Review Article

Drug Discovery Tools and *In Silico* Techniques: A Review

Yogesh S. Wankhede[1*], Vaishnavi V. Khairnar[2], Amol R. Patil[3], Avinash B. Darekar[4]

[1]Department of Pharmaceutical Chemistry, K.V.N. Naik Institute of Pharmaceutical Education and Research, Nashik, Maharashtra, India.
[2]Department of Pharmaceutical Chemistry, MET's Institute of Pharmacy, Bhujbal Knowledge City Adgaon, Nashik, Maharashtra, India.
[3]Department of Pharmaceutics, K.V.N. Naik Institute of Pharmaceutical Education and Research, Nashik, Maharashtra, India.
[4]Department of Pharmaceutics, K.V.N. Naik Institute of Pharmaceutical Education and Research, Nashik, Maharashtra, India.
***Corresponding author's E-mail:** yogwankhede111@gmail.com

Received: 11-04-2024; **Revised:** 25-06-2024; **Accepted:** 03-07-2024; **Published on:** 15-07-2024.

ABSTRACT

Traditional drug discovery methods have proved effective in developing new medications, however the process from leads identification to clinical studies can take over 12 years and cost over $1.8 billion USD on average. To evaluate the safety of drugs, including toxicity and side effects, in vivo and in vitro methods are usually utilized. ADME-Tox evaluations have increased because to recent developments in in vitro models, such as organ-on-chip technology. These techniques are still expensive, tedious and time-intensive, though. In silico techniques have gained popularity for their ability to save time, labor, and costs associated with drug development. Computational approaches have effectively led to the development of several novel drugs. Finally, we provide effective instances of antibacterial, antiviral, and anticancer drug discoveries made utilizing computational approaches. This review outlines the general processes and techniques involved in in silico drug discovery, such as target protein identification, chemical library screening, and machine learning-based toxicity assessment. It also provides an overview of the databases and prediction tools that are currently available.

Keywords: Drug Discovery, Computational approaches, ADME-tox evaluation, In-silico techniques, Ligand-based drug discovery, Structure-based drug discovery.

1. INTRODUCTION

The methods involved in conventional drug research and development, which include preclinical and clinical trials, lead molecule discovery and optimization, target identification and validation, are time-consuming and toxic. A new medicine's anticipated cost to market has risen to $1.8 billion USD in recent years, and up to 96% of drug ideas are lost to attrition. Inadequate medication effectiveness and poor drug absorption, distribution, metabolism, and excretion (ADME-Tox) are the causes of this high attrition rate. To evaluate the safety of drugs, including toxicity and side effects, in vivo and in vitro methods are usually utilized. ADME-Tox evaluations have increased because to recent developments in in vitro models, such as organ-on-chip technology. These techniques are still expensive, tedious and time-intensive, though. Using automated assays, high-throughput screening (HTS) techniques have been developed to quickly identify chemical compounds that are pharmacologically active among a huge number of molecules. The requirement for human participation is lessened by autonomous HTS systems, although the scope of HTS is still small in comparison to the diversity of chemical structures. Furthermore, automated instruments tend to be costly. Since they can help with the size, time, and expense challenges that traditional experimental techniques confront, computer-aided drug discovery (CADD) approaches have been receiving a growing amount of interest recently. As part of CADD, prospective therapeutic targets are computationally identified, huge chemical libraries are virtually screened for promising drug candidates, candidate compounds are further optimized, and the potential toxicity of each molecule is evaluated in silico. To increase the precision and effectiveness of CADD procedures, a number of approaches have been created and combined with machine learning techniques[1]. Two distinct methods are used in CADD: structure-based drug discovery (SBDD) [2] and ligand-based drug discovery (LBDD)[3]. The availability of target protein structural data is a prerequisite for choosing an appropriate CADD strategy. The target protein's structural details are needed to use the SBDD technique, and these may often be discovered experimentally via X-ray crystallography or nuclear magnetic resonance[2]. The 3D structure of the target protein can be predicted using in silico prediction techniques like homology modeling[4] or ab initio modeling[5] when neither is available. Molecular docking and structure-based virtual screening are acceptable when the structure is determined. The LBDD technique is frequently used as an alternate method when the structure is unknown and in silico approaches cannot predict a high-quality structure. Since many compounds have been found to cure illnesses and are listed in public databases, unless the target is unique, this strategy requires previous knowledge of the known active molecules of the target protein[6-8].

This review outlines the general processes and techniques involved in in silico drug discovery, such as target protein identification, chemical library screening, and machine learning-based toxicity assessment. It also provides an overview of the databases and prediction tools that are currently available.

2. BIOLOGICAL DATA ON CHEMICAL MOLECULES FOR DRUG DISCOVERY

Biological screening has produced large-scale data on hundreds of thousands of smaller molecules over the past few decades. This data is collected in easily accessible the web libraries for scientific research purposes. For instance, large-scale experiments with more than a million compounds have been produced as a result of improvements in HTS methodologies[9].

Furthermore, this biological test data has been organized into databases from chemical libraries, and as chemical synthesis and HTS techniques develop, correspondingly increases an abundance amount of data. Modern in silico drug discovery has been made easier by the gathering data and its public accessibility, which have allowed machine learning models to be developed. Initially in the drug development process, traditional prediction techniques like quantitative structure activity relationship (QSAR) models may be used to rank drug candidates according to their pharmacological characteristics and possible side effects[10]. Many machines learning-based prediction techniques have recently been created to predict drug-target interactions [11], compounds permeability across the blood-brain barrier[12], and the ADMET-Tox characteristics of therapeutic candidates[13,14]. This development has been facilitated by an increase in public resources. Using machine learning algorithms in conjunction with data collection might open up new possibilities for CADD method.[1,15]. Tables 1 and 2 provide an overview of the public databases that are available.

Table 1: Target prediction web servers.

Name	Description
Harmonizome [84]	Harmonizome is an extensive and curated collection of information on genes and proteins that was compiled from more than 70 prominent web sites. - It makes it possible to identify brand-new connections and functional pairings between biological elements (genes and proteins).
Open Targets Platform	The Open Targets Platform is a knowledge-based resource that allows the discovery and prioritization of pharmacological targets and offers proof of the correlation between established therapeutic targets and illnesses.
TargetHunter[85]	TargetHunter produces target predictions by applying the TAMOSIC algorithm, which is effective in predicting the biological targets of chemicals that have been searched.
Similarity Ensemble Approach (SEA)[86]	Based on the chemical similarity of ligands, SEA assigns a ranking to target proteins. - Human protein target groups are allocated 65,000 ligands. A similarity score is computed using ligand topology.
SwissTarget Prediction[87]	SwissTargetPrediction analyzes for similarities between proteins to determine which ones could be prospective drugs targets. "- Along with 3068 macromolecular targets, the revised edition includes 376,342 experimentally active molecules.
SuperPred	SuperPred is a linear regression model that predicts the target proteins of chemicals by training it using ECFP4 fingerprints.
MuSSeL[89]	MuSSeL employs a multifingerprint similarity search technique to anticipate small compounds' possible therapeutic targets.
DisGenNET[90]	DisGenNET gives details regarding genes and variations linked to disorders in humans.
HitPick [92]	HitPick employs three methods to identify potential drug targets from hit compounds: a modified naïve Bayesian model, a one-nearest-neighbor similarity search, and the B-score technique.
MolTarPred[91]	MolTarPred contains a list of prospective pharmacological targets and related molecules.

Table 2: Protein target databases.

Name	Description
DrugBank[93]	13,857 drug entities total, of which 2661 authorized drug molecules and 1425 approved biologics (vaccines, peptides, and proteins)
ChEMBL[94]	1.9 million chemical compounds and 13,382 drug targets in total
ChemBank[95]	Data about millions of micro medicinal molecules and hundreds of biochemical tests
Therapeutic Target Database (TTD)[96]	Experimental validation data regarding 3419 therapeutic targets and 37,316 medicinal compounds
Comparative Toxicogenomics Database	Details regarding 51,300 genes, 5500 phenotypes, 7200 disorders, and 45 million toxicogenomic interactions of 16,300 chemical substances
SuperTarget[97]	Details about 6219 drug targets and 195,770 small drug compounds
ChemSpider[98]	Text and structural data on more than 67 million chemical compounds
The Toxin and Toxin (T3DB)[99]	Details on 2073 targets for toxins, 3678 toxins, and 42,374 linkages between toxins and targets
Promiscuous[100]	Data on 2,727,520 drug-target reactions, 9430 drug targets, and 991,805 small compounds

3. IDENTIFYING THE TARGET

A biological entity that has the ability to modify disease phenotypes—typically a protein—is referred to as a drug or therapeutic target[16]. Therefore, the first and most crucial stage in the drug development process is identifying promising drug targets.

Experimental methods are used to find therapeutic targets by conventional means, such as comparing genes expressed differently in both normal and abnormal tissues or cells, as well as identifying proteins that have a strong correlation with proteins linked to the disease.

3.1. Experimental methods

Studies of the molecular and biochemical pathophysiology of diseases are necessary for conventional research methods for target identification.

Such research broadens our understanding of diseases, but it might take a while to identify potential therapeutic targets. Target identification has been sped up recently with the development of genome-scale screening tools including target deconvolution, stable isotope labeling by amino acids in cell culture (SILAC), and haploinsufficiency profiling (HIP).

By sensitizing cells to chemicals and detecting gene products linked to the survival of disease cell lines, HIP is a genome-wide screening test for finding potential therapeutic targets[17]. The HIP test has the benefit of evaluating hundreds of genes at once and without requiring prior understanding of the development of disease. The complex pathophysiology of many diseases makes it challenging to find therapeutic targets in conventional drug development techniques. In this case, an opposite approach may be used: substances that can alter disease phenotypes can be identified by screening, and matching target proteins.[17]

Various methods are used in targeted, such as protein microarrays, biochemical inhibition and affinity chromatography.[18] SILAC is an effective, reproducible assay that can unbiasedly, completely and robustly identify small molecule probes drug binding protein target.[19,20]

This technology has recently been combined with mass spectrometry based proteomics and affinity chromatography to better elucidate drug-protein interactions[19]. Despite its advantages, SILAC has several disadvantages that hinder its widespread use and applicability: (i) isotopic labeling is expensive, (ii) requires the use of complex equipment such as solution mass spectrometry, and (iii) generation and verification of theimmobilized solution. Biological activity takes a long time.[21]

3.2 Identification of targets computationally

Testing methods are very expensive and are often done at low scale due to their complexity. To overcome these problems, in silico methods have been developed to identify potential drug targets[22].

Target proteins can be predicted based on experimental d ata [22,24], data mining [25], or inference from protein network s[26]. Descriptors and fingerprints are frequently utilized in the creation of prediction models because they provide a quantitative depiction of a compound's physical and chemical properties[27]. Using a subtractive method might aid in improving projected targets. For instance, by eliminating redundant enzymes, homologous enzymes to those of humans or gut flora, extracellular enzymes, non-essential proteins, and other materials from the H. pylori proteome, possible therapeutic targets to cure Helicobacter pylori infection can be found[28].

3.3 Confirmation of targets

Upon identification of a target, the subsequent course of action involves verifying if the alteration of the target's biological function impacts the illness phenotype[29]. Predicted targets may be evaluated and biological functions can be modulated using different techniques. The most popular of these techniques is the use of small interfering RNAs (siRNAs)[30], which imitate the actions of drugs by limiting translation and temporarily suppressing the target protein [31,32] With siRNAs, target inhibition may be studied without the need for inhibitors or prior protein structural information[31]. However, the degree of suppression by siRNAs may impact cellular physiologies differently and may thus result in conflicting outcomes for disorders with complicated pathophysiology, such as neurological diseases[33]. In these circumstances, animal models with altered or deleted target genes may provide more useful information for target confirmation.

4. DRUG SCREENING TECHNIQUES USING IN SILICO METHODS

Identifying small compounds that can alter the function of a target protein that has been discovered and, in turn, alter the disease phenotype is the aim of drug development. Additionally, finding tiny compounds with minimal toxicity and effective pharmacokinetic features is essential. The process of finding new drugs is a protracted, costly, and hazardous series of intricate procedures, such as pharmacokinetics, preclinical toxicity assessments, candidate validation, and drug candidate identification. Conventional medication development and research (R&D) is costly and takes a long period. A medicine typically takes 10 to 12 years to reach the market, and each successful drug's discovery is said to have cost between $800 million and $1.8 billion USD[2,34].

Identifying pharmacologically effective chemical compounds is the initial challenge in the drug discovery process. Experimental HTS typically has a success rate of 0.01% to 0.14%[35]. Approximately 40–60% of medication failures in the latter phases are caused by deficiencies in ADME-Tox, which is another major obstacle[36,37]. For a long time, the pharmaceutical sector has benefited greatly from in silico drug discovery technologies[38–39]. Cost and time efficiency are the key advantages of in silico drug discovery. Furthermore, it may be implemented at any step of the

drug development process, from preclinical and clinical to drug screening [40] significantly lowering the probability of failure.

4.1 Ligand-based drug Design

LBDD methods make use of existing information about active pharmaceuticals, including their structural, physical, and chemical characteristics, to forecast novel pharmacological molecules with comparable biological effects [41]. (Figure 1). Based on the idea that compounds with high structural and physicochemical similarities are more likely to have similar biological activity, drug compounds are predicted based on the similarity of features (e.g., aromaticity, hydrogen bond acceptors, hydrogen bond donors, hydrophobicity, anion, and cation residues) between chemical compounds [42].

When the target protein's 3D structure is unknown, LBDD is often used. When protein structure is unknown, techniques like pharmacophore modeling and QSAR can yield valuable insights into target-ligand interactions [43].

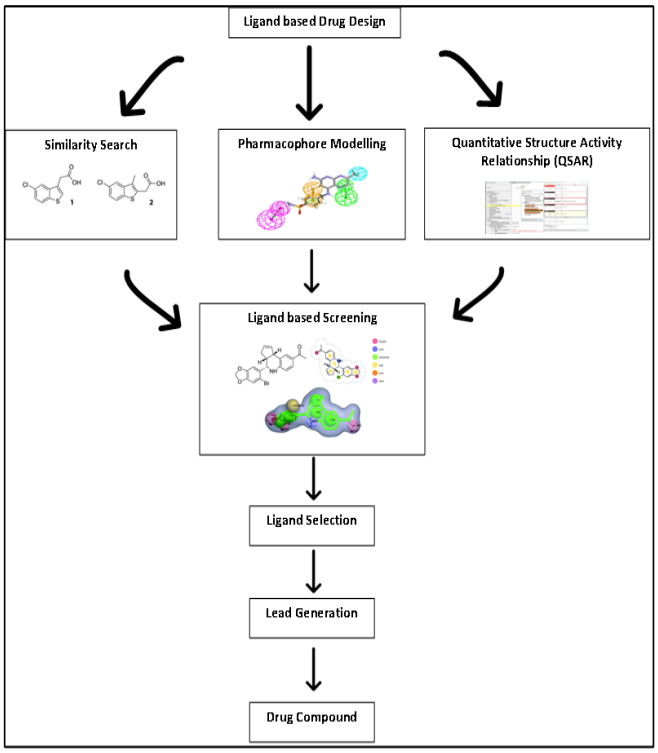

Figure 1: Ligand Based Drug Discovery methodology

4.1.1: Screening for similarity

Discovering novel compounds that resemble well-known chemical compounds can be accomplished through the use of popular and efficient compound similarity searches.

The underlying premise of these techniques is that molecules with comparable physical-chemical characteristics are more likely to have comparable biological activity[44,45]. A similarity search strategy has been used recently to identify numerous powerful molecules, such as agonists for a G-protein-coupled receptor (GPR30) that have been discovered using this technique[46].

4.1.2. Pharmacophore modeling

Compound libraries are screened using pharmacophore models as a query to find compounds with comparable structural and physical-chemical characteristics. Structurally varied active ligands are computed to produce energetically stable conformations in order to find pharmacophores. Their structures are then ordered and stacked in order to locate comparable functional groups that are shared by the active ligands.

These pharmacophore-containing chemical compounds may represent novel therapeutic prospects. In order to find more effective therapeutic compounds, pharmacophore modeling has been used[47,48]. As an effective example, new inhibitors of the bacterial type II topoisomerase bacterial DNA gyrase B.

4.1.3. Quantitative structure-activity relationships

Mathematical models that link a compound's physical and structural characteristics to its biological activity are produced by QSAR techniques. Originally created in 1962 [49] by Hansch and Fujita, QSAR is a well-known technique in drug development. This technique uses QSAR models to predict the biological activity of given chemicals in order to find novel drug compounds or optimize lead molecules. Molecular descriptors, which capture the structural and chemical properties of compounds, are utilized to train the models[50]. Chemical compounds having established biological activity are gathered for the purpose of building QSAR models [51], and these compounds are then employed for model training and assessment. As a recent development to get around the drawbacks of the conventional QSAR techniques, 3D-QSAR approaches have been created.

For QSAR modeling, there are a number of online servers and tools available, such as QSAR-Co[52] and Open3DQSAR[53] (Table 3).

Table 3: QSAR modeling tools

QSAR Tools	Description
QSAR-Co	Programs for creating reliable multi-target classification-based QSAR models using either linear discriminate analysis or random forest approach
Open3DQSAR	Utilizing partial least square chemometric technique for pharmacophore discovery, 3D-QSAR model generating software
SYBYL-X	Lead identification and optimization, macromolecular modeling, and small molecular modeling
McQSAR	A QSAR model-generating extension of a genetic algorithm
QSAR ToolBox	Compounds with comparable structural properties may be identified using a toolkit that combines computational techniques, theoretical knowledge, and experimental data from many sources.

4.2 Structure-based drug discovery

Using their respective structures, ligand and target protein binding affinities, or binding pockets, are computed using Structure-based drug design methods as opposed to ligand-based drug discovery[54] (Fig. 2). The prediction of binding affinity is achieved using a combination of fragment-based docking, molecule docking, and molecular dynamic modeling[55-57]. Several drugs approved by the FDA as well as those undergoing clinical trials were effectively developed through the use of SBDD techniques[58]. The first HIV-1 protease medications authorized by the FDA were amprenavir and saquinavir, which were created using SBDD techniques[59,60].

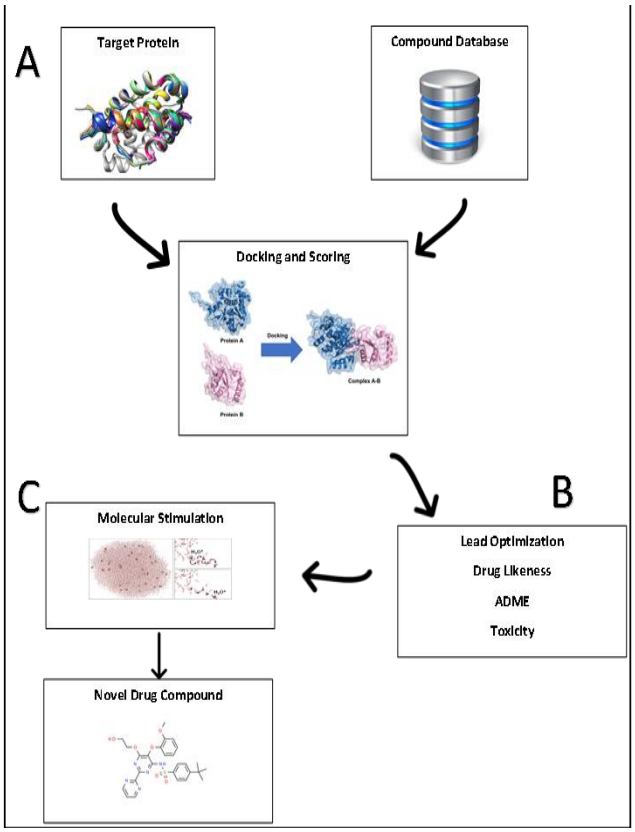

Figure 2: Structure based drug Discovery

4.2.1. Generation of target protein structure

Obtaining a high-resolution 3D structure of the target protein—which may be found in the Protein Data Bank (PDB)[61]—is the first step in the SBDD process. If the structure hasn't been solved yet, it can be predicted from scratch or by using precedent structures with comparable sequences. Proteins with high sequence identity are assumed to have comparable 3D structural conformations and functions in homology modeling. For homology modeling, a variety of tools and internet resources are accessible;

The ability to locate the most thermodynamically stable conformation using an accurate energy function and an effective search strategy that narrows down the pool of potential conformations to the lowest energy state are two key components that contribute to accurate structure predictions.

4.2.2. Predicting the binding site

A concave area or tiny pocket on a protein is called a binding site, and it is here that a ligand molecule attaches to provide the intended effect (inhibition, modulation, or activation)[62]. For SBDD, the structure of co-crystallized ligands with a protein can offer useful insights. In silico techniques can be used to forecast possible binding pockets in the event that structural information on binding pockets is not available[63]. While these instruments are essential for identifying potential binding sites, a number of variables, like pocket size and template similarity, affect how accurate these predictions are. Table 5 contains a list of many binding site prediction methods available today.

Table 4: Tools for Predicting Ligand Binding Sites

Name	Description
3D Ligand Site	A protein's sequence or three-dimensional structure can be supplied by the user. This program finds homolog structures from the PDB that have bonded ligands by predicting the sequence's three-dimensional structure. The binding sites are predicted by superimposing the homolog and query structures.
CASTp 3.0	A protein's surface pockets and internal cavities may be predicted through the use of CASTp, which also offers a thorough description of every atom involved in the development of the pockets.
F pocket	A quick and effective way to anticipate pockets for substantial proteins is to use Fpocket. It offers two programs: (i) dpocket, which extracts the pocket description, and (ii) tpocket, which tests the user's own scoring function.
Pocket Depth	With a 96% accuracy rate, PocketDept is a geometry- and depth-based clustering technique that forecasts binding pockets.
Site Map	SiteMap predicts ligand binding sites based on the whole protein sequence, and then assigns a SiteScore score to each potential binding site, indicating the site's likelihood of ligand binding.

4.2.3. Molecular docking

The process of finding ligands with a high binding affinity by molecular docking begins when the 3D structure of a target protein has been established.

By using electrostatic and van der Waals interactions, molecular docking algorithms determine the optimal orientation of a specific ligand within the binding pocket of a target protein and determine their affinity [64,65]. It is possible to visually screen a huge number of ligands to identify those that have a high binding affinity to the target protein using docking algorithms. In addition to helping to anticipate protein-protein interactions and assess

complicated affinities, docking techniques can also help to improve our knowledge of signaling networks. Since biological processes are mediated by protein-protein interactions, our comprehension of the functional mechanisms and functions of protein-protein docked complexes in the cell can be aided by their prediction [66]. Table 5 contains a list of common docking tools.

Table 5: Molecular docking software.

Name	Description
AutoDock Vina	Approved as a quick and precise docking tool, AutoDock Vina is extensively utilized. - The docking optimization process is accelerated by the employment of many stochastic global optimization techniques, such as genetic algorithms, particle swarm optimization, and simulated annealing. Additionally, it permits the flexible treatment of receptor side chains during docking.
DOCK	Multiple functions are made available by this docking program, including molecular dynamic simulation, ligand and receptor desolvation, ligand conformational entropy correction, Hawkins–Cramer–Truhlar GB/SA solvation, and receptor flexibility during docking analysis.
pyDOCK	Using the sophisticated pyDock scoring method, pyDOCK is a quick and effective web server for rigid-body docking prediction.
Discovery Studio	Molecular dynamic/quantum mechanics/molecular mechanics simulations, macromolecule design, structure- and ligand-based drug discovery, pharmacophore and QSAR modeling are all supported by Discovery Studio, an integrated drug discovery platform.
MOE	With MOE, you can efficiently create QSAR models, conduct virtual screening, develop structure-based drugs, and model molecules through integrated drug discovery tools.
Surflex-dock	The platform Surflex-dock may be used for a number of tasks, including ligand modeling, protein structure alignment and preparation, molecular docking for virtual screening, and molecular conversion from 2D to 3D.
Glide	- The Glide fast-docking approach ranks anticipated ligand binding conformations in the binding cavity of a receptor using three distinct scoring functions (SP, XP, and HTV) and a sequence of hierarchical filters.
ClusPro	ClusPro applied a quick Fourier transform-based docking technique to predict peptide protein docking accurately and quickly.
GEMDOCK	Using its empirical scoring function, GEMDOCK offers a very accurate way to predict the shape and orientation of ligands within a receptor's binding region.

4.2.4 Docking based on fragments

There are structural components (fragments) in drug compounds. Some of these fragments, like the pharmacophore, are necessary to demonstrate biological activity, while others are just structurally necessary to put substructures together. The full structures of chemical compounds are used in conventional molecular docking methods to determine the compounds' binding affinities with binding pockets. While the affinity of fragments detected by fragment-based docking techniques is often lower than that of complete ligand complexes, it is nevertheless tolerable [67]. Subsequently, functional groups are added to the fragments or they are combined with additional fragments to maximize their binding affinity [68]. The initial stage in fragment-based docking is to create a structurally varied library of fragments [69]. When building druggable fragments, the "rule of three" is often followed[70]: a molecular weight <300 Da, a cLogP ≤3, hydrogen bond donors ≤3, and hydrogen bond acceptors ≤3 [68-70]. The next step is screening potent fragments according to the binding affinity that was calculated using standard molecular docking techniques. The screened segments usually have weak affinities because they contain important substructures like pharmacophores. Therefore, screened fragments are altered by adding functional groups or additional fragments to increase their effectiveness.

4.2.5. Molecular dynamic simulation

Through the use of MD simulations, chemical compounds may be virtually screened for potential drugs by gaining a knowledge of the structural characteristics of proteins and the stability of protein-ligand complexes. More pharmacological molecules with greater efficacy can be designed as a result of its assistance in identifying other binding sites that can be drugged, such as allosteric sites[74-75].

Proteins are flexible, and their flexibility is important in ligand binding, but prediction of the motions of protein binding pockets and ligands involves high computational cost due to the complex atomic interactions between the target protein and ligand molecule. Molecular dynamics (MD) simulation was first introduced in the 1970s to overcome this limitation [71]. It involves solution of Newton's equation of motion to simulate atomic motions and to reduce the calculation complexity [72,73].

The best-docked complexes in computational drug discovery are often submitted to MD simulations to verify their binding. In summary, standard parameters are used to build protein and ligand topologies.

5. ADME-TOX EVALUATION

The process of evaluating pharmacokinetic characteristics, such as ADME-Tox, accompanies the discovery of potential new drugs. Computing techniques can also be used to forecast ADME-Tox because of developments in machine learning algorithms and gathered datasets.

Preclinical studies are thought to remove 40–60% of medication candidates due to ADME-Tox issues [36]. In order to achieve their pharmacological effects, drug molecules need to pass across a number of physiological barriers, including the blood-brain barrier, the gastrointestinal barrier, and microcirculatory barriers.

They could need to be metabolically converted in order to be activated, or they might transform into a hazardous substance with unfavorable consequences [76].

For ADME-Tox evaluations, conventional experimental techniques are still time-consuming and expensive. Lipinski's "rule of five"–a molecular weight <500 Da, lipophilicity <5, number of rotatable bonds <10, hydrogen bond donors <5, and hydrogen bond acceptors <10–is an easier guideline for determining a chemical compound's drug-likeness. More sophisticated prediction techniques are being employed more frequently these days to forecast drug-likeness in terms of ADME-Tox characteristics, as opposed to relying just on this basic guideline. Many models based on machine learning have been created to forecast the pharmacokinetic characteristics of chemical substances.

Table 6: provides a list of online servers and tools that are accessible for ADME-Tox predictions.

Name	Description
SwissADME	Provides a user-friendly environment to compute physicochemical descriptors and ADME parameters
ADMETlab	Computes ADME and toxicity features of compounds
PreADMET 2.0	Provides numerical information on the ADME and toxicity of chemical compounds
LightBBB	Predicts blood-brain barrier permeability of compounds
ToxinPred	Predicts and designs toxic and non-toxic peptides
ProTox-II	Predicts the toxicity profile of compounds

6. EFFECTIVE USES OF IN SILICO DRUG DEVELOPMENT

The process of creating new therapeutic medications is costly and laborious. In the modern pharmaceutical business, in silico technology has become indispensable since it can cut down on the time and resources needed for drug development. The drug discovery process now includes computational prediction tools at every step thanks to developments in computational algorithms and knowledge databases. A wide range of disorders, including cancer [77,78], diabetes, viral [79,80], and bacterial infections[81,82,84], have been effectively treated using therapeutic molecules designed and identified by computational drug discovery approaches.

CONCLUSIONS

The efficiency and accuracy of in silico drug target and therapeutic drug identification has increased during the last few decades. The accumulation of publicly available biological data and the quick development of computer techniques have recently led to an acceleration in in silico drug discovery. The biological activities of targets are clarified by chemical biology, and the discovery of prospective drug candidates is facilitated by CADD approaches, which employ structural information of either the drug target (structure-based) or ligands with known bioactivity (ligand-based). Because CADD approaches may expedite drug discovery by utilizing current information on ligand receptor interactions, structural optimization, and synthesis, they are now a crucial component of the drug development process.

Author contributions

YS Wankhede: conceptualization, data curation, VV Khairnar and AR Patil: methodology, manuscript writing. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement: The Authors are thankful to K.V.N. Naik Institute of Pharmaceutical Education and Research, Nashik, Maharashtra, India for providing necessary facilities.

Source of Support: The author(s) received no financial support for the research, authorship, and/or publication of this article

REFERENCES

1. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18:463–477.
2. Jhoti H, Leach AR. Structure-based Drug Discovery. Springer; 2007.
3. Vidal D, Garcia-Serna R, Mestres J. Ligand-based approaches to in silico pharmacology. In: Chemoinformatics and Computational Chemical Biology. Springer; 2011. p. 489–502.
4. Cavasotto CN, Phatak SS. Homology modeling in drug discovery: current trends and applications. Drug Discov Today. 2009;14:676–683.
5. Wu S, Skolnick J, Zhang Y. Ab initio modeling of small proteins by iterative TASSER simulations. BMC Biol. 2007;5:1–10.
6. Tatum NJ, Liebeschuetz JW, Cole JC, Frita R, Herledan A, Baulard AR, et al. New active leads for tuberculosis booster drugs by structure-based drug discovery. Org Biomol Chem. 2017;15:10245–10255.
7. Evers A, Klabunde T. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. J Med Chem. 2005;48:1088–1097.
8. Talele TT, Khedkar SA, Rigby AC. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. Curr Top Med Chem. 2010;10:127–141.
9. Brandish PE, Chiu C-S, Schneeweis J, Brandon NJ, Leech CL, Kornienko O, et al. A cell-based ultra-high-throughput screening assay for identifying inhibitors of D-amino acid oxidase. J Biomol Screen. 2006;11:481–487.

10. Tropsha A. QSAR in Drug Discovery, Drug Design: Structure-And Ligand-Based Approaches. 2010: p. 1.
11. Nascimento AC, Prudêncio RB, Costa IG. A multiple kernel learning algorithm for drug-target interaction prediction. BMC Bioinf. 2016;17:1–16.
12. Wang W, Kim MT, Sedykh A, Zhu H. Developing enhanced blood–brain barrier permeability models: integrating external bio-assay data in QSAR modeling. Pharm Res. 2015;32:3055–3065.
13. Schyman P, Liu R, Desai V, Wallqvist A. vNN web server for ADMET predictions. Front Pharmacol. 2017;8:889.
14. Liu R, Tawa G, Wallqvist A. Locally weighted learning methods for predicting dose-dependent toxicity with application to the human maximum recommended daily dose. Chem Res Toxicol. 2012;25:2216–2226.
15. Hochreiter S, Klambauer G, Rarey M. Machine Learning in Drug Discovery. ACS Publications; 2018.
16. Katsila T, Spyroulias GA, Patrinos GP, Matsoukas M-T. Computational approaches in target identification and drug discovery. Comput Struct Biotechnol J. 2016;14:177–184.
17. Giaever G, Shoemaker DD, Jones TW, Liang H, Winzeler EA, Astromoff A, et al. Genomic profiling of drug sensitivities via induced haploinsufficiency. Nat Genet. 1999;21:278–283.
18. Kubota K, Funabashi M, Ogura Y. Target deconvolution from phenotype-based drug discovery by using chemical proteomics approaches. Biochim Biophys Acta Proteins Proteom. 2019;1867:22–27.
19. Ong S-E, Schenone M, Margolin AA, Li X, Do K, Doud MK, et al. Identifying the proteins to which small-molecule probes and drugs bind in cells. Proc Nat Acad Sci. 2009;106:4617–4622.
20. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteom. 2002;1:376–386.
21. Chan JN, Nislow C, Emili A. Recent advances and method development for drug target identification. Trends Pharmacol Sci. 2010;31:82–88.
22. Jenkins JL, Bender A, Davies JW. In silico target fishing: predicting biological targets from chemical structure. Drug Discov Today Technol. 2006;3:413–421.
23. Ezzat A, Wu M, Li XL, Kwoh CK. Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. Brief Bioinform. 2019;20:1337–1357.
24. Cichonska A, Ravikumar B, Parri E, Timonen S, Pahikkala T, Airola A, et al. Computational-experimental approach to drug-target interaction mapping: a case study on kinase inhibitors. PLoS Comput Biol. 2017;13:e1005678.
25. Zheng S, Dharssi S, Wu M, Li J, Lu Z. Text mining for drug discovery. Method Mol Biol. 2019:231–252.
26. Agamah FE, Mazandu GK, Hassan R, Bope CD, Thomford NE, Ghansah A, Chimusa ER. Computational/in silico methods in drug target and lead prediction. Brief Bioinform. 2020;21:1663–1675.
27. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A. The Harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database. 2016;2016.
28. Gao K, Nguyen DD, Sresht V, Mathiowetz AM, Tu M, Wei GW. Are 2D fingerprints still valuable for drug discovery? Phys Chem Chem Phys. 2020;22:8373–8390.
29. Ibrahim KA, Helmy OM, Kashef MT, Elkhamissy TR, Ramadan MA. Identification of potential drug targets in Helicobacter pylori using in silico subtractive proteomics approaches and their possible inhibition through drug repurposing. Pathogens. 2020;9:747.
30. Drews J. Drug discovery: a historical perspective. Science. 2000;287:1960–1964.
31. Wermuth CG. The Practice of Medicinal Chemistry. Academic Press; 2011.
32. Lindsay MA. Target discovery. Nat Rev Drug Discov. 2003;2:831–838.
33. Na D, Rouf M, O'Kane CJ, Rubinsztein DC, Gsponer J. NeuroGeM, a knowledgebase of genetic modifiers in neurodegenerative diseases. BMC Med Genom. 2013;6:1–14.
34. Myers S, Baker A. Drug discovery—an operating model for a new era. Nat Biotechnol. 2001;19:727–730.
35. Zhu T, Cao S, Su P-C, Patel R, Shah D, Chokshi HB, et al. Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis: Miniperspective. J Med Chem. 2013;56:6560–6572.
36. Kennedy T. Managing the drug discovery/development interface. Drug Discov Today. 1997;2:436–444.
37. Venkatesh S, Lipper RA. Role of the development scientist in compound lead selection and optimization. J Pharm Sci. 2000;89:145–154.
38. Noori HR, Spanagel R. In Silico Pharmacology: Drug Design and Discovery's Gate to the Future. Springer; 2013.
39. Jorgensen WL. The many roles of computation in drug discovery. Science. 2004;303:1813–1818.
40. Na D. User Guides for Biologists to Learn Computational Methods. Springer; 2020.
41. Wadood A, Ahmed N, Shah L, Ahmad A, Hassan H, Shams S. In-silico drug design: an approach which revolutionized the drug discovery process. Drug Des Devel Ther. 2013;1:3.
42. Martin YC, Kofron JL, Traphagen LM. Do structurally similar molecules have similar biological activity? J Med Chem. 2002;45:4350–4358.
43. Prada-Gracia D, Huerta-Yépez S, Moreno-Vargas LM. Aplicación de métodos computacionales para el descubrimiento, diseño y optimización de fármacos contra el cáncer. Bol Méd Hosp Infant Méx. 2016;73:411–423.
44. Bender A, Jenkins JL, Scheiber J, Sukuru SCK, Glick M, Davies JW. How similar are similarity searching methods? A principal component analysis of molecular descriptor space. J Chem Inf Model. 2009;49:108–119.
45. Johnson MA, Maggiora GM. Concepts and Applications of Molecular Similarity. Wiley; 1990.
46. Bologa CG, Revankar CM, Young SM, Edwards BS, Arterburn JB, Kiselyov AS, Parker MA, Tkachenko SE, Savchuck NP, Sklar LA. Virtual and biomolecular screening converge on a selective agonist for GPR30. Nat Chem Biol. 2006;2:207–212.
47. Kubinyi H. Success Stories of Computer-Aided Design. In: Computer Applications in Pharmaceutical Research and Development. Wiley-Interscience; 2006. p. 377–424.
48. Mustata G, Follis AV, Hammoudeh DI, Metallo SJ, Wang H, Prochownik EV, Lazo JS, Bahar I. Discovery of novel Myc–Max heterodimer disruptors with a three-dimensional pharmacophore model. J Med Chem. 2009;52:1247–1250.
49. Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. Nature. 1962;194:178–180.
50. Leo A, Hoekman D. Exploring QSAR. American Chemical Society; 1995.
51. Xie L, Evangelidis T, Xie L, Bourne PE. Drug discovery using chemical systems biology: weak inhibition of multiple kinases may contribute

to the anti-cancer effect of nelfinavir. PLoS Comput Biol. 2011;7:e1002037.

52. Ambure P, Halder AK, Gonzalez Díaz H, Cordeiro MND. QSAR-Co: an open-source software for developing robust multitasking or multitarget classification-based QSAR models. J Chem Inf Model. 2019;59:2538–2544.

53. Tosco P, Balle T. Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. J Mol Model. 2011;17:201–208.

54. Rognan D. Structure-based approaches to target fishing and ligand profiling. Mol Inform. 2010;29:176–187.

55. Kaya S, Tüzün B, Kaya C, Obot IB. Determination of corrosion inhibition effects of amino acids: quantum chemical and molecular dynamic simulation study. J Taiwan Inst Chem Eng. 2016;58:528–535.

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

57. Supuran CT. Advances in structure-based drug discovery of carbonic anhydrase inhibitors. Expet Opin Drug Discov. 2017;12:61–88.

58. Hardy LW, Abraham DJ, Safo MK. Structure-based drug design. Burger Med Chem Drug Discov. 2003:417–469.

59. Craig J, Duncan I, Hockley D, Grief C, Roberts N, Mills J. Antiviral properties of Ro 31-8959, an inhibitor of human immunodeficiency virus (HIV) proteinase. Antivir Res. 1991;16:295–305.

60. Kim E, Baker C, Dwyer M, Murcko Ma, Rao B, Tung R, Navia M. Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. J Am Chem Soc. 1995;117:1181–1182.

61. Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, Costanzo LD, Duarte JM, Dutta S, Feng Z, et al. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. Nucleic Acids Res. 2016;45:D271–D281.

62. Kalyaanamoorthy S, Chen Y-PP. Structure-based drug design to augment hit discovery. Drug Discov Today. 2011;16:831–839.

63. Laurie ATR, Jackson RM. Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. Curr Protein Peptide Sci. 2006;7:395–406.

64. Combs AP. Structure-based drug design of new leads for phosphatase research. Idrugs. 2007;10:112–115.

65. Gohlke H, Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed. 2002;41:2644–2676.

66. Waugh DF. Protein-protein interactions. Adv Protein Chem. 1954;9:325–437.

67. Chen Y, Shoichet BK. Molecular docking and ligand specificity in fragment-based inhibitor discovery. Nat Chem Biol. 2009;5:358–364.

68. Kumar A, Voet A, Zhang K. Fragment based drug design: from experimental to computational approaches. Curr Med Chem. 2012;19:5128–5147.

69. Scott DE, Coyne AG, Hudson SA, Abell C. Fragment-based approaches in drug discovery and chemical biology. Biochem. 2012;51:4990–5003.

70. Jhoti H, Williams G, Rees DC, Murray CW. The 'rule of three' for fragment-based drug discovery: where are we now? Nat Rev Drug Discov. 2013;12:644.

71. Karim M, Islam MN, Jewel GNA. In Silico Identification of Potential Drug Targets by Subtractive Genome Analysis of Enterococcus Faecium DO. BioRxiv. 2020.

72. Hansson T, Oostenbrink C, van Gunsteren W. Molecular dynamics simulations. Curr Opin Struct Biol. 2002;12:190–196.

73. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature. 1977;267:585–590.

74. Grant BJ, Lukman S, Hocker HJ, Sayyah J, Brown JH, McCammon JA, Gorfe AA. Novel allosteric sites on Ras for lead generation. PloS One. 2011;6:e25711.

75. Nair PC, Malde AK, Drinkwater N, Mark AE. Missing fragments: detecting cooperative binding in fragment-based drug design. ACS Med Chem Lett. 2012;3:322–326.

76. Gardiner SJ, Begg EJ. Pharmacogenetics, drug-metabolizing enzymes, and clinical practice. Pharmacol Rev. 2006;58:521–590.

77. Reddy R, Mutyala R, Aparoy P, Reddanna P, Reddy MR. Computer aided drug design approaches to develop cyclooxygenase based novel anti-inflammatory and anti-cancer drugs. Curr Pharm Des. 2007;13:3505–3517.

78. Cordeiro M, Speck-Planche A. Computer-aided drug design, synthesis and evaluation of new anti-cancer drugs. Curr Topics Med Chem. 2012;12:2703.

79. Sham HL, Kempf DJ, Molla A, Marsh KC, Kumar GN, Chen CM, Kati W, Stewart K, Lal R, Hsu A. ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. Antimicrob Agents Chemother. 1998;42:3218–3224.

80. Doyon L, Tremblay S, Bourgon L, Wardrop E, Cordingley MG. Selection and characterization of HIV-1 showing reduced susceptibility to the non-peptidic protease inhibitor tipranavir. Antivir Res. 2005;68:27–35.

81. Njogu PM, Guantai EM, Pavadai E, Chibale K. Computer-aided drug discovery approaches against the tropical infectious diseases malaria, tuberculosis, trypanosomiasis, and leishmaniasis. ACS Infect Dis. 2016;2:8–31.

82. Xiong M, Guo Z, Han B, Chen M. Combating multidrug resistance in bacterial infection by targeting functional proteome with natural products. Nat Prod Res. 2015;29:1624–1629.

83. Krohn A, Redshaw S, Ritchie JC, Graves BJ, Hatada MH. Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)- hydroxyethylamine isostere. J Med Chem. 1991;34:3340–3342.

84. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A. The Harmonizome: a Collection of Processed Datasets Gathered to Serve and Mine Knowledge about Genes and Proteins. Database. 2016;2016.

85. Wang L, Ma C, Wipf P, Liu H, Su W, Xie X-Q. TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. AAPS J. 2013;15:395–406.

86. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, Jensen NH, Kuijer MB, Matos RC, Tran TB. Predicting new molecular targets for known drugs. Nature. 2009;462:175–181.

87. Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V. SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res. 2014;42:W32–W38.

88. Dunkel M, Günther S, Ahmed J, Wittig B, Preissner R. SuperPred: drug classification and target prediction. Nucleic Acids Res. 2008;36:W55–W59.

89. Alberga D, Trisciuzzi D, Montaruli M, Leonetti F, Mangiatordi GF, Nicolotti O. A new approach for drug target and bioactivity prediction: the multifingerprint similarity search algorithm (MuSSeL). J Chem Inf Model. 2018;59:586–596.

90. Pinero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48:D845–D855.

91. Peon A, Li H, Ghislat G, Leung KS, Wong MH, Lu G, Ballester PJ. MolTarPred, A web tool for comprehensive target prediction with reliability estimation. Chem Biol Drug Des. 2019;94:1390–1401.

92. Liu X, Vogt I, Haque T, Campillos M. HitPick: a web server for hit identification and target prediction of chemical screenings. Bioinformatics. 2013;29:1910–1912.

93. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res. 2008;36:D901–D906.

94. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B. PubChem 2019 update: improved access to chemical data. Nucleic Acids Res. 2019;47:D1102–D1109.

95. Seiler KP, George GA, Happ MP, Bodycombe NE, Carrinski HA, Norton S, Brudz S, Sullivan JP, Muhlich J, Serrano M. ChemBank: a small-molecule screening and cheminformatics resource database. Nucleic Acids Res. 2007;36:D351–D359.

96. Wang Y, Zhang S, Li F, Zhou Y, Zhang Y, Wang Z, Zhang R, Zhu J, Ren Y, Tan Y. Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. Nucleic Acids Res. 2020;48:D1031–D1041.

97. Hecker N, Ahmed J, von Eichborn J, Dunkel M, Macha K, Eckert A, Gilson MK, Bourne PE, Preissner R. SuperTarget goes quantitative: update on drug–target interactions. Nucleic Acids Res. 2012;40:D1113–D1117.

98. Pence HE, Williams A. ChemSpider: an Online Chemical Information Resource. ACS Publications; 2010.

99. Wishart D, Arndt D, Pon A, Sajed T, Guo AC, Djoumbou Y, Knox C, Wilson M, Liang Y, Grant J. T3DB: the toxic exposome database. Nucleic Acids Res. 2015;43:D928–D934.

100. Von Eichborn J, Murgueitio MS, Dunkel M, Koerner S, Bourne PE, Preissner R. PROMISCUOUS: a database for network-based drug-repositioning. Nucleic Acids Res. 2010;39:D1060–D1066.