European Journal of Pharmaceutical Sciences 181 (2023) 106324

Contents lists available at ScienceDirect

European Journal of Pharmaceutical Sciences

journal homepage: www.elsevier.com/locate/ejps

CADD, AI and ML in drug discovery: A comprehensive review

Divya Vemula, Perka Jayasurya, Varthiya Sushmitha, Yethirajula Naveen Kumar, Vasundhra Bhandari *

National Institute of Pharmaceutical Education and Research- Hyderabad, India

ARTICLE INFO

Keywords:
Computer-aided drug design
Artificial intelligence
Machine learning
Deep learning
drug discovery and development

ABSTRACT

Computer-aided drug design (CADD) is an emerging field that has drawn a lot of interest because of its potential to expedite and lower the cost of the drug development process. Drug discovery research is expensive and time-consuming, and it frequently took 10–15 years for a drug to be commercially available. CADD has significantly impacted this area of research. Further, the combination of CADD with Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) technologies to handle enormous amounts of biological data has reduced the time and cost associated with the drug development process. This review will discuss how CADD, AI, ML, and DL approaches help identify drug candidates and various other steps of the drug discovery process. It will also provide a detailed overview of the different *in silico* tools used and how these approaches interact.

1. Introduction

The inventive process of developing novel drugs based on understanding a biological target is known as drug design (Zhou and Zhong, 2017). Computer-aided drug design (CADD) combines various computer tools to identify and develop a promising drug development lead. CADD includes computational chemistry, molecular modeling, molecular design, and rational drug design (Muegge et al., 2017). In today's big data environment, having access to massive data is no guarantee of getting applicable predictive models (Schneider, 2018). To forecast therapeutic efficacy and side effects, techniques must be developed that systematically address high volume, multidimensional, and sparse data sources (Zhu, 2020; Zhang et al., 2017). The highlights of the review are the description of recent studies using traditional CADD, artificial intelligence (AI), machine learning (ML) and innovative deep learning technologies (DL).

1.1. Computer-aided drug design (CADD)

CADD utilizes two different techniques based on the availability of either 3D structures of a protein or ligands (Fig. 1). They are known as structure-based drug design (SBDD) and ligand-based drug design (LBDD) (Fig. 1). In some cases, the integration of both techniques has shown good accuracy in finding the lead molecules (Batool et al., 2019).

a) **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 (Kawato et al., 2015). A new era of SBDD in drug discovery and design has begun by disclosing many biological molecules' three-dimensional (3D) structures (Middleton, 2007; Hosfield et al., 2003; Lavecchia and Di, 2013). SBDD has emerged as a possible means of generating and optimizing ligands in the pharmaceutical industry (Park et al., 2012; Jorgensen, 2004; Gurung et al., 2021). Preparation of the target, identification of the binding site, molecular docking, virtual screening and molecular dynamics are the basic steps of SBDD.

1.1.1. Overview of the steps involved in SBDD

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 elucidation techniques, 3D structures of proteins deposited in the Protein Data Bank (PDB) are readily accessible (Vyas et al., 2012). When the target proteins' 3D structures are unavailable, computational methods like comparative or homology modeling (Lemer et al., 1995), threading

* Corresponding author at: Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research, Hyderabad, Balanagar, Hyderabad-500037, Telangana, India
E-mail address: Vasundhra.b@niperhyd.ac.in (V. Bhandari).

https://doi.org/10.1016/j.ejps.2022.106324
Received 24 August 2022; Received in revised form 26 October 2022; Accepted 3 November 2022
Available online 5 November 2022

(Lee et al., 2017), and *ab initio* modeling (Lesk and Chothia, 1980) have been successful in determining the structures of proteins from their sequences.

a) **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 (Muhammed and Aki-Yalcin, 2019). 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.

b) **Fold recognition or threading methods:** are used to find proteins with comparable folds but no sequence similarity (K. Mizuguchi, 2004; Jones et al., 1992). The structure is taken into account 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 (Yan et al., 2009; S. Wu and Zhang, 2008). This process is repeated for each database's empirically determined 3D structures, providing the structure that best matches the query sequence (Table 2) (K Mizuguchi, 2004). It is employed in SBDD research.

c) ***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 (Simons et al., 1997), 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 (Xu and Zhang, 2012).

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. 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 (Liang et al., 1998; An et al., 2005). Tools like POCKET, SURFNET, Q-SITE FINDER, DoGSite Scorer server (Zhang et al., 2011), CASTp (Sahu et al., 2017), NSiteMatch (Sun and Chen, 2017), metapocket (Huang, 2009), DEPTH (Tan et al., 2013), LISE (Xie et al., 2013) and MSpocket (Zhu and Pisabarro, 2011) are the *in silico* tools used to predict the binding sites of a target protein.

After the binding site is found, tools or servers such as Epock (Laurent et al., 2015), TRAnsient Pockets in Proteins (TRAPP) (Stank et al., 2017), and POVME (Wagner et al., 2017) are used to determine the volume of the binding pocket (Table 4).

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 are then ranked using scoring techniques. Many biological processes, such as signal transmission, cell control, and other

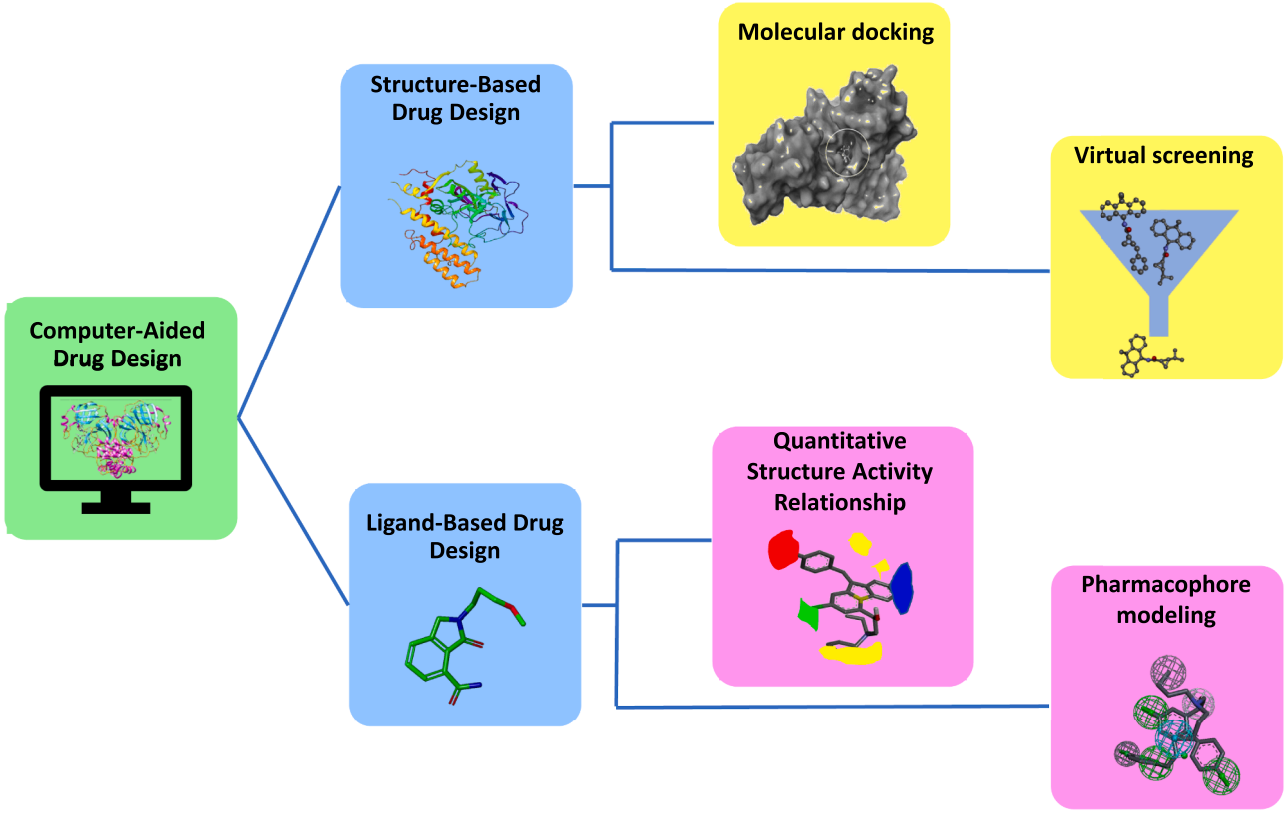


Fig. 1. Overview of CADD.

Table 1
Homology modeling popular structure prediction tools & their methods of prediction (Söding et al., 2005; Bates et al., 2001; J. Peng and Xu, 2011; Källberg et al., 2012; Kelley and Sternberg, 2009; Kelley et al., 2015; Schwede et al., 2003; Bowie et al., 1991).

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.expasy.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
6	RaptorX	Single/multi-template threading, alignment quality prediction	http://raptorx.uchicago.edu/
7	ESyPred3D	Template detection, alignment & 3D modeling	https://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/
8	MOE (Molecular Operating Environment)	Template identification, use of multiple templates and accounting for other environments, loop modeling, rotamer libraries for side chain conformations.	https://www.chemcomp.com/Products.htm
9	Yasara	Detection of templates, alignment including ligands and oligomers, hybridization of model fragments	http://www.yasara.org/
10	FoldX	Energy calculations & protein design	https://foldxsuite.crg.eu/
11	BhageerathH	Combination of *ab initio* & homology models	http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp

Table 2
Tools for threading methods (Jones, 1999; Zhang, 2008; J. Peng and Xu, 2011; S. Wu and Zhang, 2008).

S. NO	TOOL	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
3	I-TASSER	Iterative template fragment assembly	https://zhanggroup.org/I-TASSER/
4	DescFold	SVM-based machine learning algorithms in protein fold recognition	http://202.112.170.199/DescFold/index.html

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 (S.Y. Huang and Zou, 2010; Jiang and Kim, 1991; Ferrari et al., 2004) (Table 6).

Table 3
Tools for *ab initio* modeling.

S. NO	TOOL	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/

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
6	BiteNet	large-scale detection of protein binding sites	https://github.com/i-Molecule/bitenet
7	Metapocket	Combines 8 pocket methods	http://metapocket.eml.org
8	DEPTH	A web server to compute depth and predict small-molecule binding cavities in proteins and predict the pKa of ionizable residues in proteins	http://cospi.iiserpune.ac.in/depth
9	LISE	Ligand-binding site prediction using ligand-interacting and binding site-enriched protein triangles	http://lise.ibms.sinica.edu.tw
10	MSpocket	An orientation-independent algorithm for the detection of ligand binding pockets	http://appserver.biotec.tudresden.de/MSPocket/
11	Epock	Rapid analysis of protein pocket dynamics	http://epock.bitbucket.org
12	TRAnsient Pockets in Proteins (TRAPP)	A Tool for Analysis of Transient Binding Pockets in Proteins	https://trapp.h-its.org/
13	POVME	An Enhanced Tool for Determining Pocket Shape and Volume Characteristics	https://github.com/POVME/POVME
14	POOL	Machine learning application for functional site prediction in proteins	http://www.pool.neu.edu./
15	MetalDetector	Predicts metal binding sites in proteins using sequence information alone. Prediction is limited to transition metals	http://metaldetector.dsi.unifi.it/

1.1.2. Classification methods for protein-ligand docking

A **Protein Flexibility**: Current methods to account for protein flexibility are grouped into four categories:

i **Soft Docking:** In docking simulations, it relaxes the interatomic Vander Waals contacts, allowing for a slight overlap between the ligand and the protein.

ii **Side-Chain Flexibility:** One of the early studies was Leach's ligand docking approach, which uses a rotamer library to combine discrete side-chain flexibility (Leach, 1994). Since then, a slew of new techniques for adding continuous or discrete side-chain flexibility into ligand docking has been proposed (R. Abagyan et al., 1994; Desmet et al., 1997; Schaffer and Verkhivker, 1998; Schnecke and Kuhn, 2000; Frimurer et al., 2003; Zavodszky and Kuhn, 2005).

iii **Molecular Relaxation:** The third method considers protein flexibility by using rigid-body docking to introduce the ligand into the binding site and then relax the protein backbone and nearby side-chain atoms. The initial rigid-body docking allows for atomic conflicts between the protein and the inserted ligand orientations/conformations to accommodate protein conformational differences. The complexes are relaxed or minimised using Monte Carlo (MC) simulations, Molecular Dynamic simulations, or other methods (Apostolakis et al., 1998; Davis and Baker, 2009).

iv **Protein Ensemble Docking:** The most extensively used methods for adding protein flexibility involve a collection of protein structures to reflect various conformational variations (Carlson, 2002; Carlson and McCammon, 2000; Teodoro and Kavraki, 2003; Teague, 2003; Cozzini et al., 2008; Totrov and Abagyan, 2008). Osterberg et al. improved the approach with AutoDock (G. M. Morris et al., 1998), a bigger ensemble of 21 HIV-1 protease conformations (Osterberg et al., 2002).

B **Ligand Sampling:** there are two ligand sampling algorithms: a systematic search and stochastic algorithms (Table 5).

i **Systematic Search**: For flexible-ligand docking, systematic search techniques are typically utilized, which create all potential ligand binding conformations by exploring all of the ligand's degrees of freedom.

ii **Stochastic Algorithms**: Ligand binding orientations and conformations are sampled in stochastic algorithms by making random modifications to the ligand at each step in both the conformational space and the translational/rotational space of the ligand.

C Scoring function**:**

The docking score is a calculation based on scoring functions that evaluate the complex's energetic affinity. These scoring functions can be found on molecular mechanics, empirical data, expertise, or consensus dock. For example, it considers binding energetics using the AMBER force field, whereas SURFLEX utilizes an empirical function. Consensus scoring is a method for predicting the binding affinities of compounds for a specific target by combining the results of several scoring algorithms. Three basic categories according to their methods of derivation: force field, empirical, and knowledge-based scoring functions (S.Y. Huang and Zou, 2010).

i **Force field (FF) scoring functions** (G.M. Morris et al., 1998; Huang et al., 2006; Meng et al., 1992) are based on the decomposition of the ligand binding energy into individual interaction terms such as van der waals (VDW) energies, electrostatic energies, bond stretching/bending/torsional energies, etc., using a set of derived force-field parameters such as AMBER (Weiner and Kollman, 1981) or CHARMM (Nilsson and Karplus, 1986; Brooks et al., 1983) force fields**.**

ii **Empirical scoring functions**: A complex's binding energy score is derived by adding up several weighted empirical energy terms such as VDW energy, electrostatic energy, hydrogen bonding energy, desolvation term, entropy term, hydrophobicity term etc. (S.Y. Huang and Zou, 2010).

iii **Knowledge-based scoring functions** are generated directly from structural information in protein-ligand complexes that have been determined experimentally (Tanaka and Scheraga, 1976; Miyazawa and Jernigan, 1985; Sippl, 1990; Verkhivker et al., 1995) The potential of mean force (S.-Y. Huang and Zou, 2010), described by the inverse Boltzmann relation (P.D. Thomas and Dill, 1996; Koppensteiner and Sippl, 1998; P.D. Thomas and Dill, 1996; Liu and Wang, 2015) provides the basis for knowledge-based scoring functions.

IV **Virtual screening (VS):** *In silico* methods for choosing promising compounds from chemical databases is known as virtual screening (Lyne, 2002), and can be considered the computerized equivalent of experimental biological evaluation methods like high-throughput screening (HTS) (Stahura and Bajorath, 2004). One of the most common methodologies in drug development is to use vast and chemically varied compound libraries for computational and biological screening (Kodadek, 2011). VS is divided into two categories: (i) structure-based virtual screening (SBVS) (Oprea, 2004) and (ii) ligand-based virtual screening (LBVS) (Fig. 2).

i **Structure-based virtual screening (SBVS):** This is a computer-based method for searching a chemical compound library for novel bioactive compounds against a specific therapeutic target in early-stage drug development projects (Wu et al., 2017). The compound database in SBVS is docked into a predetermined target binding site (Lionta et al., 2014). In addition to predicting binding mode, SBVS assigns ranking to the docked molecules. This rating can be used as a sole criterion for selecting promising molecules or combined with other evaluation methodologies. Experiments are conducted to determine the biological activity of the indicated medications on the molecular target under investigation (Gangwal et al., 2015). SBVS comprises of four steps: (i) molecular target preparation, (ii) compound database selection, (iii) molecular docking, and (iv) post-docking analysis.

Programs for graphical display of molecular docking results (Table 7) (Pettersen et al., 2004; Sayle and Milner-White, 1995; Moll et al., 2006; Lill and Danielson, 2011; Humphrey et al., 1996).

Table 5
Software/ Tools that uses stochastic or systematic search algorithms (Morris et al., 1996; Jones et al., 1997; Baxter et al., 1998; Grosdidier et al., 2007; Venkatachalam et al., 2003; R. Abagyan et al., 1994; Zsoldos et al., 2007; McGann, 2012; Jain, 2003; Ewing et al., 2001; Friesner et al., 2004; Pang et al., 2001; Sochacka, 2014; Wu et al., 2003; Rarey et al., 1996; Welch et al., 1996; Miller et al., 1994; Schnecke and Kuhn, 1999; Mizutani et al., 1994).

Random/Stochastic Search	Systematic Search
AutoDock	eHiTS
Gold	FRED
PRO_LEADS	Surflex-Dock
EADock	DOCK
LigandFit	GLIDE
ICM	EUDOC
Molegro Virtual Docker	FlexX
CDocker	Hammerhead
GlamDock	Flog
PLANTS	SLIDE
MolDock	ADAM

Table 6
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/
AutoDock Vina Extended	2018	OneAngstrom	Genetic Algorithm	Commercial	https://github.com/ccsb-scripps/AutoDock-Vina
BetaDock	2011	Hanyang University	Based on Voroni Diagram	Freeware	http://voronoi.hanyang.ac.kr/software.htm
Blaster	2009	University of California San Francisco	Anchor and grow based docking program	Freeware	https://blaster.docking.org/
DARWIN	2000	The Wistar Institute	genetic algorithm	Freeware	https://darwin.cirad.fr/product.php
DOCK	1988	University of California-San Francisco	Shape Matching	Freeware for academic use	https://dock.compbio.ucsf.edu/
DockVision	1992	DockVision	Based on Monte Carlo, genetic algorithm, and database screening docking algorithms	Commercial	http://dockvision.sness.net/
DOLINA	2014	University of Basel	Local induced-fit algorithm	Academic	
EADock	2007	Swiss Institute of Bioinformatics	Based on evolutionary algorithms	Freeware	https://aurelien.latitude77.org/projects/eadock/gettingstarted/index.html
FlexX	2001	BioSolveIT	Incremental Construction	Commercial	https://www.biosolveit.de/download/?product=flexx
FlexAID	2015	University of Sherbrooke	Target side-chain flexibility and soft scoring function, based on surface complementarity	Open source (Apache License)	https://www.biosolveit.de/download/?product=flexx
FLIPDock	2007	Scripps Research Institute	Genetic algorithm-based docking program using FlexTree data structures to represent a protein-ligand complex	Freeware for academic use	https://www.scripps.edu/sanner/FLIPDock/oldIndex.html
GalaxyPepDock	2018	Seoul National University	Based on interaction similarity & energy optimization	Open source (GNU GPL) (standalone application) Freeware for academic use (web server)	https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=PEPDOCK
GEMDOCK	2004	National Chiao Tung University	Generic Evolutionary Method for molecular docking	Freeware	http://gemdock.life.nctu.edu.tw/
Glide	2004	Schrödinger	Exhaustive search-based docking program	Commercial	https://www.schrodinger.com/products/glide
GOLD	1995	Collaboration between the University of Sheffield, GlaxoSmithKline plc and CCDC	Genetic algorithm based, flexible ligand, partial flexibility for protein	Commercial	https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/
GPCRautomodel	2012	INRA	Automates the homology modeling of mammalian olfactory receptors (ORs) based on the six three-dimensional (3D) structures of G protein-coupled receptors (GPCRs) available so far and performs the docking of odorants on these models	Freeware for academic use	http://genome.jouy.inra.fr/GPCRautomdl/cgi-bin/welcome.pl
idTarget	2012	National Taiwan University	Predicts possible binding targets of a small chemical molecule via a divide-and-conquer docking approach	Freeware	http://idtarget.rcas.sinica.edu.tw/
LeDock	2016	Lephar	Incremental construction	Freeware for academic use	http://www.lephar.com/software.htm
LightDock	2018	Barcelona Supercomputing Center	Protein-protein, protein-DNA, protein-peptide docking using different scoring functions, backbone flexibility modeled by ANM and written in Python3 and Rust(programming language)	Open source (GNU GPL)	https://lightdock.org/
MedusaDock 2.0	2019	Dokholyan Laboratory	Rapid flexible docking using a stochastic rotamer library of ligands.	Free to use	https://dokhlab.med.psu.edu/cpi/#/MedusaDock
MOLS 2.0	2016	University of Madras	This algorithm uses mutually orthogonal Latin squares (MOLS) and uses a variant of mean field theory for analysis	Open source (GNU LGPL)	https://sourceforge.net/projects/mols2–0/files/
SwissDock	2011	Swiss Institute of Bioinformatics	Predictive web service for protein-small molecule ligand interactions.	Free to use webservice for academic usage	http://www.swissdock.ch/

ii **Ligand-Based Virtual Screening (LBVS):** By employing a computational technique called ligand-based virtual screening, a model of the target protein can be generated based on the information of the ligands that bind to the target effectively. Following that, the likelihood that the new ligand will bind to the target is predicted using this model (Jayaraj and Jain, 2019). LBVS is the sole approach without target protein 3D structure (Hamza et al., 2012; Geppert et al., 2010). LBVS tries to identify structurally diverse molecules with similar attributes using known active chemicals as input information.

V **Molecular Dynamic (MD) Simulation**:

This sophisticated physical technique is based on Newton's equations of motion guiding interatomic interactions. It is used to forecast the positions of each atom in a the molecular system with respect to time. Molecular dynamics (MD) simulations are crucial in investigating protein behavior. In MD simulations, chemical bonds and bond angles are

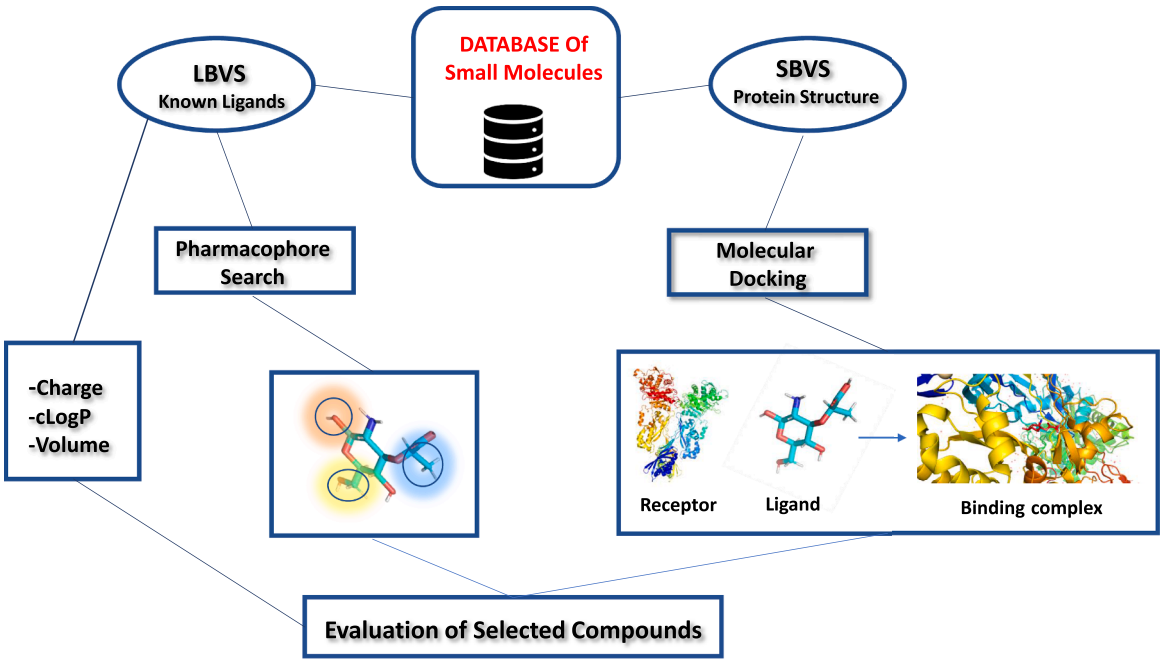


Fig. 2. Structure-based and Ligand-based Virtual Screening overview.

Table 7
Programs for graphical display of SBVS and molecular docking results.

S. NO	PROGRAMS	Description	Link
1	UCSF Chimera	Program for the interactive visualization and analysis of molecular structures and related data, including density maps, trajectories, and sequence alignments.	https://www.cgl.ucsf.edu/chimera/
2	JSmol	JSmol is a molecular viewer that is free and open source. It can search databases for structural data on molecules and map them in three dimensions.	http://wiki.jmol.org/index.php/JSmol
3	Jmol	An open-source Java viewer for chemical structures in 3D with features for chemicals, crystals, materials and biomolecules	http://jmol.sourceforge.net/
4	RasMol	RasMol is a molecular graphics program intended for the visualization of proteins, nucleic acids and small molecules.	http://www.openrasmol.org/
5	BALL	An extensible tool for visualizing and modeling bio-molecular structures.	https://ball-project.org/visualization/
6	Pymol	A cross-platform molecular graphic tool and has been widely used for 3D visualization of macromolecules.	https://pymol.org/2/
7	VMD	A molecular visualization program for displaying, animating & analyzing large biomolecular systems using 3-D graphics & built-in scripting.	https://www.ks.uiuc.edu/Research/vmd/

depicted using simple virtual springs, whereas dihedral angles are handled using a sinusoidal function. Originally designed to speed up video games, GPUs are now being used to accelerate molecular dynamics simulations significantly (Liu et al., 2008; Shan et al., 2011). Molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) (Hou et al., 2010), linear interaction energy (LIE) (Hansson et al., 1998) and free energy perturbation methods (FEP) are some of the MD applications used for free energy calculations to correlate experimental and calculated binding affinities of small molecules to proteins. Molecular dynamics simulations can calculate a trajectory of conformations as a function of time using Newtonian physics and force fields like Amber (Wang et al., 2004) or CHARMM (Vanommeslaeghe et al., 2010) OPLS, GROMOS, and coarse-grained force fields (Wang and O'Mara, 2021) (Table 8).

i AMBER (Assisted Model Building with Energy Refinement): AMBER is a group of molecular simulation programmes and a set of molecular mechanical force fields for simulating biomolecules. Peter Kollman's group developed this at the University of California, San Francisco, in the late 1970s to investigate diverse molecules like proteins, DNA, RNA, carbohydrates, organic molecules, peptidomimetics, lipids and fluorinated aromatic amino acids. The standard water models used in the AMBER force fields are TIP3P, SPC/E, TIP4PEW, and OPC [http://ambermd.org/index.php]. To impose molecular symmetry, partial charges are assigned with an electrostatic surface potential in the all-atom atomistic Assisted Model Building with Energy Refinement (AMBER) family of force fields. (Cornell et al., 1995).

ii CHARMM (Chemistry at Harvard Macromolecular Mechanics): The program CHARMM was initially developed by Professor Martin Karplus group, from Harvard University, which coordinates efforts for empirical force field parametrization. It has specific force fields parameterized for a wide range of molecules. The partial charges in the CHARMM force fields are typically fit to scale energies from *ab initio* calculations (MacKerell et al., 1998).

iii OPLS (optimized potentials for liquid simulations): The OPLS force field showed accuracy in prediction of structural and thermodynamic properties obtained by Monte Carlo simulations when compared with experimental findings.

iv The OPLS-AA force field, which was created to reproduce the quantum mechanical conformational energy profiles of small molecules. Additionally, it obtained several bonded parameters from AMBER. (Jorgensen and Tirado-Rives, 1988)

v GROMOS (Groningen Molecular Simulation): GROMOS is a multipurpose computer simulation tool for molecular dynamics used to research biomolecular systems. It also has a built-in force field that includes proteins, nucleotides, sugars, and other molecules. It may be

Table 8
List of force fields used in MD simulation.

S. NO	Force Field	Proteins	Carbohydrates	Nucleic acids (DNA/RNA)	Lipids	Organic molecules
1	AMBER	ff19SB (Tian et al., 2020), ff14SBonlysc, ff14SB, ff15ipq-m, ff15ipq	GLYCAM-06j (Kirschner et al., 2008)	OL15 (Galindo-Murillo et al., 2016), OL3 (Zgarbova et al., 2011; Bergonzo and Cheatham, 2015)	LIPID14 LIPID21,	gaff2
2	CHARMM	CHARMM 36 m, CHARMM 36 & CHARMM 22/CMAP	CHARMM36	C36 DNA, C36 RNA & C27 RNA (Hart et al., 2012; Denning et al., 2011; Foloppe and MacKerell, 2011) and DNA	CHARMM36 lipids, CHARMM22 & CHARMM27 lipids (Schlenkrich et al., 1996; Feller et al., 1997)	CHARMM General FF (CGenFF)
3	OPLS	OPLS-UA, OPLS-AA, OPLS3, OPLS3e, OPLS4	OPLS-AA-SEI (Kony et al., 2002)	OPLSAA/M (Robertson et al., 2019)	OPLS-AA/Berger (Maciejewski et al., 2014)	OPLS-AA, OPLS3, OPLS3e, OPLS4 (Dodda et al., 2017)
4	GROMOS	GROMOS87 &GROMOS96	GROMOS53A6GLYC (Pol-Fachin et al., 2012), GROMOS 45A4,	GROMOS 43A1, GROMOS 45A3 & GROMOS 45A1	GROMOS 54A8 (Marzuoli et al., 2019)	GROMOS96, GROMOS 53A5 & GROMOS 53A6 (Oostenbrink et al., 2005)
5	Coarse grained	MARTINI, OPEP (Kalimeri et al., 2015), UNRES (Liwo et al., 1997) & PaLaCe (Pasi et al., 2013)	MARTINI (Gautieri et al., 2010)	SIRAH (Uusitalo et al., 2015)	MARTINI and ELBA (Siani et al., 2016)	Martini 3 (Alessandri et al., 2022)

used to model various chemical and physical systems, including glasses, liquid crystals, polymers, crystals, and solutions of biomolecules.

vi Coarse-grained force field (CG): The CG force fields lower the computing cost of calculations by reducing the number of degrees of freedom in the model, allowing bigger systems to be simulated for longer periods. There are two general approaches to coarse-grained (CG) model: bottom-up and top-down. The partitioning of free energy between the polar and nonpolar phases of numerous chemical compounds served as the foundation for the CGMartini force field. The Martini force field has also been developed in close collaboration with atomistic models, particularly with regard to bound interactions (Barnoud and Monticelli, 2015).

b) 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 (Aparoy et al., 2012). LBDD methods use a known antibiotic (ligands) as a target to established a structure-activity relationship (SAR) between their physiochemical properties and antibiotic activities, which can improve existing drugs or guide the development of new drugs with enhanced activity (Yu and MacKerell, 2017). QSAR (quantitative structure-activity relationship) 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 9). Ligand similarity search methods are effective because structurally related compounds have comparable binding properties

I Quantitative structural activity relationship (QSAR)

A Quantitative structural activity relationship (QSAR) is a computerized statistical tool for explaining the observed variation in structure changes produced by replacement (Table 10). 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– 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 constant, Hyperconjugation correction, Molar volume, Molar refractivity, MR substituent constants, Molecular weight, Van der Waals radii Interatomic distances.

Table 9
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	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
ChEMBLdb	ChEMBL is a manually curated database of bioactive molecules with drug-like properties	Free	https://www.ebi.ac.uk
ChemSpider	ChemSpider is a database of chemicals	Free	http://www.chemspider.com

Table 10
Tools For QSAR.

TOOLS	DESCRIPTION	AVAILABILITY	URL
OECD QSAR Toolbox	The OECD QSAR Toolbox is a free software for screening and assessing chemical substances that uses 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
PhramQSAR	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/
AutoQSAR	AutoQSAR automates the creation of high-quality, predictive QSAR models and makes their application trivially simple	COMMERCIAL	https://www.schrodinger.com
GUSAR	GUSAR software was developed to create QSAR/QSPR models on the basis of the appropriate training sets represented as SDfile contained data about chemical structures and endpoint in quantitative terms.	FREE	http://www.way2drug.com/gusar/index.html

Substructural effects: Three-dimensional geometry Fragment and molecular properties.

Steps involved in QSAR (Fig. 3) (Gerstmeier et al., 2019; Acharya et al., 2011):

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, assess the statistical equation's stability. To anticipate the biological activity of a novel chemical, use a statistical model.

II PHARMACOPHORE MODELING

According to the IUPAC definition, Pharmacophore is "an ensemble of steric and electronic properties required to achieve optimal supramolecular interactions with a given biological target and to trigger or prevent it's biological response" (Gao et al., 2010). A pharmacophore is an abstract description of the structural properties required for a biological macromolecule to recognize a ligand (Table 11) (Zhao et al., 2021).

Steps Involved in Pharmacophore modeling (Fig. 4.) (Gao et al., 2010)**:**

a **Choose a training set of ligands** - For the pharmacophore model development, choose a structurally diverse group of compounds. The list of molecules should include both active and inactive compounds, as a pharmacophore model must be able to distinguish between molecules with and without bioactivity.
b **Analysis of Conformation** - Create a list of low-energy conformations for each selected compound that will likely include the bioactive conformation.
c **Molecular superimposition** - superimpose all possible combinations of the molecules' low-energy conformations. Functional groups that are similar in all of the molecules in the collection could be fitted (e.g., phenyl rings or carboxylic acid groups). The active conformation is assumed to be the set of conformations that results in the best fit.
d **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.'
e **Validation** - A pharmacophore model is a hypothesis that explains the pharmacological actions of a group of compounds that bind to the same biological target.

APPLICATIONS OF CADD IN DRUG DISCOVERY (Table 12).

1.2. Artificial intelligence in drug design and drug discovery

The 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

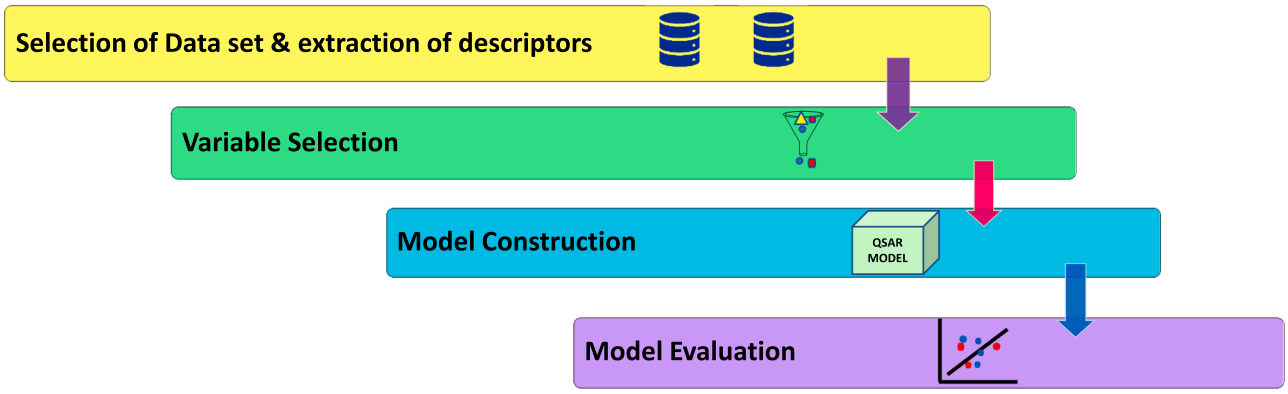


Fig. 3. Quantitative structure-activity relationship Workflow.

Table 11
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
Catalyst	catalyst is a versatile product for pharmacophore establishment, structural alignment, activity prediction and 3D database generation	commercial
CATSlight2	CATSlight2 is available as a web-server for molecular similarity	Free
PharmMapper	PharmMapper Server is a freely accessed web-server designed to identify potential target candidates for the given probe small molecules	Free
Pharmer	Pharmer is a new computational approach to pharmacophore search those scales with the breadth and complexity of the query, not the size of the compound library being screened.	Free
Phase	Phase is a complete, user-friendly pharmacophore modeling solution designed to maximize performance in virtual screening and lead optimization	commercial
ZincPharmer	ZINCPharmer is free pharmacophore search software for screening the purchasable subset of the ZINC database	Free

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 (Becker, 2019). 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. AI can identify hit and lead compounds, validate the drug target faster, and optimize drug structure design (Mak and Pichika, 2019). It can also assist in the 3D structure prediction of a targeted protein, protein-protein interactions, drug activity, and *de novo* drug design (Table 13) (Paul et al., 2021).

AI Techniques

The fundamental AI techniques are Heuristics, Support Vector Machines, Artificial Neural Networks, the Markov Decision Process, and Natural Language Processing (Fig. 5).

I **Heuristics:** A heuristic search strategy is an artificial intelligence (AI) search that looks for a good but necessarily perfect solution from the available possibilities. Leon et al. (2021) used a semi-exhaustive approach and heuristic search algorithm for a Fragment-based drug design (FBDD) strategy to design HSP90 inhibitors. They used a heuristic-based technique to find new high-quality ligands from the sub-ligands obtained via the semi-exhaustive approach's deconstruction process in less time. This procedure can be used in place of the semi-exhaustive reconstruction process. The goal of heuristic search is to reduce the time spent on the reconstruction process, allowing for the rapid creation of high-quality novel ligands (León et al., 2021). In the case of drug interaction prediction, Olayan et al. suggested a heuristic method for similarity selection (Olayan et al., 2018).

II **Support vector machines (SVMs): These** are supervised machine learning methods for classification, regression and outliers detection. SVMS generally uses a hyperplane to separate classes. In the drug discovery process, SVMs were combined with QSAR/QSPR and virtual screening techniques over the last 20 years (Gertrudes et al., 2012). There are two types of SVMs: linear and nonlinear. When the data can be perfectly classified into two classes using a single straight line, it is called a Linear SVM, whereas a nonlinear SVM is used when the data cannot be classified into two classes by a single line. In this case, kernel techniques can be used to classify them. The non-linear data can be classified using some quadratic functions called kernels which allow us to make decision boundaries to divide the data points. SVMs are used in drug discovery projects like Prediction of hERG Liability (Using SVM Classification, Bootstrapping and Jackknifing) (Sun et al., 2017). Using a multiple linear regression

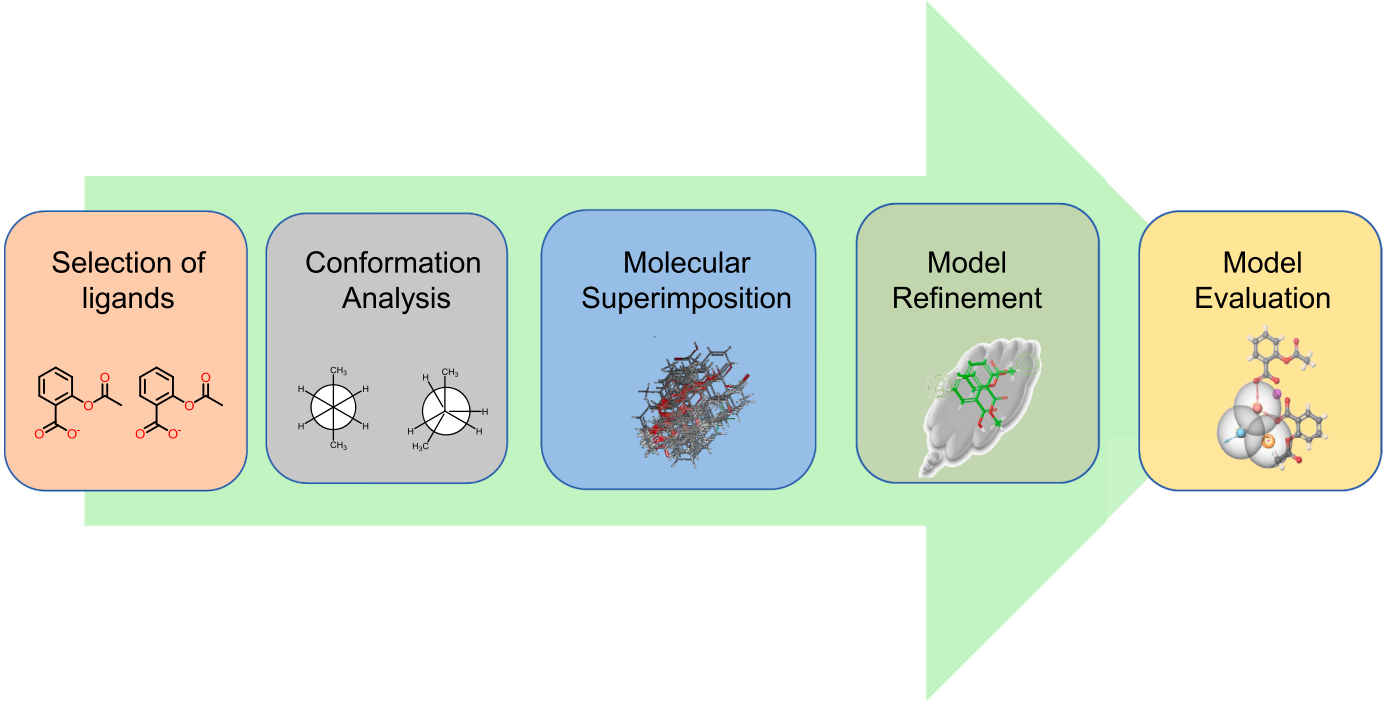
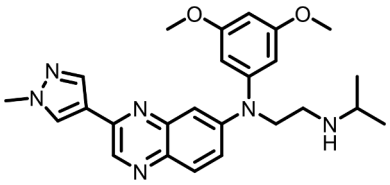
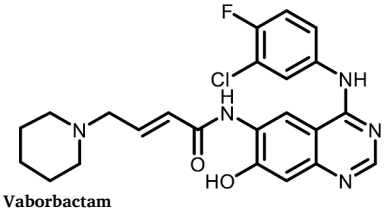
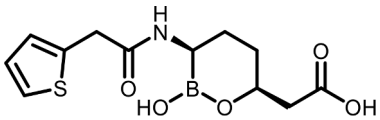
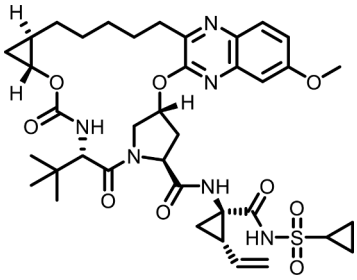
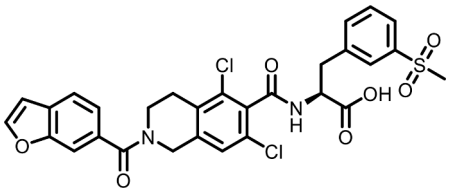
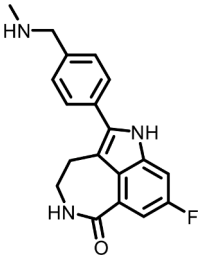
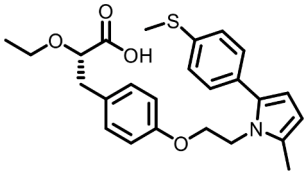
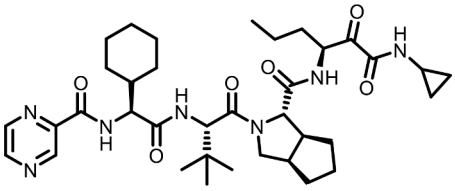


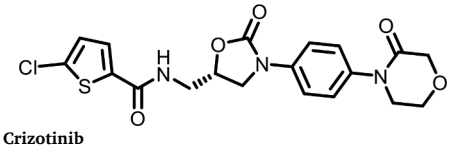
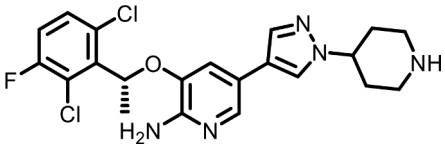
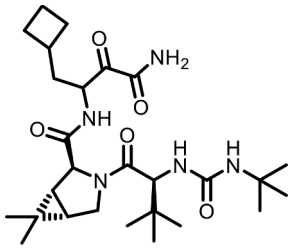
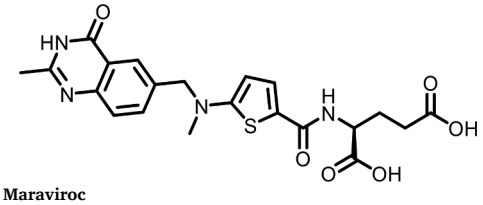
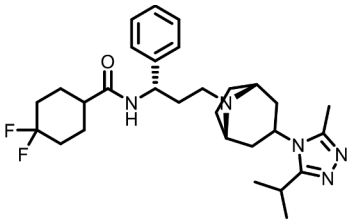
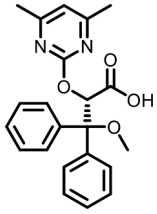
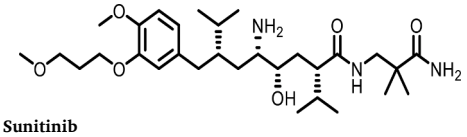
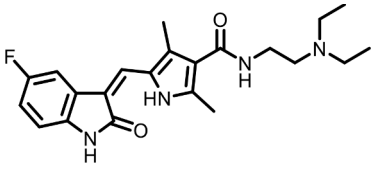
Fig. 4. Steps involved in Pharmacophore modeling.

Table 12
Applications of CADD in drug discovery.

DRUG	TARGET	THERAPEUTICUSE	YEAR OF FDA APPROVAL	REFERENCE
Erdafitinib 	Fibroblast growth factor receptors (FGFR)	Urothelial carcinoma	2019	(Murray et al., 2019)
Dacomitinib 	Multi-kinase	Non-Small-Cell Lung Cancer (NSCLC)	2018	(Reed and Smaill, 2016)
Vaborbactam 	beta-lactamase	Bacterial infections	2017	(Hecker et al., 2015)
Grazoprevir 	NS3/4 serine protease	Chronic Hepatitis C (HCV)	2016	(Harper et al., 2012)
Lifitegrast 	LFA-1/ICAM-1 (leukocyte function-associated antigen-1/ Intercellular Adhesion Molecule-1)	Dry eye disease	2016	(Abidi et al., 2016)
Rucaparib 	Poly (ADP-ribose) polymerase (PARP-1)	Prostate Cancer	2016	(White et al., 2000)
Saroglitazar 	Peroxisome Proliferator-Activated Receptor (PPAR)	Diabetic D yslipidemia	2013	(Agrawal, 2014)
Telaprevir 	NS3/4A protease	Chronic Hepatitis C	2011	(Rao et al., 2015)

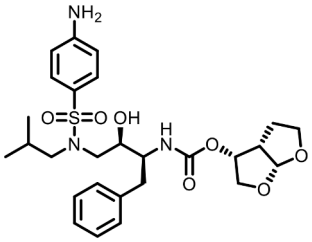
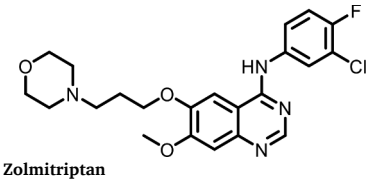
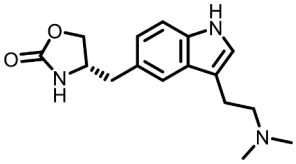
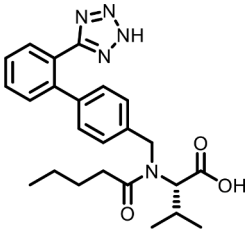
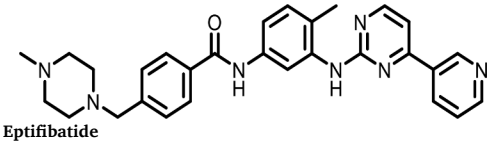
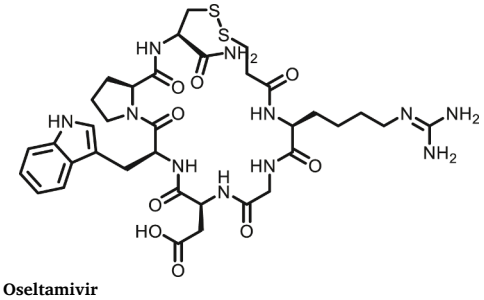
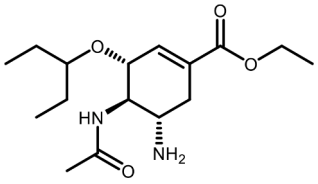
(*continued on next page*)

Table 12 (*continued*)

DRUG	TARGET	THERAPEUTICUSE	YEAR OF FDA APPROVAL	REFERENCE
Rivaroxaban 	Clotting Factor Xa	Deep Venous Thrombosis (DVT)	2011	(Perzborn et al., 2011)
Crizotinib 	Anaplastic Lymphoma Kinase (ALK) & ROS Proto-Oncogene 1 (ROS1)	Non-Small Cell Lung Cancer (NSCLC)	2011	(Cui et al., 2011)
Boceprevir 	Hepatitis C virus (HCV)	Chronic hepatitis C	2011	***(Njoroge et al., 2008)
Tomudex 	Thymidylate synthase	Colorectal cancer	2009	(Rutenber and Stroud, 1996)
Maraviroc 	C-C chemokine receptor type 5/ Envelope glycoprotein GP120 (CCR5/gp120)	Human Immunodeficiency Virus (HIV)	2007	(Veljkovic et al., 2015)
Ambrisentan 	Endothelin-A	Pulmonary Arterial Hypertension	2007	(Rivera-Lebron and Risbano, 2017)
Aliskiren 	Angiotensinogen	Blood pressure	2007	(Cohen, 2007)
Sunitinib 	VEGF-R2 kinase	kidney cancer	2006	(Sun et al., 2009)

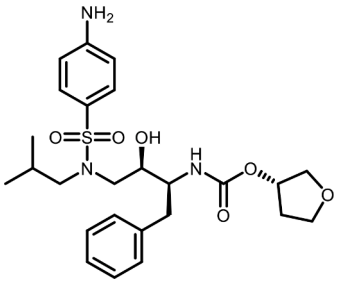
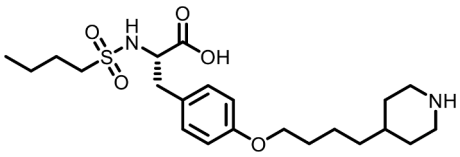
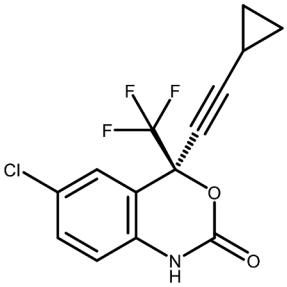
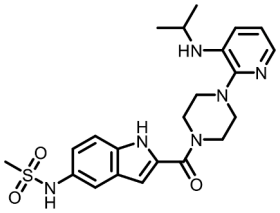
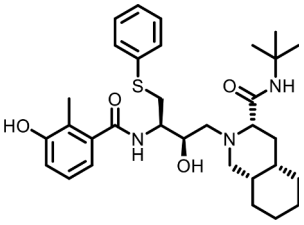
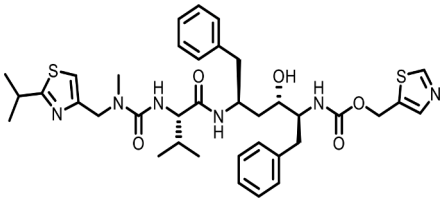
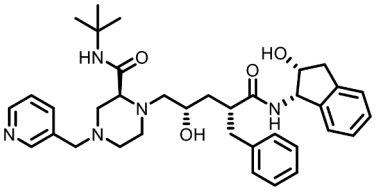
(*continued on next page*)

Table 12 (*continued*)

DRUG	TARGET	THERAPEUTICUSE	YEAR OF FDA APPROVAL	REFERENCE
Darunavir 	Nonpeptidic HIV-1 protease	Human Immunodeficiency Virus (HIV) infection	2006	(Ghosh et al., 2007)
Gefitinib 	GFRv tyrosine kinase	Non-Small Cell Lung Cancer	2003	(Barker and Andrews, 2013)
Zolmitriptan 	5-hydroxytryptamine (5HT)1B/ 1D/(1F) receptor	Migraine	2003	(Fischer and Ganellin, 2015)
Valsartan 	Angiotensin II receptor	Hypertension	2002	(Aulakh et al., 2007)
Imatinib 	Abl tyrosine kinase	Acute lymphoblastic leukemia	2001	(Schindler et al., 2000)
Eptifibatide 	Glycoprotein IIb/IIIa protein	Myocardial infarction	2001	(Goa and Noble, 2039)
Oseltamivir 	Influenza A & B neuraminidase	in the treatment of the infection caused by the flu virus (influenza A and influenza B)	1999	(Lew et al., 2000)

(*continued on next page*)

Table 12 (*continued*)

DRUG	TARGET	THERAPEUTICUSE	YEAR OF FDA APPROVAL	REFERENCE
Amprenavir 	Human Immunodeficiency Virus (HIV) protease	Human Immunodeficiency Virus (HIV) infection	1999	(Lyle, 2007)
Tirofiban 	Integrin (GP) IIb/IIIa and Fibrinogen receptor	Heart attack	1999	(Lynch et al., 1995)
Efavirenz 	Non-Nucleoside Reverse Transcriptase protein	Human Immunodeficiency Virus (HIV) infection	1998	(Best and Goicoechea, 2008)
Delavirdine 	HIV reverse transcriptase protein	Human Immunodeficiency Virus (HIV) infection	1997	(Adams et al., 2002)
Nelfinavir 	HIV-1 protease protein	Human Immunodeficiency Virus (HIV) infection	1997	(Aruksakunwong et al., 2007)
Ritonavir 	HIV-1 protease inhibitor	to treat human immunodeficiency virus (HIV) infection	1996	(Kim et al., 1995)
Indinavir 	HIV-1 protease	Human Immunodeficiency Virus (HIV) infection	1996	(Wlodawer and Vondrasek, 1998)

(*continued on next page*)

Table 12 (*continued*)

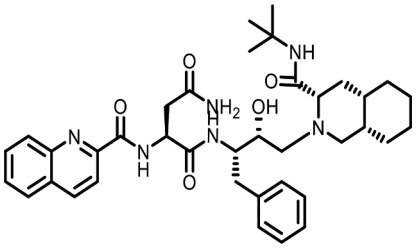
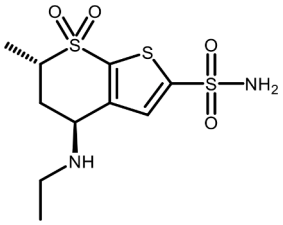
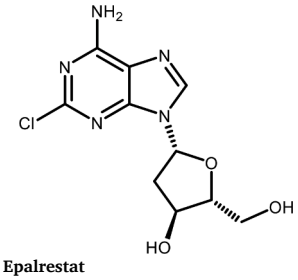
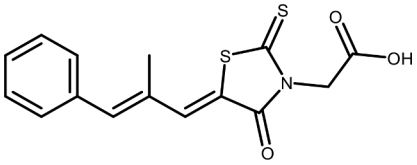
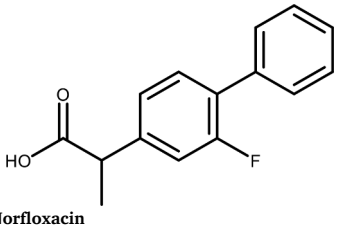
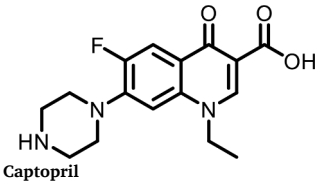
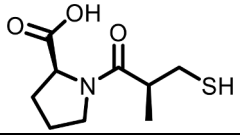
DRUG	TARGET	THERAPEUTICUSE	YEAR OF FDA APPROVAL	REFERENCE
Saquinavir 	HIV-1 protease	Human Immunodeficiency Virus (HIV) infection	1995	(De Clercq, 2009)
Dorzolamide 	Carbonic anhydrase	Glaucoma and cystoid macular edema	1994	(Supuran, 2012)
Cladribine 	Adenosine deaminase	hairy cell leukemia	1993	(Warnke et al., 2010)
Epalrestat 	Aldose Reductase	Diabetic neuropathy	1992	(Zhu, 2013)
Flurbiprofen 	cyclooxygenase-2	Nonsteroidal Anti-Inflammatory agent (NSAID)	1988	(Ashraf et al., 2016)
Norfloxacin 	Topoisomerase II and IV	Urinary Tract Infections	1986	(Takahashi et al., 2003)
Captopril 	Angiotensin Converting Enzyme (ACE)	Hypertension	1981	(Cushman and Ondetti, 1991)

Table 13
Artificial Intelligence Tools used in drug discovery.

S. NO	NAME OF TOOL	DESCRIPTION	AVAILABILITY	WEBSITE LINK
1	AlphaFold	The tertiary structure of a protein is predicted using a deep neural network.	Free	https://deepmind.com/blog/alphafold
2	Chemputer	Provides a step-by-step guide to creating a compound.	Commercial	https://zenodo.org/record/1481731
3	Chemical VAE	Chemical design automation using a variational autoencoder (VAE))	Free	https://github.com/aspuru-guzik-group/chemical_vae
4	DeltaVina	Predict small molecule binding affinity with medication using a mix of random forest (RF) and the AutoDock score method.	Free	https://github.com/chengwang88/deltavina
5	Hit Dexter	Predict small molecule binding affinity with medication using a mix of random forest (RF) and the AutoDock score method.	Free	http://hitdexter2.zbh.uni-hamburg.de
6	InnerOuterRNN	Inner- and outer recursive neural networks are used to predict physical, chemical, and biological features.	Free	https://github.com/Chemoinformatics/InnerOuterRNN
7	JunctionTree VAE	Use a combination of random forest (RF) and the AutoDock scoring method to predict small molecule binding affinity with drugs.	Free	https://github.com/wengong-jin/icml18-jtnn
8	NNScore	Using a neural network-based scoring function, predict protein–ligand interaction affinity.	Free	http://www.nbcr.net/software/nnscore
9	ORGANIC	ML algorithm for de novo design of organic molecules and polymers	Free	https://github.com/aspuru-guzik-group/ORGANIC
10	Open Drug Discovery Toolkit (ODDT's)	Using the random forest score (RF)-Score and the NNScore, a chemoinformatics pipeline has been developed.	Free	https://github.com/oddt/oddt
11	PPB2	Using the nearest neighbor and machine learning algorithms, predict the target of the query molecule.	Free	https://ppb2.gdb.tools/
12	QML	Quantum machine learning Python toolkit	Free	https://www.qmlcode.org/
13	REINVENT	RNN (recurrent neural network) and RL (recurrent learning) are used to create a new molecule (reinforcement learning)	Commercial	https://github.com/MarcusOlivecrona/REINVENT
14	XenoSite	Predictor of Metabolism and reactivity of small molecules	Free	https://swami.wustl.edu/xenosite
16	SMARTCyp	Site of metabolism prediction for CYP450 enzymes	Free	https://smartcyp.sund.ku.dk/
18	DIA-NN	Tool for proteomic data processing	Free	https://github.com/vdemichev/DiaNN/releases/tag/1.8

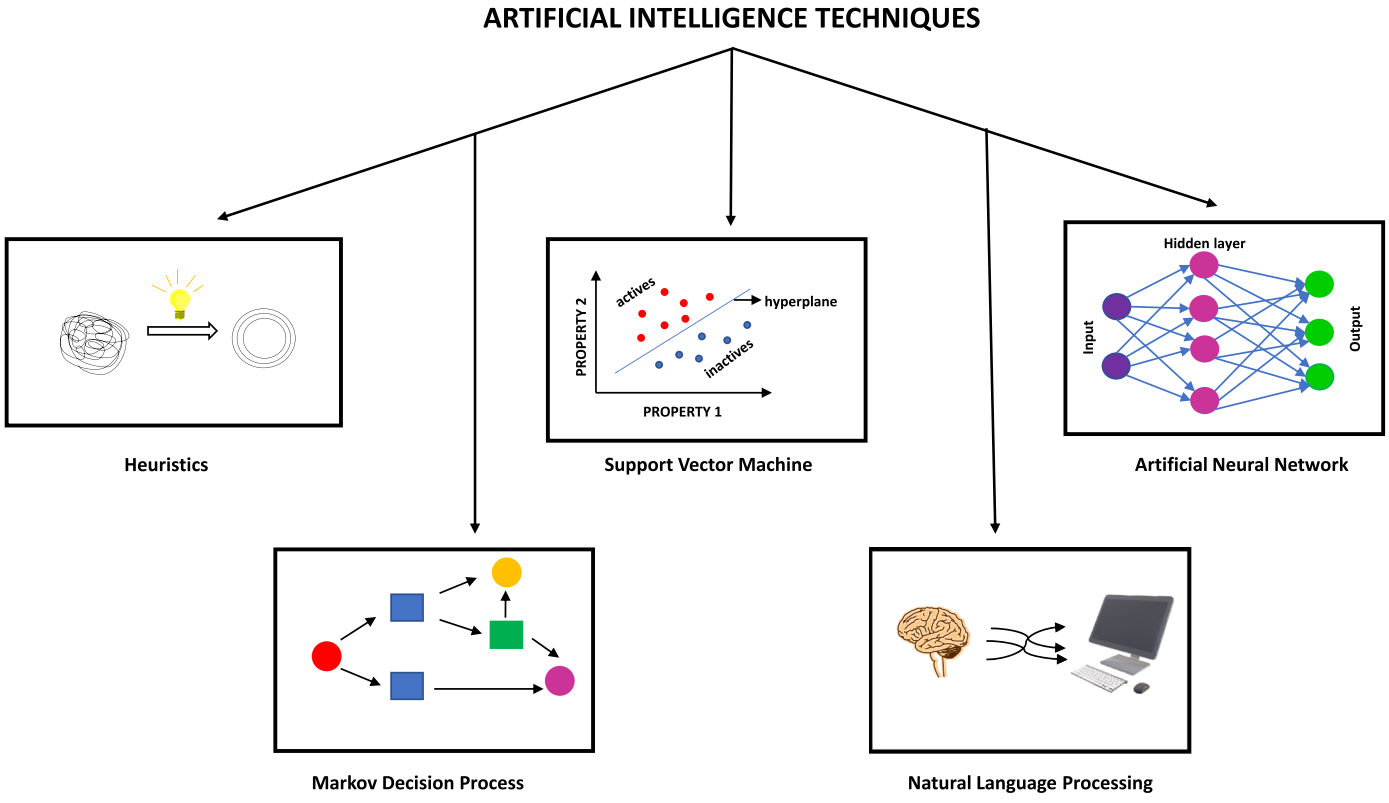


Figure 5. Techniques of AI

Fig. 5. Techniques of AI.

(MLR) and a support vector machine (SVM) method, researchers investigated quantitative structure-activity relationship (QSAR) models using various descriptor sets and training/test set selection methods to predict the bioactivity of hepatitis C virus (HCV) NS3/4A protease inhibitors. A combination of the best sub- and whole dataset SVM models can be employed in a drug development pipeline as reliable lead designing tools for novel NS3/4A protease inhibitor scaffolds (Qin et al., 2017). A pharmacological data set including 32 derivatives of 4- aminopyrimidine-5-carbaldehyde oxime was utilized to create linear and nonlinear QSAR-based models to predict VEGFR-2 inhibitory effects successfully. The built models clearly show good relationships between the structure and activity of the molecules under investigation. The Genetic algorithm (GA) was integrated with multiple linear regression and support vector machine to create these robust models, termed GA-MLR and GA-SVM, respectively (Nekoei et al., 2015). In their study, Wei et al. used a multistage virtual screening technique that included SVM, shape-based screening, pharmacophore modeling, and molecular docking to target the HIV-1 protease. They discovered that, compared to single virtual screening approaches, the multistage virtual screening approach can significantly improve virtual screening efficiency and accuracy by reducing computer time while also increasing hit rates and enrichment factors (Wei et al., 2015). Qiao et al. presented a feasible combinatorial ligand-based approach for identifying possible ACAT-2 (Acetyl-CoA acetyltransferase-2) selective inhibitors from Chinese herbs, which could aid in the screening and development of ACAT-2 selective inhibitors in the future. For the discovery of ACAT-2 inhibitors, a selective SVM model and a bioactive SVR model were created (Qiao et al., 2016).

III **Artificial Neural Networks:** Artificial Neural Networks (ANN) are processing devices modelled roughly after the brain's neural structure. The most significant difference between the two is that the ANN may include hundreds or thousands of neurons, but the neural structure of an animal or human brain may contain billions. The ANN's neurons can be organized into numerous layers. This network has an input layer that receives, processes, and converts all inputs into outputs for the subsequent layers. One or more layers of neurons travel via inputs and outputs to form the hidden layers. Finally, the output layer takes the inputs from the last hidden layer and turns them to the user's output (Aybars, 2021). Based on architecture and connectivity patterns, ANNs are classified into five types (Mandlik et al., 2016):

A **Multilayered Perceptron/Backpropagation Networks:** This is the most commonly used supervised neural network that back propagates the iteration produced in the network, preventing possible errors. The back propagation neural network was used to predict ADMET properties of unknown ligands by matching their 2D descriptors with that of known ligands (Puri et al., 2015).

B **Kohonen Neural Network:** This is an unsupervised neural network which projects the points into a 2D plane from a multi-dimensional space. By sampling the points on the surface and then constructing a 2D feature map, kohenen maps or self-organizing were utilized to predict molecular surface attributes (Zupan and Gasteiger, 1999). Polanski et al. trained SOMs to predict the pharmacophoric properties of metabotropic glutamate receptor 5 (mGlu5) negative allosteric modulators (Polanski, 2003).

C **Counter Propagation (CPG) Networks:** These supervised networks use Kohenen's training algorithm. In CPG networks, the input layer is used to calculate distance, whereas the weights of both the input and output layers are changed during the adaptation steps. Liu et al. presented a Dependency-based Convolutional Neural Network (DCNN) for drug-drug interaction extraction.

DCNN is a text-mining technique for predicting DDIs from unstructured biomedical literature and existing knowledge sets (Liu et al., 2016). Schuster et al. used counter propagation neural networks and SOMs to perform virtual screening to explore acetylcholinesterase inhibitors (Rupp et al., 2010).

A **Bayesian Neural Networks:** To represent weight uncertainty, the Bayesian technique learns weight distributions we can sample to build output for a given input, rather than learning precise weight (and bias) values as in a traditional neural network. Bayesian models were used to develop QSAR models on MHC class II binding affinity of peptides (Winkler and Burden, 2004). Peter et al. used bayesian neural networks for the classification of a set of pyrazines with different aromatic properties (Klocker et al., 2002).

B **Recurrent Neural Networks:** This is a special artificial neural network specifically designed to work with time series data or data containing sequences. RNNs feature a concept of 'memory', which allows them to store the states or information of prior inputs to construct the sequence's subsequent output. The ability to handle sequence data, handle inputs of various time points and store or 'memorize' past knowledge are all key advantages of RNNs. Based on a different architecture, the RNNs are classified into four types:

i. One-to-one network: Conventional neural networks employ this method where single input produces a single output.
ii. One to many networks: In this case, a single input produces multiple outputs.
iii. Many to one network: The inputs from different time intervals produce a single output.
iv. Many to many networks: In this case, two inputs may result in three or more outputs.

RNNs have many applications in drug design and the drug discovery process. Yasonik et al. (2020) used recurrent neural networks to rank the molecules based on molecular properties to improve the results of optimization of molecules as a part of his multi-objective *denovo* drug design in conjunction with no sorting algorithm, which usually relies on abundant known molecules (Yasonik, 2020). Li et al. used RNN-based generated models to explore the chemical space near and far from the explored space to develop potential kinase inhibitors. However, the neighbor exploited space indicates a lack of innovation. They used RNN-based generative models and virtual screening to solve this challenge (Li et al., 2020). RNNs were used to design fragment-based drug discovery.

IV **Markov Decision Process (MDP)**: It is a decision-making modeling framework in which the outcome is partly random and partially reliant on the decision maker's input in some scenarios. Optimal planning is another instance where MDP is applied. MDP's primary purpose is to find a policy for the decision maker that specifies what specific action should be taken at when the point in time. The following components comprise an MDP model: A collection of possible states, such as the states of a door (open or closed). A set of possible actions: a predetermined set of actions, such as closing or opening a door. Transition probabilities: What is the likelihood that the door will be locked after the step of closing the door has been completed? Rewards - are utilized to guide the planning process (Givan and Parr, 2001). Egbhali-Zarch et al. utilized the Markov decision model for decision-making to select the type-2 diabetic patients effective

treatment by considering adverse drug reaction effects of the medication (Eghbali-Zarch et al., 2019).

V **Natural Language Processing**: Natural Language Processing (NLP) encompasses various tasks, from speech detection to language production, which necessitate diverse approaches. Part-of-Speech tagging, Named Entity Recognition, and Parsing are three of the fundamental approaches. NLP methodology facilitates the processing and exploitation of biochemical text, while enabling a deeper understanding of biochemical language to illuminate bimolecular recognition principles. Biochemical and biological knowledge is being enhanced through natural language processing technologies to accelerate drug discovery for improving human health (Öztürk et al., 2020).

1.3. 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 human 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 insights in data mining. Machine learning is categorized into four groups based on the methodologies and methods of learning: Supervised, Unsupervised, Semi- supervised and Reinforcement learning. The AI algorithms significant tasks in an extensive dataset include classification, regression, grouping, and pattern recognition (Carracedo-Reboredo et al., 2021). 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 (Dara et al., 2021). As seen by 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 (Klambauer et al., 2019).

A **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 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 of the examples of regression algorithms are Simple linear regression algorithm, Multivariate regression algorithm, Decision tree algorithm and Lasso regression.

B **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 the similarities, patterns, and differences**.** Supervised machine learning is further classified into two categories:

Clustering: This algorithm is used to obtain the data's inherent groups. It's a method of grouping things into a cluster so that the objects with the most similarities stay in one group while the objects from other groups have fewer or no similarities. Some examples of clustering algorithms are K- Means Clustering algorithm, Mean-shift algorithm, DBSCAN Algorithm, Principal Component Analysis and Independent Component Analysis.

i **Association:** This is an unsupervised learning method for uncovering interesting relationships between variables in a large dataset. The primary goal of this learning method is to determine which data items depend on which other items and map those variables accordingly to maximize profit. Some examples of Association rule learning are Apriori Algorithm, Eclat, and FP-growth algorithm.

A **Semi-Supervised Machine learning:** This machine learning algorithm falls between supervised and unsupervised learning. It uses a combination of labelled and unlabeled datasets throughout the training phase. It constitutes a middle ground between supervised (with labelled training data) and unsupervised (with no labelled training data) learning processes. Semi- supervised learning's fundamental goal is to make good use of all accessible data rather than just labelled data like supervised learning does. Similar data is first clustered using an unsupervised learning technique, which aids in labeling unlabeled data into labelled data. It's because acquiring labelled data is more expensive than acquiring unlabeled data.

B **Reinforcement Learning:** Training machine learning models to make a series of judgments is known as reinforcement learning. In an uncertain, potentially complex environment, the agent learns to achieve a goal. Artificial intelligence meets a game-like circumstance in reinforcement learning. The computer uses trial and error to find a solution to the problem. Artificial intelligence is given either rewards or penalties for the acts it takes to get it to accomplish what the programmer desires. Its purpose is to increase the total prize as much as possible. Reinforcement learning is categorized into two types:

i **Positive reinforcement learning** indicates adding something to increase the likelihood of the specified behavior occurring again. It strengthens and influences the agent's behavior.

ii **Negative Reinforcement learning** has the exact opposite effect as positive RL. Avoiding the unfavorable situation enhances the likelihood that the specific behavior will occur again.

2. Applications

Using pseudo amino-acid composition and the fuzzy K-nearest neighbor approach, Xiau et al. (2013) built a 2-level multi-label classifier (iAMP-2 L) to detect and specify the function(s) of an AMP (Xiao et al., 2013). Using a semi-supervised technique and a clustering algorithm, Vishnepolsky et al. (2018) created a model to predict AMP specificity active against certain Gram-negative microorganisms (Vishnepolsky et al., 2018).

The immunological checkpoint indolamine 2,3-dioxygenase (IDO) is a prospective target for cancer immunotherapy. Machine learning approaches have found three IDO inhibitors with strong efficacy (Zhang et al., 2018).

Many CNS problems, including neurodegenerative diseases and trauma, necessitate a combination of treatments to address neuroprotection, cell repair, and regeneration. Through computational techniques such as ML, knowledge acquired in neurodegenerative processes and neuroprotective treatments can be leveraged to identify medication combinations that can be reused as possible neuroprotective medicines (Romeo-Guitart et al., 2018).

Using machine learning algorithms to learn patterns in existing medication-related biological data and link them to specific diseases to be treated is a potential technique for drug repositioning (Zhao and So, 2019; Wang et al., 2019).

In order to repurpose the medications for the treatment of diabetes, Moinul et al., found crucial fingerprints necessary for the inhibition of the SGLT2 (Sodium-glucose Cotransporter-2) protein. They have accelerated the anti-SGLT2 drug discovery process by using ligand-based drug design, particularly fragment-based in silico drug design. The outcomes of investigations using Bayesian and recursive partitioning are very helpful in understanding the mechanical facets of the fingerprints that supported inhibitory mechanisms. Additionally, they developed statistically supported machine learning models to filter potential hits

from a collection of FDA-approved medications (Moinul et al., 2022).

A vast number of natural compounds were screened using the combined ML/FPL (Machine Learning/False Pulling Ligand) method in order to find acetylcholinesterase (AChE) inhibitors to treat Alzheimer's disease. With the help of molecular docking, the binding affinities of the most potent lead compounds were predicted. Atomistic simulations were then run to further confirm the ML and docking calculations (Thai et al., 2022).

2.1. Deep learning

Machine learning has a branch called deep learning. It's a data abstraction algorithm that employs numerous processing layers of complex structures or multiple non-linear transformations to abstract data (Lipinski et al., 2019). Deep learning neural networks, also known as artificial neural networks, use a combination of data inputs, weights, and biases to imitate the human brain. These pieces work together to recognize, classify, and characterize items in the data accurately. There are two types of deep learning techniques: Supervised deep learning and Unsupervised deep learning. Deep learning refers to machine learning techniques employing numerous layers to extract higher- level features from raw data (Schmidhuber, 2015). It is based on Artificial neural networks (ANNs) inspired by the human brain, which can communicate and distribute information to several parts of the body with the help of neurons (Marblestone et al., 2016). Regulators reject a significant proportion of potential drugs. Insufficient efficacy (on-target effect), unwanted interactions (off-target effects), or unexpected hazardous consequences are several reasons for failure. Deep learning has been used to anticipate biomolecular targets, off-targets, and hazardous consequences of environmental chemicals in foods, home items, and pharmaceuticals (Dahl et al., 2014).

The difference between supervised deep learning and concept learning in humans and animals is that the student in the former case is a computational network. Well- labelled data is used to train supervised deep learning frameworks (Table 14). It teaches the learning algorithm to generalize from the training data and to implement in unseen situations. Some the examples of supervised deep learning algorithms are Artificial neural networks (ANN), Convolutional neural networks (CNN) and Recurrent neural networks (RNN).

i **Artificial neural network (ANN):** A computational network that functions similarly to the human central nervous system. The neural network learns by comparing its output to a known value after training on a dataset. Error is the difference between these two values. By altering the weightage and bias of connections, a good neural network seeks to reduce error.

ii **Convolutional neural network (CNN/ConvNet):** This deep neural network is used to evaluate visual imagery in deep learning. It employs a technique known as Convolution. Convolution is a mathematical operation on two functions that results in a third function that expresses how the shape of the other changes the form of one. Multiple layers of artificial neurons make up convolutional neural networks. Artificial neurons are mathematical functions that calculate the weighted sum of various inputs and output an activation value, similar to their biological counterparts.

iii **Recurrent neural network (RNN):** This is a supervised deep learning method that considers the output of the previous step as an input for the present stage. It has a hidden state that contains the information to memorize a sequence. RNNs can transform independent inputs into dependent inputs, simplifying the process of increasing parameters and remembering each prior output by feeding each production into the next hidden layer.

3. Applications

- Stokes et al. 2020 identified Halicin and eight other potential antibiotics from the ZINC database by training a deep learning network to identify compounds active against antibiotic-resistant bacteria such as Carbapenemase-producing *Enterobacterales, Mycobacterium tuberculosis, Acinetobacter baumannii, and Clostridioides difficile* (Stokes et al., 2020).
- Zhang et al. published a paper on drug repurposing using deep learning (2020). Chemical sequences [simplified molecular-input line-entry system (SMILES) strings] and amino acid (AA) sequences were used as input in this work, which used a deep learning-based drug-target interaction model called Molecule Transformer-Drug

Table 14
Deep Learning Tools used in drug discovery.

S. NO	NAME OF TOOL	DESCRIPTION	Website Link
1	Tensorflow	It particularly focuses on artificial neural networks and interference of deep neural networks.	https://www.tensorflow.org
2	Pytorch	A machine learning framework that accelerates the path from research prototyping to production deployment.	https://pytorch.org/
3	Scikit learn	It is the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.	https://scikit-learn.org/
4	MXNet	It is a scalable deep learning framework that supports deep learning models, such as; convolutional neural networks (CNNs) and long short-term memory networks (LSTMs).	https://mxnet.apache.org/versions/1.9.0/
5	Gluon	It is a deep learning library jointly created by AWS and Microsoft that helps developers build, train and deploy machine learning models in the cloud.	https://auto.gluon.ai/
8	Deep docking	a novel deep learning platform that is suitable for docking billions of molecular structures in a rapid, yet accurate fashion	https://github.com/jamesgleave/DD_protocol
9	Deep Chem	For compound identification, an open-source Python library employs a deep learning method.	https://github.com/deepchem/deepchem
10	DeepTox	Using a deep learning algorithm, predict the toxicity of chemical substances.	www.bioinf.jku.at/research/DeepTox
11	DeepNeural Net QSAR	Using a hierarchical deep neural network, predict molecular activity (DNN)	https://github.com/Merck/DeepNeuralNet-QSAR
12	PotentialNet	Using a graph convolutional neural network, predict binding affinity (CNN)	https://pubs.acs.org/doi/full/10.1021/acscentsci.8b00507
13	Conv_qsar_fast	CNN approach to predict molecular characteristics	https://github.com/connorcoley/conv_qsar_fast
14	Neural graph fingerprint	Using CNN, predict the properties of new compounds.	https://github.com/HIPS/neural-fingerprint
15	PADME	utilizes feed-forward neural networks for predicting drug target interactions	https://www.padme.ai/
16	Tox_(R)CNN	a deep CVNN method evaluated the cytotoxicity of drugs	https://github.com/nielintos/Tox-CNN

Target Interaction (MT-DTI). This research led to the discovery of several commercially available antiviral medicines that may interact with the SARS-CoV-2 proteins (Zhang et al., 2020; Beck et al., 2020).

- In prediction investigations utilizing DL models, pharmacokinetic and toxicity factors such as water solubility (Lusci et al. 2013) (Qiao et al., 2016) and particular toxicities (Unterthiner et al. 2015; Xu et al. 2015; Capuzzi et al. 2016) were also discovered as target-properties (Lusci et al., 2013; Unterthiner et al., 2015; Capuzzi et al., 2016).
- Tsubaki et al. (2018) proposed a method for predicting protein-ligand interactions that uses protein sequences and molecular fingerprints of ligands (from convolutional neural networks and graph neural networks, respectively) as vector input. The proposed method outperformed some techniques, including kNN, random forest, logistic regression, and SVM. Finally, the authors found that the vectors generated by the models accurately identified key amino acid residues at the binding site involved in drug-target interactions (Tsubaki et al., 2018).

4. Conclusion

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 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 aforementioned CADD techniques are widely accepted as being far from perfect and omnipotent in all cases. For current computational approaches to be employed effectively, considerable constraints must be overcome. Artificial intelligence, machine learning, and deep learning approaches were used with fundamental 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, we have explained in detail how computational tools and techniques have been applied to drug discovery and development so far. We have also described their success stories and the list of current tools and software used in the drug discovery and development process. Consequently, researchers and students can use the above information in their research to design novel drugs against specific targets.

5. 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 the most significant challenge when using computational approaches. Therefore, the problem in this situation 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.

Data availability

No data was used for the research described in the article.

References

Abagyan, R., Totrov, M., Kuznetsov, D., 1994. ICM—A new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. J. Comput. Chem. 15, 488–506. https://doi.org/10.1002/jcc.540150503.

Abidi, A., Shukla, P., Ahmad, A., 2016. Lifitegrast: a novel drug for treatment of dry eye disease. J. Pharmacol. Pharmacotherapeutics 7 (4), 194–198. https://doi.org/10.4103/0976-500X.195920.

Acharya, C., Coop, A., E Polli, J., D MacKerell, A., 2011. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Curr. Comput. Aided Drug Des. 7 (1), 10–22. https://doi.org/10.2174/157340911793743547.

Agrawal, R., 2014. The first approved agent in the Glitazar's class: saroglitazar. Curr. Drug Targets 15 (2), 151–155.

Alessandri, R., Barnoud, J., Gertsen, A.S., Patmanidis, I., de Vries, A.H., Souza, P.C., Marrink, S.J., 2022. Martini 3 Coarse-Grained Force Field: small Molecules. Adv. Theory Simul. 5 (1), 2100391 https://doi.org/10.1002/adts.202100391.

An, J, Totrov, M, Abagyan, R, 2005. Pocketome via comprehensive identification and classification of ligand binding envelopes. Mol. Cell Proteomics 4 (6), 752–761. https://doi.org/10.1074/mcp.M400159-MCP200.

Aparoy, P., Kumar Reddy, K., Reddanna, P., 2012. Structure and ligand based drug design strategies in the development of novel 5-LOX inhibitors. Curr. Med. Chem. 19 (22), 3763–3778. https://doi.org/10.2174/092986712801661112.

Apostolakis, J., Pluckthun, A., Caflisch, A., 1998. Docking small ligands in flexible binding sites. J. Comput. Chem. 19, 21–37. https://doi.org/10.1002/(SICI)1096-987X(19980115)19. :1%3C21::AID-JCC2%3E3.0.CO;2-0.

Aybars, O., 2021. Using neural nets to recognize handwritten digits (Doctoral dissertation. Zakład Systemów Przetwarzania Informacji).

Barnoud, J., Monticelli, L., 2015. Coarse-grained force fields for molecular simulations. Molecular Modeling of Proteins. Humana Press, New York, NY, pp. 125–149.

Bates, P.A., Kelley, L.A., MacCallum, R.M., Sternberg, M.J., 2001. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins: Struct. Funct. Genet 45 (S5), 39–46. https://doi.org/10.1002/prot.1168.

Batool, M., Ahmad, B., Choi, S., 2019. A structure-based drug discovery paradigm. Int. J. Mol. Sci. 20 (11), 2783. https://doi.org/10.3390/ijms20112783.

Baxter, C.A., Murray, C.W., Clark, D.E., Westhead, D.R., Eldridge, M.D., 1998. Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins 33, 367–382. https://doi.org/10.1002/(SICI)1097-0134(19981115)33, 3%3C367::AID-PROT6%3E3.0.CO;2-W.

Beck, B.R., Shin, B., Choi, Y., Park, S., Kang, K, 2020. A drug-target interaction deep learning model predicts commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2). Comput. Struct. Biotec. 18, 784–790. https://doi.org/10.1016/j.csbj.2020.03.025.

Becker, A., 2019. Artificial intelligence in medicine: what is it doing for us today? Health Policy Technol 8 (2), 198–205. https://doi.org/10.1177/1747493018778713.

Bergonzo, C., Cheatham III, T.E., 2015. Improved force field parameters lead to a better description of RNA structure. J. of Chem. Theory Comput. 11, 3969–3972. https://doi.org/10.1021/acs.jctc.5b00444.

Bowie, J.U., Lüthy, R., Eisenberg, D., 1991. A method to identify protein sequences that fold into a known three-dimensional structure. sci. 253 (5016), 164–170. https://doi.org/10.1126/science.1853201.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M., 1983. CHARMM – a programm for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4, 187–217. https://doi.org/10.1002/jcc.540040211.

Capuzzi, S.J., Politi, R., Isayev, O., Farag, S., Tropsha, A., 2016. QSAR modeling of Tox21 challenge stress response and nuclear receptor signaling toxicity assays. Front. Environ. Sci. 4, 3. https://doi.org/10.3389/fenvs.2016.00003.

Carlson, H.A., 2002. Protein flexibility is an important component of structure-based drug discovery. Curr. Pharm. Des. 8, 1571–1578. https://doi.org/10.2174/1381612023394232.

Carlson, H.A., McCammon, J.A., 2000. Accommodating protein flexibility in computational drug design. Mol. Pharmacol. 57, 213–218.

Carracedo-Reboredo, P., Liñares-Blanco, J., Rodríguez-Fernández, N., Cedrón, F., Novoa, F.J., Carballal, A., et al., 2021. review on machine learning approaches and trends in drug discovery. Comput. Struct. Biotechnol. J. 19, 4538–4558. https://doi.org/10.1016/j.csbj.2021.08.011.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., 1995. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids,and Organic Molecules. J. Am. Chem. Soc. 117, 5179–5197. https://doi.org/10.1021/ja00124a002.

Cozzini, P., Kellogg, G.E., Spyrakis, F., Abraham, D.J., Costantino, G., Emerson, A., Fanelli, F., Gohlke, H., Kuhn, L.A., Morris, G.M., et al., 2008. Target flexibility: an emerging consideration in drug discovery and design. J. Med. Chem. 51, 6237–6255. https://doi.org/10.1021/jm800562d.

Dahl, G.E., Jaitly, N. and Salakhutdinov, R, Multi-task neural networks for QSAR predictions, *arXiv preprint arXiv* (2014) 1406–1231. 10.3389/fncom.2016.00094.

Dara, S., Dhamercherla, S., Jadav, S.S., Babu, C.H., Ahsan, M.J., 2021. Machine Learning in Drug Discovery: a Review. Artif Intell Rev 1–53. https://doi.org/10.1007/s10462-021-10058-4.

Davis, I.W., Baker, D., 2009. ROSETTALIGAND docking with full ligand and receptor flxibility. J. Mol. Biol. 385, 381–392. https://doi.org/10.1016/j.jmb.2008.11.010.

Denning, E.J., Priyakumar, U.D., Nilsson, L., MacKerell Jr., A.D., 2011. Impact of 2′-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. J. Comput. Chem. 32, 1929–1943. https://doi.org/10.1002/jcc.21777.

Desmet, J., Wilson, I.A., Joniau, M., De Maeyer, M., Lasters, I., 1997. Computation of the binding of fully flexible peptides to proteins with flexible side-chains. FASEB J 11, 164–172. https://doi.org/10.1096/fasebj.11.2.9039959.

Dodda, L.S., Cabeza de Vaca, I., Tirado-Rives, J., Jorgensen, W.L., 2017. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. Nucleic Acids Res 45 (W1), W331–W336. https://doi.org/10.1093/nar/gkx312.

Eghbali-Zarch, M., Tavakkoli-Moghaddam, R., Esfahanian, F., Azaron, A., Sepehri, M.M., 2019. A Markov decision process for modeling adverse drug reactions in medication treatment of type 2 diabetes. Proceedings of the Institution of Mechanical Engineers, Part H: IEEE Eng Med Biol Mag 233 (8), 793–811. https://doi.org/10.1177/0954411919853394.

Ewing, T.J., Makino, S., Skillman, A.G., Kuntz, I.D., 2001. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J. Comput. Aided Mol. Des. 15, 411–428. https://doi.org/10.1023/a:1011115820450.

Feller, S.E., Yin, D., Pastor, R.W., MacKerell Jr, A.D., 1997. Molecular dynamics simulation of unsaturated lipid bilayers at low hydration: parameterization and comparison with diffraction studies. Biophys. J. 73 (5), 2269–2279. https://doi.org/10.1016/S0006-3495(97)78259-6.

Ferrari, A.M., Wei, B.Q., Costantino, L., Shoichet, B.K., 2004. Soft docking and multiple receptor conformations in virtual screening. J. Med. Chem. 47 (21), 5076–5084. https://doi.org/10.1021/jm049756p.

Foloppe, N., MacKerell Jr., A.D., 2011. All-Atom Empirical Force Field for Nucleic Acids: 1) Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. J. Comput. Chem. 86–104. https://doi.org/10.1002/(SICI)1096-987X(20000130)21. :2<86::AID−JCC2>3.0.CO;2-G.

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shaw, D.E., Shelley, M., et al., 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 47, 1739–1749. https://doi.org/10.1021/jm0306430.

Frimurer, T.M., Peters, G.H., Iversen, L.F., Andersen, H.S., Moller, N.P., Olsen, O.H., 2003. Ligand-induced conformational changes: improved predictions of ligand binding conformations and affinities. Biophys. J. 84, 2273–2281. https://doi.org/10.1016/S0006-3495(03)75033-4.

Galindo-Murillo, R., Robertson, J.C., Zgarbovic, M., Sponer, J., Otyepka, M., Jureska, P., Cheatham, T.E., 2016. Assessing the Current State of Amber Force Field Modifications for DNA. J. Chem. Theory Comput. 12, 4114–4127. https://doi.org/10.1021/acs.jctc.6b00186.

Gangwal, R.P., Damre, M.V., Das, N.R., Dhoke, G.V., Bhadauriya, A., Varikoti, R.A., et al., 2015. Structure-based virtual screening to identify selective phosphodiesterase 4B inhibitors. J. Mol. Graph. Model. 57, 89–98. https://doi.org/10.1016/j.jmgm.2015.01.007.

Gao, Q., Yang, L., Zhu, Y., 2010. Pharmacophore based drug design approach as a practical process in drug discovery. Curr Comput Aided Drug Des 6 (1), 37–49. https://doi.org/10.2174/157340910790980151.

Gautieri, A., Russo, A., Vesentini, S., Redaelli, A., Buehler, M.J., 2010. Coarse-grained model of collagen molecules using an extended MARTINI force field. J. Chem. Theory Comput. 6 (4), 1210–1218. https://doi.org/10.1021/ct100015v.

Geppert, H., Vogt, M., Bajorath, J., 2010. Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. J. Chem. Inf. Model. 50 (2), 205–216. https://doi.org/10.1021/ci900419k.

Gerstmeier, J., Seegers, J., Witt, F., Waltenberger, B., Temml, V., Rollinger, J.M., et al., 2019. Ginkgolic acid is a multi-target inhibitor of key enzymes in pro-inflammatory lipid mediator biosynthesis. Front. Pharmacol. 10, 797. https://doi.org/10.3389/fphar.2019.00797.

Gertrudes, JC, Maltarollo, VG, Silva, RA, et al., 2012. Machine learning techniques and drug design. Curr Med Chem. 19, 4289–4297. https://doi.org/10.2174/092986712802884259.

Givan, B., Parr, R., 2001. An Introduction to Markov decision Processes. Purdue University.

Grosdidier, A., Zoete, V., Michielin, O., 2007. EADock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization. Proteins 67, 1010–1025. https://doi.org/10.1002/prot.21367.

Gurung, A.B., Ali, M.A., Lee, J., Farah, M.A., Al-Anazi, K.M., 2021. An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19. BioMed Res. Int. https://doi.org/10.1155/2021/8853056.

Hamza, A., Wei, N.N., Zhan, C.G., 2012. Ligand-based virtual screening approach using a new scoring function. J. Chem. Inf. Model 52 (4), 963–974. https://doi.org/10.1021/ci200617d.

Hansson, T., Marelius, J., Åqvist, J., 1998. Ligand binding affinity prediction by linear interaction energy methods. J. Comput. Aided Mol. Des. 12 (1), 27–35. https://doi.org/10.1023/a:1007930623000.

Harper, S., McCauley, J.A., Rudd, M.T., Ferrara, M., DiFilippo, M., Crescenzi, B., Koch, U., Petrocchi, A., Holloway, M.K., Butcher, J.W., Romano, J.J., 2012. Discovery of MK-5172, a macrocyclic hepatitis C virus NS3/4a protease inhibitor. ACS Med. Chem. Lett. 3 (4), 332–336. https://doi.org/10.1021/ml300017p.

Hart, K., Foloppe, N., Baker, C.M., Denning, E.J., Nilsson, L., MacKerell Jr., A.D., 2012. Optimization of the CHARMM additive force field for DNA: improved treatment of the BI/BII conformational equilibrium. J. Chem. Theory Comput. 8, 348–362. https://doi.org/10.1021/ct200723y.

Hecker, S.J., Reddy, K.R., Totrov, M., Hirst, G.C., Lomovskaya, O., Griffith, D.C., King, P., Tsivkovski, R., Sun, D., Sabet, M. and Tarazi, Z., Discovery of a cyclic boronic acid β-lactamase inhibitor (RPX7009) with utility vs class A serine carbapenemases (2015). DOI: 10.1021/acs.jmedchem.5b00127.

Hosfield, D., Palan, J., Hilgers, M., Scheibe, D., McRee, D.E., Stevens, R.C, 2003. A fully integrated protein crystallization platform for small-molecule drug discovery. J. Struct. Biol. 142 (1), 207–217. https://doi.org/10.1016/s1047-8477(03)00051-0.

Hou, T, Wang, J, Li, Y, Wang, W., 2010. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J. Chem. Inf. Model. 51, 69. https://doi.org/10.1021/ci100275a.

Huang, B, 2009. MetaPocket: a meta-approach to improve protein ligand binding site prediction. OMICS 13 (4), 325–330. https://doi.org/10.1089/omi.2009.0045.

Huang, N., Kalyanaraman, C., Irwin, J.J., Jacobson, M.P., 2006. Molecular mechanics methods for predicting protein-ligand binding. J. Chem. Inf. Model. 46, 243–253. https://doi.org/10.1021/ci0502855.

Huang, S.Y., Zou, X., 2010. Advances and challenges in protein-ligand docking. Int. J. Mol. Sci. 11 (8), 3016–3034. https://doi.org/10.3390/ijms11083016.

Humphrey, W., Dalke, A., Schulten, K, 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

Jain, A.N., 2003. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J. Med. Chem. 46, 499–511. https://doi.org/10.1021/jm020406h.

Jayaraj, P.B., Jain, S., 2019. Ligand based virtual screening using SVM on GPU. Comput. Biol. Chem. 83, 107143 https://doi.org/10.1016/j.compbiolchem.2019.107143.

Jiang, F., Kim, S.H., 1991. Soft docking": matching of molecular surface cubes. J. Mol. Biol. 219 (1), 79–102. https://doi.org/10.1016/0022-2836(91)90859-5.

Jones, D.T, 1999. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J. Mol. Biol. 287 (4), 797–815. https://doi.org/10.1006/jmbi.1999.2583.

Jones, D.T., Taylort, W.R., Thornton, J.M., 1992. A new approach to protein fold recognition. Nature 358 (6381), 86–89. https://doi.org/10.1038/358086a0.

Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R., 1997. Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 267, 727–748. https://doi.org/10.1006/jmbi.1996.0897.

Jorgensen, W.L, 2004. The many roles of computation in drug discovery. Science 303 (5665), 1813–1818. https://doi.org/10.1126/science.1096361.

Jorgensen, W.L., Tirado-Rives, J., 1988. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. 110 (6), 1657–1666. https://doi.org/10.1021/ja00214a001.

Kalimeri, M., Derreumaux, P., Sterpone, F., 2015. Are coarse-grained models apt to detect protein thermal stability? The case of OPEP force field. J. Non-Cryst. Solids 407, 494–501. https://doi.org/10.1016/j.jnoncrysol.2014.07.005.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., et al., 2012. Template- based protein structure modeling using the RaptorX web server. Nat. Protoc. 7 (8), 1511–1522. https://doi.org/10.1038/nprot.2012.085.

Kawato, T., Mizohata, E., Shimizu, Y., Meshizuka, T., Yamamoto, T., Takasu, N., et al., 2015. Structure-based design of a streptavidin mutant specific for an artificial biotin analogue. J. Biochem. 157 (6), 467–475. https://doi.org/10.1093/jb/mvv004.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10 (6), 845–858. https://doi.org/10.1038/nprot.2015.053.

Kelley, L.A., Sternberg, M.J, 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4 (3), 363–371. https://doi.org/10.1038/nprot.2009.2.

Kirschner, K.N., Yongye, A.B., Tschampel, S.M., González-Outeiriño, J., Daniels, C.R., Foley, B.L., Woods, R.J., 2008. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J. Comput. Chem. 29, 622–655. https://doi.org/10.1002/jcc.20820.

Klambauer, G., Hochreiter, S., Rarey, M., 2019. Machine learning in drug discovery. J. Chem. Inf. Model. 59 (3), 945–946. https://doi.org/10.1021/acs.jcim.9b00136.

Klocker, J., Wailzer, B., Buchbauer, G., Wolschann, P., 2002. Bayesian neural networks for aroma classification. J. chem. inf. comput. sci. 42 (6), 1443–1449. https://doi.org/10.1021/ci0202640.

Kodadek, T., 2011. The rise, fall and reinvention of combinatorial chemistry. Chem. Commun. 47, 9757–9763. https://doi.org/10.1039/c1cc12102b.

Kony, D., Damm, W., Stoll, S., Van Gunsteren, W.F., 2002. An improved OPLS–AA force field for carbohydrates. J. Comput. Chem. 23 (15), 1416–1429. https://doi.org/10.1002/jcc.10139.

Koppensteiner, W.A., Sippl, M.J., 1998. Knowledge-based potentials – Back to the roots. Biochemistry Mosc 63, 247–252.

Laurent, B., Chavent, M., Cragnolini, T., Dahl, A.C., Pasquali, S., Derreumaux, P., et al., 2015. Epock: rapid analysis of protein pocket dynamics. Bioinformatics, 31 (9), 1478–1480. https://doi.org/10.1093/bioinformatics/btu822.

Lavecchia, A., Di, Giovanni, 2013. C, Virtual screening strategies in drug discovery: a critical review. Curr. Med. Chem. 20 (23), 2839–2860. https://doi.org/10.2174/09298673113209990001.

Leach, A.R., 1994. Ligand docking to proteins with discrete side-chain flexibility. J. Mol. Biol. 235, 345–356. https://doi.org/10.1016/s0022-2836(05)80038-5.

Lee, J., Freddolino, P.L. and Zhang, Y. Ab initio protein structure prediction. In *From protein structure to function with bioinformatics* (3-35). Springer, Dordrecht. (2017).

Lemer, C.M.R., Rooman, M.J., Wodak, S.J., 1995. Protein structure prediction by threading methods: evaluation of current techniques. Proteins: Struct. Funct. Genet. 23 (3), 337–355. https://doi.org/10.1002/prot.340230308.

León, R., Soto-Delgado, J., Montero, E., Vargas, M., 2021. Development of Computational Approaches with a Fragment-Based Drug Design Strategy: *in Silico* Hsp90 Inhibitors Discovery. Int. J. Mol. Sci. 22 (24), 13226. https://doi.org/10.3390/ijms222413226.

Lesk, A.M., Chothia, C., 1980. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. J. Mol. Biol. 136 (3), 225–270. https://doi.org/10.1016/0022-2836(80)90373-3.
Li, X., Xu, Y., Yao, H., Lin, K., 2020. Chemical space exploration based on recurrent neural networks: applications in discovering kinase inhibitors. J. Cheminformatics 12 (1), 1–13. https://doi.org/10.1186/s13321-020-00446-3.
Liang, J, Woodward, C, Edelsbrunner, H, 1998. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci 7 (9), 1884–1897. https://doi.org/10.1002/prot.24065.
Lill, M.A., Danielson, M.L., 2011. Computer-aided drug design platform using PyMOL. J. Comput. Aided Mol. Des. 25, 13–19. https://doi.org/10.1007/s10822-010-9395-8.
Lionta, E., Spyrou, G., Vassilatis, D.K., Cournia, Z., 2014. Structure-based virtual screening for drug discovery: principles, applications and recent advances. Curr. Top. Med. Chem. 14, 1923–1938. https://doi.org/10.2174/1568026614666140929124445.
Lipinski, C.F., Maltarollo, V.G., Oliveira, P.R., Da Silva, A.B., Honorio, K.M., 2019. Advances and perspectives in applying deep learning for drug design and discovery. Front. Robot. AI 108. https://doi.org/10.3389/frobt.2019.00108.
Liu, J., Wang, R., 2015. Classification of current scoring functions. J. Chem. Inf. Model 55 (3), 475–482. https://doi.org/10.1021/ci500731a.
Liu, S., Chen, K., Chen, Q., Tang, B., 2016. Dependency-based convolutional neural network for drug-drug interaction extraction. In: the 2016 Proc. IEEE Int. Conf. Bioinform. Biomed. Workshops (BIBM). https://doi.org/10.1109/BIBM.2016.7822671, 1074 1080.
Liu, W, Schmidt, B, Voss, G, Müller-Wittig, W., 2008. Accelerating molecular dynamics simulations using graphics processing units with CUDA. Comput. Phys. Commun 179, 634–641. https://doi.org/10.1016/j.cpc.2008.05.008.
Liwo, A, Oldziej, S, Pincus, MR, Wawak, RJ, Rackovsky, S, Scheraga, HA, 1997. A unitedresidue force fi eld for off-lattice proteinstructure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. J. Comput. Chem. 18, 849–873. https://doi.org/10.1002/(SICI)1096-987X(199705)18. :7<849::AID−JCC1>3.0.CO;2-R.
Lusci, A., Pollastri, G., Baldi, P, 2013. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. J. Chem. Inf. Model 53, 1563–1575. https://doi.org/10.1021/ci400187y.
Lyne, P.D., 2002. Structure-based virtual screening an overview. Drug Discov. Today 7, 649–657. https://doi.org/10.1016/s1359-6446(02)02483-2.
Maciejewski, A., Pasenkiewicz-Gierula, M., Cramariuc, O., Vattulainen, I., Rog, T., 2014. Refined OPLS all-atom force field for saturated phosphatidylcholine bilayers at full hydration. The J. Phys. Chem. B 118 (17), 4571–4581. https://doi.org/10.1021/jp5016627.
MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 102, 3586–3616. https://doi.org/10.1021/jp973084f.
Mak, K.-K., Pichika, M.R, 2019. Artificial intelligence in drug development: present status and future prospects. Drug Discov. Today. 24, 773–780. https://doi.org/10.1177/1747493018778713.
Mandlik, V., Bejugam, P.R. and Singh, S., Application of artificial neural networks in modern drug discovery, In *Artificial Neural Network for Drug Design, Delivery and Disposition* (2016) 123–139 Academic Press. 10.1016/B978-0-12-801559-9.
Marblestone, A.H., Wayne, G., Kording, K, 2016. Toward an integration of deep learning and neuroscience. Front Comput Neurosci 94. https://doi.org/10.3389/fncom.2016.00094.
Marzuoli, I., Margreitter, C., Fraternali, F., 2019. Lipid head group parameterization for GROMOS 54A8: a consistent approach with protein force field description. J. Chem. Theory Comput. 15 (10), 5175–5193. https://doi.org/10.1021/acs.jctc.9b00509.
McGann, M., 2012. FRED and HYBRID docking performance on standardized datasets. J. Comput. Aided Mol. Des. 26, 897–906. https://doi.org/10.1007/s10822-012-9584-8.
Meng, E.C., Shoichet, B.K., Kuntz, I.D., 1992. Automated docking with grid-based energy approach to macromolecule-ligand interactions. J. Comput. Chem. 13, 505–524.
Middleton, D.A., 2007. Solid-state NMR spectroscopy as a tool for drug design: from membrane-embedded targets to amyloid fibrils. Biochem. Soc. Trans. 35 (5), 985–990. https://doi.org/10.1042/BST0350985.
Miller, M.D., Kearsley, S.K., Underwood, D.J., Sheridan, R.P., 1994. FLOG: a system to select "quasi-flexible" ligands complementary to a receptor of known three-dimensional structure. J. Comput. Aided Mol. Des. 8, 153–174. https://doi.org/10.1007/BF00119865.
Miyazawa, S., Jernigan, R.L., 1985. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 18, 534–552.
Mizuguchi, K, 2004. Fold recognition for drug discovery. Drug Discov. Today: *Targets* 3 (1), 18–23. https://doi.org/10.1016/S1741-8372(04)02392-8.
Mizuguchi, K., 2004. Fold recognition for drug discovery. Drug Discov. Today: *Targets* 3 (1), 18–23. https://doi.org/10.1016/S1741-8372(04)02392-8.
Mizutani, M.Y., Tomioka, N., Itai, A., 1994. Rational automatic search method for stable docking models of protein and ligand. J. Mol. Biol. 243, 310–326. https://doi.org/10.1006/jmbi.1994.1656.
Moinul, M., Amin, S.A., Kumar, P., Patil, U.K., Gajbhiye, A., Jha, T., Gayen, S., 2022. Exploring sodium glucose cotransporter (SGLT2) inhibitors with machine learning approach:a novel hope in anti-diabetes drug discovery. J. Mol. Graph. Model. 111, 108106 https://doi.org/10.1016/j.jmgm.2021.108106.
Moll, A., Hildebrandt, A., Lenhof, H.P., Kohlbacher, O., 2006. BALLView: a tool for research and education in molecular modeling. Bioinformatics 22, 365–366. https://doi.org/10.1093/bioinformatics/bti818.
Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., et al., 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. 19, 1639–1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19. :14%3C1639::AID-JCC10%3E3.0.CO;2-B.
Morris, G.M., Goodsell, D.S., Huey, R., Olson, A.J., 1996. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. J. Comput. Aided Mol. Des. 10, 293–304. https://doi.org/10.1007/BF00124499.
Muegge, I, Bergner, A, Kriegl, J.M, 2017. Computer-aided drug design at Boehringer Ingelheim. J. Comput. Aided Mol. Des. 31 (3), 275–285. https://doi.org/10.1007/s10822- 016-9975-3.
Muhammed, M.T., Aki-Yalcin, E, 2019. Homology modeling in drug discovery: overview, current applications, and future perspectives. Chem Biol Drug Des 93 (1), 12–20. https://doi.org/10.1111/cbdd.13388.
Murray, C.W., Newell, D.R., Angibaud, P., 2019. A successful collaboration between academia, biotech and pharma led to discovery of erdafitinib, a selective FGFR inhibitor recently approved by the FDA. Medchemcomm 10 (9), 1509–1511. https://doi.org/10.1039/C9MD90044F.
Nekoei, M, Mohammad Hosseini, M, Pourbasheer, E, 2015. QSAR study of VEGFR-2 inhibitors by using genetic algorithm-multiple linear regressions (GA-MLR) and genetic algorithm-support vector machine (GA-SVM): a comparative approach. Med. Chem. Res 24, 3037–3046. https://doi.org/10.1007/s00044-015-1354-4.
Nilsson, L., Karplus, M., 1986. Empirical energy functions for energy minimization and dynamics of nucleic acids. J. Comput. Chem. 7, 591–616.
Olayan, R.S., Ashoor, H., Bajic, V.B., 2018. DDR: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. Bioinformatics 34 (7), 1164–1173. https://doi.org/10.1093/bioinformatics/btx731.
Oostenbrink, C., Soares, T.A., Van der Vegt, N.F., Van Gunsteren, W.F., 2005. Validation of the 53A6 GROMOS force field. European Biophysics Journal 34 (4), 273–284. https://doi.org/10.1007/s00249-004-0448-6.
Oprea, T.I., 2004. Matter, H, Integrating virtual screening in lead discovery. Curr. Opin. Chem. Biol. 8, 349–358. https://doi.org/10.1016/j.cbpa.2004.06.008.
Osterberg, F., Morris, G.M., Sanner, M.F., Olson, A.J., Goodsell, D.S., 2002. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins 46, 34–40. https://doi.org/10.1002/prot.10028.
Öztürk, H., Özgür, A., Schwaller, P., Laino, T., Ozkirimli, E., 2020. Exploring chemical space using natural language processing methodologies for drug discovery. Drug Discov. Today 25 (4), 689–705. https://doi.org/10.1016/j.drudis.2020.01.020.
Pang, Y.P., Perola, E., Xu, K., Prendergast, F.G., 2001. EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. J. Comput. Chem. 22, 1750–1771. https://doi.org/10.1002/jcc.1129.
Park, H., Chien, P.N., Ryu, S.E., 2012. Discovery of potent inhibitors of receptor protein tyrosine phosphatase sigma through the structure-based virtual screening. Bioorg. Med. Chem. Lett. 22 (20), 6333–6337. https://doi.org/10.1177/1747493018778713.
Pasi, M., Lavery, R., Ceres, .PaLaCe, 2013. A coarse-grain protein model for studying mechanical properties. J. Chem. Theory Comput. 9 (1), 785–793. https://doi.org/10.1021/ct3007925.
Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., Tekade, R.K., 2021. Artificial intelligence in drug discovery and development. Drug Discov. Today 26 (1), 80. https://doi.org/10.1016/j.drudis.2020.10.010.
Peng, J., Xu, J., 2011. A multiple-template approach to protein threading. Proteins 79 (6), 1930–1939. https://doi.org/10.1002/prot.23016.
Peng, J., Xu, J., 2011. RaptorX: exploiting structure information for protein alignment by statistical inference. Proteins: Struct. Funct. Genet 79 (S10), 161–171. https://doi.org/10.1002/prot.23175.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., et al., 2004. UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084.
Polanski, J., 2003. Self-organizing neural networks for pharmacophore mapping. Adv Drug Deliv Rev 55 (9), 1149–1162. https://doi.org/10.1016/s0169-409x(03)00116-9.
Pol-Fachin, L., Rusu, V.H., Verli, H., Lins, R.D., 2012. GROMOS 53A6GLYC, an improved GROMOS force field for hexopyranose-based carbohydratesJ. Chem. Theory Comput. 8 (11), 4681–4690. https://doi.org/10.1021/ct300479h.
Puri, M., Pathak, Y., Sutariya, V.K., Tipparaju, S., Moreno, W., 2015. Artificial Neural Network For Drug design, Delivery and Disposition. Academic Press.
Qiao, L-S, Zhang, X-B, Jiang, L, 2016. Identification of potential ACAT-2 selective inhibitors using pharmacophore, SVM and SVR from Chinese herbs. Mol. Divers 20, 933–944. https://doi.org/10.1007/s11030-016-9684-9.
Qin, Z, Wang, M, Yan, A., 2017. QSAR studies of the bioactivity of hepatitis C virus (HCV) NS3/4A protease inhibitors by multiple linear regression(MLR) and support vector machine (SVM). Bioorg Med Chem Lett. 27 https://doi.org/10.1016/j.bmcl.2017.05.001, 2931–29.
Rao, B.G., Murcko, M., Tebbe, M.J., Kwong, A.D., 2015. Discovery and development of telaprevir (Incivek™)eA protease inhibitor to treat hepatitis C infection. Successful Drug Discovery 1.
Rarey, M., Kramer, B., Lengauer, T., Klebe, G., 1996. A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol. 261, 470–489. https://doi.org/10.1006/jmbi.1996.0477.
Reed, J.E., Smaill, J.B., 2016. The discovery of dacomitinib, a potent irreversible EGFR inhibitor, in: comprehensive Accounts of Pharmaceutical Research and Development: from Discovery to Late-Stage Process Development. J. Am. Chem. Soc. 1, 207–233.

Robertson, M.J., Qian, Y., Robinson, M.C., Tirado-Rives, J., Jorgensen, W.L., 2019. Development and Testing of the OPLS-AA/M Force Field for RNA. J. Chem. Theory Comput. 15 (4), 2734–2742. https://doi.org/10.1021/acs.jctc.9b00054.
Romeo-Guitart, D, Forés, J, Herrando-Grabulosa, M, Valls, R, Leiva-Rodríguez, T, Galea, E, et al., 2018. Neuro protective drug for nerve trauma revealed using artificial intelligence. Sci Rep 8 (1), 1–15. https://doi.org/10.1038/s41598-018-19767-3.
Rupp, M., Schroeter, T., Steri, R., Zettl, H., Proschak, E., Hansen, K., Rau, O., Schwarz, O., Müller-Kuhrt, L., Schubert-Zsilavecz, M., Müller, K.R., 2010. From machine learning to natural product derivatives that selectively activate transcription factor PPARγ. ChemMedChem: Chemistry Enabling Drug Discovery 5 (2), 191–194. https://doi.org/10.1002/cmdc.200900469.
Sahu, A., Patra, P.K., Yadav, M.K., Varma, M., 2017. Identification and characterization of ErbB4 kinase inhibitors for effective breast cancer therapy. J. Recept. Signal Transduct. Res. 37 (5), 470–480. https://doi.org/10.1080/10799893.2017.1342129.
Sayle, R.A., Milner-White, E.J., 1995. RASMOL: biomolecular graphics for all. Trends Biochem. Sci. 20, 374–376. https://doi.org/10.1016/s0968-0004(00)89080-5.
Schaffer, L., Verkhivker, G.M., 1998. Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side-chain optimization. Proteins 33, 295–310. https://doi.org/10.1002/(sici)1097-0134(19981101)33.:2<295::aid− prot12>3.0.co;2-f.
Schlenkrich, M., Brickmann, J., MacKerell Jr., A.D., Karplus, M, 1996. An Empirical Potential Energy Function for Phospholipids: criteria for Parameter Optimization and Applications. In: Merz Jr., K.M., Roux, B. (Eds.), Biological Membranes: A Molecular Perspective from Computation and Experiment. Birkhauser, Boston, pp. 31–81. https://doi.org/10.1007/978-1-4684-8580-6_2.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.
Schnecke, V., Kuhn, L.A., 1999. Database screening for HIV protease ligands: the influence of binding-site conformation and representation on ligand selectivity. Proc. Int. Conf. Intell. Syst. Mol. Biol. 242–251.
Schnecke, V., Kuhn, L.A., 2000. Virtual screening with solvation and ligand-induced complementarity. Perspect. Drug Discov. Des. 20, 171–190. https://doi.org/10.1007/0- 306-46883-2_10.
Schneider, G, 2018. Automating drug discovery. Nat. Rev. Drug Discov. 17, 97–113. https://doi.org/10.1038/nrd.2017.232.
Schwede, T., Kopp, J, Guex, N, Peitsch, MC, 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 31, 3381–3385. https://doi.org/10.1093/nar/gkg520.
Shan, Y, Kim, ET, Eastwood, MP, Dror, RO, Seeliger, MA, et al., 2011. How does a drug molecule find its target binding site? J Am Chem Soc 133, 9181–9183. https://doi.org/10.1021/ja202726y.
Siani, P., de Souza, R.M., Dias, L.G., Itri, R., Khandelia, H., 2016. An overview of molecular dynamics simulations of oxidized lipid systems, with a comparison of ELBA and MARTINI force fields for coarse grained lipid simulations. Biochimica et Biophysica Acta (BBA)-Biomembranes 1858 (10), 2498–2511. https://doi.org/10.1016/j.bbamem.2016.03.031.
Simons, K.T., Kooperberg, C., Huang, E., Baker, D, 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J. Mol. Biol. 268 (1), 209–225. https://doi.org/10.1006/jmbi.1997.0959.
Sippl, M.J., 1990. Calculation of conformational ensembles from potentials of mean force. J. Mol. Biol. 213, 859–883. https://doi.org/10.1016/s0022-2836(05)80269-4.
Sochacka, J., 2014. Docking of thiopurine derivatives to human serum albumin and binding site analysis with Molegro Virtual Docker. Acta Pol. Pharm. 71, 343–349.
Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33 (suppl_2), W244–W248. https://doi.org/10.1093/nar/gki408.
Stahura, F.L., Bajorath, J., 2004. Virtual screening methods that complement HTS. Comb. Chem. High Throughput Screen. 7, 259–269. https://doi.org/10.2174/1386207043328706.
Stank, A., Kokh, D.B., Horn, M., Sizikova, E., Neil, R., Panecka, J., et al., 2017. TRAPP webserver: predicting protein binding site flexibility and detecting transient binding pockets. Nucleic Acids Res 45 (W1), W325–W330. https://doi.org/10.1093/nar/gkx277.
Stokes, J.M., Yang, K., Swanson, K., Jin, W.G., Cubillos-Ruiz, A., Donghia, N.M., et al., 2020. A deep learning approach to antibiotic discovery. Cell 180, 688–702. https://doi.org/10.1016/j.cell.2020.01.021.
Sun, H, Huang, R, Xia, M, et al., 2017. Prediction of hERG liability – using SVM classification, bootstrapping and jackknifing. Mol. Inform 36, 1600126. https://doi.org/10.1002/minf.201600126.
Sun, J., Chen, K, 2017. NSiteMatch: prediction of binding sites of nucleotides by identifying the structure similarity of local surface patches. Comput. Math Methods Med., 5471607 https://doi.org/10.1155/2017/5471607.
Tan, K.P., Nguyen, T.B., Patel, S., Varadarajan, R., Madhusudhan, M.S., 2013. Depth: A web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. Nucleic Acids Res 41, W314–W321. https://doi.org/10.1093/nar/gkt503.
Tanaka, S., Scheraga, H.A., 1976. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. Macromolecules 9, 945–950. https://doi.org/10.1021/ma60054a013.
Teague, S.J., 2003. Implications of protein flexibility for drug discovery. Nat. Rev. Drug Discov. 2, 527–541. https://doi.org/10.1038/nrd1129.
Teodoro, M.L., Kavraki, L.E., 2003. Conformational flexibility models for the receptor in structure-based drug design. Curr. Pharm. Des. 9, 1635–1648. https://doi.org/10.2174/1381612033454595.
Thai, Q.M., Pham, T.N.H., Hiep, D.M., Pham, M.Q., Tran, P.T., Nguyen, T.H., Ngo, S.T., 2022. Searching for AChE inhibitors from natural compounds by using machine learning and atomistic simulations. J. Mol. Graph. Model., 108230 https://doi.org/10.1016/j.jmgm.2022.108230.
Thomas, P.D., Dill, K.A., 1996. Statistical potentials extracted from protein structures: how accurate are they? J. Mol. Biol. 257, 457–469. https://doi.org/10.1006/jmbi.1996.0175.
Tian, C., Kasavajhala, K., Belfon, K., Raguette, L., Huang, H., Migues, A., Bickel, J., Wang, Y., Pincay, J., Wu, Q., Simmerling, C., 2020. ff19SB: amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. J. Chem. Theory Comput. 16, 528–552. https://doi.org/10.1021/acs.jctc.9b00591.
Totrov, M., Abagyan, R., 2008. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr. Opin. Struct. Biol. 18, 178–184. https://doi.org/10.1016/j.sbi.2008.01.004.
Tsubaki, M., Tomii, K., Sese, J., 2018. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics 35, 309–318. https://doi.org/10.1093/bioinformatics/bty53.
Unterthiner, T, Mayr, A, Klambauer, G, and Hochreiter, S, Toxicity prediction using deep learning, ArXiv.org, (2015)1503.01445.
Uusitalo, J.J., Ingólfsson, H.I., Akhshi, P., Tieleman, D.P., Marrink, S.J., 2015. Martini coarse-grained force field: extension to DNA. J. Chem. Theory Comput. 11 (8), 3932–3945. https://doi.org/10.1021/acs.jctc.5b00286.
Vanommeslaeghe, K, Hatcher, E, Acharya, C, Kundu, S, Zhong, S, et al., 2010. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem. 31, 671–690. https://doi.org/10.1002/jcc.21367.
Venkatachalam, C.M., Jiang, X., Oldfield, T., Waldman, M., 2003. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J. Mol. Graph. Model. 21, 289–307. https://doi.org/10.1016/s1093-3263(02)00164-x.
Verkhivker, G., Appelt, K., Freer, S.T., Villafranca, J.E., 1995. Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand- protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. Protein Eng. Des. Sel. 8 (7), 677–691. https://doi.org/10.1093/protein/8.7.677.
Vishnepolsky, B, Gabrielian, A, Rosenthal, A, et al., 2018. Predictive model of linear antimicrobial peptides active against gram-negative bacteria. J. Chem. Inf. Model. 58, 1141–1151. https://doi.org/10.1021/acs.jcim.8b00118.
Vyas, V.K., Ukawala, R.D., Ghate, M., Chintha, C, 2012. Homology modelling a fast tool for drug discovery: current perspectives. Indian J. Pharm. Sci 74 (1), 1. https://doi.org/10.4103/0250-474X.102537.
Wagner, J.R., Sørensen, J., Hensley, N., Wong, C., Zhu, C., Perison, T., et al., 2017. POVME 3.0: software for mapping binding pocket flexibility. J. Chem. Theory Comput. 13 (9), 4584–4592. https://doi.org/10.1021/acs.jctc.7b00500.
Wang, J, Wolf, RM, Caldwell, JW, Kollman, PA, Case, DA, 2004. Development and testing of a general amber force field. J. Comput. Chem. 25, 1157–1174. https://doi.org/10.1002/jcc.20035.
Wang, L., O'Mara, M.L., 2021. Effect of the force field on molecular dynamics simulations of the multidrug efflux protein P-glycoprotein. J. Chem. Theory Comput. 17 (10), 6491–6508. https://doi.org/10.1021/acs.jctc.1c00414.
Wang, Y., Yella, J., Jegga, A.G., 2019. Transcriptomic Data Mining and Repurposing For Computational Drug discovery, in Computational Methods For Drug Repurposing. Springer, pp. 73–95. https://doi.org/10.1007/978-1-4939-8955-3_5.
Wei, Y, Li, J, Chen, Z, 2015. Multistage virtual screening and identification of novel HIV-1 protease inhibitors by integrating SVM, shape, pharmacophore and docking methods. Eur. J. Med. Chem. 101, 409–418. https://doi.org/10.1016/j.ejmech.2015.06.054.
Weiner, P.K., Kollman, P.A., 1981. AMBER – assisted model building with energy refinementa general program for modeling molecules and their interactions. J. Comput. Chem. 2, 287–303. https://doi.org/10.1002/jcc.540020311.
Welch, W., Ruppert, J., Jain, A.N., 1996. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. Chem. Biol. 3, 449–462. https://doi.org/10.1016/s1074-5521(96)90093-9.
White, A.W., Almassy, R., Calvert, A.H., Curtin, N.J., Griffin, R.J., Hostomsky, Z., Maegley, K., Newell, D.R., Srinivasan, S., Golding, B.T., 2000. Resistance-modifying agents. 9. Synthesis and biological properties of benzimidazole inhibitors of the DNA repair enzyme poly (ADP-ribose) polymerase. J. Med. Chem. 43 (22), 4084–4097. https://doi.org/10.1021/jm000950v.
Winkler, D.A., Burden, F.R., 2004. Bayesian neural nets for modeling in drug discovery. Drug Discov. Today: BIOSILICO 2 (3), 104–111. https://doi.org/10.1016/S1741-8364(04)02393-5.
Wu, C.H., Arighi, C.N., Ross, K.E. (Eds.), 2017. Protein bioinformatics: from Protein Modifications and Networks to Proteomics. Humana Press.
Wu, G., Robertson, D.H., Brooks 3rd, C.L., Vieth, M., 2003. Detailed analysis of grid-based molecular docking: a case study of CDOCKER-A CHARMm-based MD docking algorithm. J. Comput. Chem. 24, 1549–1562. https://doi.org/10.1002/jcc.10306.
Wu, S., Zhang, Y., 2008. MUSTER: improving protein sequence profile–profile alignments by using multiple sources of structure information. Proteins: Struct. Funct. Genet. 772 (2), 547–556. https://doi.org/10.1002/prot.21945.
Xiao, X, Wang, P, Lin, W-Z, et al., 2013. iAMP-2L: a two-level multi-label classifier for identifying antimicrobial peptides and their functional types. Anal Biochem 436, 168–177. https://doi.org/10.1016/j.ab.2013.01.019.

Xie, Z.R., Liu, C.K., Hsiao, F.C., Yao, A., Hwang, M.J.LISE, 2013. A server using ligand-interacting and site-enriched protein triangles for prediction of ligand-binding sites. Nucleic Acids Res 41 (Web Server issue), W292–W296. https://doi.org/10.1093/nar/gkt300.

Xu, D, Zhang, Y, 2012. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins: Struct. Funct. Genet. 80 (7), 1715–1735. https://doi.org/10.1002/prot.24065.

Yan, R.X., Si, J.N., Wang, C., Zhang, Z., 2009. DescFold: a web server for protein fold recognition. BMC Bioinform 10 (1), 1–14. https://doi.org/10.1186/1471-2105-10-416.

Yasonik, J., 2020. Multiobjective de novo drug design with recurrent neural networks and nondominated sorting. J. Cheminformatics 12 (1), 1–9. https://doi.org/10.1186/s13321-020-00419-6.

Yu, W. and MacKerell, A.D., Computer-aided drug design methods. In *Antibiotics* (pp. 85-106). Humana Press, New York, NY. (2017). DOI: 10.1007/978-1-4939-6634-9_5.

Zavodszky, M.I., Kuhn, L.A., 2005. Side-chain flexibility in protein–ligand binding: the minimal rotation hypothesis. Protein Sci 14 (4), 1104–1114. https://doi.org/10.1110/ps.041153605.

Zgarbova, M., Otyepka, M., Sponer, J., Mladek, A., Banas, P., Cheatham, T.E., Jurecka, P., 2011. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. J. Chem. Theory Comput. 7, 2886–2902. https://doi.org/10.1021/ct200162x.

Zhang, H, Liu, W, Liu, Z, Ju, Y, Xu, M, Zhang, Y, et al., 2018. Discovery of indoleamine 2, 3-dioxygenase inhibitors using machine learning based virtual screening. Medchemcomm 9 (6), 937–945. https://doi.org/10.1039/c7md00642j.

Zhang, L., Tan, J., Han, D., Zhu, H, 2017. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug Discov. Today 22, 1680–1685. https://doi.org/10.1016/j.drudis.2017.08.010.

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinform 9 (1), 1–8. https://doi.org/10.1186/1471-2105-9-40.

Zhang, Y., Xi, H., Juhas, M, 2020. Biosensing detection of the SARS-CoV-2 D614G mutation. Trends Genet 37, 299–302. https://doi.org/10.1016/j.tig.2020.12.004.

Zhang, Z., Li, Y., Lin, B., Schroeder, M., Huang, B., 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics 27 (15), 2083–2088. https://doi.org/10.1093/bioinformatics/btr331.

Zhao, K., So, H.-C., 2019. Using drug expression profiles and machine learning approach for drug repurposing, Computational methods for drug repurposing. Springer 219–237. https://doi.org/10.1007/978-1-4939-8955-3_13.

Zhao, S., Li, X., Peng, W., Wang, L., Ye, W., Zhao, Y., et al., 2021. Ligand-based pharmacophore modeling, virtual screening and biological evaluation to identify novel TGR5 agonists. RSC Adv 11 (16), 9403–9409. https://doi.org/10.1039/d0ra10168k.

Zhou, S.F, Zhong, W.Z, 2017. Drug design and discovery: principles and applications. Molecules 22 (2), 279. https://doi.org/10.3390/molecules22020279.

Zhu, H, 2020. Big data and artificial intelligence modeling for drug discovery. Annu. Rev. Pharmacol. Toxicol. 60, 573–589. https://doi.org/10.1146/annurev-pharmtox-010919-023324.

Zhu, H., Pisabarro, M.T., 2011. MSPocket: an orientation-independent algorithm for the detection of ligand binding pockets. Bioinformatics 27 (3), 351–358. https://doi.org/10.1093/bioinformatics/btq672.

Zsoldos, Z., Reid, D., Simon, A., Sadjad, S.B., Johnson, A.P., 2007. eHiTS: a new fast, exhaustive flexible ligand docking system. J. Mol. Graph. Model. 26, 198–212. https://doi.org/10.1016/j.jmgm.2006.06.002.

Zupan, J., Gasteiger, J., 1999. Neural Networks in Chemistry and Drug Design. John Wiley & Sons, Inc.

Further reading

Lew, W., Chen, X., Kim, C.U., 2000. Discovery and development of GS 4104 (oseltamivir) an orally active influenza neuraminidase inhibitor. Curr. Med. Chem. 7 (6), 663–672. https://doi.org/10.2174/0929867003374886.

Aruksakunwong, O., Promsri, S., Wittayanarakul, K., Nimmanpipug, P., Lee, V.S., Wijitkosoom, A., et al., 2007. Current development on HIV-1 protease inhibitors. Curr. Comput. Aided Drug Des 3 (3), 201–213. https://doi.org/10.2174/157340907781695431.

Lyle, T. Ribonucleic Acid Viruses: antivirals for Human Immunodeficiency Virus, (2007).

Lynch, J.J., Cook, J.J., Sitko, G.R., Holahan, M.A., Ramjit, D.R., Mellott, M.J., Stranieri, M.T., Stabilito, I.I., Zhang, G., Lynch, R.J., 1995. Nonpeptide glycoprotein IIb/IIIa inhibitors. 5. Antithrombotic effects of MK-0383. J. Pharmacol. Exp. Ther. 272 (1), 20–32.

Ashraf, Z., Kanwal, M., Hassan, M., Abdullah, S., Waheed, M., Ahsan, H. and Kim, S.J., Flurbiprofen–antioxidant mutual prodrugs as safer nonsteroidal anti-inflammatory drugs: synthesis, pharmacological investigation, and computational molecular modeling. Drug design, development and therapy, 10(2016), 2401. DOI: 10.2147/DDDT.S109318.

Aulakh, G.K., Sodhi, R.K., Singh, M., 2007. An update on non-peptide angiotensin receptor antagonists and related RAAS modulators. Life Sci 81 (8), 615–639. https://doi.org/10.1016/j.lfs.2007.06.007.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A. J., 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. 19, 1639–1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19. :14%3C1639::AID-JCC10%3E3.0.CO;2-B.

Njoroge, F.G., Chen, K.X., Shih, N.Y., Piwinski, J.J., 2008. Challenges in modern drug discovery: a case study of boceprevir, an HCV protease inhibitor for the treatment of hepatitis C virus infection. Acc. Chem. Res. 41 (1), 50–59. https://doi.org/10.1021/ar700109k.

Barker, A., Andrews, D, 2013. Discovery and development of the anticancer agent gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase. Introduction to Biological and Small Molecule Drug Research and Development. Elsevier, pp. 255–281. https://doi.org/10.1016/B978-0-12-397176-0.00008-X.

Perzborn, E., Roehrig, S., Straub, A., Kubitza, D., Misselwitz, F., 2011. The discovery and development of rivaroxaban, an oral, direct factor Xa inhibitor. Nat. Rev. Drug Discov. 10 (1), 61–75. https://doi.org/10.1038/nrd3185.

Rivera-Lebron, B.N., Risbano, M.G., 2017. Ambrisentan: a review of its use in pulmonary arterial hypertension. Ther. Adv. Respir. Dis. 11 (6), 233–244. https://doi.org/10.1177/1753465817696040.

Rutenber, E.E., Stroud, R.M., 1996. Binding of the anticancer drug ZD1694 to E. coli thymidylate synthase: assessing specificity and affinity. Structure 4 (11), 1317–1324. https://doi.org/10.1016/s0969-2126(96)00139-6.

Schindler, T., Bornmann, W., Pellicena, P., Miller, W.T., Clarkson, B., Kuriyan, J., 2000. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science 289 (5486), 1938–1942. https://doi.org/10.1126/science.289.5486.1938.

Sun, C.L., Christensen, J.G., McMahon, G., 2009. Discovery and Development of Sunitinib (SU11248): a Multitarget Tyrosine Kinase Inhibitor of Tumor growth, survival, and Angiogenesis. Kinase Inhibitor Drugs. https://doi.org/10.2165/11318860- 000000000-00000.

Supuran, C.T., 2012. Structure-based drug discovery of carbonic anhydrase inhibitors. J. Enzyme Inhib. Med. Chem. 27 (6), 759–772. https://doi.org/10.3109/14756366.2012.672983.

Takahashi, H., Hayakawa, I., Akimoto, T., 2003. The history of the development and changes of quinolone antibacterial agents. Yakushigaku Zasshi 38 (2), 161–179.

Thomas, P.D., Dill, K.A., 1996. An iterative method for extracting energy-like quantities from protein structures. Proc. Natl. Acad. Sci. USA 93, 11628–11633. https://doi.org/10.1073/pnas.93.21.11628.

Abagyan, R., Totrov, M., Kuznetzov, D., 1994. ICM – a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. J. Comput. Chem. 15, 488–506. https://doi.org/10.1002/jcc.540150503.

Veljkovic, N., Vucicevic, J., Tassini, S., Glisic, S., Veljkovic, V., Radi, M., 2015. Preclinical discovery and development of maraviroc for the treatment of HIV. Expet Opin. Drug Discov. 10 (6), 671–684. https://doi.org/10.1517/17460441.2015.1041497.

Warnke, C., Wiendl, H., Hartung, H.P., Stüve, O., Kieseier, B.C., 2010. Identification of targets and new developments in the treatment of multiple sclerosis–focus on cladribine. Drug Des. Devel. Ther. 4, 117. https://doi.org/10.2147/dddt.s6627.

Wlodawer, A., Vondrasek, J., 1998. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. Annu Rev Biophys Biomol Struct 27 (1), 249–284.

Wu, S., Zhang, Y., 2008. MUSTER: improving protein sequence profile–profile alignments by using multiple sources of structure information. Proteins: Struct. Funct. Genet 72 (2), 547–556. https://doi.org/10.1002/prot.21945.

Best, B.M., Goicoechea, M., 2008. Efavirenz–still first-line king? Expert Opin Drug Metab Toxicol 4 (7), 965–972. https://doi.org/10.1517/17425255.4.7.965.

Zhu, C., 2013. Aldose reductase inhibitors as potential therapeutic drugs of diabetic complications (Vol. 2). chapter.

Cohen, N.C., 2007. Structure-based drug design and the discovery of aliskiren (Tekturna):perseverance and creativity to overcome a R&D pipeline challenge. Chem. Biol. Drug Des. 70 (6), 557–565. https://doi.org/10.1111/j.1747-0285.2007.00599.x.

Cushman, D.W., Ondetti, M.A., 1991. History of the design of captopril and related inhibitors of angiotensin-converting enzyme. Hypertension 17 (4), 589–592. https://doi.org/10.1161/01.hyp.17.4.589.

De Clercq, E., 2009. The history of antiretrovirals: key discoveries over the past 25 years. Rev. Med. Virol. 19 (5), 287–299. https://doi.org/10.1002/rmv.624.

Fischer, J., Ganellin, C.R., 2015. Analogue-based drug discovery. Chem. Int.–Newsmagazine for IUPAC, 32 (4), 12–15. https://doi.org/10.1515/ci.2010.32.4.12.

W.J. Adams, P.A. Aristoff, R.K. Jensen, W. Morozowich, D.L. Romero, W.C. Schinzer, W. G. et al., Discovery and development of the BHAP nonnucleoside reverse transcriptase inhibitor delavirdine mesylate, in: integration of Pharmaceutical Discovery and Development, Springer, (2002) 285–312. DOI: 10.1007/0-306-47384-4_13.

Ghosh, A.K., Dawson, Z.L., Mitsuya, H., 2007. Darunavir, a conceptually new HIV-1 protease inhibitor for treating drug-resistant HIV. Bioorg. Med. 15 (24), 7576–7580. https://doi.org/10.1016/j.bmc.2007.09.010.

Huang, S.Y., Zou, X., 2010. Advances and challenges in protein-ligand docking. Int. J. Mol. Sci. 11 (8), 3016–3034. https://doi.org/10.3390/ijms11083016.

Huang, S.Y., Zou, X., 2010. Advances and challenges in protein-ligand docking. Int. J. Mol. Sci. 11 (8), 3016–3034. https://doi.org/10.3390/ijms11083016.

Huang, S.-Y., Zou, X., 2010. Mean-force scoring functions for protein-ligand binding. Annu. Rep.Comput. Chem. 6, 281–296.

Kim, E.E., Baker, C.T., Dwyer, M.D., Murcko, M.A., Rao, B.G., Tung, R.D., et al., 1995. Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. J. Am. Chem. Soc. 117 (3), 1181–1182. https://doi.org/10.1021/ja00108a056.

Cui, J.J.; Tran-Dube, M.; Shen, H.; Nambu, M.; Kung, P.P.; Pairish M. et al. 2011 Structure based drug design of crizotinib (PF-02341066), a potent and selective dual inhibitor of mesenchymal epithelial transition factor (c-MET) kinase and anaplastic lymphoma kinase (ALK). DOI: 10.1021/jm2007613.

Goa, K.L. and Noble, S., 2039. Eptifibatide. Drugs, 57(3), 439–462.