L, da Silva RC, DCF A, da Junior JLSGV, Gonçalves EC (2022) In silico improvement of the cyanobacterial lectin microvirin and mannose interaction. J Biomol Struct Dyn 40:1064–1073. https://doi.org/10.1080/07391102.2020.1821782

Dearden JC, Cronin MTD, Kaiser KLE (2009) How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). SAR QSAR Environ Res 20:241–266. https://doi.org/10.1080/10629360902949567

Dolan M, Noah J, Modeling DH-H, 2011 undefined (2012) Comparison of common homology modeling algorithms: application of user-defined alignments. Springer 857:399–414. https://doi.org/10.1007/978-1-61779-588-6_18

Duke J (2020) Database of biologically active phytochemicals & their activity

Dunkel M, Fullbeck M, Neumann S, Preissner R (2006) SuperNatural: a searchable database of available natural compounds. Nucleic Acids Res 34:D678–D683. https://doi.org/10.1093/NAR/GKJ132

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

Dzemyda G, Kurasova O, Medvedev V, Dzemydaitė G (2019) Visualization of data: methods, software, and applications, 295–307. https://doi.org/10.1007/978-3-030-02487-1_18

Fedyushkina I, Reyes IR, … AL-… SSB, 2014 undefined (1990) Prediction of the action of ligands of steroid hormone receptors. Springer 8:53–58. https://doi.org/10.1134/S1990750814010041

Funahashi A, Morohashi M, … HK-, 2003 undefined CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. oww-files-public.s3.amazonaws.com

Galati S, Di Stefano M, Martinelli E, … MM-IJ of, 2022 undefined (2022) VenomPred: a machine learning based platform for molecular toxicity predictions. mdpi.com. https://doi.org/10.3390/ijms23042105

Ghosh S, Datta A, Choi H (2021) multiSLIDE is a web server for exploring connected elements of biological pathways in multi-omics data. Nat Commun 12(1):1–11. https://doi.org/10.1038/s41467-021-22650-x

Grzegorzewski J, Brandhorst J, Green K, Eleftheriadou D, Duport Y, Barthorscht F, Köller A, Ke DYJ, De Angelis S, König M (2021) PK-DB: pharmacokinetics database for individualized and stratified computational modeling. Nucleic Acids Res 49:D1358–D1364. https://doi.org/10.1093/NAR/GKAA990

Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P (2021) Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Mol Divers 25:1315–1360. https://doi.org/10.1007/S11030-021-10217-3

Hayward S, Leader DP, Al-Shubailly F, Milner-White EJ (2014) Rings and ribbons in protein structures: characterization using helical parameters and Ramachandran plots for repeating dipeptides. Proteins 82:230–239. https://doi.org/10.1002/PROT.24357

Hess B, Bekker H et al (1997) LINCS: a linear constraint solver for molecular simulations. Wiley Online Library 18:1463–1472

Honorato-Zimmer R, Reynaert B, Vergara I, Perez-Acle T (2010) Conan: a platform for complex network analysis

Horvath D (2010) Pharmacophore-based virtual screening. 261–298. https://doi.org/10.1007/978-1-60761-839-3_11

Humer C, Heberle H, Montanari F, Wolf T, Huber F, Henderson R, Heinrich J, Streit M (2022) ChemInformatics Model Explorer (CIME): exploratory analysis of chemical model explanations. J Cheminform 14:1–14. https://doi.org/10.1186/S13321-022-00600-Z/FIGURES/10

Humphrey W, Dalke A, graphics KS-J of molecular, 1996 undefined VMD: visual molecular dynamics. Elsevier

Johansson MU, Zoete V, Michielin O, Guex N (2012) Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. BMC Bioinform 13. https://doi.org/10.1186/1471-2105-13-173

Kaczanowski S, Zielenkiewicz P (2010) Why similar protein sequences encode similar three-dimensional structures? Theor Chem Accounts 125:643–650. https://doi.org/10.1007/S00214-009-0656-3/FIGURES/4

Kalani K, Yadav DK, Khan F, Srivastava SK, Suri N (2012) Pharmacophore, QSAR, and ADME based semisynthesis and in vitro evaluation of ursolic acid analogs for anticancer activity. J Mol Model 18:3389–3413. https://doi.org/10.1007/S00894-011-1327-6

Kar S, Roy K, Leszczynski J (2018) Impact of pharmaceuticals on the environment: risk assessment using QSAR modeling approach. Methods Mol Biol 1800:395–443. https://doi.org/10.1007/978-1-4939-7899-1_19

Karami Y, Guyon F, De Vries S, reports PT-S, 2018 undefined DaReUS-Loop: accurate loop modeling using fragments from remote or unrelated proteins. nature.com

Kaushik AC, Kumar A, Bharadwaj S, Chaudhary R, Sahi S (2018) Three-dimensional (3D) pharmacophore modelling-based drug designing by computational technique. SpringerBriefs Comp Sci:27–31. https://doi.org/10.1007/978-3-319-75732-2_4

Khashan R (2012) FragVLib a free database mining software for generating "fragment-based virtual library" using pocket similarity search of ligand-receptor complexes. J Cheminform 4. https://doi.org/10.1186/1758-2946-4-18

Koumakis L, Kanterakis A, Kartsaki E, Chatzimina M, Zervakis M, Tsiknakis M, Vassou D, Kafetzopoulos D, Marias K, Moustakis V, Potamias G (2016) MinePath: mining for phenotype differential sub-paths in molecular pathways. PLoS Comput Biol 12. https://doi.org/10.1371/JOURNAL.PCBI.1005187

Lagunin A, Ivanov S, Rudik A, Filimonov D, Poroikov V (2013) DIGEP-Pred: web service for in silico prediction of drug-induced gene expression profiles based on structural formula. Bioinformatics 29:2062–2063. https://doi.org/10.1093/BIOINFORMATICS/BTT322

Lagunin A, Goel R, Gawande D, … PP-N product, 2014 undefined Chemo-and bioinformatics resources for in silico drug discovery from medicinal plants beyond their traditional use: a critical review. pubs.rsc.org. https://doi.org/10.1039/c0xx00000x

Lans I, Palacio-Rodríguez K, Cavasotto CN, Cossio P (2020) Flexi-pharma: a molecule-ranking strategy for virtual screening using pharmacophores from ligand-free conformational ensembles. J Comput Aided Mol Des 34:1063–1077. https://doi.org/10.1007/S10822-020-00329-7

Laplaza R, Peccati F, A. Boto R, Quan C, Carbone A, Piquemal JP, Maday Y, Contreras-García J (2021) NCIPLOT and the analysis of noncovalent interactions using the reduced density gradient. Wiley Interdiscip Rev Comput Mol Sci 11. https://doi.org/10.1002/WCMS.1497

Lei T, Li Y, Song Y, Li D, Sun H, Hou T (2016) ADMET evaluation in drug discovery: 15. Accurate prediction of rat oral acute toxicity using relevance vector machine and consensus modeling. J Cheminform 8:1–19. https://doi.org/10.1186/S13321-016-0117-7/TABLES/10

Lewis-Atwell T, Townsend P, Tetrahedron M-G, 2021 undefined Comparisons of different force fields in conformational analysis and searching of organic molecules: a review. Elsevier

Li H, Chang Y, Lee J, … IB-N acids, 2017 undefined DynOmics: dynamics of structural proteome and beyond. academic.oup.com

Lim N, Pavlidis P (2021) Evaluation of connectivity map shows limited reproducibility in drug repositioning. Sci Reports 11(1):1–14. https://doi.org/10.1038/s41598-021-97005-z

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

Liu T, Naderi M, Alvin C et al (2017) Undefined (2017) break down in order to build up: decomposing small molecules for fragment-based drug design with eMolFrag. ACS Publications 57:627–631. https://doi.org/10.1021/acs.jcim.6b00596

Lu J, Bioinformatics HC-, 2016 undefined ChemTreeMap: an interactive map of biochemical similarity in molecular datasets. academic.oup.com

Lu XJ, Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res 31:5108–5121. https://doi.org/10.1093/NAR/GKG680

Luo W, Pant G, Bhavnasi YK, Blanchard SG, Brouwer C (2017) Pathview Web: user friendly pathway visualization and data integration. Nucleic Acids Res 45:W501–W508. https://doi.org/10.1093/NAR/GKX372

Magnotti EL, Moy J, Sleppy R, Carey A, Firdyiwek Y, Garrett RH, Grisham CM (2019) Developing and implementing a free online protein structure and function exploration project to teach undergraduate students macromolecular structure-function relationships. J Chem Educ 96:729–733. https://doi.org/10.1021/ACS.JCHEMED.8B00956

Malathi K, Engineering SR-B, G, 2018 undefined (2018) Bioinformatics approaches for new drug discovery: a review. Taylor & Francis 34:243–260. https://doi.org/10.1080/02648725.2018.1502984

Marvel SW, To K, Grimm FA, Wright FA, Rusyn I, Reif DM (2018) ToxPi Graphical User Interface 2.0: dynamic exploration, visualization, and sharing of integrated data models. BMC Bioinform 19. https://doi.org/10.1186/S12859-018-2089-2

Mosihuzzaman M, Choudhary MI (2008) Protocols on safety, efficacy, standardization, and documentation of herbal medicine (IUPAC technical report). Pure Appl Chem 80:2195–2230. https://doi.org/10.1351/PAC200880102195/HTML

Muzammil S, Neves Cruz J, Mumtaz R, Rasul I, Hayat S, Khan MA, Khan AM, Ijaz MU, Lima RR, Zubair M (2023) Effects of drying temperature and solvents on in vitro diabetic wound healing potential of Moringa oleifera leaf extracts. Molecules 28

Pearson WR (2013) An introduction to sequence similarity ("homology") searching. Curr Protoc Bioinformatics. https://doi.org/10.1002/0471250953.BI0301S42

Pirch S, Müller F, Iofinova E, Pazmandi J, Hütter CVR, Chiettini M, Sin C, Boztug K, Podkosova I, Kaufmann H, Menche J (2021) The VRNetzer platform enables interactive network analysis in virtual reality. Nat Commun 12(1):–14. https://doi.org/10.1038/s41467-021-22570-w

Probst D, Reymond JL (2018) SmilesDrawer: parsing and drawing SMILES-encoded molecular structures using client-side JavaScript. J Chem Inf Model 58:1–7. https://doi.org/10.1021/ACS.JCIM.7B00425/SUPPL_FILE/CI7B00425_SI_001.PDF

Rajan K, Zielesny A, Steinbeck C (2021) DECIMER 1.0: deep learning for chemical image recognition using transformers. J Cheminform 13:1–16. https://doi.org/10.1186/S13321-021-00538-8/TABLES/15

Rego CMA, Francisco AF, Boeno CN, Paloschi MV, Lopes JA, MDS S, Santana HM, Serrath SN, Rodrigues JE, Lemos CTL, Dutra RSS, da Cruz JN, dos Santos CBR, da S. Setúbal S, Fontes MRM, Soares AM, Pires WL, Zuliani JP (2022) Inflammasome NLRP3 activation induced by Convulxin, a C-type lectin-like isolated from Crotalus durissus terrificus snake venom. Sci Rep 12:1–17. https://doi.org/10.1038/s41598-022-08735-7

Rifaioglu AS, Atas H, Martin MJ, Cetin-Atalay R, Atalay V, Doğan T (2019) Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. Brief Bioinform 20:1878–1912. https://doi.org/10.1093/BIB/BBY061

Saeedipour S, Tai D, Fang J (2015) ChemCom: a software program for searching and comparing chemical libraries. J Chem Inf Model 55:1292–1296. https://doi.org/10.1021/CI500713S

Sahoo RN, Pattanaik S, Pattnaik G, Mallick S, Mohapatra R Review on the use of Molecular Docking as the First Line Tool in Drug Discovery and Development. ijpsonline.com 1334

Salomon-Ferrer R, … DC-WI, 2013 undefined (2012) An overview of the Amber biomolecular simulation package. Wiley Online Library 3:198–210. https://doi.org/10.1002/wcms.1121

Sander T, Freyss J, … M von K-J of chemical, 2015 undefined (2015) DataWarrior: an open-source program for chemistry aware data visualization and analysis. ACS Publications 55:460–473. https://doi.org/10.1021/ci500588j

Schyman P, Liu R, Desai V, Wallqvist A (2017) vNN web server for ADMET predictions. Front Pharmacol 8. https://doi.org/10.3389/FPHAR.2017.00889/FULL

Sellis D, Vlachakis D, Vlassi M (2009) Gromita: a fully integrated Graphical User Interface to Gromacs 4. Bioinform Biol Insights 3:99–102. https://doi.org/10.4137/BBI.S3207

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks genome.cshlp.org. https://doi.org/10.1101/gr.1239303

Singh N, Chaput L, Villoutreix BO (2021) Virtual screening web servers: designing chemical probes and drug candidates in the cyberspace. Brief Bioinform 22:1790–1818. https://doi.org/10.1093/BIB/BBAA034

Spyropoulos IC, Liakopoulos TD, Bagos PG, Hamodrakas SJ (2004) TMRPres2D: high quality visual representation of transmembrane protein models. academic.oup.com 20:3258–3260. https://doi.org/10.1093/bioinformatics/bth358

Stasinakis P, Molecular DN-B and, 2017 undefined (2017) Modeling of DNA and protein organization levels with Cn3D software. Wiley Online Library 45:126–129. https://doi.org/10.1002/bmb.20998

Stierand K, Rarey M (2010) PoseView -- molecular interaction patterns at a glance. J Cheminform 2. https://doi.org/10.1186/1758-2946-2-S1-P50

Strasser A, Wittmann H-J (2013) Construction of ligands. Modelling of GPCRs, 29–36. https://doi.org/10.1007/978-94-007-4596-4_4

Sun X, Zhang Y, Zhou Y, Lian X, Yan L, Pan T, Jin T, Xie H, Liang Z, Qiu W, Wang J, Li Z, Zhu F, Sui X (2022) NPCDR: natural product-based drug combination and its disease-specific molecular regulation. Nucleic Acids Res 50:D1324–D1333. https://doi.org/10.1093/NAR/GKAB913

Tang Q, Nie F, Zhao Q, Bioinformatics WC-B in, 2022 undefined A merged molecular representation deep learning method for blood–brain barrier permeability prediction. academic.oup.com 2022:1–10. https://doi.org/10.1093/bib/bbac357

Tao A, Huang Y, Shinohara Y, Caylor ML, Pashikanti S, Xu D (2019) EzCADD: a rapid 2D/3D visualization-enabled web modeling environment for democratizing computer-aided drug design. J Chem Inf Model 59:18–24. https://doi.org/10.1021/ACS.JCIM.8B00633/ASSET/IMAGES/LARGE/CI-2018-006338_0006.JPEG

Tarini M, Cignoni P (2006) QuteMol

Valsecchi C, Grisoni F, Motta S, … LB-T and A, 2020 undefined NURA: a curated dataset of nuclear receptor modulators. Elsevier

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701–1718. https://doi.org/10.1002/JCC.20291

Wang T, Sun J, Medicine QZ-C in B and, 2022 undefined Investigating cardiotoxicity related with hERG channel blockers using molecular fingerprints and graph attention mechanism. Elsevier

Weber JR (2009) ProteinShader: illustrative rendering of macromolecules. BMC Struct Biol 9. https://doi.org/10.1186/1472-6807-9-19

Weir H, Thompson K, Woodward A, Choi B, Braun A, Martínez TJ (2021) ChemPix: automated recognition of hand-drawn hydrocarbon structures using deep learning. Chem Sci 12:10622–10633. https://doi.org/10.1039/D1SC02957F

Wink LH, Baker DL, Cole JA, Parrill AL (2019) A benchmark study of loop modeling methods applied to G protein-coupled receptors. J Comput Aided Mol Des 33:573–595. https://doi.org/10.1007/S10822-019-00196-X

Wong YI, Kwan G (2015) Identification of protein-ligand binding site using machine learning and hybrid pre-processing techniques

Wu L, Yan B, Han J, Li R, Xiao J, … SH-NA, 2022 undefined TOXRIC: a comprehensive database of toxicological data and benchmarks. academic.oup.com

Xie T, Song S, Li S, Ouyang L, proliferation LX-C, 2015 undefined Review of natural product databases. Wiley Online Library

Xie Y, Li H, Luo X, Li H, Gao Q, Zhang L, Teng Y, Zhao Q, Zuo Z, Ren J (2022) IBS 2.0: an upgraded illustrator for the visualization of biological sequences. Nucleic Acids Res 50:W420–W426. https://doi.org/10.1093/NAR/GKAC373

Yadav DK, Meena A, Srivastava A, Chanda D, Khan F, Chattopadhyay S (2010) Development of QSAR model for immunomodulatory activity of natural coumarinolignoids. Drug Des Devel Ther 4:173. https://doi.org/10.2147/DDDT.S10875

Yuan S, Chan H, Reviews ZH-WI, 2017 undefined (2017) Using PyMOL as a platform for computational drug design. Wiley Online Library 7. https://doi.org/10.1002/wcms.1298

Zhang J, Chiodini R, Badr A, Zhang G (2011) The impact of next-generation sequencing on genomics. J Genet Genomics 38:95. https://doi.org/10.1016/J.JGG.2011.02.003

Zhou P, Shang Z (2009) 2D molecular graphics: a flattened world of chemistry and biology. Brief Bioinform 10:247–258. https://doi.org/10.1093/BIB/BBP013

Zhou G, Xia J (2018) OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. Nucleic Acids Res 46:W514–W522. https://doi.org/10.1093/NAR/GKY510