Bulletin of Pharmaceutical Research 2025;15(3):202
An Official Publication of Association of Pharmacy Professionals
ISSN: 2249-6041 (Print); ISSN: 2249-9245 (Online)
DOI: 10.21276/bpr.2025.15.3.2

REVIEW ARTICLE

REVOLUTIONIZING DRUG DISCOVERY: COMPUTATIONAL STRATEGIES FOR MODERN THERAPEUTICS

Shifa Khan[1], Shrey Dahiya[2], Priyanshu Nema[1] and Sushil Kumar Kashaw[1]*

[1]*Integrated Drug Discovery Research Laboratory, Department of Pharmaceutical Sciences, Dr. Harisingh Gour University (A Central University), Sagar-470003, Madhya Pradesh, India*
[2]*Graduate School of Technology, Touro University, New York 10036, USA*

**E-mail*: sushilkashaw@gmail.com
Tel.: +91 9425655720.

Received: Jan 11, 2025 / Revised: Dec 30, 2025 / Accepted: Dec 31, 2025

Drug discovery is a complex and resource-intensive process requiring substantial time and financial investment, often taking 10-15 years and costing billions of dollars to bring a single drug to market. Despite these efforts, success rates in clinical trials remain low, with a significant number of failures attributed to suboptimal pharmacokinetic and toxicity profiles. Computer-aided drug design (CADD) has emerged as a transformative tool in accelerating the drug discovery process by minimizing costs and failures during the later stages of development. CADD encompasses a range of computational approaches, including structure-based drug design (SBDD) and ligand-based drug design (LBDD), which leverage the structural and chemical information of therapeutic targets and ligands to identify and optimize lead molecules. These techniques, combined with advancements in artificial intelligence (AI) and machine learning (ML), have revolutionized the identification of drug candidates by enabling high-throughput virtual screening, molecular docking, quantitative structure-activity relationships (QSAR), and pharmacophore modeling. The availability of three-dimensional structures of proteins through X-ray crystallography, NMR, and computational methods like homology modeling and ab initio modeling has further facilitated target preparation for SBDD. Simultaneously, LBDD capitalizes known ligand interactions to design new molecules with enhanced activity and improved pharmacokinetic properties. This review highlights the principles, methodologies, and recent advancements in CADD, emphasizing its critical role in identifying novel drug candidates. By integrating traditional computational methods with AI and ML, CADD continues to redefine the paradigm of modern drug design, paving the way for innovative, efficient, and targeted therapeutic solutions.

Key words: Drug discovery, Computer-aided drug design, SBDD, Artificial intelligence, Machine learning.

INTRODUCTION

Drug discovery is a lengthy process that takes around 10-15 years [1] and costs up to 2.558 billion USD for a drug to reach the market [2]. It is a multistep process that begins with the identification of suitable drug targets, validation of drug targets, hit to lead discovery, optimization of lead molecules, and preclinical and clinical studies [3]. Despite the high investments and time incurred for the discovery of new drugs, the success rate through clinical trials is only 13% with a relatively high drug attrition rate [4]. In the majority of cases (40-60%), the drug failure at a later stage has been reported due to lack of optimum pharmacokinetic properties on absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox) [5]. The use of computer-aided

drug discovery (CADD) techniques in preliminary studies by leading pharmaceutical companies and research groups has helped to expedite the drug discovery and development process by minimizing the costs and failures in the final stage [6]. The application of rational drug design as an integral part of CADD provides useful insights into the understanding of the binding affinity and molecular interaction between target protein and ligand. Additionally, lead identification in pharmaceutical research has been facilitated by the availability of supercomputing facility, parallel processing, and advanced programs, algorithms, and tools [7]. Furthermore, recent advancements in artificial intelligence (AI) and machine learning methods have greatly aided in analyzing, learning, and explaining the pharmaceutical-related big data in the drug discovery process [8]. Different methods employed in the identification of new inhibitors from chemical databases include pharmacophore modeling, quantitative structure activity relationship (QSAR), molecular docking, quantum mechanics, and statistical learning methods. CADD can be broadly divided into structure-based and ligand-based drug design approaches; both have been widely used in the drug discovery process in the identification of suitable lead molecules. While the structure-based drug design relies on the three-dimensional structure of the target receptor and its active sites to understand the molecular interaction between the receptor and ligand, the ligand based-drug design depends on the knowledge of ligands interacting with the given target receptor [9]. Computer-aided drug design has a large number of successes and continues to play a vital role in the drug discovery process [10]. In this regard, the approach has been utilized in proposing drug candidates against coronavirus disease 2019 (COVID-19). COVID-19 is caused by a novel coronavirus known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which taxonomically belongs to the Beta corona virus genre and possesses high nucleotide sequence similarity with severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS CoV). The epidemiology, genome composition, pathogenesis, animal models, diagnostics, and vaccine development with references to various computational biology approaches for MERS-CoV infections have been comprehensively reviewed [11]. SARS-CoV-2 is a positive-sense single-stranded enveloped RNA virus approximately 30,000bp in length which utilizes host cellular machinery to execute various pathogenic processes such as viral entry, genomic replication, and protein synthesis [12]. Like SARS and MERS, the genome of SARS-CoV-2 encodes sixteen nonstructural proteins (nsps) such as main protease (Mpro), papain-like protease, RNA-dependent RNA polymerase (RdRp), helicase etc., four structural proteins (envelope, membrane, spike, and nucleocapsid), and other accessory proteins. While the spike glycoprotein is essential for the interaction of the virus with the host cell receptor, the nsps play a major role during the virus life cycle by engaging in the production of sub genomic RNAs [13, 14]. The nonstructural and structural proteins, therefore, offer promising targets for the design and development of antiviral agents against COVID-19 [13]. The lack of effective vaccines or drugs for the treatment of COVID-19 and the high mortality rate necessitates the rapid discovery of novel drugs [15], and computer-aided drug design is believed to be an important tool to achieve the identification of novel therapeutics. There is a possibility of the development of effective lead molecules against COVID-19 by utilizing natural lead molecules obtained through virtual screening and pharmacokinetic prediction [16]. To speed up the discovery of a potential treatment for SARS-CoV-2 infection in humans, repurposing of broad-spectrum antiviral drugs is a promising strategy due to the availability of the pharmacokinetic and pharmacodynamic data of these drugs [17]. The availability of complete genome sequence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the elucidation of the viral protein structures through X-ray crystallography, nuclear magnetic resonance (NMR), electron microscopy, and homology modelling approach have allowed the identification of inhibitor drugs against the essential therapeutic drug targets of COVID-19. This review article provides useful insights into some common *in silico* methods used in CADD and how these methods have been currently used and can be of help in the drug discovery process of COVID-19.

Drug design

A drug is a foreign molecule that affects biological processes and is used to prevent, diagnose or treat disease [18]. Drugs can be of natural origin or produced synthetically. The

ideal drug should have a specific action, be safe, non-toxic, without side effects or as few as possible, be chemically and metabolically stable, be synthetically feasible, be soluble in water in therapeutic concentrations in order to avoid precipitation in the blood stream, be soluble in lipids as well, in order to be able to cross the lipid membranes and distribute around the body, and finally, be a unique molecule [19,20]. To exert their effects, drugs interact with specific targets in the human body. As a result of these interactions, two types of effects are produced: effects of the drug on the human body, and effects of the human body on the drug. These effects are considered by Pharmacodynamics and Pharmacokinetics, respectively. [21]. Pharmacodynamics focuses on the mechanisms of drug action, the relationships between drug concentration and effect, and adverse reactions. Pharmacokinetics studies the absorption, distribution, metabolism and excretion of the drug over time, the so-called ADME processes or ADME properties of drugs. The process of drug discovery and development consists of three main stages: drug discovery, pre-clinical development and clinical trials. The drug discovery starts with the finding of a hit molecule. A hit is a molecule that elicits a desired activity in a screening assay [22,23]. Then, the structure of this molecule is optimized in terms of improving affinity and selectivity, reducing toxicity, improving water and lipid solubility, improving ADME properties in general and converting the hit molecule into a lead molecule. The further optimization of the lead molecule delivers the drug candidate. Next, preclinical studies are focused on clarifying the mode of action of the drug candidate, its pharmaco-kinetics in animals such as bioavailability, toxic metabolites, if any, routes of excretion, efficacy on animals, drug formulation and stability tests of this formulation.

The clinical trials are the longest and the most expensive stage of the process, consisting of three phases. In the first phase, up to 100 healthy volunteers are involved. The aim of this phase is to evaluate the safety of the drug on humans, its pharmacokinetics in the human body and immediate side effects if there are any. In the second phase, the drug is administered to several hundred patients suffering from the target disease.

At second phase, the efficacy of the drug and its short-term safety are tested. In the third phase, several thousand patients from several clinical centres around the world are involved. The aim of this phase is to collect sufficient data for the efficacy and safety of the drug. If the drug passes this phase successfully, it is ready for registration and marketing. However, the surveillance of the drug does not stop here. The drug continues to be observed for safety and side effects. This last phase is known as a post-marketing surveillance and it is practically endless, it continues until the drug is on the market (**Figure 1**).

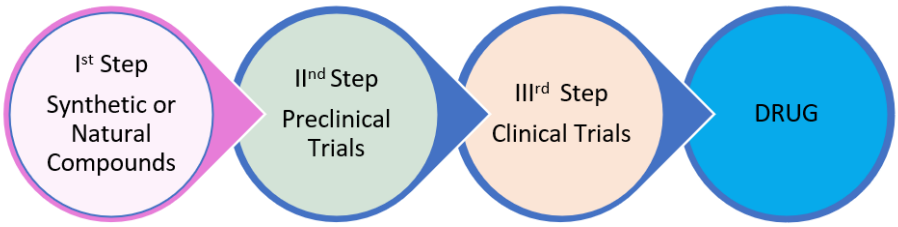

Fig. 1. Traditional methods of drug discovery

History of drug design

There were several major achievements in the science of drug design that made it the main approach in the current and the future drug discovery [19]. The first of them is the understanding of drug–receptor recognition. In the early 1890s, Emil Fisher compared the drug–receptor interaction to the key and lock interplay. He considered that both the drug and the receptor interact as solid bodies without changing their conformations. Lately, Daniel Koshland suggested that both molecules do undergo conformational changes during interaction and adopt the most suitable conformation to connect each other. This hypothesis has been proven many times by X-ray structures and *in silico* simulations and now it is known that ligands indeed change their conformations during interaction and adopt conformations that optimally fit the contact surfaces. The target macromolecule is an internal molecule which is involved in disease. Affecting the functions of target macromolecule could change the etiology of the disease, the

pathophysiology or simply improve the symptoms. Of the 20,000 protein-coding genes in the human genome, about 3000 have been estimated to be part of the so-called druggable genome or druggable proteins [24]. These are proteins that have the ability to bind drug-like molecules. Oprea et al. [25] divided the whole proteome into four groups according to the target development level, i.e., how much we know about a given protein (**Figure 2**).

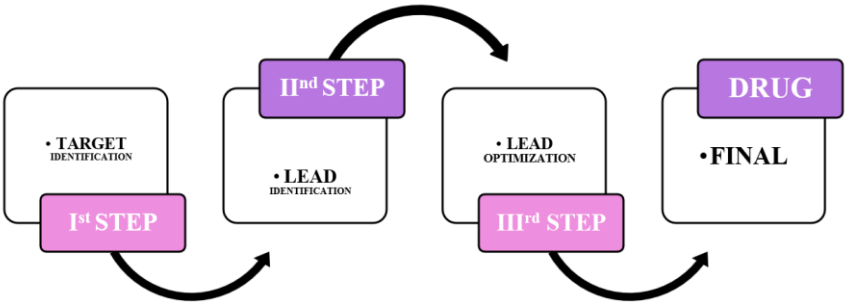

Fig. 2. Modern drug discovery

Clinically known targets include macromolecules linked to at least one approved drug. Their number is 659, i.e. 3% of the human proteome. Of them, 25% are enzymes, 21% are ion channels, 16% are gamma-protein-coupled receptors, 9% are different types of kinases, 4% are transporter proteins, 3% are nuclear receptors and the remaining 22% include other protein families and orphan receptors. Chemically known targets include proteins known to bind with high potency small molecules that are not yet drugs. They constitute 6% of the human proteome. Biologically known targets refer to proteins that have a link to any disease but have not been studied for binding to small molecules. More than half of the human proteome (53%) belongs here. The "dark" targets include the unstudied proteins. They are 38% of the whole proteome. Obviously, there is a huge field for future discoveries. Since 2014, there has been an initiative funded by the NIH aiming to illuminate the druggable genome. All new data discovered are collected on the Pharos website (pharos.nih.gov, accessed on 20 February 2022) [26]. The 3D structures of proteins are resolved by X-ray crystallography, NMR spectroscopy and more recently by the modern cryogenic electron microscopy and collected in the Protein Data Bank (PDB). PDB currently contains more than 180 thousand structures [27]. Most of them are single proteins, in apo-form, without a ligand. However, some of them are in complexes with ligands and this information is more valuable. It is located on protein-binding site. The major achievement in the modern drug design is the development of *in silico* modelling technologies, applied for virtual screening, compound design, energy calculations, SAR and QSAR analysis, ADME modelling, modelling of drug–target interactions. In order to apply all these advanced technologies, the molecular structure should be encoded numerically [28]. The encoded structures can be analyzed, searched, visualized and compared with other structures. The structures could be encoded by binary strings, smiles strings, 2D graphs and 3D structures. The 3D molecular modelling and visualization also are considered among the greatest achievements in drug design technologies. The molecular properties also should be encoded. Different types of descriptors have been developed over the years. The 1D descriptors are derived from the 1D structure such as element composition and molecular weight. The 2D descriptors are generated from the molecular graph such as the number of hydrogen bond acceptors and donors, the partition coefficients log P and log D, water solubility, etc. The 3D descriptors are extracted from the 3Dstructures. These are molecular surface area or polar surface area, molecular volume etc. Carracedo-Reboredo et al. define 4D descriptors, providing information about the interactions between ligands and protein-binding sites [29]. Apart from single numbers, descriptors could be composed of several numbers. These are the descriptor sets which 5 of 10 describe a complex property or a set of properties. Once there is a quantitative description of a set of structures and a quantitative measurement of their activities such as IC_{50}, EC_{50}, K_d, etc, then different types of quantitative analyses can be applied. This was consucted for the first time in 1964 when correlating the antimicrobial activity of penicillin derivatives with descriptors relating to the

hydrophobic and electronic properties of the molecules [30]. The correlations between molecular descriptors of a series of compounds and their activities are assigned as Quantitative Structure Activity Relationships (QSAR). Corwin Hansch (1918–2011) is considered the father of drug design [31]. The major achievement in the modern drug design is the development of in silico modelling technologies, applied for virtual screening, compound design, energy calculations, SAR and QSAR analysis, ADME modelling, modelling of drug–target interactions. In order to be applied all these advanced technologies, the molecular structure should be encoded numerically [28]. The encoded structures can be analyzed, searched, visualized and compared with other structures. The structures could be encoded by binary strings, smiles strings, 2D graphs and 3D structures. The 3D molecular modelling and visualization also are considered among the greatest achievements in drug design technologies. Lately, in the analyses, methods such as principal component analysis [32] and partial least squares [33] have been involved. Chemometrics evolved into chemo informatics including molecular modelling, chemical information and tools and algorithms to handle the enormous amount of data coming from combinatorial chemistry and HTS [34]. At the same time, bioinformatics also emerged to organize and analyze biological data, the amount of which increased dramatically, especially after deciphering the human genome. The molecular properties also should be encoded. Different types of descriptors have been developed over the years. The 1D descriptors are derived from the 1D structure such as element composition and molecular weight. The 2D descriptors are generated from the molecular graph such as the number of hydrogen bond acceptors and donors, the partition coefficients log P and log D, water solubility, etc. The 3D descriptors are extracted from the 3D structures. These are molecular surface area or polar surface area, molecular volume, etc. Carracedo-Reboredo et al. define 4D descriptors, providing information about the interactions between ligands and protein-binding sites [29]. Apart from single numbers, descriptors could be composed of several numbers. These are the descriptor sets which describe a complex property or a set of properties. Nowadays, the structure–activity relationships are analyzed by machine learning methods such as random forest, decision trees, neural networks, k nearest neighbors, etc. [35,36]. Many authors classify these methods as "black box" methods because they are not able to distinguish immediately between relevant and irrelevant for the activity descriptors as the regression methods do [36]. However, they are very good in predictions and classifications, significantly better than the statistical analyses and regression models. Both types of methods do not replace each other, nor contradict. They are complimentary and most effective when are used in combination.

Computer-aided drug design

Computer-aided drug design (CADD) is one of the pivotal approaches to contemporary pre-clinical drug discovery, and various computational techniques and software programs are typically used in combination, in a bid to achieve the desired outcome. Several approved drugs have been developed with the aid of CADD [37]. Computer-aided drug design (CADD) can be defined as computational approaches that are used to discover, develop, and analyze drug and active molecules with similar biochemical properties [9,41] Homology modelling, molecular docking, virtual screening (VS) or virtual high-throughput screening (HTVS), quantitative structure activity relationship (QSAR) and three-dimensional (3D) pharmacophore mapping generally, are the main constituents of CADD [38-41]. Among these techniques, it seems that virtual screening is the major contributor to CADD and it has become somewhat a proven and well-appreciated computational method, that stands as a contemporary to the experimental high-throughput screening for hit identification and optimization [42-44].

CADD can be broadly divided into structure-based and ligand-based drug design approaches, both have been widely used in the drug discovery process in the identification of suitable lead molecules. While the structure-based drug design relies on the three-dimensional structure of the target receptor and its active sites to understand the molecular interaction between the receptor and ligand, the ligand based-drug design depends on the knowledge of ligands interacting with the given target receptor [9]. CADD utilizes two different techniques based on the availability of either 3D structures of a protein or ligands. They are known as structure-based drug design (SBDD) and ligand-based drug design (LBDD). In some cases, the integration of

both techniques has shown good accuracy in finding the lead molecules [45] (**Figures 3**, **4**).

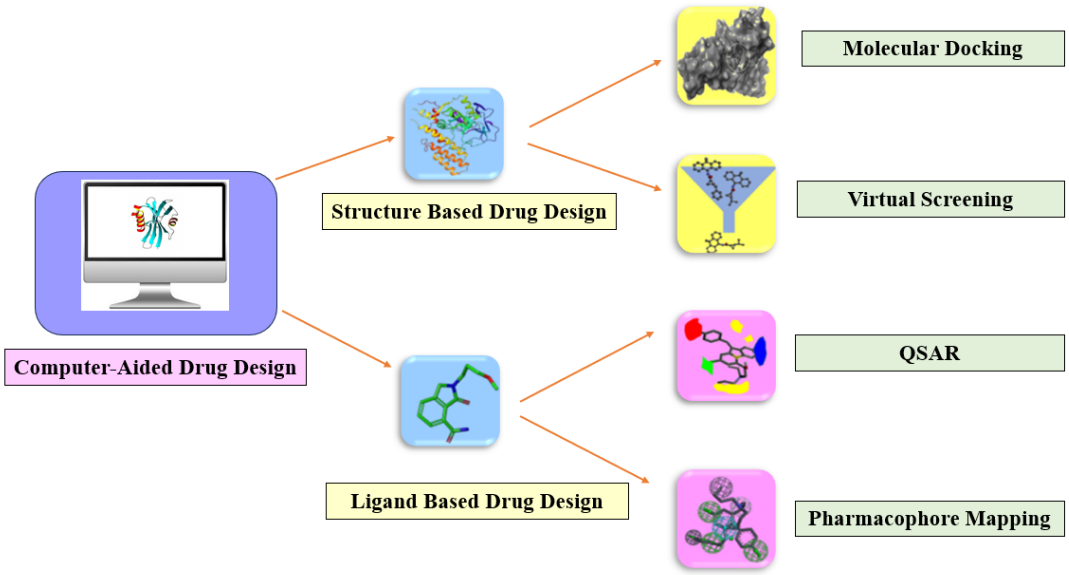

Fig. 3. Overview of CADD

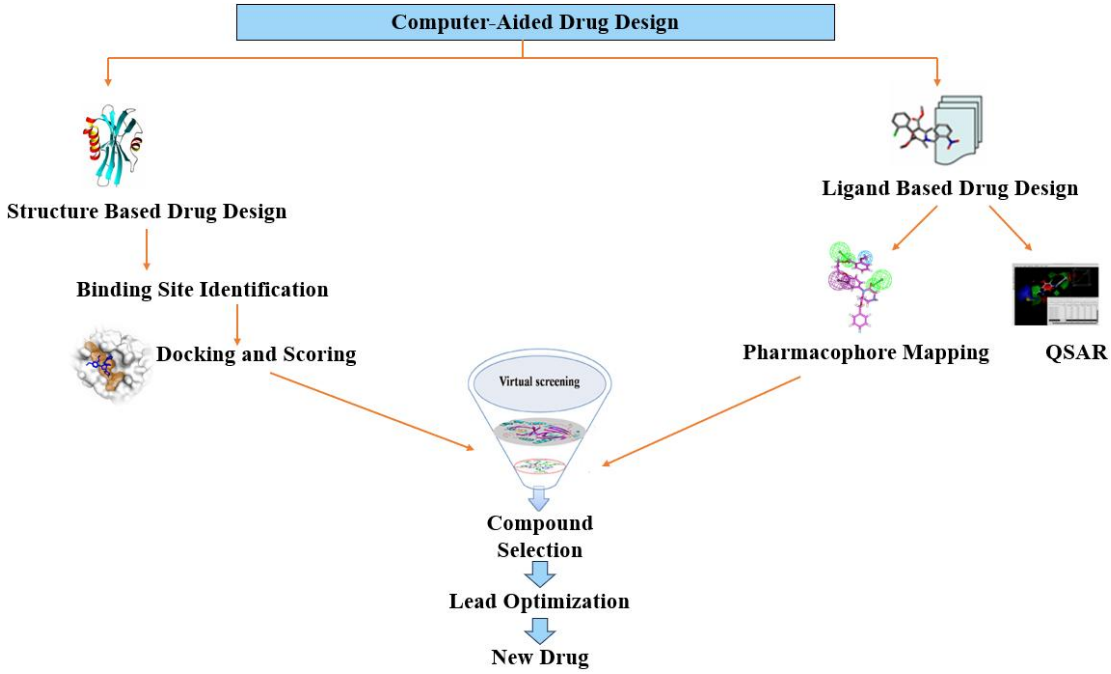

Fig. 4. Discovery of a new drug using CADD

Structure-Based Drug Design (SBDD)

The underlying principles of structure- based drug design are the accessibility of the therapeutic target protein's three- dimensional structures and the characterization of the binding site cavity [46]. A new era of SBDD in drug discovery and design has begun by disclosing many biological molecules three-dimensional (3D) structures [47-49]. SBDD has emerged as a possible means of generating and optimizing ligands in the pharmaceutical industry [50-52]. Preparation of the target, identification of the binding site, molecular docking, virtual screening and molecular dynamics are the basic steps of SBDD (**Figure 5**).

I. Target preparation

Preparing the target macromolecule structure is the most crucial step in SBDD. Due to the rapid advancement of X-ray and NMR structure

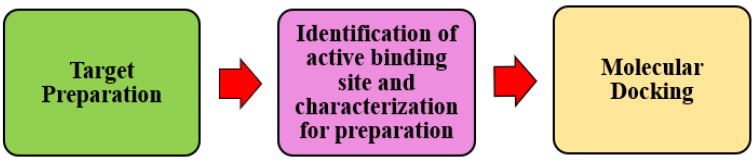

Fig. 5. Overview of the steps involved in SBDD

elucidation techniques, 3D structures of proteins deposited in the Protein Data Bank (PDB) are readily accessible [53]. When the target proteins' 3D structures are unavailable, computational methods like comparative or homology modelling [54], threading [55], and ab initio modeling [56] have been successful in determining the structures of proteins from their sequences.

Homology modeling or Comparative modeling

Protein 3D structure can be deduced from its amino acid sequence using various computational structure prediction techniques, including homology modeling (**Table 1**). It is regarded as the computational structure prediction method with the highest level of accuracy. There are several simple and easy-to-follow steps in it [57]. It involves finding a structural template protein with a similar sequence, aligning their sequences, using aligned region coordinates, predicting missing atom coordinates of the target, model building, and refinement. The NCBI Basic Local Alignment Search Tool (BLAST) is one of the most widely used bioinformatics sequence alignment tools for sequence similarity searches.

Fold recognition or threading methods

These methods are used to find proteins with comparable folds but no sequence similarity [58,59]. The structure is considered by proteins (pairwise interaction). The sequence of a known protein structure is replaced by the query sequence of the target of interest, for which the structure is unknown. The resulting "threaded" structure is then evaluated using various scoring systems [60,61]. This process is repeated for each database's empirically determined 3D structures, providing the structure that best matches the query sequence (**Table 2**) [58]. It is employed in SBDD research.

Ab initio or de novo modeling

Ab initio or *de novo* modeling is performed when there is not enough homogeneity in the structure to perform comparison modeling (**Table 3**). If the target protein does not have any template structures in current biological databases, ab initio modeling is a computational technique that should be applied. A prominent *de novo* structure prediction technique is ab initio structure prediction [62], which is implemented in Rosetta. Rosetta protein structure prediction methods are effective, according to studies from the Critical Assessment of Protein Structure Prediction (CASP) study. In recent CASP experiments, QUARK, the Zhang group's ab initio structure prediction server, has also shown encouraging results [63].

II. Identification of active binding site and characterization for preparation

Drug activity necessitates the interaction of protein and ligand. It's only possible if high-affinity binding sites can be found. The development of novel approaches in a structure-based drug discovery method highly depends on identifying druggable cavities or pockets on a target protein.

Table 1. Homology modeling popular structure prediction tools and their methods of prediction [64-71]

S. No.	Tools	Method	Link
1	MODELLER	Satisfaction of spatial restraints	https://salilab.org/modeller
2	Swiss model	Fragment-based assembly and local similarity	https://swissmodel.ex pasy.org
3	Phyre & Phyre2	Remote template detection, alignment, 3D modeling, multi-templates, *ab initio*	http://www.sbg.bio. ic.ac.uk/phyre2/html/page.cgi? id=index
4	3D-JIGSAW	Fragment-based assembly	http://bmm.crick.ac.uk
5	HHpred	Pairwise comparison of profile HMMs	https://toolkit. tuebingen.mpg. de/tools/hhpred

Table 2. Tools for threading methods [72-74,61]

S. No.	Tools	Method	Link
1	MUSTER	Profile-profile alignment with multiple structural information	https://zhanggroup.org/MUSTER/
2	GenTHREADER	Sequence alignment, threading evaluation by neural networks	http://www.ebisu.co.uk/chemogenomix.com/chemogenomix/GenThreader.html https://zhanggroup.org/I-TASSER/
3	I-TASSER	Iterative template fragment assembly	http://202.112.170.199/DescFold/index.html
4	DescFold	SVM-based machine learning algorithms in protein fold recognition	

Table 3. Tools for *ab initio* modeling

S. No.	Tools	Method	Link
1	QUARK	Replica-exchange MC and optimized knowledge-based force field	https://zhanggroup.org/C-QUARK/
2	Rosetta/Robetta	Fragment assembly, simulated annealing	https://www.hsls.pitt.edu/obrc/index.php?page=URL1098813924
3	I-TASSER	Fragment assembly	https://zhanggroup.org/I-TASSER/
4	CABS- FOLD	User provided distance restraints from sparse experimental data	http://biocomp.chem.uw.edu.pl/CABSfold/
5	EVfold	Calculate evolutionary variation by co-evolved residue pairs	http://cbio.mskcc.org/foldingproteins/

The term "binding sites" (BSs) refers to protein surface cavities that can vary significantly in size and shape, whether they contain a ligand or not [75,76]. Tools like POCKET, SURFNET, Q-SITE FINDER, DoG Site Scorer server [77], CASTp [78]. NSiteMatch [79], meta pocket [80], DEPTH [81], LISE [82] and MSpocket [83] are *in silico* tools used to predict binding sites of a target protein. After the binding site is found, tools or servers such as Epock [84], TRAnsient Pockets in Proteins (TRAPP) [85], and POVME [86] are used to determine binding pocket volume (**Table 4**).

Table 4. Binding site prediction tools

S. No.	Tools	Method	Link
1	FPOCKET	Open-source protein pocket (cavity) detection algorithm based on Voronoi tessellation	https://bioserv.rpbs.univ-paris-diderot.fr/services/fpocket/
2	SURFNET	Calculation of clefts in protein surfaces	https://www.ebi.ac.uk/thornton-srv/software/SURFNET/
3	Q-SITEFINDER	Web server for identifying protein-binding sites	https://swmath.org/software/35551
4	DoGSite Scorer server	Binding site prediction, druggability assessment	https://bio.tools/dogsitescorer
5	CASTp	Binding sites and active sites of proteins and DNAs	http://sts.bioe.uic.edu/castp/index.html?1ycs

III. Molecular docking

Molecular docking is a technique for determining the conformation and orientation (together referred to as "position") of ligand molecules in the binding site of a macromolecular target. Search algorithms are used to produce poses and are then ranked using scoring techniques. Many biological processes, such as signal transmission, cell control, and other macromolecular assemblies, rely on molecular recognitions such

as enzyme-substrate, drug-protein, drug-nucleic acid, protein-nucleic acid, and protein- protein interactions. Sampling and scoring are two crucial components of a protein- ligand docking method. Ligand sampling and protein flexibility are two facets of sampling, which refer to creating possible ligand binding orientations/conformations near a protein's binding site. Scoring predicts binding tightness for individual ligand orientations/conformations using physical or empirical functions [87-89] (**Table 5**).

Table 5. List of docking tools/software

Software/ Tools	Year	Company name	Search algorithm	Free/Commercial	Link
AutoDock	1990	The Scripps Research Institute	Lamarkian Genetic Algorithm	Open source (GNU GPL)	https://autodock.scripps.edu/
AutoDock Vina	2010	The Scripps Research Institute	Genetic Algorithm	Open source (Apache License)	https://vina.scripps.edu/
BetaDock	2011	Hanyang University	Based on Voroni Diagram	Freeware	http://voronoi.hanyang.ac.kr/software.htm
DOCK	1988	University of California-San Francisco	Shape Matching	Freeware for academic use	https://dock.compbio.ucsf.edu/
FlexX	2001	BioSolveIT	Incremental Construction	Commercial	https://www.biosolveit.de/download/?product=flexx

Ligand-Based Drug Design (LBDD)

In the absence of 3D information about the receptor, ligand-based drug design is used. The technique relies on knowledge of molecules that bind to the biological target of interest [90]. LBDD methods use a known antibiotic (ligands) as a target to establish a structure-activity relationship (SAR) between their physio-chemical properties and antibiotic activities, which can improve existing drugs or guide the development of new drugs with enhanced activity [91]. QSAR and pharmacophore are two methods of LBDD. The chemical fingerprint of known ligands that bind to a target is utilized in molecular similarity approaches to identify compounds with similar fingerprints using molecular libraries for screening (**Table 6**). Ligand similarity search methods are effective because structurally related compounds have comparable binding properties.

Table 6. Small molecule databases

Tool name	Description	Availability	Url
DrugBank	It is a comprehensive, freely accessible, online database containing information on drug and drug targets	Free	http://www.drugbank.ca/downloads
PubChem	PubChem is a database of chemical molecules and their activities against biological assay	Free	https://pubchem.ncbi.nlm.nih.gov/
BindingDB	BindingDB is a public, web- accessible database of measured binding affinities, focusing chiefly on the interactions of protein considered to be drug- targets with small, drug-like molecules	Free	http://www.bindingdb.org/
BindingMOAD	Binding MOAD (Mother of All Databases) is a database of 9836 protein-ligand crystal structures	Free	https://bindingmoad.org/

I. Quantitative structural activity relationship

Quantitative structural activity relationship (QSAR) is a computerized statistical tool for explaining the observed variation in structure changes produced by replacement (**Table 7**). These models mathematically demonstrate how the structural properties of a ligand affect the activity response of a target that binds it. Molecular parameters that can be used for building QSAR models may include electronic, hydrophobic, steric, and sub-structural effects. In detail parameters are discussed below.

Electronic effects: Ionization constants, Sigma substituent constant, distribution constant, resonance effect, field effect, molecular orbital indices, atomic/electron net charge, nucleophilic super delocalizability, electrophilic super delocalizability, free radical super delocalizability, energy of the lowest empty molecular orbital and highest occupied molecular orbital, frontier atom polarizability, intermolecular coulombic interaction energy, electric field created at point (A) by a set of charges on a molecule. Hydrophobic parameters: Partition coefficients, Pi substituent constants, Rm value in liquid–liquid chromatography, Elution time in high-pressure liquid chromatography (HPLC), Solubility, Solvent partition coefficients. Steric effects: Intramolecular steric effects, Steric substituent constants, Hyperconjugation correction, Molar volume, Molar refractivity, MR substituent constants, Molecular weight, Van der Waals radii Interatomic distances. Substructural effects: Three-dimensional geometry Fragment and molecular properties. Steps involved in QSAR (**Figure 6**) [92,93] included i) Preparing molecules for QSAR experiment: Obtain a congeneric set of ligands tested in a similar biological assay and shown a wide range of action. ii) Selection for descriptors in training set: Identify and determine the molecular descriptors related to the compounds' physiochemical properties. iii) Calculate values for descriptors in training set: Split the molecules randomly into two groups: training and test sets. Using the training set, identify and calculate the correlation coefficient that can explain the relationship between descriptor values and biological activity. iv) Evaluation of internal and external validation: Using the test set molecules, assessing the statistical equation's stability. To anticipate the biological activity of a novel chemical, use a statistical model.

Table 7. Tools for QSAR

Tools	Description	Availability	Url
OECD QSAR Toolbox	The OECD QSAR Toolbox is free software for screening and assessing chemical substances that use computational methods as an alternative to animal testing	Free	https://qsartoolbox.org
CORAL	QSPR/QSAR analysis for substances represented by Simplified Molecular Input-Line Entry System (SMILES) by the Monte Carlo method.	Free	http://www.insilico.eu/coral
PharmQSAR	PharmQSAR is a 3D (QSAR) software package that builds statistical models (CoMFA, CoMSIA and HyPhar) based on data obtained from experimental assays	Commercial	https://pharmacelera.com/pharmqsar
GUSAR	GUSAR software was developed to create QSAR/QSPR models on the basis of the appropriate training sets represented as SD file containing data about chemical structures and endpoint in quantitative terms.	Free	http://www.way2drug.com/gusar/index.html

II. Pharmacophore modeling

According to the IUPAC definition, pharmacophore is "an ensemble of steric and electronic properties required to achieve optimal supra molecular interactions with a given biological target and to trigger or prevent its biological response" [94]. A pharmacophore is an abstract description of the structural properties required for the biological macro-molecules to recognize a ligand (**Table 8, Figure 7**) [95].

Choose a training set of ligands

For the pharmacophore model development, a structurally diverse group of compounds is chosen. List of molecules should include active and inactive compounds, as models must be able to clarify molecules with/without bioactivity.

Analysis of conformation

A list of low-energy conformations for each selected compound is created that will include the bioactive conformation.

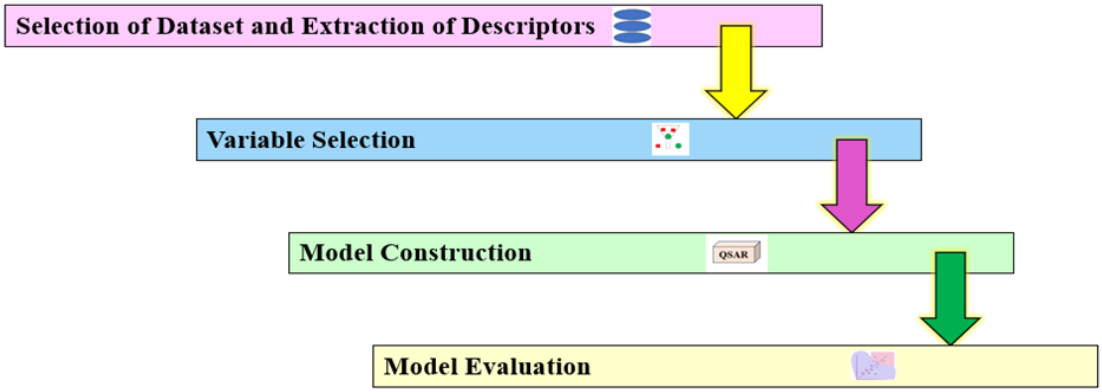

Fig. 6. Quantitative structure-activity relationship workflow.

Table 8. Tools for pharmacophore modeling

Tools	Description	Availability
MolSign	MolSign is a complete module for pharmacophore identification and modeling	Commercial
LigandScout	LigandScout is computer software that allows creating (3D) pharmacophore models from structural data of macromolecule-ligand complexes, or from training and test sets of organic molecules	Commercial
CATSlight2	CATSlight2 is available as a web-server for molecular similarity	Free

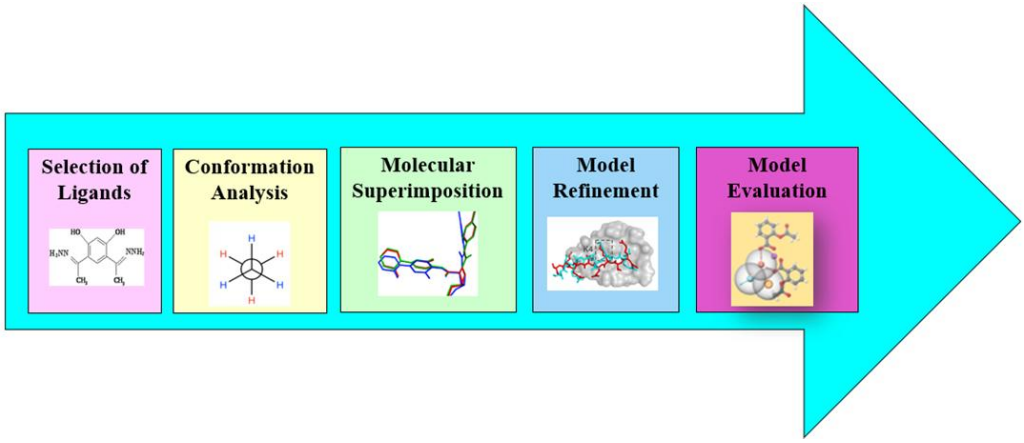

Fig. 7. Steps involved in pharmacophore modeling [94]

Molecular superimposition
All possible combinations of the molecules' low-energy conformations were superimposed. Functional groups that are similar in all the molecules in the collection could be fitted (e.g. phenyl rings). The active conformation is assumed to be the set of conformations that results in the best fit.

Abstraction
Create an abstract representation of the overlaid molecules. Superimposed phenyl rings, for example, could be referred to as an 'aromatic ring' pharmacophore element in a more conceptual sense. Similarly, hydroxyl groups could be labelled as a pharmacophore element that acts as a 'hydrogen-bond donor/ acceptor'.

Validation
A pharmacophore model is a hypothesis that explains the pharmacological actions of a group of compounds that bind to -736 same biotarget.

Applications of computational drug design

Several CADD studies have been reported in the past years. Here, we briefly illustrate selected studies that apply virtual screening tools in drug discovery. In 2014, a study [96] highlighted the structure based rational drug design against Tip60 histone acetyl transferase, an attractive target for cancer drug discovery. In their study, the 3D structure of the human Tip60 acetyl transferase domain was retrieved from PDB for modeling experiments. However, since several key residues were missing in the structure, homology modeling was employed using the human Tip60 sequence for the target and the incomplete human Tip60 crystal structure as the template. They used Alpha Site Finder to find the binding site into which pentamidine (PNT, known inhibitor) and acetyl-CoA (natural substrate) were later docked. Subsequently, PNT derivatives were computationally generated and docked to the Tip60 model to determine their binding affinities. MD simulations were also implemented to account for explicit water molecules and flexibility. Finally, they identified and experimentally validated TH1834 as a potential lead for the inhibition of Tip60 activity. As for studies focused on ligand-based techniques, our laboratory performed 3D QSAR analyses of hetero cyclic quinones to investigate their inhibitory activity on vascular smooth muscle cell (SMC) proliferation [97] and their cytostatic activity [98]. In both studies, we used genetic algorithm with linear assignment of hyper molecular alignment of database (GALAHAD) to generate and refine the pharmacophore model for molecule alignment, and CoMFA and CoMSIA to relate the molecular properties of the compounds with their observed activities.

In the first study, our group determined the inhibitory activity of known heterocyclic quinone inhibitors for SMC proliferation and utilized 3D-QSAR to obtain and study their 3D molecular contour maps. The information acquired from this research can then be employed for the design of more potent SMC proliferation inhibitors [97]. For the second study, we examined the cytostatic activity of heterocyclic quinones by utilizing semi-empirical calculations and the resulting LUMO energy to optimize the structures and compute the potential for the one-electron reduction of quinones, respectively. Our study provided reasonable correlation between the cytotoxic activity of the compounds with their calculated reduction potential, yielding 3D QSAR models and contour maps that can be used to design compounds with reduced cytotoxic activity [98]. While both SBDD and LBDD were successfully utilized individually in the studies mentioned above, relying solely on one approach greatly limits the probability of finding a plausible lead. Integration of these two methods in a drug discovery study can provide better and more extensive information in the modeling of innovative drug candidates against various diseases. This can be observed in a recent computational study involving the transient receptor potential vanilloid type 1 (TRPV1), wherein the combi nation of homology modeling, pharmacophore filtering, and molecular docking identified several potential antagonists. Initially, they generated a pharmacophore model from known TRPV1 antagonists and used it to filter the NCI database before performing docking experiments against the hTRPV1 homology model. From the docking results, they selected the top-scoring compounds for further experimental testing, yielding novel antagonists for hTRPV1 [99]. In addition to assisting in the selection of potential leads for a given target, computational tools can also provide viable interaction hypotheses for experimentally validated inhibitors of a target disease. As an example, our group conducted molecular modeling studies of potent DNA methyltransferase (DNMT) inhibitors, SGI-1027 and CBC12, to propose the binding modes of these compounds and give a possible explanation for their observed activity in vitro. Remarkably, the binding scores obtained for SGI-1027 were in excellent agreement with the published experimental results, validating the docking models. Moreover, the docking result of CBC12 corroborates the proposed inhibitory mechanism for DNMTs which suggests the use of ‘‘long’’ scaffolds in the design of DNMT inhibitors [100].

Artificial intelligence in drug design and drug discovery

AI is mainly used to determine the severity of disease and to predict whether the treatment will be effective for a particular patient even before it is administered, to avoid or resolve issues that may arise during treatment, and as a patient-assistance technology during treatment procedures or operations. It determines how specific instruments or drugs are used throughout therapy. It invents or extrapolates

new uses for those instruments or medicines to increase safety and efficacy [101].

Artificial Intelligence (AI) is the simulation of human intelligence by machines or computers. They usually work by consuming massive volumes of pre-trained models, analyzing the information for correlations and patterns, and then using these patterns for prediction. Artificial intelligence can identify hit and lead compounds, validate the drug target faster, and optimize drug structure design [102]. It can also assist in the 3D structure prediction of a targeted protein, protein-protein interactions, drug activity, and de novo drug design (**Figure 8**) [103]. To achieve these objectives, several specialized computational tools and software platforms as shown in **Figure 8**, are utilized. For instance, AlphaFold has revolutionized structural biology by predicting 3D protein structures with atomic accuracy directly from amino acid sequences [104]. DeltaVina and NNScore are advanced scoring functions; DeltaVina improves the evaluation of protein-ligand binding affinities, while NNScore utilizes artificial neural networks to predict binding poses more accurately than traditional empirical scoring methods [105]. Quantum Machine Learning (QML) merges quantum chemistry calculations with machine learning algorithms, while tools like ORGANIC (Objective-Reinforced Generative Adversarial Networks for Inverse Chemistry) apply generative adversarial frameworks to design novel, synthetically feasible molecules with desired chemical properties [106]. Lastly, standard computer-aided drug design (CADD) platforms act as the foundational pipeline integrating these diverse tools.

AI techniques

The fundamental AI techniques are Heuristics, Support Vector Machines, Artificial Neural Networks, the Markov Decision Process, and Natural Language Processing (**Figure 9**). Each of these core techniques serves a unique purpose in computational pharmacology. Heuristics are rule-based, practical problem-solving methodologies used to rapidly filter out structurally undesirable compounds from massive chemical libraries. Support Vector Machines (SVMs) function as robust classifiers that separate active compounds from inactive ones based on chemical descriptors [107]. Artificial Neural Networks (ANNs) mimic biological brain structures to process non-linear biological data, mapping complex relationships between molecular structures and their pharmacological activities [107]. The Markov Decision Process (MDP) provides a mathematical framework for modeling decision-making in settings where outcomes are partly random, which is highly useful in optimizing multi-step de novo drug synthesis routes. Finally, Natural Language Processing (NLP) is applied to mine vast biomedical literature, clinical trial registries, and electronic health records to extract hidden associations between drugs, targets, and side effects.

Machine learning in drug design and drug discovery

Machine learning is a branch of AI and computer science that focuses on utilizing the data and algorithms to mimic the way of learning process of humans by gradually increasing its accuracy. It is an emerging field of data science which uses statistical approaches and trained algorithms to generate classifications or predictions revealing critical in sights in data mining.

Machine learning is categorized into four groups based on the methodologies and methods of learning: Supervised, Unsupervised learning, Semi-supervised learning, and reinforcement learning [107]. Semi-supervised learning utilizes a small amount of labeled data alongside a large pool of unlabeled data to improve prediction accuracy, whereas reinforcement learning trains algorithms through a system of rewards and penalties to design optimal molecular paths [107]. The AI algorithms significant tasks in an extensive dataset include classification, regression, grouping, and pattern recognition. Machine learning techniques increase decision-making in pharmaceutical data across various applications, including QSAR analysis, hit discoveries, and *de novo* drug designs, allowing for more accurate results. As seen by the recent breakthroughs in this sector, machine learning, particularly deep learning, has become an ever more significant and productive basis for computational approaches in drug development [108] (**Figure 10**).

Supervised machine learning

In this technique, a labelled dataset (some specifications are given as input which enroute the machines to provide us with the output) is used to train the machines. Based on the training, the device predicts the outcome. Supervised machine learning is further classified

Fig. 8. AI tools used in drug discovery

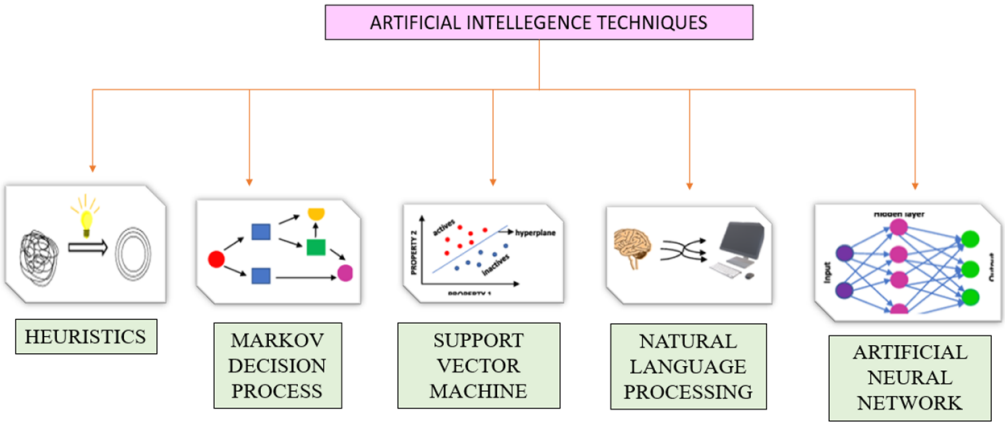

Fig. 9. Techniques of AI

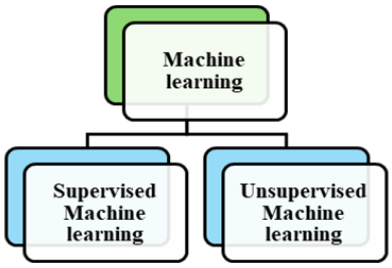

Fig. 10. Types of machine learning

into two categories: i) Classification: This algorithm predicts a dataset's categories and solves the classification problems. Examples of classification algorithms are Random Forest Algorithm, Decision Tree Algorithm, Logistic Regression Algorithm and Support Vector Machine Algorithm. ii) Regression: This algorithm is used to predict continuous output variables and addresses regression problems in which the input and output variables have a linear relationship. Some examples of regression algorithms are simple linear regression algorithm, multivariate regression algorithm, decision tree algorithm, lasso regression. In a practical pharmaceutical context, regression models are routinely deployed to predict quantitative variables, such as the exact half-maximal inhibitory concentration (IC_{50}) of a drug candidate, its solubility, or its binding affinity values [109].

Unsupervised machine learning

In this technique, the machine is trained with an unlabeled dataset without any supervision. The main aim of the unsupervised learning algorithm is to group or categorize the unsorted dataset according to similarities, patterns, & differences. Common algorithms used in this category include K-means clustering, hierarchical clustering, and Principal Component Analysis (PCA) [108]. In pharmaceutical research, these unsupervised techniques are widely applied to group compounds with similar chemical scaffolds together, compress high-dimensional genomic and proteomic data into interpretable formats, and identify novel, unexpected disease sub-types from raw patient data [108,109].

CONCLUSION

The computational techniques are crucial and appreciated tools in both academia and industry, and they undoubtedly speed up the drug

discovery process. This study focuses on the state of the art of CADD techniques in the drug design and discovery framework during the last few decades. Enormous progress has been made in drug development employing *in silico* approaches, resulting in an unambiguously good sector image. However, the CADD techniques are widely accepted as being far from perfect and omnipotent in all cases.

For current computational approaches to being employed effectively, considerable constraints must be overcome. AI intelligence, machine learning, and deep learning approaches were used with funda mental CADD procedures to provide more accurate and exact results. In the Big Data Era, these hybrid methods produce successful results by dealing with massive amounts of data while allowing quick decision-making. Because these methods are rapid and accurate, the biggest problem arises during model development. If there is no active supervision, faulty results may ensue. In our review, it is explained in detail how computational tools and techniques have been applied to drug discovery and development.

Challenges and perspectives

Approaches based on artificial intelligence and machine learning are preferable to the conventional CADD approach employed in drug discovery, particularly when handling vast amounts of input data. Utilizing input data from publicly accessible databases of poor quality and unbalanced presents most significant challenge when using the computational approaches. Therefore, the problem is selecting specific, manually curated resources with high-quality input data. The complexity of models developed increases the risk of overfitting when there is a lack of data, which is another concern. Therefore, it is necessary to create more adaptable models to solve this issue.

REFERENCES

1. Song CM, Lim SJ, Tong JC. Recent advances in computer-aided drug design. *Brief Bioinform*. 2009; 10(5):579-91. doi:10.1093/bib/bbp023
2. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ*. 2016;47:20-33. doi:10.1016/j.jhealeco.2016.01.012
3. Vohora D, Singh G. Pharmaceutical medicine and translational clinical research. Academic Press: Elsevier, USA, 1st edition, 2017.
4. Zhong F, Xing J, Li X, et al. Artificial intelligence in drug design. *Sci China Life Sci*. 2018;61(10):1191-1204. doi:10.1007/s11427-018-9342-2
5. Hou T, Xu X. Recent development and application of virtual screening in drug discovery: An overview. *Curr Pharm Des*. 2004;10(9):1011-33. doi:10.2174/1381612043452721
6. Yu W, MacKerell Jr AD. Computer-aided drug design methods. *Methods Mol Biol*. 2017;1520:85-106. doi:10.1007/978-1-4939-6634-9_5
7. Macalino SJY, Gosu V, Hong S, Choi S. Role of computer-aided drug design in modern drug discovery. *Arch Pharm Res*. 2015;38(9):1686-1701. doi:10.1007/s12272-015-0640-5
8. Duch W, Swaminathan K, Meller J. Artificial intelligence approaches for rational drug design and discovery. *Current Pharm Des*. 2007;13(14):1497-1508. doi:10.2174/138161207780765954
9. Huang HJ, Yu HW, Chen CY, et al. Current developments of computer-aided drug design. *J Taiwan Inst Chem Engrs*. 2010;41(6):623-35. doi:10.1016/j.jtice.2010.03.017
10. Hassan Baig M, Ahmad K, Roy S, et al. Computer aided drug design: success and limitations. *Curr Pharm Des*. 2016;22(5):572-81. doi:10.2174/1381612822666151125000550
11. Skariyachan S, Challapilli SB, Packirisamy S, Kumargowda ST, Sridhar VS. Recent aspects on the pathogenesis mechanism, animal models and novel therapeutic interventions for middle east respiratory syndrome coronavirus infections. *Front Microbiol*. 2019;10:569. doi:10.3389/fmicb.2019.00569
12. Amin SA, Jha T. Fight against novel coronavirus: A perspective of medicinal chemists. *Eur J Med Chem*. 2020;201:112559. doi:10.1016/j.ejmech.2020.112559
13. Goyal B, Goyal D. Targeting the dimerization of the main protease of coronaviruses: A potential broad-spectrum therapeutic strategy. *ACS Comb Sci*. 2020; 22(6):297-305. doi:10.1021/acscombsci.0c00058
14. Dai W, Zhang B, Jiang XM, et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science*. 2020;368(6497):1331-5. doi:10.1126/science.abb4489
15. Tu YF, Chien CS, Yarmishyn AA, et al. A review of SARS-CoV-2 and the ongoing clinical trials. *Int J Mol Sci*. 2020;21(7):2657. doi:10.3390/ijms21072657
16. Gopal D, Skariyachan S. Recent perspectives on COVID 19 and computer-aided virtual screening of natural compounds for the development of therapeutic agents towards SARS-CoV-2. In: Roy K. (eds) In silico Modeling of Drugs against Coronaviruses. Methods in Pharmacology and Toxicology. Humana Press, New York, 2020, pp. 433-71. doi:10.1007/7653_2020_44
17. Pillaiyar T, Meenakshisundaram S, Manickam M. Recent discovery and development of inhibitors targeting coronaviruses. *Drug Discov Today*. 2020; 25(4):668-88. doi:10.1016/j.drudis.2020.01.015
18. Drugs@FDA Glossary of Terms; US Food & Drug Administration: Silver Spring, MD, USA, 2022.
19. Molecular Conceptor Learning Series; Synergix Ltd.: Singapore, 2012.

20. Young DC. (ed) Properties that make a molecule a good drug. In: Computational Drug Design; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2009; pp. 9-39. doi:10.1002/9780470451854.ch2
21. Rowland M, Tozer TN. (eds) Fundamental concepts and terminology. In: Clinical Pharmacokinetics and Pharmacodynamics, 4th edition, Lippincott Williams & Wilkins: Baltimore, MD, USA, 2011; pp. 17-45.
22. Sinha S, Vohora D. (eds) Drug discovery and development: An overview. In: Pharmaceutical Medicine and Translational Clinical Research; Elsevier Inc.: Amsterdam, The Netherlands, 2018.
23. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. *Br J Pharmacol*. 2011;162(6):1239-49. doi:10.1111/j.1476-5381.2010.01127.x
24. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov*. 2002;1(9):727-30. doi:10.1038/nrd892
25. Oprea TI, Bologa CG, Brunak S, et al. Unexplored therapeutic opportunities in the human genome. *Nat Rev Drug Discov*. 2018;17:317-32.
26. Sheils TK, Mathias SL, Kelleher KJ, et al. TCRD and Pharos 2021: Mining the human proteome for disease biology. *Nucleic Acids Res*. 2021;49:D1334-46. doi:10.1093/nar/gkaa993
27. Berman H, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data. *Nucleic Acids Res*. 2007;35:D301-3. doi:10.1093/nar/gkl971
28. David L, Thakkar A, Mercado R, et al. Molecular representations in AI-driven drug discovery: A review and practical guide. *J Cheminform*. 2020;12(1):56. doi:10.1186/s13321-020-00460-5
29. Carracedo-Reboredo P, Liñares-Blanco J, Rodríguez-Fernández N, et al. A review on machine learning approaches and trends in drug discovery. *Comput Struct Biotechnol J*. 2021;19:4538-58. doi:10.1016/j.csbj.2021. 08.011
30. Hansch C, Steward AR. The use of substituent constants in the analysis of the structure-activity relationship in penicillin derivatives. *J Med Chem*. 1964;7(6):691-4. doi:10.1021/jm00336a001
31. Martin Y, Stouch T. In tribute to Corwin Hansch, father of QSAR. *J Comput Aided Mol Des*. 2011;25(6):491. doi:10.1007/s10822-011-9449-6
32. Wold S, Dunn III WJ. Multivariate quantitative structure-activity relationships (QSAR): Conditions for their applicability. *J Chem Inf Comput Sci*. 1983;23(1):6-13. doi:10.1021/ci00037a002
33. Wold S, Gelada P, Esbensen K, et al. Multiway principal components and PLS-analysis. *J Chemometr*. 1987;1(1):41-56. doi:10.1002/cem.1180010107
34. Xu J, Hagler A. Chemoinformatics and drug discovery. *Molecules*. 2002;7(8):566-600. doi:10.3390/70800566
35. Mitchell JBO. Machine learning methods in chemoinformatics. *Comput Mol Sci*. 2014;4(5):468-81. doi:10.1002/wcms.1183
36. Lo YC, Rensi SE, Torng W, et al. Machine learning in chemoinformatics and drug discovery. *Drug Discov Today*. 2018;23:1538-46. doi:10.1016/j.ejmech.2021.113705
37. Surabhi S, Singh BK. Computer aided drug design: An overview. *J Drug Deliv Therapeut*. 2018;8(5):504-9. doi:10.22270/jddt.v8i5.1894
38. Veselovsky A, Ivanov AS. Strategy of computer-aided drug design. *Curr Drug Targets Infect Disord*. 2003;3(1):33-40. doi:10.2174/1568005033342145
39. Hassan Baig M, Ahmad K, Roy S, et al. Computer aided drug design: success and limitations. *Curr Pharm Des*. 2016;22(5):572-81. doi:10.2174/1381612822666151125000550
40. Kore PP, Mutha MM, Antre R, et al. Computer-aided drug design: An innovative tool for modeling. *Open J Med Chem*. 2012;02(04):139-48. doi:10.4236/ojmc.2012.24017
41. Sun H. Pharmacophore-based virtual screening. *Curr Med Chem*. 2008;15(10):1018-24. doi:10.2174/092986708784049630
42. Shoichet BK. Virtual screening of chemical libraries. *Nature*. 2004;432(7019):862-5. doi:10.1038/nature03197
43. Stumpfe D, Ripphausen P, Bajorath J. Virtual compound screening in drug discovery. *Future Med Chem*. 2012;4(5):593-602. doi:10.4155/fmc.12.19
44. Batool M, Ahmad B, Choi S. A structure-based drug discovery paradigm. *Int J Mol Sci*. 2019;20(11):2783. doi:10.3390/ijms20112783
45. Kawato T, Mizohata E, Shimizu Y, et al. Structure-based design of a streptavidin mutant specific for an artificial biotin analogue. *J Biochem*. 2015;157(6):467-75. doi:10.1093/jb/mvv004
46. Middleton DA. Solid-state NMR spectroscopy as a tool for drug design: From membrane-embedded targets to amyloid fibrils. *Biochem Soc Trans*. 2007;35(5):985-90. doi:10.1042/BST0350985
47. Hosfield D, Palan J, Hilgers M, et al. A fully integrated protein crystallization platform for small-molecule drug discovery. *J Struct Biol*. 2003;142(1):207-17. doi:10.101 6/s1047-8477(03)00051-0
48. Lavecchia A, Di Giovanni C. Virtual screening strategies in drug discovery: A critical review. *Curr Med Chem*. 2013;20(23):2839-60. doi:10.2174/09298673113209990001
49. Park H, Chien PN, Ryu SE. Discovery of potent inhibitors of receptor protein tyrosine phosphatase sigma through the structure-based virtual screening. *Bioorg Med Chem Lett*. 2012;22(20):6333-7. doi:10.1177/174749301877 8713
50. Jorgensen WL. The many roles of computation in drug discovery. *Science*. 2004;303(5665):1813-8. doi:10.1126/science.1096361
51. Gurung AB, Ali MA, Lee J, et al. An updated review of computer-aided drug design and its application to COVID-19. *Biomed Res Int*. 2021;2021:8853056. doi:10.1155/2021/8853056
52. Vyas VK, Ukawala RD, Ghate M, et al. Homology modelling a fast tool for drug discovery: current perspectives. *Indian J Pharm Sci*. 2012;74(1):1-17. doi:10.4103/0250-474X.102537
53. Lemer CMR, Rooman MJ, Wodak SJ. Protein structure prediction by threading methods: Evaluation of current techniques. *Proteins Struct Funct Genet*. 1995;23(3):337-55. doi:10.1002/prot.340230308
54. Lee J, Freddolino PL, Zhang Y. Ab initio protein structure prediction. In: Rigden DJ (ed) From protein structure to function with bioinformatics. Springer,

Dordrecht; 2017, pp. 3-35. doi:10.1007/978-94-024-1069-3_1
55. Lesk AM, Chothia C. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J Mol Biol*. 1980; 136(3):225-70. doi:10.1016/0022-2836(80)90373-3
56. Muhammed MT, Aki-Yalcin E. Homology modeling in drug discovery: overview, current applications, and future perspectives. *Chem Biol Drug Des*. 2019;93(1): 12-20. doi:10.1111/cbdd.13388
57. Mizuguchi K. Fold recognition for drug discovery. *Drug Discov Today Targets*. 2004;3(1):18-23. doi:10.1016/S1741-8372(04)02392-8
58. Jones DT, Taylort WR, Thornton JM. A new approach to protein fold recognition. *Nature*. 1992;358(6381):86-9. doi:10.1038/358086a0
59. Yan RX, Si JN, Wang C, et al. DescFold: A web server for protein fold recognition. *BMC Bioinform*. 2009;10(1):1-14. doi:10.1186/1471-2105-10-416
60. Wu S, Zhang Y. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins*. 2008;72(2):547-56. doi:10.1002/prot.21945
61. Simons KT, Kooperberg C, Huang E, et al. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol*. 1997;268(1): 209-25. doi:10.1006/jmbi.1997.0959
62. Xu D, Zhang Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins*. 2012;80(7): 1715-35. doi:10.1002/prot.24065
63. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 2005;33:W244-8. doi:10.1093/nar/gki408
64. Bates PA, Kelley LA, MacCallum RM, et al. Enhancement of protein modeling by human intervention in applying the automatic programs 3D- JIGSAW and 3D-PSSM. *Proteins*. 2001;S5:39-46. doi:10.1002/prot.1168
65. Peng J, Xu J. A multiple-template approach to protein threading. *Proteins*. 2011;79(6):1930-9. doi:10.1002/prot.23016
66. Kallberg M, Wang H, Wang S, et al. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*. 2012;7(8):1511-22. doi:10.1038/nprot.2012.085
67. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*. 2009;4(3):363-71. doi:10.1038/ nprot.2009.2
68. Kelley LA, Mezulis S, Yates CM, et al. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10(6):845-58. doi:10.1038/nprot.2015.053
69. Schwede T, Kopp J, Guex N, et al. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res*. 2003;31:3381-5. doi:10.1093/nar/gkg520
70. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*. 1991;253(5016):164-70. doi:10.1126/science.1853201
71. Jones DT. GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol*. 1999;287(4):797-815. doi:10.1006/jmbi.19 99.2583
72. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinform*. 2008;9(1):1-8. doi:10.1186/1471-2105-9-40
73. Peng J, Xu J. RaptorX: Exploiting structure information for protein alignment by statistical inference. *Proteins*. 2011;79(S10):161-71. doi:10.1002/prot.23175
74. Liang J, Woodward C, Edelsbrunner H. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci*. 1998;7(9):1884-97. doi:10.1002/prot.24065
75. An J, Totrov M, Abagyan R. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol Cell Proteomics*. 2005; 4(6):752-61. doi:10.1074/mcp.M400159-MCP200
76. Zhang Z, Li Y, Lin B, et al. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics*. 2011;27(15):2083-8. doi:10.1093/bioinformatics/btr331
77. Sahu A, Patra PK, Yadav MK, et al. Identification and characterization of ErbB4 kinase inhibitors for effective breast cancer therapy. *J Recept Signal Transduct Res*. 2017;37(5):470-80. doi:10.1080/10799893.2017.1342129
78. Sun J, Chen K. NSiteMatch: Prediction of binding sites of nucleotides by identifying the structure similarity of local surface patches. *Comput Math Methods Med*. 2017;5471607. doi:10.1155/2017/5471607
79. Huang B. MetaPocket: a meta-approach to improve protein ligand binding site prediction. *OMICS*. 2009; 13(4):325-30. doi:10.1089/omi.2009.0045
80. Tan KP, Nguyen TB, Patel S, et al. Depth: A web server to compute depth, cavity sizes, detect potential small-molecule ligand- binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic Acids Res*. 2013; 41:W314-21. doi:10.1093/nar/gkt503
81. Xie ZR, Liu CK, Hsiao FC, et al. A server using ligand-interacting and site-enriched protein triangles for prediction of ligand-binding sites. *Nucleic Acids Res*. 2013;41:W292-6. doi:10.1093/nar/gkt300
82. Zhu H, Pisabarro MT. MSPocket: an orientation-independent algorithm for the detection of ligand binding pockets. *Bioinformatics*. 2011;27(3):351-8. doi:10.1093/bioinformatics/btq672
83. Laurent B, Chavent M, Cragnolini T, et al. Epock: Rapid analysis of protein pocket dynamics. *Bioinformatics*. 2015;31(9):1478-80.
84. Stank A, Kokh DB, Horn M, et al. TRAPP webserver: predicting protein binding site flexibility and detecting transient binding pockets. *Nucleic Acids Res*. 2017; 45(W1):W325-30. doi:10.1093/nar/gkx277
85. Wagner JR, Sørensen J, Hensley N, et al. POVME 3.0: software for mapping binding pocket flexibility. *J Chem Theory Comput*. 2017;13(9):4584-92. doi:10.1021/acs.jctc.7b00500
86. Huang SY, Zou X. Advances and challenges in protein-ligand docking. *Int J Mol Sci*. 2010;11(8):3016-34. doi:10.3390/ijms11083016

87. Jiang F, Kim SH. Soft docking": matching of molecular surface cubes. *J Mol Biol*. 1991;219(1):79-102. doi:10.1016/0022-2836(91)90859-5
88. Ferrari AM, Wei BQ, Costantino L, et al. Soft docking and multiple receptor conformations in virtual screening. *J Med Chem*. 2004;47(21):5076-84. doi:10.1021/jm049756p
89. Aparoy P, Kumar Reddy K, Reddanna P. Structure and ligand based drug design strategies in the development of novel 5-LOX inhibitors. *Curr Med Chem*. 2012; 19(22):3763-78. doi:10.2174/092986712801661112
90. Yu W, MacKerell Jr AD. Computer-aided drug design methods. 2017:1520:85-106. doi:10.1007/978-1-4939-6634-9_5
91. Gerstmeier J, Seegers J, Witt F, et al. Ginkgolic acid is a multi-target inhibitor of key enzymes in pro-inflammatory lipid mediator biosynthesis. *Front Pharmacol*. 2019;10:797.
92. Acharya C, Coop A, Polli E, et al. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Curr Comput Aided Drug Des*. 2011;7(1):10-22. doi:10.2174/157340911793743547
93. Gao Q, Yang L, Zhu Y. Pharmacophore based drug design approach as a practical process in drug discovery. *Curr Comput Aided Drug Des*. 2010;6(1):37-49. doi:10.2174/157340910790980151
94. Zhao S, Li X, Peng W, et al. Ligand-based pharmacophore modeling, virtual screening and biological evaluation to identify novel TGR5 agonists. *RSC Adv*. 2021;11(16):9403-9. doi:10.1039/d0ra10168k
95. Gao C, Bourke E, Scobie M, et al. Rational design and validation of a Tip60 histone acetyltransferase inhibitor. *Sci Rep*. 2014;4:5372. doi:10.1038/srep05372
96. Ryu CK, Lee Y, Park SG, et al. 3D-QSAR studies of heterocyclic quinones with inhibitory activity on vascular smooth muscle cell proliferation using pharmacophore-based alignment. *Bioorg Med Chem*. 2008;16(22):9772-9. doi:10.1016/j.bmc.2008.09.062
97. Lee Y, Kim S, Rhee HK, et al. 3D-QSAR studies of cytotoxic heterocyclic quinones using calculated reduction potential. *Drug Dev Ment Res*. 2009;70(6): 438-44. doi:10.1002/ddr.20320
98. Feng ZW, Pearce LV, Xu XM, et al. Structural insight into tetrameric hTRPV1 from homology modeling, molecular docking, molecular dynamics simulation, virtual screening, and bioassay validations. *J Chem Inf Model*. 2015;55(3):572-88. doi:10.1021/ci5007189
99. Yoo J, Choi S, Medina-Franco JL. Molecular modeling studies of the novel inhibitors of DNA methyltransferases SGI 1027 and CBC12: Implications for the mechanism of inhibition of DNMTs. *PLoS One*. 2013; 8(4):e62152. doi:10.1371/journal.pone.0062152
100. Becker A. Artificial intelligence in medicine: what is it doing for us today? *Health Policy Technol*. 2019;8(2): 198-205. doi:10.1177/1747493018778713
100. Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects. *Drug Discov Today*. 2019;24:773-80. doi:10.1177/1747493018778713
101. Paul D, Sanap G, Shenoy S, et al. Artificial intelligence in drug discovery and development. *Drug Discov Today*. 2021;26(1):80. doi:10.1016/j.drudis.2020.10.010
102. Carracedo-Reboredo P, Linares-Blanco J, Rodríguez-Fernandez N, et al. review on machine learning approaches and trends in drug discovery. *Comput Struct Biotechnol J*. 2021;19:4538-58.
103. Dara S, Dhamercherla S, Jadav SS, et al. Machine Learning in Drug Discovery: a Review. *Artif Intell Rev*. 2021;1-53. doi:10.1007/s10462-021-10058-4
104. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-9.
105. Durrant JD, McCammon JA. NNScore: a neural-network-based scoring function for characterization of protein-ligand complexes. *J Chem Inf Model*. 2010; 50(10):1865-71. doi:10.1021/ci100244v
106. Guimaraes GL, Sanchez-Lengeling B, Outeiral C, Farias PLC, Aspuru-Guzik A. Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models. arXiv:1705.10843, 2018. doi:10.48550/arXiv.1705.10843
107. Lavecchia A. Machine-learning approaches in drug discovery: methods and applications. *Drug Discov Today*. 2015;20(3):318-31. doi:10.1016/j.drudis.2014.10.012
108. Klambauer G, Hochreiter S, Rarey M. Machine learning in drug discovery. *J Chem Inf Model*. 2019;59(3):945-6. doi:10.1021/acs.jcim.9b00136
109. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. *Drug Discov Today*. 2018;23(6):1241-50. doi:10.1016/j.drudis.2018.01.039

*****