Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Chapter 2

Fundamental considerations in drug design

Manoj Kumar Mahapatra\textsuperscript{a} and Muthukumar Karuppasamy\textsuperscript{b}
\textsuperscript{a}Kanak Manjari Institute of Pharmaceutical Sciences, Rourkela, Odisha, India, \textsuperscript{b}YaAn Pharmaceutical and Medical Communications, Sivakasi, Tamil Nadu, India

Chapter outline

1 Fundamentals of rational drug design (RDD) \hfill 17
1.1 Rational drug design 17
1.2 Structure-based drug design (SBDD) 18
1.3 Ligand-based drug design (LBDD) 19

2 Concepts of physicochemical properties \hfill 20
2.1 Structural properties and stereoechemistry 20
2.2 Drug receptors and receptor theories 21
2.3 Pharmacokinetics and pharmacodynamics 22
2.4 SARs and QSARs 23
2.5 Prodrugs and drug metabolism 23
2.6 Metabolite antagonism and enzyme inhibition 24
2.7 Nucleic acid-based drug design 25
2.8 Lead compounds 25
2.9 Peptidomimetics and analog design 26
2.10 Reverse pharmacology and drug repurposing strategies 26

3 Fundamentals of computer-aided drug design (CADD) \hfill 27
3.1 Structure-based drug design (SBDD) 28
3.2 Ligand-based drug design (LBDD) 36
3.3 Virtual screening techniques 39
3.4 ADME analysis and measures of drug-likeness 45

4 Conclusion \hfill 47

References \hfill 48

1 Fundamentals of rational drug design (RDD)

1.1 Rational drug design

Drug discovery, drug development, and its associated processes in pharma industries and academic research institutions have been quite cumbersome, tedious, and demanding. However, the advent of novel computational and
synthetic approaches has somewhat reduced the concerns of drug discovery and development. Since the discovery of penicillin for the treatment of several bacterial infections in 1928 by Alexander Fleming, drugs have been a critical part of human life.\textsuperscript{1,2} Mankind has always been on the receiving end of a constant threat from microorganisms such as bacteria, virus, protozoan, fungus, etc.; therefore, drugs with less adverse effects and significant pharmacological activities or safe and potent drugs would be much appreciated. The cost for inventing a new drug for any particular disease condition would be around 2–3 billion dollars and time spent in every drug discovery program would be approximately 12–15 years.\textsuperscript{3,4} To lessen drug discovery duration and expenses through the narrow down of drug-like compounds in the discovery pipeline, a rational drug design strategy has been in practice for a few decades. In general, drug discovery consists of the following stages:

- Identification, characterization, and validation of targets.
- Development, optimization, and troubleshooting of assays.
- Identification and optimization of lead molecules.
- Absorption, distribution, metabolism, extraction, and toxicity studies.
- Pharmacodynamic and pharmacokinetic studies.
- Clinical trials.

Rational drug design approaches have been employed to explore and invent novel molecules against diseases or any dysfunction in the human body.\textsuperscript{5} Two widely used methods in rational drug design, structure-based drug design and ligand-based drug design, are currently in the drug discovery domain to get rid of the typical trial and error approach.\textsuperscript{6,7} Rational drug design mainly focuses on target discovery of hits and lead compounds and optimization of drug-like candidates where thorough data on biochemical and structural characteristics of a protein target would be incorporated. Techniques such as nuclear magnetic resonance (NMR) and X-ray crystallography assist in investigating the structural features of protein targets that could be very beneficial in the target discovery.\textsuperscript{8}

\subsection*{1.2 Structure-based drug design (SBDD)}

SBDD is based on the three-dimensional biological structure of the target of interest and its relevant data via computational approach (homology modeling) or experiments, which may be proteins, receptors, and enzymes. SBDD precisely monitors the binding abilities of chosen ligand molecules to the desired target region and envisages the possible indispensable binding pocket sites present in the target, followed by the affinity of ligands to their respective biomolecular target. The resulting information of the target, as well as ligands, can be useful in designing high potent ligand molecules with a desired pharmacological and toxicological profile.\textsuperscript{9–11} SBDD method has the following key steps\textsuperscript{12,13}:
1. Preparation of protein structure
2. Identification of a binding site in the protein of interest
3. Preparation of ligands
4. Docking and scoring functionalities.

The challenges and limitations in the SBDD include the following: (1) The flexibility of the target throughout the molecular docking and modeling should be taken into account where the static nature of the target has been still used in some SBDD methods. Avoiding the target flexibility leads to inaccurate and false ligands in the designed program. (2) Considering water molecules in the target environment would be essential where it possesses and facilitates hydrogen bonding interaction and free energy in the protein-ligand complex. Nevertheless, the availability of a copious number of water molecules would create difficulty in either choosing an exact water molecule or neglecting an unwanted water molecule. (3) The role of solvation for solvated drug molecules in molecular docking is very important, but including the solvation effect would be very difficult; hence, the ultimate scoring functions of the lead compounds will be largely affected.

1.3 Ligand-based drug design (LBDD)

LBDD is employed for several processes in drug discovery, such as lead identification and lead optimization in the absence of three-dimensional structures of the target of interest. Advanced techniques such as pharmacophore modeling and three-dimensional quantitative structure-activity relationship (3D QSAR) are the familiar tools used in LBDD, which could yield predictive QSAR models for lead molecules optimization and possible key interactions between both ligands and drug targets. LBDD is typically an indirect method to expedite the drug development of therapeutically active drug molecules through the potential interaction of drug-like candidates with the target of interest. LBDD also indicates computer-aided drug design (CADD). The crucial part of CADD is to reduce the time duration for novel drug molecules’ identification, characterization, and structural optimization. The advantages or features of CADD could be implemented in the prodrugs where the bioavailability and specificity of parent molecules would be escalated by the administration of prodrugs.

2 Concepts of physicochemical properties

When the drug molecule enters the human body, it forms an interaction or bonding with membrane proteins or receptors found on the surface of cells. The interaction of the drug with the membrane receptors is based on the polarity, reactivity, and nature of chemical moieties present in the drug compounds. The drug and protein interaction may be of weak forces, including hydrogen
bonds, ionic bonds, van der Waals forces, dipole-dipole forces, and dipole-ion forces; only a few cases show the presence of covalent bonds that elicit unwanted bonding of drug candidates and adverse effects.\textsuperscript{24} The physicochemical properties of drugs would determine the efficacy and safety profile. The indispensable physicochemical properties are as follows:

1. Water or lipid solubility profile
2. Acid-base attributes
3. Physical state
4. Nature of hydrogen bonding
5. Isosterism and bioisosterism
6. Redox potential
7. Ionization capacity
8. Dissociation constant
9. Partition coefficient
10. Complexation
11. Protein binding
12. Surface activity

2.1 Structural properties and stereochemistry

The stereochemistry of chemicals and drug molecules begins with the topic “chirality,” four different substituents attached to a carbon atom make a molecule chiral, and the carbon is referred to as the chiral center. The definition of isomerism or isomers is “compounds having similar molecular formula but different structural arrangements.” Isomers are divided into two categories such as constitutional (structural) and stereoisomers.\textsuperscript{25,26} In constitutional isomers, the arrangement of atoms will be different, whereas they show similar molecular formulas, however, chemically different compounds. Of note, constitutional isomers demonstrate identical pharmacological efficacies. Stereoisomers are molecules that consist of similar sets of atoms; they are constructed in the same positions, but the arrangement of atoms is spatially different. Stereoisomers can be of two types, optical isomers or enantiomers and geometrical isomers.\textsuperscript{27–29} Two types of orientations for chiral molecules would be possible where these two forms of the same compound are known as enantiomers. Nonsuperimposable mirror images of one another can be found in enantiomers. There are two different types of enantiomers, levo or (−) and dextro, based upon their abilities to rotate the plane-polarized light. In contrast, racemate molecules, composed of an identical ratio of enantiomers (50% of levo and 50% of dextro), do not exhibit the characteristic property of rotating the plane-polarized light. The mechanism of racemate on rotating plane-polarized light is opposite, and equal opposite effects of enantiomers terminate each other’s effects. In general, enantiomers possess similar physicochemical properties but different pharmacological or biological activities. The difference in pharmacological activity of
enantiomers is due to the orientation of various substituents attached to the chiral center atom, which may favorably allow for the beneficial interaction with the target protein over its counterpart. Many clinically available drug molecules are chiral, and they are dispensed as racemate form. With regard to pharmacokinetic and pharmacodynamic attributes, racemates differ from enantiomers.

2.2 Drug receptors and receptor theories

Drug receptors are the macromolecules, which are proteins, and get the bonding or interaction with ligands/drug compounds to proceed with an effector response. The presence of drug receptors ranges from the cell membrane and cytosolic regions to intranuclear areas of the cell. The effector response could be a therapeutic effect or adverse effect. Drug receptors often reveal distinguished characteristics depending upon the concentration of availability of the drug molecules. Receptor selectivity is another important parameter in the therapeutic efficacy of active drug candidates, and it is mainly decided how the ligand binds with a particular receptor than other receptors. The affinity of ligands attached or interacted with the receptors would contribute to the pharmacokinetic as well as pharmacodynamic activity of the same drug. For instance, receptors with greater affinity for a specific drug molecule need a lesser concentration of drug for a complete saturation level.

A drug with full agonist property demonstrates maximum biological effect or conformational modification at the site of action and great efficacy at enough concentration. However, antagonists show zero efficacy with no effect, only producing a blockade at the receptors by replacing agonist compounds. In other words, antagonists cannot create a conformational modification. Another class of compounds known as inverse agonist or partial agonist molecules exhibits a low efficacy and biological effect at a given concentration. Partial agonists unveil a mixture of pharmacological activities of agonist and antagonist. Modest to low efficacy has resulted from the submaximal consequences of partial agonists. Researchers Ehrlich and Langley began the work on receptor theory; later on, the discovery of propranolol (β-adrenergic antagonist) convinced the receptor theory and existence of drug receptors. The drug and receptor interactions are assigned and designated as lock-and-key models. The receptor is considered a lock that contains a region with active sites and different types of pockets, whereas the drug candidate as a key can bind with the receptors to make a conformational change, followed by therapeutic effects.

2.2.1 Occupation theory

The specific receptors and their concept were published by Langley where the study of antagonistic activity of atropine in pilocarpine-triggered salivary flow in cat models was examined. Based on Langley’s research work, the
occupation theory was proposed by many researchers. Clark (1926)\textsuperscript{37} revealed that the interaction of receptors and ligands through a mathematical association could be equivalent to the interaction of oxygen and carbon dioxide along with hemoglobin. In short, the communication between ligand molecules and receptors was more like Langmuir’s isotherm reaction.\textsuperscript{38} Ariens proposed two critical parameters such as affinity and intrinsic activity to rectify the deficiency in Clark’s research statement. Then, Stephenson had presented a few ideas about the occupation theory to improve earlier research findings.

2.2.2 Rate theory

Rate theory is considered as an alternative approach to occupation theory and its modified versions. Of note, in 1961, Paton came up with the receptor rate theory of drug action. As per rate theory, the pharmacological response is directly proportional to the rate at which the active drug molecule binds with the receptors. He also proposed that association and dissociation rate constants were very essential in establishing the affinity and a resulting therapeutic activity of ligands. The proportion of dissociation and association rate constants would decide the key parameter “affinity.” The triggered activity of the drug molecule will be determined by the dissociation rate constant. More importantly, the intrinsic activity employed in occupation rate theory is replaced by the dissociation rate constant.\textsuperscript{39,40}

2.2.3 Induced fit theory

The main attributes of the induced fit theory are as follows: (1) The availability of defined orientation of catalytic moieties is necessary for enzymatic activities; (2) significant modification in the three-dimensional structure of the amino acids present in the active site would be caused by the substrate itself; and (3) appropriate alignment of catalytic groups would be found as a result of the modification in the protein structure which is triggered by the substrate. On the contrary, a non-substrate would not cause this effect on catalytic groups.\textsuperscript{41}

2.3 Pharmacokinetics and pharmacodynamics

Pharmacology is the study of drug molecules and their therapeutic and adverse effects on human beings. Pharmacokinetics and pharmacodynamics are the branches of pharmacology where the action of the body on the drug is described as pharmacokinetics, whereas the action of the drug on the body is called pharmacodynamics. Pharmacology also deals with the study of legal and illicit compounds, endogenous and exogenous molecules, prescription and over-the-counter medications, and natural, semisynthetic, and synthetic drug molecules. Pharmacodynamics defines the interaction between drug molecules and targets including receptors and enzymes. To study the therapeutic profile of any drug candidate, in vitro and in vivo studies could be performed, and it is referred to as pharmacokinetics.\textsuperscript{42–45}
2.4 SARs and QSARs

Structure-activity relationship or SAR is a method to investigate the qualitative association between pharmacophore or chemical moieties or functional groups present in the active ligand compound and their desired pharmacological activity. Quantitative structure-activity relationship (QSAR) is defined as the quantification approach to quantify the relationship between the chemical structures or pharmacophore and biological activity.\textsuperscript{46,47} As per QSAR protocol, a library of small molecules would be identified that exhibit the desired therapeutic activity. Subsequently, a quantitative association is found between pharmacophoric groups in the lead molecules and the pharmacological potential. This QSAR model would be considered as a tool for the optimization of active novel molecules to increase the therapeutic efficacies of the selected compounds. The predicted molecules will be then evaluated for their pharmacological activity through experiments. The common protocol of QSAR is comprised of the following steps:

1. Identification of ligands with experimental and biological measurement. The chosen ligands must be from diverse chemical libraries.
2. Establishment of chemical descriptors that are related to different structural and physicochemical properties of drug candidates.
3. Investigation of the association between molecular descriptors and pharmacological activity.
4. Evaluation of predictive power and statistical stability of the QSAR model.

The selection of molecular descriptors and the capability of creating a suitable mathematical association between molecular descriptors and pharmacological efficacy are the integral steps for developing a QSAR model. The statistical methods, including partial least square analysis, principal component analysis, and multivariable linear regression analysis, are currently employed in the QSAR.\textsuperscript{48} Molecular descriptors are further classified into various categories as follows:\textsuperscript{49}

- 1D—molecular formula considerations in the calculation.
- 2D—employing connection details of compounds and atoms.
- i3D—internal descriptors, generally representing three-dimensional coordinates information of molecules.
- x3D—external three-dimensional descriptors that utilize 3D information of every molecule together with a complete frame of the reference compound.

2.5 Prodrugs and drug metabolism

Prodrugs are the bioinactive compounds that actively undergo chemical modification or enzymatic catalysis to form or release the parent drug (active candidate for the particular disease condition). Later, the released parent drug would exhibit its therapeutic activity. Prodrugs is the concept, which is used very frequently in drug discovery and development, to overcome the concerns
in pharmacokinetic and physicochemical attributes of drug molecules. At the moment, approximately 4%-8% of drug candidates launched in the market are prodrugs. It is currently a well-established approach in drug discovery and development where the efficacy and chances of entering into the market would be higher in the prodrug strategy. The prodrug approach essentially regulates and reduces a copious number of hindrances including low aqueous solubility, lack of oral absorption, fast presystemic metabolism, insufficient penetration in the brain, chemical instability, toxic concerns, and irritation of the compound locally. Besides the above advantages, it also ameliorates drug targeting so that unwanted adverse effects could be minimized.

A multitude of parameters must be taken into account while designing the prodrugs; the fact is that the prodrug could affect efficacy, toxicity, and distribution of the parent molecule. The parameters under investigation in designing the prodrugs are as follows:

1. Selection of the functional groups that can be modified for prodrug development.
2. Protecting groups and their safety and toxicity profile, selection should be based on the dose of the drug, disease condition, and treatment duration.
3. Pharmacokinetic characteristics such as absorption, distribution, metabolism, excretion, and toxicity of prodrug and parent compound must be examined in detail.
4. A study on degradation derivatives and by-products must be conducted. These formed degradative by-products may have an effect on physical and chemical stability and eventually result in the production of various metabolic derivatives.

The familiar functional groups which are responsive to prodrug development are carbonyl, hydroxyl, carboxylic, amine, phosphonate, and phosphate.

The pharmaceutical applications of prodrugs include oral absorption improvement, enhanced aqueous solubility, lipophilicity enhancement, improved parenteral and topical administration, and sustained duration of drug activity.

### 2.6 Metabolite antagonism and enzyme inhibition

Research work from Wood and Fildes had proven the notion of metabolite antagonism; structurally similar drug compounds to that of the essential metabolite could disrupt the activities of the particular metabolite. Wood and Fildes reported the case studies of p-aminobenzoic acid and its antagonistic activities on sulfanilamide. The presence of structural features in any pharmacophore is very mandatory to exhibit typical characteristics of metabolites antagonism; for instance, specific functional groups and certain atomic distances between such mentioned groups are appreciated. There are two main classes of metabolite antagonists such as amino acid antagonists and purine and pyrimidine antagonists. The
molecules that possess the antagonizing activities of metabolic pathways are called antimetabolites. Their mechanism of action is either replacing any particular entity in the metabolic pathways or inhibiting the receptor.57

Enzymes are biochemical catalysts that participate in biological reactions. Our human body consists of several thousands of various types of enzymes, which always act as a group to maintain homeostasis and regulate biological functions. In general, the inactive state of a particular enzyme or infection-associated entry of a foreign enzyme may cause plenty of disease conditions; however, blocking the activities of a certain enzyme is an uphill task. Of note, normal functions of cells happen via a network of enzymatic pathways; therefore, inhibiting a specific enzyme or designing a drug candidate that can target the enzyme would be tedious. In biochemical reactions, enzymes (E) combine with the substrate (S) to produce an enzyme-substrate complex (E-S). The formed E-S complex further undergoes catalytic reactions to form the enzyme-product complex (E-P); finally, dissociation occurs to yield the product (P) and release the free state of the enzyme. The enzymatic inhibition is of two types such as reversible and irreversible inhibitions. Reversible enzymatic inhibition is further divided into competitive and noncompetitive inhibitions.58

2.7 Nucleic acid-based drug design
Nucleic acid-based drugs (NABDs) such as ribozymes, antisense oligonucleotides, RNA interference or gene silencing approach (siRNA, miRNA), transcription factor decoys, and DNAzymes have been recently discovered for their pharmacological and clinical applications through modulation of gene expression. With assistance from the human genome program and expression data analysis using the transcriptome approach, NABDs have made a big stride in the drug discovery paradigm. The notable case in the application of NABDs would be the limited number of approvals by USFDA since the confounding factors, including off-target effects, acute therapeutic activities, drug delivery issues, and intrinsic adverse effects, could make them unfavorable.59 Since the approval of the first antisense oligonucleotide molecule, Vitravene, in the treatment of cytomegalovirus retinitis in AIDS patients, an abundant number of NABDs are in the clinical trials at the moment.60

2.8 Lead compounds
The lead compounds either derived from natural resources or synthetic routes possess potent activity with less adverse effects, so they can be used directly without any further modification. However, this may happen on a few occasions and most of the time we may confront a lot of concerns such as unwanted off-target effects, low efficacy, and harmful adverse events. In that case, structural modification is necessary. For instance, the objective is to develop the orally active compound; one may consider following Lipinski’s rule of five or Veber’s
parameters. In general, the recommended properties for a lead compound are a ClogP value of 1–3 and a molecular weight of 100–350 amu. With regard to molecular weight and ClogP values, research studies revealed that an 80 amu molecular weight increase in pharmacophore/lead compound would result in an increase of 1 in ClogP during the conversion of a lead compound to the final drug candidate. Typically, a lead compound should have a smaller number of hydrogen bond acceptors and aromatic rings compared with the final molecule. There is a method to measure the ligand or binding efficiency of lead compounds, which could be determined by dividing the free energy of binding for each compound by the number of nonhydrogen atoms available in the pharmacophore. Similarly, in the fragment-based lead discovery approach, a rule of three has been followed. A ClogP value of 3, a molecular weight less than 300, no more than three hydrogen bond acceptors, hydrogen bond donors, and rotatable bonds and a polar surface area of 60 Å² are the requirements of the rule of three to explore for the fragments.\textsuperscript{61}

2.9 Peptidomimetics and analog design

Peptides and proteins as lead compounds in the drug design of novel molecules have been considered very significant; currently, they include protease inhibitors, matrix metalloproteinase inhibitors, and renin inhibitors. Peptides have been considered as lead candidates in drug discovery research because they are involved in many biological functions as enzyme substrates and receptor ligands. The lead compounds derived from this category exhibit the attributes of peptides, yet the pharmacokinetic properties make them inappropriate classes. They are referred to as peptidomimetics; their pharmacophore features mimic the peptides and proteins in three-dimensional space and capability to have an interaction with targets of interest, generating similar biological activities. In addition, peptidomimetics are mainly meant to lessen the issues, including poor bioavailability and low half-life due to proteolysis reaction. Replacement of peptide bonds (which are enzymatically and chemically vulnerable) takes place with either a stable functional group to peptidase enzymes or low binding capacities to active sites. Alkene could replace peptide bonds, whereas the activity retained by the alkene group can be known as bioisosteric property. Moreover, alkene mimics the peptide double bond and is not a substrate for the peptidase enzymes family. Peptidomimetics are hydrophobic in nature; as a result, the disadvantages, including poor aqueous solubility and oral absorption, occur. Increasing the polarity of residues could minimize the poor aqueous solubility profile of peptidomimetics.\textsuperscript{62,63}

2.10 Reverse pharmacology and drug repurposing strategies

Reverse pharmacology is a science that changes the conventional drug discovery process, reversing the lab-to-clinics practice to clinics-to-lab practice. It is
based upon the integration of very well-documented clinical experiences and experiential observations into lead molecules using transdisciplinary exploratory studies, finally developing drug candidates via preclinical and clinical studies. In this reverse pharmacology, safety is the focal aspect, whereas efficacy is being considered as a validating point. A huge number of postmarketing withdrawals, attrition rates, and expenses have resulted in reverse pharmacology. (Reserpine and artemisinin are the prime examples of reverse pharmacology approach.) A major advantage of reverse pharmacology is the link between knowledge obtained from traditional medicine and modern technology to yield the leads with better safety and efficacy profile.64,65

Drug repurposing is the approach that is focused on new patient populations with well-known drug compounds, which was earlier intended for the different disease condition. Experimental screening methods and in silico techniques, including real-world data and omics-based repurposing, are the approaches employed in drug repurposing.66

3 Fundamentals of computer-aided drug design (CADD)

The application of computers and computational methods in the field of drug design and discovery process is referred to as computer-aided drug design (CADD). It is beneficial in the hit-to-lead discovery, lead optimization which drastically reduced the time and cost factor involved in the new drug discovery process. Modeling three-dimensional structures of ligand and protein, simulation, prediction of binding interactions and energy is a challenging job in the field of drug design. Most of the molecular modeling methods are based on molecular mechanics or quantum mechanics, although both the methods generate equations for calculating total energy of the system but differ from each other in some fundamental aspects, which are illustrated in Table 2.1.

| Characteristics                  | Molecular mechanics                  | Quantum mechanics                |
|----------------------------------|--------------------------------------|----------------------------------|
| State of the molecule            | Frozen at 0K in vacuum               | Neither frozen nor in vacuum     |
| Electronic nature of atom        | Not taken into consideration          | Considered                       |
| Structural vibrations            | Not taken into consideration          | Considered                       |
| Influence of the medium          |                                      |                                  |

Continued
Broadly, CADD has been studied under two different classes: structure-based drug design and ligand-based drug design.

### 3.1 Structure-based drug design (SBDD)

A high-resolution protein structure or a prepared homology model of the protein is a vital need of structure-based designing. Protein structural information like binding sites, cavities, secondary binding sites, etc., is highly useful for the discovery of small-molecule binding agents which can modulate biological activity. This information is required for ascertaining the molecular interactions of the ligand within the binding cavity. The main target of SBDD is to design and discover ligand molecules with high binding affinity and of complementary features. SBDD includes docking, molecular dynamics, and pharmacophore modeling.

#### 3.1.1 Docking

Docking is an in silico method which predicts the interactions and probable binding conformation of ligand or drug inside the binding cavity of receptor structure. Docking involves two critical steps: (i) search algorithm: to search the conformational space and generate all probable conformations or poses of the ligand within the binding site; (ii) scoring function: to identify the most suitable pose and rank them.67

*Search algorithm:* A search algorithm explores the conformational space of all the probable poses of ligands within the binding site. During ligand-receptor complex formation, it involves conformational changes in ligand and in receptor along with any of the ions or water molecules associated with them. It leads...
to generating a multitude of degrees of freedom which is difficult to handle in terms of computational cost. Therefore, most of the docking programs impart full flexibility to the ligands, but the receptor is kept rigid.\textsuperscript{68,69} Based on the nature of searching algorithms, conformational sampling techniques for introducing ligand flexibility are of three types: (i) systematic search; (ii) stochastic search; and (iii) deterministic search.

**Systematic search:** With a gradual rotation of all available rotatable bonds (rotational degrees of freedom) of ligand, several structures can be generated covering all possible conformations. But it is quite difficult in terms of time and massive calculations required to handle such a huge number of structures. A rapid search of the conformational space could be achieved in an economical way by restricting rotational degrees of freedom. It is to be overcome either by applying a set of termination criteria to restrict the number of generated conformations or by adopting incremental reconstruction of the ligand within the binding site. Incremental reconstruction may be carried out in two ways: (i) Ligand is segregated into a rigid core fragment which should be docked first followed by addition of other flexible side chains successively and with gradual docking; (ii) ligand is broken into several flexible fragments which are docked individually and then reconnected within the binding site to build the final conformation. ADAM, DOCK, DOCK 6, FlexX, FLOG, GLIDE, Hammerhead, and LUDI are examples of some docking programs that use systematic search algorithm.

**Stochastic search:** This method randomly varies the orientation of rotatable bonds and translational motion of the ligand to create several conformations. The binding affinity of each of the conformations is estimated for its evaluation. Stochastic search can use any of the following methods: (i) Monte Carlo simulation, (ii) genetic algorithm, and (iii) tabu search method.

**Monte Carlo simulations:** A random conformation of the drug molecule or ligand is generated, which undergoes a variation of either dihedral angle and rotation or translation of the ligand (one parameter at a time) to generate new conformations and estimate their energy. The low-energy conformers are accepted, and the ones with high energy are rejected based on the Metropolis algorithm which works on the principle of Boltzmann probability distribution. The Monte Carlo method is often combined with simulated annealing, in which a docking procedure is initiated with high temperature to cross the energy barriers and can explore the conformational space. The temperature is gradually reduced along with the conformational freedom of the ligand, so that it can be trapped in a low-energy minimum. For example, DockVision, GLIDE, ICM, MCDOCK, and MOE programs use this method.

**Genetic algorithms:** Genetic algorithm concept is inspired from Darwin’s evolution theory. A random population of a single ligand structure (each member is different in terms of conformation or pose) is generated, which is referred to as seed or initial generation. A ligand pose represents “chromosome,” whereas its properties like dihedral angle and rotational or translational degrees
of freedom represent “gene.” More diversity can be introduced into the generated structures with the help of genetic operators like mutation operator (new structures are generated by changing the rotational or translational degrees of freedom of the initial generation ligand) or crossover operator (new structures are generated by merging ligands of two different poses). Each of the new structures undergoes an evaluation of its fitness by estimating its binding affinity with the receptor. The ligand pose can survive the evaluation if it retains most of the interactions with the binding site, to become a parent ligand of another new generation. This process can be terminated when it reaches a predefined number of generations or by applying RMSD criteria. This method cannot guarantee achieving a global minimum. For example, AutoDock 4, Darwin, and Gold programs use this method.

Tabu search methods: It is a memory-based scoring function which keeps a record of all previous conformations, so that a single conformation is not revisited again. New conformations are accepted or rejected based on a comparison of its RMSD with other conformations. This function can be combined with Monte Carlo or genetic algorithm methods to cross the local energy minimum. For example, PRO_LEADS program uses this method.

Deterministic search: This method is also known as a simulation technique where the ligand either passes through a trajectory-like path signifying the biological behavior of the system with respect to time and conformational space or reconfigures itself to attain a stable state. This method suffers from the drawback that the system gets easily trapped in the local energy minima. Applying simulation temperature to overcome the energy barrier is often seen with molecular dynamics and energy minimization methods. Because of the time-consuming process, generally these are used for refining the outcomes of genetic algorithms or Monte Carlo simulation techniques.

Scoring function: Scoring functions are utilized for determining binding affinity between the receptor’s active site and the ligand by using various mathematical models. Based on these binding affinities, different poses or conformations of ligands can be ranked to separate or filter active compounds. It can also help in the search for an allosteric binding site inside the receptor and optimization of lead compounds. Some of the important factors like solvation model, ions, cofactors, protein flexibility, and the training set structures used for optimizing the scoring method must be taken care of, for getting accurate results. Generally, many of these factors are neglected to accelerate the docking process, which makes the scoring functions less successful in predicting the binding affinity accurately or distinguishing actives from inactives. Various scoring functions are force field, empirical, knowledge-based, consensus, machine learning, and hybrid methods.

Force field: A set of mathematical functions, which helps in estimating the total potential energy of the system, is known as a force field. During ligand-receptor complex formation, various intermolecular interactions are exhibited between the ligand and receptor. These interactions can be classified as bonded
interactions (due to stretching, bending, torsion) or nonbonded interactions (due to electrostatic force, van der Waals force, H-bonding). The force field equation is based on classical physics or molecular mechanics concept where the electrostatic and van der Waals interactions are described by Coulombic energy and Lennard-Jones potential, respectively.

Total potential energy of a system \( E_{\text{Total}} \) is often represented as

\[
E_{\text{Total}} = \sum E_{\text{Stretching}} + \sum E_{\text{Bending}} + \sum E_{\text{Torsion}} + \sum E_{\text{vdw}} + \sum E_{\text{coulombic}}
\]

where \( E_{\text{Stretching}} \) is the bond stretching energy; \( E_{\text{Bending}} \) is the bond energy due to change in bond angle; \( E_{\text{Torsion}} \) is the change in dihedral angle due to change in conformation; \( E_{\text{vdw}} \) is the energy due to van der Waals force; and \( E_{\text{coulombic}} \) is the energy due to electrostatic forces.

The nonbonded interaction energy (comprised of \( E_{\text{vdw}} \) and \( E_{\text{coulombic}} \)) can be represented as

\[
E = \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{rec}} \left[ \frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^6} + \frac{332 q_i q_j}{D r_{ij}} \right]
\]

(2.1)

where \( i \) and \( j \) are two nonbonded atoms situated at a distance of \( r \) from each other; \( A_{ij} \) is the van der Waals repulsive force among the atoms \( i \) and \( j \); \( B_{ij} \) is the van der Waals attractive force among the atoms \( i \) and \( j \); \( q_i \) = charge on atom \( i \); \( q_j \) = charge on atom \( j \); \( D \) = dielectric function; 332 = factor for conversion of electrostatic energy into Kcal/mol.

The key disadvantage of the force field-based method is the absence of a solvation model and entropy term in the estimation of binding energy. Therefore, it is not suitable for complex interactions involving the making or breaking of covalent bonds between the ligand and receptor, e.g., Goldscore and Sybyl/D-Score.

**Empirical:** These functions measure the binding free energy based on weighted structural parameters like polar interactions (hydrogen bonding, ionic interactions) and nonpolar interactions (hydrophobic interactions, desolvation, entropy). A group of ligand-receptor complexes having known binding affinity (experimentally determined) and known binding conformation are used as training set to generate an equation. The multiple linear regression method is utilized to optimize the weight constants as coefficients which on summation can be used for estimation of the binding energy of the complex. Each coefficient or function describes the contribution of individual structural parameters toward total energy. Biasness of training set data hinders empirical scoring function from filtering false actives \( ^{69,75} \) and also hampers the performance for proteins other than those used in training set data. Another drawback is that the quantification of desolvation as well as entropy terms is very difficult with this method. AutoDock, DOCK 6, ChemScore, GLIDE, SCORE, Sybyl-X/F-score are some of the programs using empirical scoring function.
Knowledge based: This method utilizes the pairwise energy potential (interatomic interactions between ligand and receptor), taken from a large database of ligand-receptor complex, to obtain a general scoring function. Ligand-receptor interactions which occur more frequently are assigned with interaction-free energy, which provides a favorable contribution toward binding affinity. The final score is determined by summation of all the interaction free energy obtained from different types of interatomic interactions.

\[
A(r) = -k_B T \ln g_{ij}(r)
\]

\(A(r)\) is the interaction free energy which occurs more frequently; \(k_B\) = Boltzmann constant; \(T\) = absolute temperature; \(i\) and \(j\) are atoms of protein and ligand, respectively, present at a distance \(r\); \(g_{ij}(r)\) = atom pair distribution function for ligand-receptor complex, which is calculated from the database of a protein-ligand complex.

The accuracy of this scoring function depends on the diverse types of interactions present in the database of complexes. One of the advantages of this scoring function is that the implicit treatment of desolvation and entropy terms is possible. DrugScore\textsuperscript{76} and ParaDockS\textsuperscript{77} are examples of knowledge-based scoring function.

Consensus: This method combines several scoring functions to increase the accuracy of a docking result. It can reduce errors arising from the use of a single scoring function and also improves the chances of selecting true actives. However, the use of correlated scoring functions tends to amplify errors and incorrect results. Many studies have reported greater reliability of consensus scoring over a single scoring function.\textsuperscript{78} FlexX and X-CSCORE are examples of consensus scoring functions.

Machine learning methods: These methods have been successfully used in QSAR study for reliable prediction of various physicochemical and druggable parameters. It leads to the development of statistical QSAR models to estimate the binding energy of ligand-receptor complex, e.g., CScore, ID-Score, NNScore 2.0, RFSScore-VS, SFScscoreRF, SVR-EP, and SVR-KB.

Hybrid method: This method uses a combination of two or more scoring functions into a single function to increase the performance for screening inactives, e.g., GalaxyDock BP2 (combination of physics-based, empirical, and knowledge-based scores),\textsuperscript{79} iScore (combination of force field-based and empirical scoring function),\textsuperscript{80} and SMoG2016 (combination of empirical and knowledge-based scoring function).\textsuperscript{81}

Physics-based scoring is an alternative scoring function that can implicitly treat the solvation effect. This method has obtained improved accuracy in determining binding energy, e.g., MM-PBSA (molecular mechanics/Poisson-Boltzmann solvent-accessible surface area) and MM-GBSA (molecular mechanics/generalized Born solvent-accessible surface area).\textsuperscript{82}

A typical docking procedure consists of four steps: ligand setup/preparation, protein setup/preparation, docking, and postdocking analysis.
**Ligand preparation:** The input file format of a chemical structure is very important because it represents the atomic coordinates, bond types, and bond order of a ligand. Docking results along with the molecular interactions may get affected by any of the parameters of the ligand, i.e., protonation state, tautomer, conformer, etc. Therefore, ligand structure is converted into 3D conformation and then refined by energy minimization using molecular mechanics protocol.

**Protein preparation:** A high-resolution X-ray crystallographic structure of a protein is usually preferred over other structures. In case of nonavailability of X-ray structure, the protein structure can be generated from homology modeling also. When several X-ray structures of the protein are available, holoenzyme (with ligand) complex structure is preferred over apoenzyme (without ligand). If many holostructures are available, the one with a cocry stallized ligand at the binding site is preferred. Hydrogens are added to the protein structure, especially polar hydrogens for optimizing the hydrogen bonding network. Missing side chains or residues should be checked and corrected before the final refinement of protein, in which it undergoes minimization to remove any clashes. The binding site information is obtained from the cocry stallized ligand complex at the binding site of the enzyme. It helps in confining the 3D space of the binding site into a grid box, where a suitable binding pose of the ligand is searched. In the case of apoenzyme structures or when the binding site information is unknown, a time-consuming blind docking is performed where the entire protein surface is scanned for a suitable binding site. Currently, many programs are available which can detect the binding sites within an apoenzyme structure based on various pharmacophoric features.

**Docking:** The pioneering work of Kuntz et al. has led to the development of many open-source as well as commercial docking programs. It is comprised of a search algorithm and a scoring function that critically determines the speed and accuracy of docking. It searches and ranks various poses of ligands inside the conformational space available within the binding site of the receptor. Validation of the docking program is carried out by satisfying various benchmarking parameters to prove its speed, accuracy of prediction, and ability to distinguish actives from inactives.

**Postdocking analysis:** The top poses are ranked by the least binding energy score (usually with a negative sign). Poses which show steric or electrostatic clashes may be screened by applying topological filters. Energy minimization of a ligand pose inside the binding site can be performed by another program and analyzed by machine learning methods. Scaffold enrichment may be applied on the hit molecules as an alternative refinement method to recover false negatives, if they share a common structural framework with true positive ligands.

Based on the flexibility of ligands and receptors, docking can be of two types: rigid docking and flexible docking. When both the receptor and ligand are kept rigid, a limited search space is available encompassing only three
rotational and three translational degrees of freedom. A predefined set of ligand conformations can be used to address the ligand flexibility. Various scoring functions like Monte Carlo simulations, simulated annealing, evolutionary, and genetic algorithm methods have been used to incorporate ligand flexibility, whereas the receptor is kept rigid.

Molecular dynamics (MD) simulations are used to treat receptor flexibility, but they consume much computational resources and time.

### 3.1.2 Molecular dynamics (MD)

During molecular interactions, the receptor and especially its binding site can undergo conformational changes, which affects the binding energy as well as stabilization of the ligand-receptor complex. Receptor flexibility has been overlooked by docking methods to achieve speed but compromising with accuracy. Molecular dynamics has been successfully employed for simulation of ligand-receptor binding, conformational sampling, and accurate prediction of the energetics of the system. For simulating the movement of each and every atom of the ligand and receptor, we need to quantify the velocity as well as the force acting on the atoms. Initial potential energy gives information about the coordinates, energy, and velocity for each atom of the system. Applying a force on each atom for a very short span of time (approximately in femtoseconds), we can determine the acceleration from the Newtonian equation of motion, and subsequently velocity and coordinates at a new position for each atom. This process is repeated to get new positions of atoms with respect to the applied force and gradually takes the shape of a trajectory. Once the trajectory is defined, we can simulate the motion of an atom a short time into the future. The new position of the atom at a specific time in the future can be determined from its initial position coordinates and velocity by solving the equation known as Taylor’s expansion. Then, energy minimization of these structures is carried out by molecular mechanics which also helps in the study of conformation and energetics. Two important factors like the temperature of simulation and time steps are critical for carrying out MD simulations. Application of high temperature during simulation helps in overcoming the energy barriers, so that the conformations do not get trapped in local minima and can reach global minima. The selection of a small yet proper time step is essential for searching all the possible conformations. In short, molecular dynamics is referred to as solving the Newtonian equation of motion for all the atoms of the system as a function of time. A common protocol for MD simulation involves the following steps:

**Target structure:** A 3D receptor structure (good resolution), determined by the NMR method or X-ray crystallography, is preferred for only receptor simulation. It can be downloaded from the protein data bank (http://www.pdb.org). In the case of ligand-receptor complex simulation, usually a docking output file is used as input for MD simulations where the ligand is present at the binding site.

**Input structure:** Topology parameters are generated for all the atoms present in the system which contains the necessary information about the atoms, bond
connectivity, angles, their coordinates and velocities, etc. Required hydrogen atoms are added, and protonation states, terminal residues, and disulfide bridges are checked.

**Setting the simulation environment:** A simulation box (similar to the grid box in docking) is created around the binding site and is immersed into a periodic box of water molecules to solvate the protein. The water molecules are represented by various in silico models like simple point charge (SPC) or extended simple point charge (SPC/E), or three-point (TIP3P), or four-point (TIP4P). A proper dielectric constant and force field is selected along with the addition of required counterions to set the simulation environment.

**Energy minimization:** Before MD simulation, a short energy minimization (usually 500 iterations with the steepest descent algorithm) is carried out. It helps in relaxing the structure and also removes high-energy artifacts like broken hydrogen bonds which can distort the entire system.

**Heating up the system:** A high temperature of 1000K is applied for 1–25 picoseconds to stabilize or equilibrate the core structure.

**MD simulation:** The simulation period is set up as per the protein size and the availability of computational resources. The time interval is also set up, at which the output coordinates of the system are recorded for further analysis. For example, a simulation is run for a period of 200 picoseconds with a sampling time of 1 picosecond. Different conformations were recorded at 1 picosecond interval and confined the total number of structures to 200 frames.

**Trajectory analysis:** All the conformers were retrieved, and energy was minimized to rank the structures with the lowest energy binding modes. The stability, as well as the structural integrity of the system, is determined by measuring the root-mean-square deviation of all heavy atoms with respect to the parent structure. Free energy of binding can also be calculated to compare the stability of the complex before and after the simulation.85

Molecular dynamics simulations are helpful in the identification of cryptic binding sites, allosteric binding sites, binding pose of ligand, and accurate estimation of binding affinity. MD simulations have also been used in virtual screening for allowing receptor flexibility during screening. This method is known as a relaxed complex scheme (RCS) where MD simulation is run on the receptor structure to obtain multiple conformations, with which the potential hit candidates can be docked. Therefore, every hit candidate is associated with a series of docking scores and can be ranked based on the average docking score over a receptor. The development of new force fields, conjugation of quantum mechanics, and upgraded computational resources have significantly improved the performance and applications of MD simulation.86

### 3.1.3 Structure-based pharmacophore modeling

As per IUPAC, pharmacophore is very well defined as “A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to
trigger (or to block) its biological response." Various pharmacophoric features include hydrogen bond donor/acceptor, charged center (positive or negative), hydrophobic and aromatic region, metal-binding region, distance, angle, and dihedral angles. When these pharmacophoric features along with their 3D distributions are assigned within the binding site, it can reveal the structural information and interactions responsible for making a compound bioactive. This method can also retrieve structures with diverse bioisosteric scaffolds, which is difficult to explore by systematic derivatization of known compounds. Analysis of various favorable interactions between ligand and the receptor-binding site can be classified under hydrogen bonds (donors/acceptors), electrostatic charge centers (positive/negative), and hydrophobic contacts to generate pharmacophore models. Several pharmacophoric models can be aligned together to produce a common pharmacophore hypothesis or model.

Pharmacophore modeling has been utilized for searching databases, virtual screening, fragment designing, and scaffold hopping and for forecasting bioactivity of hypothetical compounds.

### 3.2 Ligand-based drug design (LBDD)

When protein structure is unknown, the drug designing process takes the lead from available known inhibitors against a specific disease. LBDD is based on the principle that structures having similar chemical features or properties also show similar biological properties too. The 2D or 3D structure of a set of known ligands is analyzed and mapped onto various structural features which can be used to search for new ligands with structural similarity. It can also be correlated with their biological activity to generate a QSAR model to estimate the biological activity of unknown compounds. It includes similarity search, QSAR model, and pharmacophore modeling.

#### 3.2.1 Similarity search

The structure of a single active compound is the minimum information we need for similarity searching. When the structural or physicochemical properties of a molecule are expressed in numerical form, it is called a descriptor. The structure of a known active compound is taken as a template or reference, based on which a library of compounds is screened to find compounds having a similar structure to that of the template. It can be performed by using either a molecular alignment algorithm or molecular descriptors/fingerprint algorithm.

**Molecular alignment algorithm:** A database is filtered and ranked based on its similarity with the template structure. When a compound from the database is aligned with the template, both the molecules may remain rigid or else, can be made flexible by applying incremental construction or genetic algorithm or fragment-based technique.
Molecular alignment is a long-time-taking process, due to which new methods were developed based on rapid comparison of topological and physicochemical descriptors of the database compounds (generated on the fly) with the template. The descriptors may be classified into 1D (molecular weight, octanol/water partition coefficient, molar refractivity), 2D (atom pair, molecular connectivity indices), 3D (polar surface area, nonpolar surface area), etc., based on the dimension of information. Generally, 1D descriptors do not provide information regarding structural or physicochemical properties of compounds and are therefore not suitable for filtering database of compounds. When descriptors of higher dimensions are used, more accuracy can be achieved but at the cost of a long computational time.

**Molecular descriptors/fingerprint algorithm:** A bit string representation of structural and physicochemical parameters of a structure is referred to as molecular fingerprint. It is expressed as a numerical value (signifies the degree of similarity) for a rapid comparison with other molecules. Two different approaches have been developed for generating molecular fingerprints, i.e., keyed representation and hashed representation.

**Keyed representation:** Each of the individual bits of a string is set to be on/off to indicate the presence/absence of a specific functional group. Depending on the ligand structure, each of the bits is either turned on or turned off in a bit-string map. For example, MACCS, a program for virtual screening, uses this method.

**Hashed representation:** It generates a specific pattern for each atom, its neighboring atom, and the connecting bonds. Thus, a unique pattern, similar to human fingerprints, is created for a structure which is used as a filter to screen the database.

**Measures of similarity:** The degree of resemblance or similarity between two structures can be quantified either by Tanimoto/Dice/Cosine coefficient (direct similarity measures) or by Euclidean/Hamming distance (distance similarity measures). Tanimoto coefficient, ranging between 0 (no similarity) and 1 (maximum similarity), is the most popular and widely used similarity measure in different screening programs.

**Data fusion:** When different similarity measures are combined to rank compounds in a database, it is referred as data fusion. This is another similarity search-based screening approach which can significantly enhance search efficiency. It may be carried out either by combining similarity search results for one template with different descriptors (similarity fusion) or by combining the results of a set of templates with one descriptor (group fusion).

### 3.2.2 QSAR modeling

The conventional approach of structure-activity relationship (SAR) in drug design and discovery has brought many successes but not without a great degree of luck. QSAR method correlates various quantifiable physicochemical
properties with biological activity. Usually, this relationship takes the form of an equation which also helps in eliminating the luck factor from the drug design process. Classical QSAR approaches like Free-Wilson and Hansch analysis have correlated the biological activity with certain structural and physicochemical parameters, respectively.  

**Free-Wilson analysis:** Free and Wilson developed a mathematical equation to correlate certain structural features (like the presence/absence of chemical substituents) with bioactivity. It can predict the activity of only those compounds having known substituents that have been included for developing the equation.  

**Hansch analysis:** Hansch followed an extra-thermodynamic approach to develop a model in the form of an equation. He proposed that biological activity can be correlated with various physicochemical factors by a mathematical model. Drug action involves two steps:

1. Transport of drug to the site of action which mainly depends on lipophilic parameters like partition coefficient and substituent hydrophobicity constant.
2. Binding of a drug to the target receptor which mainly depends on electronic (Hammett’s constant) and steric parameters (Taft’s constant, Verloop’s steric parameter).

A typical Hansch equation looks like:

$$\log \frac{1}{C} = k_1 P + k_2 \sigma + k_3 E_s + k_4,$$

where biological activity is expressed as \( \log \frac{1}{C} \) because of the very small value of \( c \); \( P \) is partition coefficient; \( \sigma \) is Hammett’s constant; \( E_s \) is Taft’s steric factor; and \( k_1, k_2, k_3, k_4 \) are constants. Hansch mathematical equation helps in the prediction of new or unknown compounds; it also provides information regarding the mechanism of the drug. However, the accuracy of this model depends on various factors like accuracy of biological activity data, inclusion of large data set, and choice of appropriate parameters. A mixed approach of both Free-Wilson and Hansch model has also been developed to widen the applicability of both the methods. However, three-dimensional parameters cannot be taken into consideration by any of the above two models.  

Usually, the dataset is divided into training and test set molecules. A QSAR model is built on a training set containing diverse chemical structures including active and inactive molecules. The test set molecules are used for testing the validity, predictive capacity, and accuracy of the developed QSAR model. Currently, various multidimensional QSAR models like 3D, 4D, 5D, and 6D QSAR have already been established based on multidimensional descriptors.

### 3.2.3 Ligand-based pharmacophore modeling

In case of nonavailability of protein structure, ligand-based pharmacophore modeling can be employed. It has two essential steps: (1) pharmacophoric
features in training set compounds are analyzed; (2) aligning all the active conformations of ligands in training set based on chemical features or molecular field descriptors. Ligand-based pharmacophore models are of two types:

**Qualitative models:** A set of active ligands (no explicit biological activity data required) of diverse structures are used to generate a common feature-based pharmacophoric hypothesis.

**Quantitative models:** A set of known active compounds (activity data expressed in $K_i$ or $IC_{50}$) are utilized to create QSAR-based predictive pharmacophoric models.

**Examples:** Catalyst includes two alternative algorithms like HypoGen and HipHop for building pharmacophore-based models. HypoGen assigns a certain weighting factor to each of the chemical features of the ligand responsible for bioactivity and constitutes a pharmacophore model. In this way, several pharmacophore hypotheses can be prepared and ranked as per ability to correlate the bioactivity. HipHop explores the surface accessibility of the active ligands suitable for interactions with the receptor to determine their absolute coordinates. Pharmacophore models are prepared based on the chemical features assigned to the absolute coordinates in different conformations. A number of pharmacophore hypotheses can be generated and ranked based on their ability to explain bioactivity.

Disco program adopts a different approach of breaking the pharmacophore into ligand points (hydrogen bonds, charged centers, hydrophobic region) and binding pocket interaction sites (complementary regions within the receptor and is mapped by the coordinates of heavy atoms of ligand). Like catalyst, here also a set of predefined conformations limit the ability to explore the entire conformational space of the ligand.

GASP program utilizes genetic algorithm to search the conformational space to generate different models. A ligand with a minimum number of common chemical features is considered as a reference or template structure. All other compounds (present in the training set) are aligned to this template to evaluate the fitness of a specific pharmacophore model based on similarity, overlaid features, and volume integral of the overlay. Unlike Catalyst and Disco, here overall shape along with any steric clashes between the ligands is taken into consideration during the generation of the final model.

### 3.3 Virtual screening techniques

Drug discovery programs have been considered as challenging and slow processes with a high failure rate. To reduce the burden of cost and time, pressure has been mounting on researchers to identify and separate unsuitable drug candidates in early drug discovery phases. Although high-throughput screening equipped with combinatorial synthesis has been a front step of “hit to lead” identification, reducing a substantial amount of time, still it suffers from drawbacks like consuming valuable resources and time. In the past two decades, virtual screening has been rapidly developed to make the drug discovery...
process more fast, cheap, and reliable. Virtual screening is an in silico method which uses various scoring and ranking functions to screen a large number of databases or yet-to-be synthesized chemical structures against a specific biological target. The concept of virtual screening has been developed from the pioneer works of Kuntz et al. and Desjarlais et al. However, Horvath coined the term “virtual screening” in his research paper based on trypanothione reductase inhibitors. This led to the evolution of a new concept in the field of computational drug design for searching new bioactive agents from a database of compounds. These agents are screened based on the structural parameters, predicted to be complementary to a specific molecular target or enzyme. Advancements in the field of computer hardware and algorithms led to the progress and widespread use of virtual screening as a computational method in the drug discovery process. The significance of virtual screening assisted with other in silico tools can be realized from the fact that more than 50 drug candidates have got green signal to proceed through clinical trials, and some of them also got approved for clinical use.

**Chemical space:** It is the hypothetical space containing all the possible chemical structures, which probably ranges from $10^{18}$ to $10^{180}$ molecules. Analyzing the infinite chemical space for searching bioactive compounds is like finding a needle in the haystack. As the drug candidates belong to various diverse sources (synthetic, natural, marine, peptides, microorganism, etc.), instead of entire chemical space-specific regions a relevant biological target has to be searched. Hence, putting biological activity as a filter enables virtual screening to find the specific biologically active regions of chemical space. Several bioactive compounds failed to become successful drug candidates because of unfavorable physicochemical properties, which adversely affect their absorption, distribution, metabolism, and elimination (ADME). This led to the evolution of drug-likeness and ADME parameters, which facilitated the virtual screening procedure and its success rate in finding the active medicinal space.

**Database:** An essential part of virtual screening includes preparation of compounds database, where compounds can be stored in 2D (SMILES) or 3D chemical structure formats (MDL SD, Sybyl mol2, CML, PDB, XYZ). Another open-source format developed by IUPAC is International Chemical Identifier (InChI) which can encode chemical structures and is able to identify various protomeric and tautomeric states. These structures are usually annotated with other information like molecular weight, synthetic source, amount available, stereochemistry, tautomer, conformers, and protonation state. These structural data along with various physicochemical as well as biological properties help in screening a database by removing the undesirable compounds, which in turn enriches the database with desirable compounds. Researchers showed keen interest in annotated compounds databases, which have information regarding both chemical structure and its possible biological activity. Therefore, it has gradually succeeded over traditional compound databases.
containing information on chemical structures only.\textsuperscript{110} Drug molecules collected from various sources have been broken into individual fragments using the retrosynthetic principle and again combined in every possible way to create a library of virtual compounds. Virtual combinatorial libraries have a tremendous impact in extending the diverse range of chemical space, which are now available for screening.\textsuperscript{111} A few examples of databases (proteins, nucleic acids, complexes as well as ligands) available in the public domain for free as well as commercial use are as follows: AntiBase, BindingDB, BraMMT (Brazilian Malaria Molecular Targets), ChEMBL, ChemSpider, CMNPD, COlleCtion of Open NatUral producTs (COCONUT), DrugBank, DrugSpaceX, EDULISS, eMolecules, GOSTAR, MCDB, MDDR (MDL Drug Data Reports), MMsINC, OOMT (Our Own Molecular Targets), PubChem, and ZINC.

Identification of a bioactive compound against a specific biological target having minimal adverse effects is the primary goal of a drug discovery program. But the presence of promiscuous compounds, frequent hits, and screening artifacts can overwhelm the actual active compounds, which poses a great challenge before the researchers.\textsuperscript{112–114} Pharmacological promiscuous compounds, which act on multiple biological targets, often come out as successful hits in virtual screening, but later on they are found to be nondrug-like. They have a noncompetitive mechanism and poor selectivity as well as a structure-activity relationship. When certain compounds interfere with the assay method and give a false-positive result, they are called artifacts in virtual screening. Various filters are applied to identify and remove such type of nuisance compounds for improving the efficiency of virtual screening.

\textit{Classification:} Depending on the knowledge of biological target structure, virtual screening techniques can be broadly categorized into two types: structure-/target-based virtual screening and ligand-based virtual screening.

### 3.3.1 Structure- or target-based virtual screening

It involves the ranking of ligands as per their affinity with the biological target, as evinced by the nature of interactions during the formation of the ligand-receptor complex.\textsuperscript{115} The ligands may be screened or categorized based on their affinity toward different biological targets.\textsuperscript{116} Knowledge about 3D structure of receptors is essential, which are developed either by X-ray crystallography, NMR spectroscopy, electron microscopy or by homology modeling.\textsuperscript{73,117} The protein structure must be checked for structural disorder or missing residues which can be rectified. Identification of a binding site within the protein structure is another prerequisite. A binding site may be associated with metal ions or water molecules which play an important role in ligand binding, and this information is essential during setting up a virtual screening. A number of binding site detection algorithms are being used by various computational tools for the identification of binding sites inside the protein structure.\textsuperscript{118} Docking and receptor-based pharmacophore modeling are the two elementary methods for carrying out structure-based virtual screening.
Docking: Since the development of the molecular docking technique, it has been proposed to be used as a filter in virtual screening. Docking is an in silico method which predicts the interactions and probable binding conformation of ligand molecules inside the binding cavity of receptor structure. Prediction of interactions, less computational time, and cost make docking a preferred method for executing virtual screening.

A typical docking-based virtual screening consists of four steps: ligand setup/preparation, protein setup/preparation, docking, and postdocking analysis, which have been discussed already. The large database of compounds should be downsized before docking for removing false positives and unsuitable structures by applying filters like 2D or 3D pharmacophoric features, drug-likeness properties, chemical reactivity, etc.

Application of constraints in virtual screening: Various constraints which are applied in virtual screening can be divided into three classes. Covalent interaction-based constraints can filter out ligands having specific covalent interactions with the receptor. Conformational space-based constraints can screen ligands occupying a specific region in the conformational space of the binding site. Pharmacophoric or scaffold-based constraints may be applied as either predocking filters to filter out unsuitable structures, or as postdocking filters to select similar binding poses, which satisfies the pharmacophoric criteria. Alternatively, shape and similarity-based constraints and motif-based constraints may also be utilized in virtual screening to ensure a similar binding manner of different ligands. A flowchart of docking-based virtual screening is presented in Fig. 2.1.

![Flowchart of docking-based virtual screening](image-url)
**Demerits:** Docking methods could not perform simulations with sufficient receptor flexibility or on protein with an induced-fit mechanism. Simulations of ligand and receptor in a polar medium, illustrating the effect of metal ions and assigning correct protonation state to the atoms, could not be achieved by docking.  

**Receptor-based pharmacophore modeling:** Structure- or receptor-based pharmacophore modeling can be applied in virtual screening for getting new leads from a database of compounds. This method can describe the molecular interactions within the binding site and emerge as an alternative method of virtual screening by overcoming the barriers observed with the docking method. Various pharmacophoric features are mapped onto the ligand structure to derive structural information and interactions responsible for making a compound bioactive. It helps to retrieve structures with diverse bioisosteric scaffolds, which is difficult to explore by systematic derivatization of known compounds. Several pharmacophoric models can be aligned together to create a common pharmacophore hypothesis or model. This hypothesis can be used as a filter for screening the database to find hits with novel scaffolds, e.g., FLAP, GBPM, GRID, LigandScout, MOE (Molecular Operating Environment; http://www.chemcomp.com/), and Unity (Tripos; http://www.tripos.com/).

**Ligand-based virtual screening:** This method is used when the protein structure is not known. The structure of known active and inactive compounds is used as templates, based on which algorithms search for new compounds having structural similarity with the templates. Ligand-based virtual screening can be performed with the help of three methods: similarity search, ligand-based pharmacophore modeling, and machine learning method.

**Similarity search:** The structure of a single active compound is the minimum information we need for similarity searching. The selection of appropriate descriptors is very important in carrying out similarity search-based screening. A compound with known activity is taken as a template or reference, based on which a library of compounds can be screened and ranked accordingly. It can be performed by using either a molecular alignment algorithm or molecular descriptors/fingerprint algorithm.

**Ligand-based pharmacophore modeling:** When receptor structure is not known, ligand-based pharmacophore modeling can be used for virtual screening. A single/group of compounds(s), with known activity against a specific target, can be analyzed to identify different chemical features from its structure. Different conformations are generated for each molecule and then aligned together to map the corresponding features. However, two different approaches can be used for the generation of ligand-based pharmacophore models: (i) A database with predefined conformations for each ligand can be used as a filter to speed up the screening, which needs a huge storage facility for handling a huge number of conformations; and (ii) a single conformation of a known active compound can be used for generating different conformations, followed by aligning these conformations with the database structures to generate models, which can be used as a filter to screen the
database. Although it does not need much storage facility, it is very slow. Catalyst, Disco, and GASP are some of the programs that utilize the ligand-based pharmacophore method.

**Machine learning method:** It is an application of artificial intelligence to create a model, built on a group of experimentally determined actives and inactives. It can predict the activity of an unknown compound against a specific target and also can distinguish active compounds from the inactive ones. Regression models can be built by using training set compounds which correlate activity with the structural information. The machine learning method also utilizes the information of inactive compounds to harvest structure-activity relationship among the dataset compounds. These generated models may be used as a filter in screening large databases. Various predicted ADME and other properties can be utilized along with the machine learning method to downsize the hit list. These are of two types: unsupervised and supervised.

**Unsupervised methods:** These methods utilize the descriptor information to correlate the biological activity with the dataset structures. It helps in identifying a specific region of the dataset containing predominantly active or inactive compounds. Since few parameters are used to build robust models, overfitting does not happen with this method, e.g., principal component analysis (PCA), K-means clustering, and self-organizing map.

**Supervised methods:** A group of compounds or a subset of the total dataset molecules with known actives and inactives are selected to form a training set to build a model. The remaining dataset compounds (test set) are used to examine the predictive capacity of the model, known as cross-validation, which is essential to avoid overfitting. Several models are built by taking different training and test set compounds, out of which a single best cross-validated model is chosen, e.g., decision tree (recursive partitioning), K-nearest neighbor, artificial neural networks, and support vector machines.

### 3.3.2 Successful applications of virtual screening

A ligand-based virtual screening of approximately 718,000 commercially available compounds was carried out based on three known glucocorticoid receptor antagonists as query structures. A 3D molecular similarity-based filter and clustering technique was utilized to downsize the database, which followed by lead identification found a compound with good activity ($K_i = 16 \text{nM}$). Further optimization led to the discovery of CORT118335, phase II clinical candidate for management of nonalcoholic steatohepatitis and schizophrenia.

Virtual screening was carried out on the AZ corporate database of about 1 million compounds based on 10 known fibrinolysis inhibitors as query molecules. A 3D electrostatic and shape-based similarity approach was utilized to obtain an active compound 4-PIOL, which on optimization led to the clinical trial candidate AZD6564 for treatment hemorrhage.
Liang et al. have successfully screened the covalent natural products database using herb-based mapping to identify the active compounds baicalein and baicalin showing PLK-1 inhibitory activity.\textsuperscript{127}

Burggraaff et al. have successfully carried out a statistical and structure-based virtual screening for the discovery of several RET kinase inhibitors.\textsuperscript{128}

Rollinger et al. discovered two novel acetylcholinesterase inhibitors (scopolin and scopoletin) by structure-based pharmacophore screening of 110,000 natural products database by using protein structure and a known inhibitor.\textsuperscript{129}

Advanced computational resources are being implemented in virtual screening to search lead compounds and to assist hit finding procedure by preselecting compounds for biological evaluation. Although the positive hits obtained from the virtual screening must undergo experimental screening, it saves time and valuable resources which get wasted on the synthesis and activity of random compounds. Undoubtedly, virtual screening makes the drug discovery process fast, efficient, and more economic.

3.4 ADME analysis and measures of drug-likeness

Usually, 9 out of 10 research projects in drug discovery course face end-stage failure. These 10 projects involve the synthesis of about 10,000–20,000 molecules, followed by their activity studies. The failure rate of the drug discovery process in the pharmaceutical industry is too high, usually about 99.99\%.\textsuperscript{130} These failures come with a huge price of approximately $500 million and $2 billion.\textsuperscript{131} Many candidates undergo failure in late-stage clinical studies due to poor ADME (absorption, distribution, metabolism, elimination) properties. Therefore, various computational models were developed to predict ADME properties before clinical studies. It helps in the preselection of good drug-like candidates for synthesis and activity studies, reduces the failure rate and cost involved in a clinical trial by removing compounds with bad ADME profile, and improves the understanding to correlate experimental and predicted ADME parameters.\textsuperscript{132} Gradually, various drug transport models like Caco-2 cell permeability representing intestinal absorption were also incorporated. Eventually, the toxicity profile gets coupled with other ADME-related properties to develop ADMET parameters which have been successfully used in virtual screening procedures for filtering large databases to select hit molecules. Data obtained from high-throughput in vitro screening assays are used by numerous computational methods and descriptors to build ADMET models.\textsuperscript{133} Various molecular property-based descriptors like polar surface area, hydrogen bonding network, octanol/water partition coefficient, or semiempirical-based descriptors can help in the quantification of pharmacokinetic or ADME properties, which is further correlated with its 3D structure by suitable models.\textsuperscript{134}

*Measures of drug-likeness:* Qualitative evaluation of basic descriptors is the most common way for examining the ADME profile of a molecule, as proposed by Lipinski’s “rule of five.” A compound is considered to have poor absorption
if it violates any two of the proposed parameters (molecular weight not more than 500; number of H-bond acceptors and donors not more than 10 and 5, respectively; calculated log$P$ not more than 5). Researchers subclassified the descriptors based on oral and nonoral drugs, different target diseases to examine their effects on the ADME parameters. Vieth et al. evaluated 1729 marketed drugs and reported a statistically significant difference between injectable drugs (high molecular weight, more polar) and oral drugs (low molecular weight, less polar). Moreover, the pharmacokinetic parameters for injectable drugs were more flexible as compared to oral drugs. It led to the conclusion that pharmacokinetic parameters need biased property distribution as per different targets and routes of administration.

**Aqueous solubility and lipophilicity:** Aqueous solubility is an essential parameter to be predicted for molecules targeting the oral route of delivery. Good solubility is highly necessary for in vitro, in vivo assays and for predicting the absorption in the gastrointestinal tract. Poor solubility negatively affects absorption and assay results and increases the development cost. Therefore, various quantitative structure-property relationship (QSPR) models have been established for the prediction of aqueous solubility using numerous molecular descriptors. Since the composition of gastrointestinal fluids is not taken into account, aqueous solubility cannot be considered as an optimal model for predicting solubility. Yalkowsky and Jain have developed an in silico model called “general solubility equation” to predict aqueous solubility with good accuracy. Lipophilicity of a drug helps itself in getting dissolved in lipid phase, and thus it can pass through the bilayer lipid membranes in the gastrointestinal tract, which can be predicted by the descriptors: Log$P$ (octanol/water partition coefficient), Log$D$ (distribution coefficient), and chromatographic hydrophobicity index.

**Ionization state:** Represented by p$K_a$, it affects the solubility, lipophilicity, and permeability of a compound and thus becomes essential for good absorption of oral drug candidates. Hammett and Taft’s approach, semiempirical methods, and density functional theory are some of the methods which are used for the prediction of ionization state.

**Permeability:** Transport of a drug across the membrane by a passive method is referred to as permeability. Apart from the in vitro determination of partition coefficient and distribution coefficient, several cell line-based assays have been developed. Caco-2 (human colorectal carcinoma) cell line can be modeled to predict intestinal permeability of drugs. This model has been utilized in the early stages of drug discovery for ranking compounds based on absorption and permeability. Similarly, MDCK (Madin-Darby canine kidney) cell-based assay can also be modeled to predict permeability as well as drug-receptor interactions.

**Blood-brain barrier:** Both the aqueous solubility and lipophilicity of a compound determine its capability to penetrate the blood-brain barrier by passive diffusion. It was measured by the parameter LogBB, determined by the ratio
of concentration of drug in brain to concentration of drug in blood. Currently, several models have been developed for the accurate prediction of LogBB by using various machine learning methods.

**Distribution:** Volume of distribution can be predicted by various in silico models, which correlates the lipophilicity and solubility descriptors with free and bound fraction of drug with plasma proteins. Amo et al. have established a model to estimate the volume of distribution whose accuracy was comparable with commercial counterpart Volsurf+.141

**Metabolism:** Various in silico models can predict the site of metabolism along with its substrate nature against a specific metabolic enzyme. A vast dataset of diverse chemicals can be taken to generate models and then converted into online prediction tools for determining pharmacokinetic parameters related to metabolism, e.g., fast metabolizer, SMARTCyp. 142

**Excretion:** Kusama et al. have developed a chemoinformatic-based model based on molecular weight, lipophilicity, charge, and protein-bound fraction in plasma. It helped in the prediction of major clearance pathways of 141 drugs with good accuracy.142

However, lack of a larger experimental dataset is a major obstacle in the development of more accurate ADME models. Yet, computational chemists strive to develop good predictive ADME models to assist the drug discovery project.

### 4 Conclusion

CADD has immensely helped medicinal chemistry researchers to bypass or hasten multiple procedures in the drug design and discovery to find out potent clinical candidates in a short period of time. CADD is very useful in critical steps such as the hit-to-lead discovery and lead optimization; therefore, it paves the way for time as well as cost deduction. Discovery through creating three-dimensional structures of ligand and protein, simulation, prediction of binding interactions and energy would be a very tedious and time-consuming task. However, as compared to conventional drug design and discovery, CADD has several advantages and is classified into SBDD and LBDD. Docking, molecular dynamics, and pharmacophore modeling are the essential steps in SBDD. Similarity search, QSAR model, and pharmacophore modeling are part of LBDD. In the drug discovery paradigm, establishing the drug-likeness with the assistance of Lipinski’s rule of five or Veber’s parameters or rule of three could be a key approach, which determines the drug-like candidates in a reasonable quick timeline. The different in silico models predict various parameters of lead compounds, aqueous solubility, lipophilicity, ionization state, permeability, distribution, metabolism, and excretion. Hence, a detailed pharmacokinetic profile obtained from in silico methods would facilitate a robust approach in drug design, discovery, and development. Nonetheless, we have many innovative techniques in medicinal chemistry, and the discovery of advances must be encouraged to further reduce cost and time duration.
References

1. Schwardt O, Kolb H, Ernst B. Drug discovery today. Curr Top Med Chem 2003;3(1):1–9. https://doi.org/10.2174/1568026033392642.
2. Tan SY, Tsutsumura Y. Alexander fleming (1881–1955): discoverer of penicillin. Singapore Med J 2015;56(7):366–7. https://doi.org/10.11622/smmedj.2015105.
3. Adam M. What to expect from rational drug design. Expert Opin Drug Discovery 2007;2(6):773–6. https://doi.org/10.1517/17460441.2.6.773.
4. Congreve M, Murray CW, Blundell TL. Keynote review: structural biology and drug discovery. Drug Discov Today 2005;10(13):895–907. https://doi.org/10.1016/S1359-6446(05)03484-7.
5. Blundell TL. Structure-based drug design. Nature 1996;384(6604):23–6. https://doi.org/10.1038/384023a0.
6. Amzel LM. Structure-based drug design. Curr Opin Biotechnol 1998;9(4):366–9. https://doi.org/10.1016/S0958-1669(98)80009-8.
7. Whittle PJ, Blundell TL. Protein structure-based drug design. Annu Rev Biophys Biomol Struct 1994;23:349–75. https://doi.org/10.1146/annurev.bb.23.060194.002025.
8. Khanna I. Drug discovery in pharmaceutical industry: productivity challenges and trends. Drug Discov Today 2012;17(19–20):1088–102. https://doi.org/10.1016/j.drudis.2012.05.007.
9. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Molecules 2015;20(7):13384–421. https://doi.org/10.3390/molecules200713384.
10. Llanos GG, Araujo LM, Jiménez IA, et al. Structure-based design, synthesis, and biological evaluation of withaferin A-analogues as potent apoptotic inducers. Eur J Med Chem 2017;140:52–64. https://doi.org/10.1016/j.ejmech.2017.09.004.
11. Mandal S, Moudgil M, Mandal SK. Rational drug design. Eur J Pharmacol 2009;625(1–3):90–100. https://doi.org/10.1016/j.ejphar.2009.06.065.
12. Kumar A, Zhang KYJ. Investigation on the effect of key water molecules on docking performance in CSARdock exercise. J Chem Inf Model 2013;53(8):1880–92. https://doi.org/10.1021/ci400052w.
13. Salsbury FR. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. Curr Opin Pharmacol 2010;10(6):738–44. https://doi.org/10.1016/j.coph.2010.09.016.
14. Kroemer RT. Structure-based drug design: docking and scoring. Curr Protein Pept Sci 2007;8(4):312–28. https://doi.org/10.2174/138920307781369382.
15. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 2004;3:935–49. https://doi.org/10.1038/nrd1549.
16. Anderson AC. The process of structure-based drug design. Chem Biol 2003;10(9):787–97. https://doi.org/10.1016/j.chembiol.2003.09.002.
17. Kalyaanamoorthy S, Chen YPP. Structure-based drug design to augment hit discovery. Drug Discov Today 2011;16(17–18):831–9. https://doi.org/10.1016/j.drudis.2011.07.006.
18. Wang X, Song K, Li L, Chen L. Structure-based drug design strategies and challenges. Curr Top Med Chem 2018;18(12):998–1006. https://doi.org/10.2174/156802618666180813152921.
19. Acharya C, Coop A, Polli JE, Mackerell AD. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Curr Comput Aided Drug Des 2011;7(1):10–22. https://doi.org/10.2174/157340911793743547.
20. Bacilieri M, Moro S. Ligand-based drug design methodologies in drug discovery process: an overview. Curr Drug Discov Technol 2006;3(3):155–65. https://doi.org/10.2174/157016306780136781.
21. Marrone TJ, Briggs JM, McCammon JA. Structure-based drug design: computational advances. Annu Rev Pharmacol Toxicol 1997;37:71–90. https://doi.org/10.1146/annurev.pharmtox.37.1.71.
22. Gane PJ, Dean PM. Recent advances in structure-based rational drug design. Curr Opin Struct Biol 2000;10(4):401–4. https://doi.org/10.1016/S0959-440X(00)00105-6.
23. Jhoti H, Leach AR. Structure-based drug discovery. United Kingdom: Springer Netherlands; 2007. p. 1–249. https://doi.org/10.1007/1-4020-4407-0.
24. Chang C, Ekins S, Bahadduri P, Swaan PW. Pharmacophore-based discovery of ligands for drug transporters. Adv Drug Deliv Rev 2006;58(12–13):1431–50. https://doi.org/10.1016/j.addr.2006.09.006.
25. Akamatsu M. Current state and perspectives of 3D-QSAR. Curr Top Med Chem 2002;2(12):1381–94. https://doi.org/10.2174/1568026023392887.
26. Ekins S. Computer applications in pharmaceutical research and development. United States: John Wiley and Sons; 2006. p. 1–805. https://doi.org/10.1002/0470037237.
27. Brocks DR, Jamali F. Stereochemical aspects of pharmacotherapy. Pharmacother: J Human Pharmacol Drug Ther 1995;15(5):551–64. https://doi.org/10.1002/j.1875-9114.1995.tb02863.x.
28. Jamali F, Mehrar R, Pasutto FM. Enantioselective aspects of drug action and disposition: therapeutic pitfalls. J Pharm Sci 1989;78(9):695–715. https://doi.org/10.1002/jps.2600780902.
29. Verma RP, Hansch C. Camptothecins: a SAR/QSAR study. Chem Rev 2009;109(1):213–35. https://doi.org/10.1021/cr0703723.
30. Mather LE, Rutten AJ. Stereochemistry and its relevance in anaesthesiology. Curr Opin Anaesthesiol 1991;4(4):473–9. https://doi.org/10.1097/00003495-199108000-00002.
31. Hutt AJ, Tan SC. Drug chirality and its clinical significance. Drugs 1996;52(5):1–12. https://doi.org/10.2165/00003495-199600525-00003.
32. Ariëns EJ. Stereochemistry: a source of problems in medicinal chemistry. Med Res Rev 1986;6(4):451–66. https://doi.org/10.1002.med.2610060404.
33. Crom WR. Effect of chirality on pharmacokinetics and pharmacodynamics. Am J Hosp Pharm 1992;49(9):S9–S14. https://doi.org/10.1093/ajhp/49.9_suppl_1.s9.
34. Wainer I, Marcotte A. Stereochimistry analytical methods and pharmacology. 2nd ed. Marcel Dekker; 1993. p. 25–34.
35. Frank A, Peirce. 06—Fundamental psychopharmacology. Compr Clin Psychol 1998;11:173–206.
36. Langley JN. On the reaction of cells and of nerve-endings to certain poisons, chiefly as regards the reaction of striated muscle to nicotine and to curari. J Physiol 1905;33(4–5):374–413. https://doi.org/10.1113/jphysiol.1905.sp001128.
37. Clark AJ. The antagonism of acetylcholine by atropine. J Physiol (Lond) 1926;61:547–66.
38. Langmuir I. The adsorption of gases on plane surfaces of glass, mica and platinum. J Am Chem Soc 1918;40(9):1361–403. https://doi.org/10.1021/ja02242a004.
39. Stephenson RP. A modification of receptor theory. Br J Pharmacol Chemother 1956;11(4):379–93. https://doi.org/10.1111/j.1476-5381.1956.tb00006.x.
40. Koshy DE. Application of a theory of enzyme specificity to protein synthesis. Proc Natl Acad Sci 1958;98–104. https://doi.org/10.1073/pnas.44.2.98.
42. Ariens EJ. Affinity and intrinsic activity in the theory of competitive inhibition. I. Problems and theory. Arch Int Pharmacodyn Ther 1954;99(1):32–49.
43. Kenakin T. Principles: receptor theory in pharmacology. Trends Pharmacol Sci 2004;25(4):186–92. https://doi.org/10.1016/j.tips.2004.02.012.
44. Stella VI, Borchardt RT, Kerns EH, Hageman MJ, Thakker DR. Optimizing the “drug-like” properties of leads in drug discovery. Biotechnol: Pharm Aspects 2006. https://doi.org/10.1007/978-0-387-44961-6_10.
45. Tietgen H, Walden M. Physicochemical properties. In: Vogel HG, Maas J, Hock FJ, Mayer D, editors. Drug discovery and evaluation: safety and pharmacokinetic assays. Berlin, Heidelberg: Springer; 2013. p. 1125–38.
46. Beaumont K, Webster R, Gardner I, Dack K. Design of ester prodrugs to enhance oral absorption of poorly permeable compounds: challenges to the discovery scientist. Curr Drug Metab 2003;4(6):461–85. https://doi.org/10.2174/1389200033489253.
47. Ettmayer P, Amidon GL, Clement B, Testa B. Lessons learned from marketed and investigational prodrugs. J Med Chem 2004;47(10):2393–404. https://doi.org/10.1021/jm0303812.
48. Testa B. Prodrug research: futile or fertile? Biochem Pharmacol 2004;68(11):2097–106. https://doi.org/10.1016/j.bcp.2004.07.005.
49. Stella VJ, Nti-Addae KW. Prodrug strategies to overcome poor water solubility. Adv Drug Deliv Rev 2007;59(7):677–94. https://doi.org/10.1016/j.addr.2007.05.013.
50. Rang DM, Ritter J, Rang FR. Dale’s. Pharmacology 2008;3–112.
51. Rautio J, Kumpulainen H, Heimbach T, et al. Prodrugs: design and clinical applications. Nat Rev Drug Discov 2008;7(3):255–70. https://doi.org/10.1038/nrd2468.
52. Waller DG, Renwick A, Hillier K. Medical pharmacology and therapeutics. UK: Oxford/Saunders (W.B.) Co Ltd; 2006. p. 3–85.
53. Wolverton SE. Basic principles of pharmacology. Elsevier BV; 2021. p. 1–11. https://doi.org/10.1016/b978-0-323-61211-1.00001-2.
54. Bryant B, Knights K, Salerno E. Pharmacology for health professionals. Elsevier; 2007. p. 1–215.
55. Golan DE, Tashjian AH, Armstrong EJ. Principles of pharmacology: the pathophysiologic basis of drug therapy. Lippincott Williams & Wilkins; 2011. p. 1–79.
56. Greenberg DM, Schulman MP. Application of “metabolite antagonism” to cancer research. Science 1947;106(2751):271–2. https://doi.org/10.1126/science.106.2751.271.
57. Shive W, Skinner CG. Metabolic antagonists. Annu Rev Biochem 1958;27(3):643–78. https://doi.org/10.1146/annurev.bi.27.070158.003235.
58. Lemke TL, Williams DA, Roche VF, Zito SW. Foye’s principles of medicinal chemistry. 7th ed. United States: Wolters Kluwer Health Adis (ESP); 2012. p. 1–1500. http://www.lww.com/Product/9781609133450.
59. Sun LQ, Wong JP. Frontiers in nucleic acid-based drug research and development. Future Med Chem 2015;7(13):1619–21. https://doi.org/10.4155/fmc.15.117.
60. Wu-pong S, Rojanasakul Y. Biopharmaceutical drug design and development. Springer; 2008. p. 196–221.
61. Hefti FF. Requirements for a lead compound to become a clinical candidate. BMC Neurosci 2008;9(3). https://doi.org/10.1186/1471-2202-9-S3-S7.
62. Ripka AS, Rich DH. Peptidomimetic design. Curr Opin Chem Biol 1998;2(4):441–52. https://doi.org/10.1016/S1367-5931(98)80119-1.
63. Vagner J, Qu H, Hruby VJ. Peptidomimetics, a synthetic tool of drug discovery. Curr Opin Chem Biol 2008;12(3):292–6. https://doi.org/10.1016/j.cbpa.2008.03.009.
64. Patwardhan B, Mashelkar RA. Traditional medicine-inspired approaches to drug discovery: can Ayurveda show the way forward? Drug Discov Today 2009;14(15–16):804–11. https://doi.org/10.1016/j.drudis.2009.05.009.
65. Surh YJ. Reverse pharmacology applicable for botanical drug development—inspiration from the legacy of traditional wisdom. J Tradit Complement Med 2011;1(1):5–7. https://doi.org/10.1016/S2225-4110(16)30051-7.

66. Cha Y, Erez T, Reynolds JJ, et al. Drug repurposing from the perspective of pharmaceutical companies. Br J Pharmacol 2018;175(2):168–80. https://doi.org/10.1111/bph.13798.

67. Barakat KH, Mane JY, Tuszyński JA. Virtual screening: an overview on methods and applications. In: Handbook of research on computational and systems biology: interdisciplinary applications. Canada: IGI Global; 2011. p. 28–60. https://doi.org/10.4018/978-1-60960-491-2.ch002.

68. Abagyan R, Totrov M. High-throughput docking for lead generation. Curr Opin Chem Biol 2001;5(4):375–82. https://doi.org/10.1016/S1367-5931(00)00217-9.

69. Bissantz C, Fokkers G, Rognan D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. J Med Chem 2000;43(25):4759–67. https://doi.org/10.1021/jm0010441.

70. Ruiz-Tagle B, Villalobos-Cid M, Dorn M, Inostroza-Ponta M. Evaluating the use of local search strategies for a memetic algorithm for the protein-ligand docking problem. In: 2017-Proceedings - international conference of the Chilean Computer Science Society, SCCC, 2017: 2018. p. 1–12. https://doi.org/10.1109/SCCC.2017.8405141.

71. Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD. Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins Struct Funct Genet 1998;33(3):367–82. https://doi.org/10.1002/(SICI)1097-0134(19981115)33:3<367::AID-PROT6>3.0.CO;2-W.

72. Li GB, Yang LL, Wang WJ, Li LL, Yang SY. ID-score: a new empirical scoring function based on a comprehensive set of descriptors related to protein-ligand interactions. J Chem Inf Model 2013;53(3):592–600. https://doi.org/10.1021/ci300493w.

73. Schneider G, Böhm H-J. Virtual screening and fast automated docking methods. Drug Discov Today 2002;7(1):64–70. https://doi.org/10.1016/s1359-6446(01)02091-8.

74. Ash S, Cline MA, Homer RW, Hurst T, Smith GB. SYBYL line notation (SLN): a versatile language for chemical structure representation. J Chem Inf Comput Sci 1997;37(1):71–9. https://doi.org/10.1021/ci960109j.

75. Reddy AS, Pati SP, Kumar PP, Pradeep HN, Sastry GN. Virtual screening in drug discovery—a computational perspective.Curr Protein Pept Sci 2007;8(4):329–51. https://doi.org/10.2174/138920307781369427.

76. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 2000;295(2):337–56. https://doi.org/10.1006/jmbi.1999.3371.

77. Meier R, Pippel M, Brandt F, Sippl W, Baldauf C. ParaDockS: a framework for molecular docking with population-based metaheuristics. J Chem Inf Model 2010;50(5):879–89. https://doi.org/10.1021/ci900467x.

78. Terp GE, Johansen BN, Christensen IT, Jørgensen FS. A new concept for multidimensional selection of ligand conformations (multiselect) and multidimensional scoring (multiscore) of protein-ligand binding affinities. J Med Chem 2001;44(14):2333–43. https://doi.org/10.1021/jm001090l.

79. Baek M, Shin WH, Chung HW, Seok C. GalaxyDock BP2 score: a hybrid scoring function for accurate protein–ligand docking. J Comput Aided Mol Des 2017;31(7):653–66. https://doi.org/10.1007/s10822-017-0030-9.

80. Geng C, Jung Y, Renaud N, Honavar V, Bonvin AMJJ, Xue LC. IScore: a novel graph kernel-based function for scoring protein-protein docking models. Bioinformatics 2020;36(1):112–21. https://doi.org/10.1093/bioinformatics/btz496.
81. Debroise T, Shakhnovich EI, Chéron N. A hybrid knowledge-based and empirical scoring function for protein-ligand interaction: SMoG2016. J Chem Inf Model 2017;57(3):584–93. https://doi.org/10.1021/acs.jcim.6b00610.

82. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112(16):6127–9. https://doi.org/10.1021/ja00172a038.

83. Mizutani MY, Tomioka N, Itai A. Rational automatic search method for stable docking models of protein and ligand. J Mol Biol 1994;243(2):310–26. https://doi.org/10.1006/jmbi.1994.1656.

84. Aguayo-Ortiz R, Fernández-De GE. Overview of computer-aided drug design for epigenetic targets. In: Epi-informatics: discovery and development of small molecule epigenetic drugs and probes. Mexico: Elsevier Inc.; 2016. p. 21–52. https://doi.org/10.1016/B978-0-12-802808-7.00002-2.

85. Lindahl E. Molecular dynamics simulations. Methods Mol Biol 2015;1215:3–26. https://doi.org/10.1007/978-1-4939-1465-4_1.

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

87. Wermuth CG, Ganellin CR, Lindberg P, Mitscher LA. Chapter 36. Glossary of terms used in medicinal chemistry (IUPAC recommendations 1997). In: Annual reports in medicinal chemistry, vol. 33. France: Academic Press Inc.; 1998. p. 385–95. https://doi.org/10.1016/S0065-7743(08)61101-X.

88. Sun H. Pharmacophore-based virtual screening. Curr Med Chem 2008;15(10):1018–24. https://doi.org/10.2174/092986708784049630.

89. Hoffren M, Murray M, Hoffmann R. Structure-based focusing using pharmacophores derived from the active site of 17beta-hydroxysteroid dehydrogenase. Curr Pharm Des 2001;7(7):547–66. https://doi.org/10.2174/1381612013397870.

90. Lemmen C, Lengauer T. Computational methods for the structural alignment of molecules. J Comput Aided Mol Des 2000;14(3):215–32. https://doi.org/10.1023/A:1008194019144.

91. Willett P, Barnard JM, Downs GM. Chemical similarity searching. J Chem Inf Comput Sci 1998;38(6):983–96. https://doi.org/10.1021/ci9800211.

92. Whittle M, Gillet VJ, Willett P, Loesel J. Analysis of data fusion methods in virtual screening: similarity and group fusion. J Chem Inf Model 2006;46(6):2206–19. https://doi.org/10.1021/ci0496144.

93. Kubinyi H. QSAR and 3D-QSAR in drug design. Part 1: methodology. Drug Discov Today 1997;2(11):457–67. https://doi.org/10.1016/S1359-6446(97)01079-9.

94. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational methods in drug discovery. Pharmacol Rev 2014;66(1):334–95. https://doi.org/10.1124/pr.112.007336.

95. Thomas G. The SAR and QSAR approaches to drug design. Fundamentals of medicinal chemistry. Wiley; 2008. p. 71–92.

96. Chandrasekaran A, Kaushik S. In: Ranganathan S, editor. Encyclopedia of bioinformatics and computational biology. Elsevier; 2019. p. 677–87.

97. Barnum D, Greene J, Smellie A, Sprague P. Identification of common functional configurations among molecules. J Chem Inf Comput Sci 1996;36(3):563–71. https://doi.org/10.1021/ci950273r.

98. Martin YC, Bures MG, Danaher EA, DeLazzer J, Lico I, Pavlik PA. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. J Comput Aided Mol Des 1993;7(1):83–102. https://doi.org/10.1007/BF00141577.
99. Jones G, Willett P, Glen RC. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. J Comput Aided Mol Des 1995;9(6):532–49. https://doi.org/10.1007/BF00124324.

100. Walters WP, Stahl MT, Murcko MA. Virtual screening—an overview. Drug Discov Today 1998;3(4):160–78. https://doi.org/10.1016/s1359-6446(97)01163-x.

101. Lahana R. How many leads from HTS? Drug Discov Today 1999;4(10):447–8. https://doi.org/10.1016/s1359-6446(99)01393-8.

102. Wilton D, Willett P, Lawson K, Mullier G. Comparison of ranking methods for virtual screening in lead-discovery programs. J Chem Inf Comput Sci 2003;43(2):469–74. https://doi.org/10.1021/ci025586i.

103. DesJarlais RL, Seibel GL, Kuntz ID, et al. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. Proc Natl Acad Sci U S A 1990;87(17):6644–8. https://doi.org/10.1073/pnas.87.17.6644.

104. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982;161(2):269–88. https://doi.org/10.1016/0022-2836(82)90153-X.

105. Horvath D. A virtual screening approach applied to the search for trypanothione reductase inhibitors. J Med Chem 1997;40(15):2412–23. https://doi.org/10.1021/jm9603781.

106. Jorgensen WL. The many roles of computation in drug discovery. Science 2004;303(5665):1813–8. https://doi.org/10.1126/science.1096361.

107. Gorse AD. Diversity in medicinal chemistry space. Curr Top Med Chem 2006;6(1):3–18. https://doi.org/10.2174/156802606775193310.

108. Hert J, Irwin JJ, Laggner C, Keiser MJ, Shoichet BK. Quantifying biogenic bias in screening libraries. Nat Chem Biol 2009;5(7):479–83. https://doi.org/10.1038/nchembio.180.

109. Kavitha C, Mahalekshmi T. Chemical file format conversion tools: an overview. Int J Eng Res Technol 2014;03(02).

110. Oprea TI, Tropsha A. Target, chemical and bioactivity databases—integration is key. Drug Discov Today Technol 2006;3(4):357–65. https://doi.org/10.1016/j.ddtec.2006.12.003.

111. Lewell XQ, Judd DB, Watson SP, Hann MM. RECAP—retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. J Chem Inf Comput Sci 1998;38(3):511–22. https://doi.org/10.1021/ci970429i.

112. Hüser J, Lohrmann E, Kalthof B, Burkhardt N, Brüggemeier U, Bechem M. High-throughput screening for targeted lead discovery. Wiley; 2006. p. 15–36. https://doi.org/10.1002/9783527609321.ch2.

113. McGovern SL, Caselli E, Grigorieff N, Shoichet BK. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. J Med Chem 2002;45(8):1712–22. https://doi.org/10.1021/jm010533y.

114. Roche O, Schneider P, Zuegge J, et al. Development of a virtual screening method for identification of “frequent hitters” in compound libraries. J Med Chem 2002;45(1):137–42. https://doi.org/10.1021/jm010934d.

115. Maia EHB, Assis LC, de Oliveira TA, da Silva AM, Taranto AG. Structure-based virtual screening: from classical to artificial intelligence. Front Chem 2020;8. https://doi.org/10.3389/fchem.2020.00343.

116. Liu S, Alnammi M, Ericksen SS, et al. Practical model selection for prospective virtual screening. J Chem Inf Model 2019;59(1):282–93. https://doi.org/10.1021/acs.jcim.8b00363.

117. Lyne PD. Structure-based virtual screening: an overview. Drug Discov Today 2002;7(20):1047–55. https://doi.org/10.1016/S1359-6446(02)02483-2.
118. Cole JC, Korb O, Olsson TSG, Liebeschuetz J. The basis for target-based virtual screening: protein structures. Wiley; 2011. p. 87–114. https://doi.org/10.1002/9783527633326.ch4.

119. Kuntz ID. Structure-based strategies for drug design and discovery. Science 1992;257(5073):1078–82. https://doi.org/10.1126/science.257.5073.1078.

120. Ford M, Livingstone D, Dearden W. Designing drugs and crop protectants: processes, problems and solutions. In: Cole J, editor. Onward development of GOLD: selective dockings using known protein properties, in EuroQSAR2002 Blackwell; 2003. p. 140–3.

121. Fradera X, Knegtel R, Mestres J. Similarity-driven docking. Proteins 2000;40:4.

122. Hindle SA, Rarey M, Bunin C, Lengauer T. Flexible docking under pharmacophore type constraints. J Comput Aided Mol Des 2002;16(2):129–49. https://doi.org/10.1023/A:1016399411208.

123. Muthas D, Sabnis YA, Lundborg M, Kärnlén A. Is it possible to increase hit rates in structure-based virtual screening by pharmacophore filtering? An investigation of the advantages and pitfalls of post-filtering. J Mol Graph Model 2008;26(8):1237–51. https://doi.org/10.1016/j.jmgm.2007.11.005.

124. Carpenter KA, Cohen DS, Jarrell JT, Huang X. Deep learning and virtual drug screening. Future Med Chem 2018;10(21):2557–67. https://doi.org/10.4155/fmc-2018-0314.

125. Koeppen H, Kriegi J, Lessel U, Tautermann C, Wellenzohn B. Chapter-3 Ligand-based virtual screening. In: Virtual screening: principles, challenges, and practical guidelines. Wiley-VCH; 2011. p. 61–86.

126. Clark DE. Virtual screening: is bigger always better? Or can small be beautiful? J Chem Inf Model 2020;60(9):4120–3. https://doi.org/10.1021/acs.jcim.0c00101.

127. Liang H, Liu H, Kuang Y, Chen L, Ye M, Lai L. Discovery of targeted covalent natural products against PLK1 by herb-based screening. J Chem Inf Model 2020;60(9):4350–8. https://doi.org/10.1021/acs.jcim.0c00074.

128. Burggraaf L, Lenselink EB, Jespers W, et al. Successive statistical and structure-based modeling to identify chemically novel kinase inhibitors. J Chem Inf Model 2020;60(9):4283–95. https://doi.org/10.1021/acs.jcim.9b01204.

129. Rollinger JM, Hornick A, Langer T, Stuppner H, Prast H. Acetylcholinesterase inhibitory activity of scopolin and scopoletin discovered by virtual screening of natural products. J Med Chem 2004;47(25):6248–54. https://doi.org/10.1021/jm049655s.

130. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 2004;711–716. https://doi.org/10.1038/nrd1470.

131. Adams CP, Brantner VV. Estimating the cost of new drug development: is it really $802 million? Health Aff 2006;420–8. https://doi.org/10.1377/hlthaff.25.2.420.

132. Merz K, Ringe D, Reynolds CH, Egan W. Predicting ADME properties in drug discovery; 2010. p. 165–78.

133. Taft CA, Da Silva VB, CHT DPDS. Current topics in computer-aided drug design. J Pharm Sci 2008;97(3):1089–98. https://doi.org/10.1002/jps.21293.

134. van de Waterbeemd H, Lennermans H, Artursson P. In: Drug bioavailability—estimation of solubility, permeability, absorption and bioavailability. Wiley-VCH; 2003. p. 439–521.

135. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 1997;23(1–3):3–25. https://doi.org/10.1016/S0169-409X(96)00423-1.
136. Vieth M, Siegel MG, Higgs RE, et al. Characteristic physical properties and structural fragments of marketed oral drugs. J Med Chem 2004;47(1):224–32. https://doi.org/10.1021/jm030267j.

137. Wan H, Holmén AG. High throughput screening of physicochemical properties and in vitro ADME profiling in drug discovery. Comb Chem High Throughput Screening 2009;12(3):315–29. https://doi.org/10.2174/138620709787581701.

138. Jain N, Yalkowsky SH. Estimation of the aqueous solubility I: application to organic nonelectrolytes. J Pharm Sci 2001;90(2):234–52. https://doi.org/10.1002/1520-6017(200102)90:2<234::AID-JPS14>3.0.CO;2-V.

139. Krüger A, Gonçalves Maltarollo V, Wrenger C, Kronenberger T. ADME profiling in drug discovery and a new path paved on silica., 2020IntechOpen, https://doi.org/10.5772/intechopen.86174.

140. Volpe DA. Drug-permeability and transporter assays in Caco-2 and MDCK cell lines. Future Med Chem 2011;3(16):2063–77. https://doi.org/10.4155/fmc.11.149.

141. del Amo EM, Ghemtio L, Xhaard H, et al. Applying linear and non-linear methods for parallel prediction of volume of distribution and fraction of unbound drug. PLoS ONE 2013. https://doi.org/10.1371/journal.pone.0074758, e74758.

142. Kusama M, Toshimoto K, Maeda K, et al. In silico classification of major clearance pathways of drugs with their physiochemical parameters. Drug Metab Dispos 2010;38(8):1362–70. https://doi.org/10.1124/dmd.110.032789.