ISBN: 978-81-970675-0-1

3. Computational Chemistry

Dr. Pradeep Kumar M. R., Deepak S. Shanawad,

Basavaraja Pujar, Prateek A. Angadi

Department of Pharmaceutical Chemistry,
KLE College of Pharmacy,
Vidyanagar, Hubli, Karnataka.

3.1 Introduction:

Computational chemistry is at the vanguard of current scientific inquiry, using computers to understand the complexities of chemical systems and processes. Through the application of quantum, molecular, and statistical mechanics, computational chemistry provides a virtual laboratory where scientists may investigate the behavior of atoms and molecules with previously unheard-of precision and detail. This field allows for the computational simulation and analysis of complicated chemical events, ranging from the kinetics of biochemical reactions to the creation of molecular bonds. Computational chemists advance our understanding of chemical structures, characteristics, and interactions by deciphering the underlying principles governing molecular behavior through complex algorithms and numerical approaches. Furthermore, in many other domains where traditional experimental approaches have difficulties with cost, time, and complexity, such as drug development, materials research, and environmental chemistry, computational chemistry is an indispensable tool. Computational chemistry quickens the pace of innovation by forecasting chemical properties and creating new compounds, pointing experimentalists in the direction of promising discoveries and facilitating the creation of ground-breaking technology.

The cover story for Fortune magazine, "The Next Industrial Revolution: Designing Drugs by Computer at Merck," appeared on October 5, 1981. According to others, this marked the beginning of a fervent interest in the possibilities of computer-aided drug design (CADD). While CADD was making headway, high-throughput screening (HTS) was starting to emerge as a more promising method for discovering new treatments, whereas these two are lead methods of computational chemistry in novel drug designing. This brute force method uses automation to sort through a large number of molecules and find the ones that cause the desired physiological reaction. The technique offers the benefit of requiring little prior knowledge or complicated design, and the tools needed to screen big libraries have grown more productive. But whereas conventional HTS frequently yields several hit compounds, some of which may be further developed into a lead and eventually a novel therapy, the hit rate for HTS is frequently incredibly low. The use of HTS has been restricted to research programmes that are able to screen sizable compound libraries due to its poor hit rate.

Over the last ten years, CADD has come back into favour as a means of drastically reducing the number of compounds that must be screened while maintaining the same degree of lead compound discovery. Potential instance, pharma trial researchers screened potential inhibitors of the diabetes-related enzyme tyrosine phosphatase-1B using CADD technologies. Of the chemicals displayed on their virtual screen, 127 of which had a hit rate of around 35% and demonstrated effective suppression. This group executed a conventional HTS against the same target concurrently.

Just 0.021% of the 400,000 chemicals that were examined exhibited inhibition. This example illustrates the effectiveness of CADD in a comparison manner. This introduction delves into the exciting field of computational chemistry, examining its methods, uses, and revolutionary influence on science and technology development.

Now here in this chapter we go for a line of reasoning of computational chemistry in drug discovery/designing (CADD).

3.2 Computer Aided Drug Design (CADD):

3.2.1 History:

An Overview of CADD's Past The receiver and lock-key concept was introduced in 1900 by P. Ehrrich (1909) and E. Fisher. The idea of quantitative structure-activity relationships (QS-AR) was developed in the 1970s, but it had certain drawbacks.

2-Dimensional, retrospective analysis: The era of molecular modelling began in the 1980s with the development of computer graphics, multidimensional NMR, X-ray crystallography, and CADD Molecular Biology. Along with combinatorial chemistry and high throughput screening, other contemporary methods like human genome bioinformatics were brought to the innovative field of medical science in the 1990s.

CADD – There are mainly two types of computer-based drug design, they are;

- Structure Based Drug Design
- Ligand Based Drug Design

3.2.2 Structure Based Drug Design:

Structure-based drug designing is the most effective and potent technique that oversees the complete drug discovery process (SBDD). The most effective resources for quickening the drug discovery process include data regarding small molecule targets, genetic information along with sequences, binding information, cytotoxicity, absorption, metabolism, excretion (ADMET) data, and other significant biological information.

Many R&D, pharmaceutical businesses are using this promising computational technique. Structure prediction, visualisation, characterization of binding sites, virtual screening and molecular docking, docked complex structure visualisation and stability analysis, ADMET screening, and binding-free energy-MM PBSA are all part of the pipeline.The 3D structures of some human and pathogenic proteins have been clarified as a result of the widespread application of biophysical methods including NMR spectroscopy and X-ray crystallography. For instance, the PDB contains about 81,000 protein structures, whereas ligand-protein cocrystal structures are only found in 129 (as of 2003) and 5,671 (as of PDBBIND) data bases and protein ligand data base, respectively. Target structure knowledge has accelerated drug discovery efforts and resulted in the development of multiple therapeutic medicines. Being able to identify possible binders to the target of biological interest quickly is a need for the drug discovery process.

Steps or Process of Structure-Based CADD:

A. Determination of Protein:

structure Several methods, including X-ray crystallography, NMR spectroscopy, and electronmicroscopy, are currently used to determine the 3D structure of a protein, and each method has its uniqueness and limitations.

- X-ray crystallography - the protein is purified and crystallized under suitableconditions, and then subjected to an intensive X-ray beam. The diffraction of an X-ray beam by the protein crystals into one or other patterns is examined to determine the distribution of electron density. Finally, a map of the electron density is generated and interpreted to determine the location of each atom.
- NMR spectroscopy - is used to determine the 3D structure of molecules.
- Electron microscopy - is also used to determine the 3D structures of large molecular assemblies, often referred to as 3DEM. A beam of electrons and an electron lense system are used to directly image the biomolecule.

B. Protein Structure Prediction:

Three computational methods widely used for protein structure prediction are:

(a) homology modeling,
(b) fold recognition,
(c) ab initio method.

Several tools for protein structure predictions are available, which utilizes different approaches and methods for modelling and refinement of the protein structure

C. Identification of binding pocket:

- A cavity on the surface or in the interior of a protein that possesses suitable properties for binding a ligand is usually referred to as a binding pocket. The set of amino acid residues around a binding pocket determines its physicochemical characteristics and, together with its shape and location in a protein, defines its functionality.
- The dynamics of protein binding pockets are crucial for their interaction specificity. Structural flexibility allows proteins to adapt to their individual molecular binding partners and facilitates the binding process.

D. Scoring function:

In drug development and other molecular modelling applications, scoring functions are commonly employed. Among them are:

- Virtual small molecule screening databases of potential ligands to find new, tiny compounds that bind to a desired protein target and are therefore helpful places to start when looking for drugs.
- De novo design of new small molecules that attach to a protein target (design "from scratch")
- Lead optimisation - screening hits to maximise their affinity and selectivity.

There are 4 types of scoring function, those are:

- Force field: affinities are calculated by applying a force field to the sum of the electrostatic and van der Waals forces between each atom in the two molecules that make up the complex. It is also common to include the intramolecular energies (also known as strain energies) of the two binding partners.
- Empirical: based on enumerating the different kinds of interactions that can occur between the two binding partners.
- Knowledge-based: Deriving statistical "potentials of mean force" from statistical observations of intermolecular close interactions in sizable 3D databases (such the Protein Data Bank or Cambridge Structural Database) is the basis of this approach.
- Machine-learning: In contrast to these classical scoring functions, machine-learning scoring functions are distinguished by not supposing a predefined functional form for the connection between the structural characteristics of the protein-ligand complex and binding affinity. In this manner, the functional form is deduced straight from the information. It has been repeatedly observed that machine-learning scoring functions perform better at predicting the binding affinities of various protein-ligand complexes than do conventional scoring systems.

E. Protein-ligand docking algorithms:

- Generally speaking, in a docking method, the bigger molecule is referred to as the receptor and the smaller one as the ligand. Finding the receptor and ligand in the most interactive conformation is the goal of docking techniques.
- Prior to docking, the 3D structures of ligands and receptors are necessary. Predicted structures could be employed if experimentally derived structures are unavailable. The two main distinctions between the various docking algorithms are that they are all comprised of objective energy functions and searching algorithms. Theoretically, proper receptor and ligand docking will result in the minimum of an energy function.
- Given the variability in the modelling of protein-ligand interaction among scoring systems, it is possible to see varying levels of performance when utilising a single scoring function for all docking activities.
  For example - AutoDock4 uses a physics-based force field scoring function that includes van der Waals, electrostatic, and directional hydrogen-bond potentials developed from an early version of the AMBER force field. Furthermore, a basic conformational entropy penalty and a pairwise-additive desolvation term based on partial charges are incorporated. The Lennard-Jones VDW terms and electrostatic terms make up the scoring function. where Aij and Bij are the VDW parameters, rij refers to the distance between the protein atom i and the ligand atom j, and qi and qj are atomic charges.

$$E = \sum i \sum j \left(\frac{Aij}{r_{ij}^{12}} - \frac{Bij}{r_{ij}^{6}} + \frac{qiqj}{\in (r_{ij}) r_{ij}} \right)$$

F. virtual screening:

- A computational method called virtual screening (VS) is used in the search for new drugs by looking through libraries of small molecules to find the structures that have the highest probability of binding to an enzyme or protein receptor, which is the target of the medicine.
- Virtual screening is characterised as "automatically evaluating very large libraries of compounds" using computer programmes. As this statement implies, the main focus of VS has been on how to reduce the vast chemical space of over
- 1060 possible molecules to a manageable number that can be acquired, evaluated, and synthesised. More realistic virtual screening (VS) scenarios centre on developing and optimising selected combinatorial libraries and enriching libraries of existing compounds from vendor offerings or internal compound repositories, even though scanning the entire chemical universe may be an intriguing task in theory. Virtual screening has become a crucial step in the drug discovery process as the method's accuracy has grown.

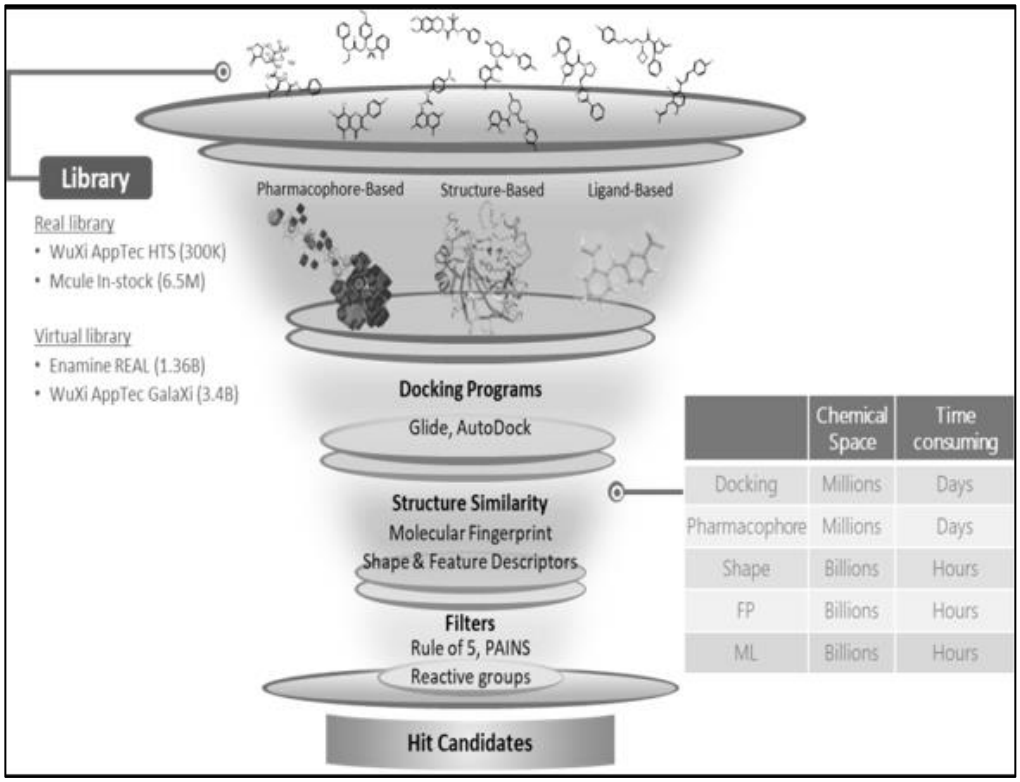

Figure 3.1: virtual screening

- Two types of VS they are:

i. Structure based virtual screening:

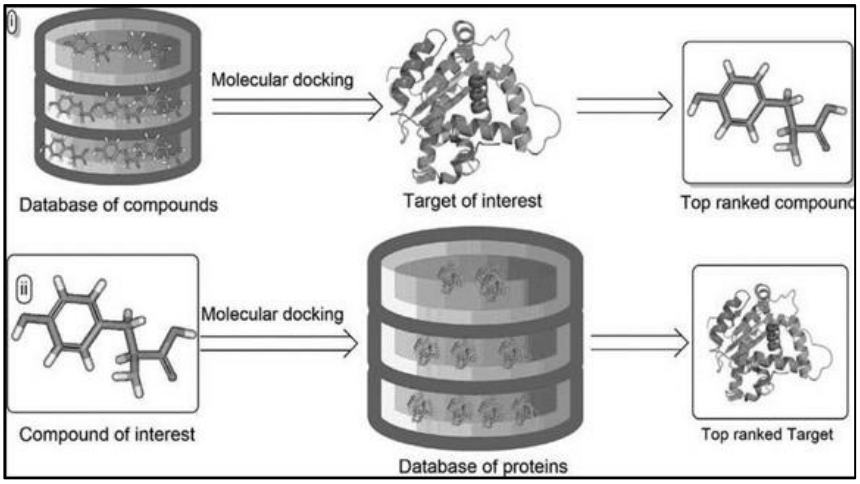

Figure 3.2: Structure based virtual screening

ii. Ligand based virtual screening:

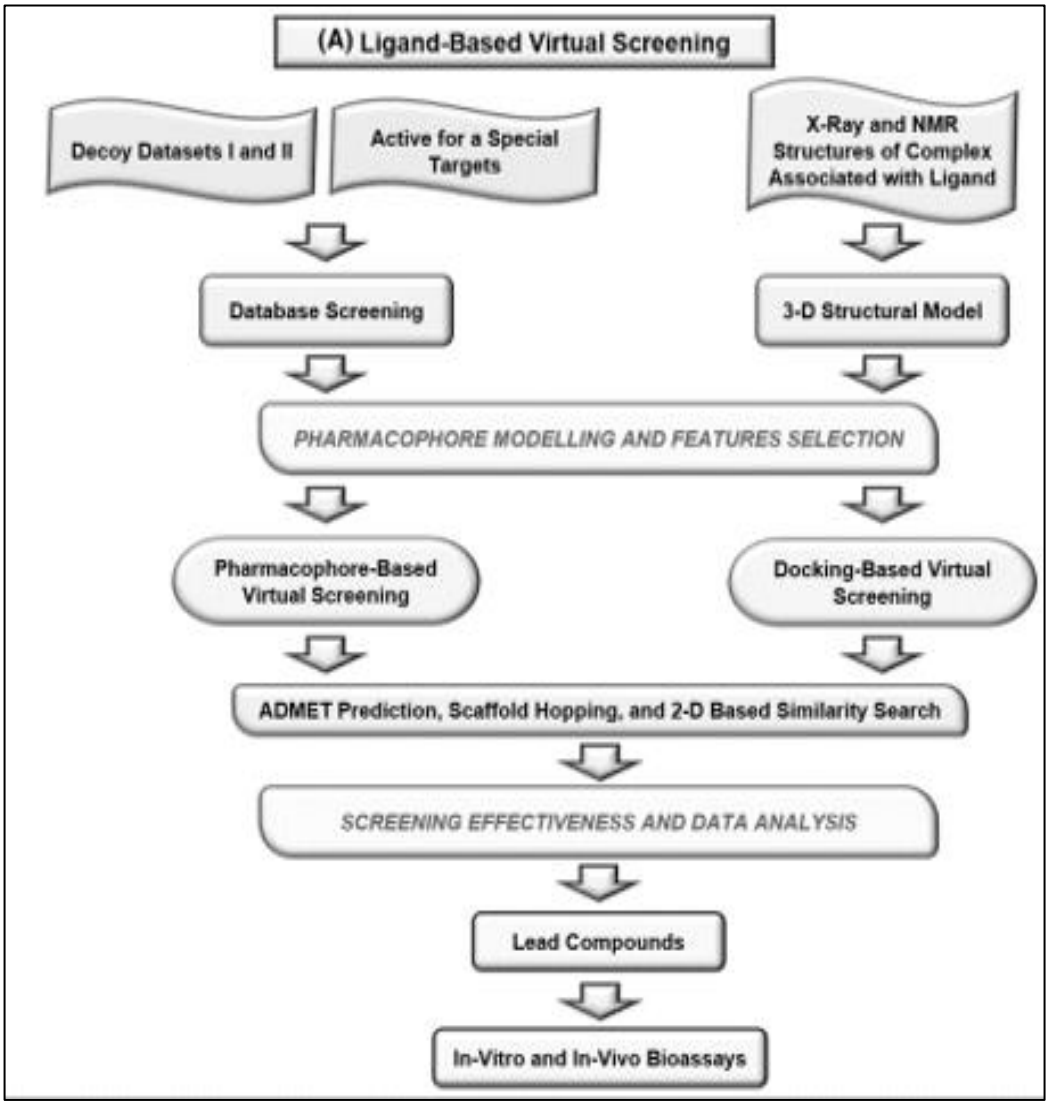

Figure 3.3: Ligand Based Virtual Screening.

A. Visualisation of protein ligand interaction diagram:

Molecular visualisation techniques often focus on the movement of single ligand molecules; yet, seeing simply one molecule can only provide an indication of the system's overall function. In order to address this problem, we do not concentrate on visualising the local effects of particular ligands on the influence of a protein and its general motion. Our proposed method decouples data preprocessing and visualisation because the simulations needed to examine these issues can involve millions of time steps.

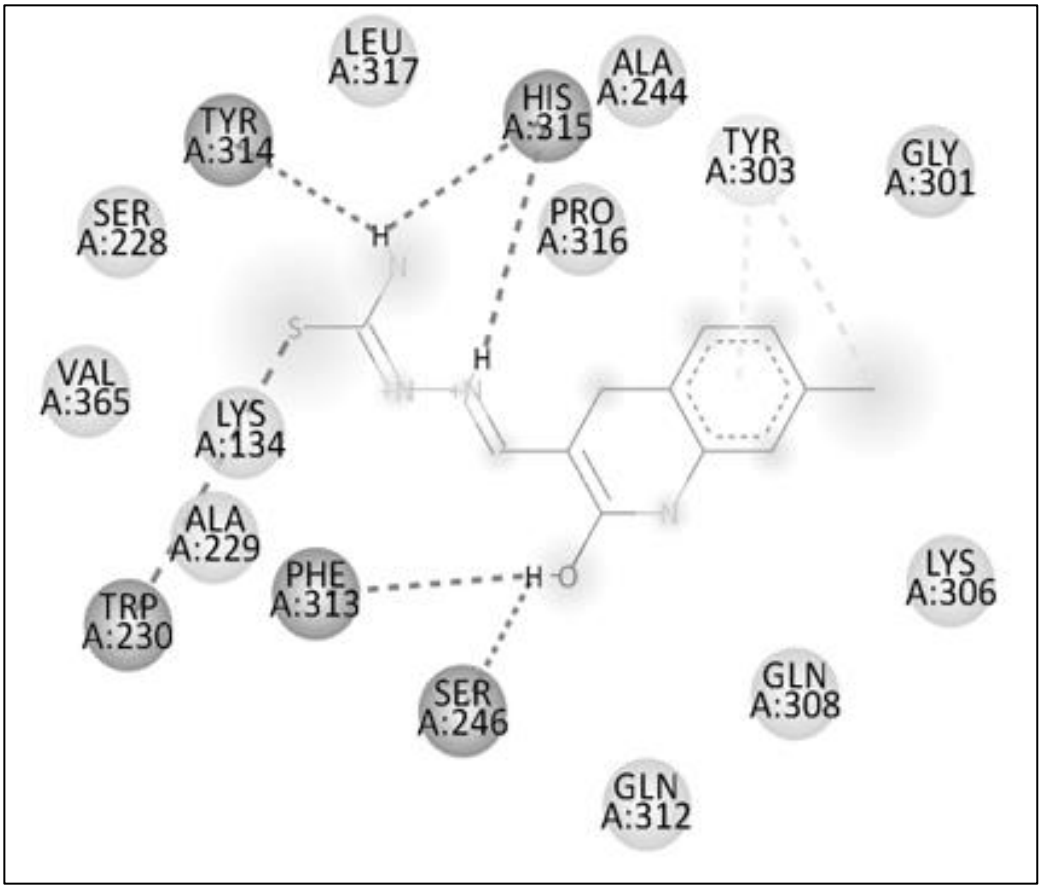

Figure 3.4: This is example for visualisation of protein ligand interaction 2D diagram.

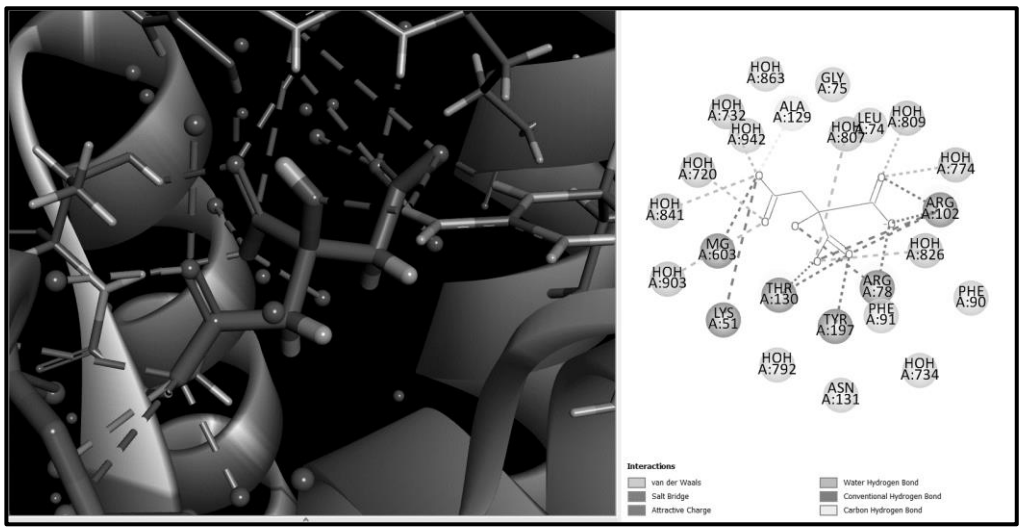

Figure 3.5: 3D and 2D View of protein and ligand interaction.

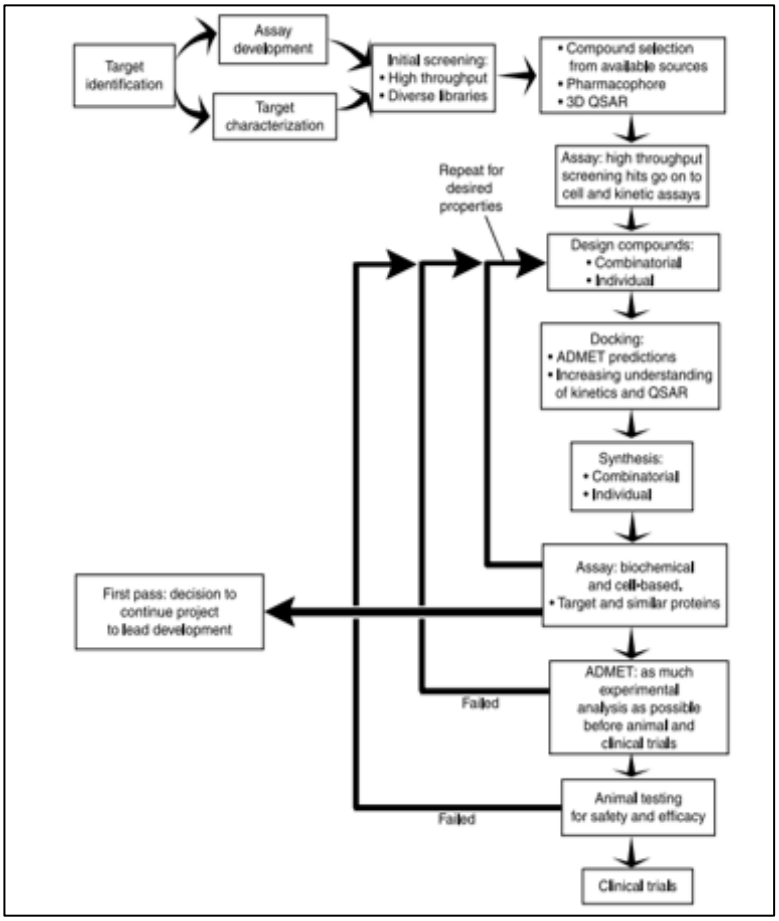

Figure 3.6: Flow Chart of SBDD

3.3 Ligand Based Drug Design:

Ligand-based computer-aided drug design (LB-CADD) is a powerful approach used in drug discovery that focuses on the chemical properties and structures of small molecules (ligands) rather than the target protein. This method is particularly useful when the three-dimensional structure of the target protein is unknown or difficult to determine. In LB-CADD, the process typically starts with the identification of a known ligand that binds to the target protein of interest. This ligand serves as a reference molecule, and its structural features and chemical properties are used to design or screen for new compounds with similar characteristics.

Key techniques and methods employed in LB-CADD include:

- Pharmacophore modeling: This involves identifying the essential structural and chemical features (pharmacophore) of the ligand that are responsible for its

interaction with the target protein. Pharmacophore models can then be used to screen databases of chemical compounds to identify molecules that match the desired pharmacophoric features.

- Quantitative structure-activity relationship (QSAR) analysis: QSAR models correlate the chemical structure of ligands with their biological activity against the target protein. By analyzing the relationship between structural descriptors (such as molecular size, shape, and electrostatic properties) and biological activity, QSAR models can predict the activity of new compounds and guide the design of optimized ligands with improved potency and selectivity.
- Similarity searching: This approach involves comparing the chemical structure of a known ligand with a database of compounds to identify molecules with similar structural features. Similarity searching can help identify structurally related compounds that may exhibit similar biological activity against the target protein.
- Ligand-based virtual screening: In this method, computational algorithms are used to screen large databases of chemical compounds to identify molecules with similar chemical properties or structural features to known ligands. Virtual screening can rapidly prioritize compounds for experimental testing based on their predicted likelihood of binding to the target protein.

Ligand-based drug design (LBDD) involves a series of steps aimed at identifying, designing, and optimizing small molecules (ligands) that bind to a target protein or receptor with high affinity and selectivity, these stages are almost as similar to the stages in SBDD type of CADD.

Here are the key steps typically involved in LBDD:

A. Target Identification and Validation:

- Define the biological target or pathway implicated in the disease of interest.
- Validate the target's relevance to the disease through experimental evidence and literature review.
- Establish the biological and pharmacological criteria for the desired ligand interaction with the target.

B. Selection of Reference Ligands:

- Identify known ligands or lead compounds that bind to the target protein with desired activity.
- Gather information on the structural and pharmacological properties of the reference ligands.

C. Pharmacophore Generation:

- Analyze the structural features and pharmacological properties of the reference ligands.
- Generate a pharmacophore model representing the essential molecular features required for binding and activity against the target protein.
- Consider features such as hydrogen bond donors/acceptors, hydrophobic regions, aromatic rings, and steric constraints.

D. Database Screening or Ligand Generation:

- Screen chemical databases or generate virtual compound libraries to identify or design new ligands.
- Utilize virtual screening, similarity searching, or de novo ligand design methods to identify ligands that match the pharmacophore model or possess desired structural features.

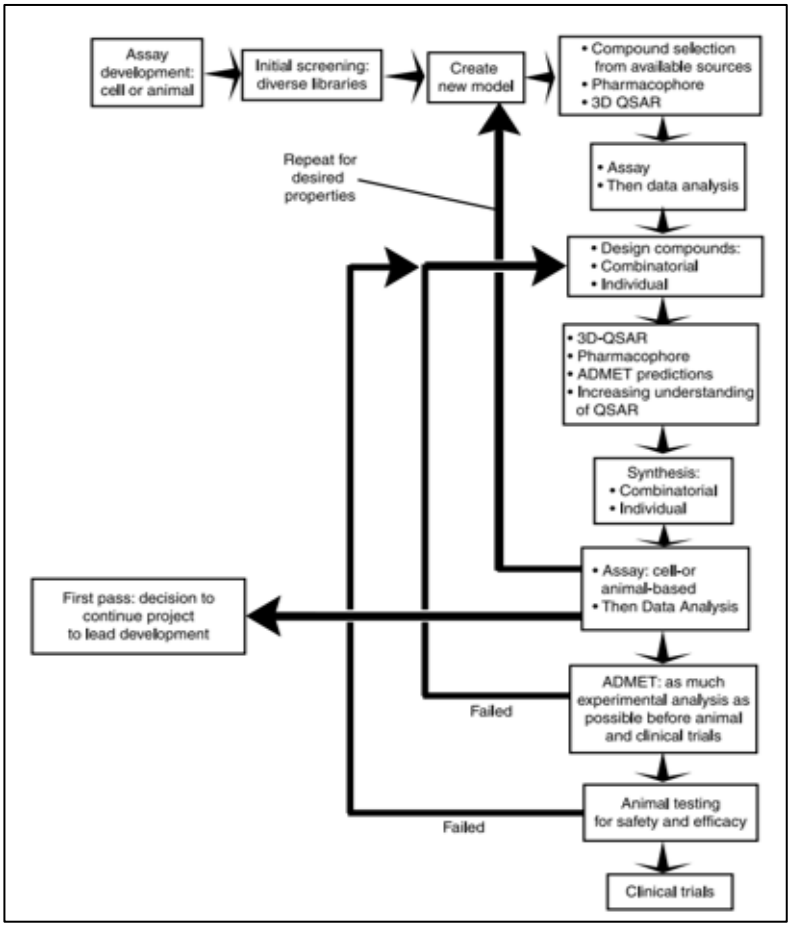

Figure 3.7: Flow chart of LBDD

E. Molecular Docking and Scoring:

- Perform molecular docking simulations to predict the binding modes and orientations of ligands within the target binding site.
- Utilize scoring functions to evaluate the binding affinity and energy of ligand-receptor interactions.
- Rank ligands based on their docking scores and prioritize those with the highest predicted binding affinity.

F. Structural Optimization:

- Refine and optimize the chemical structures of selected ligands through computational methods or chemical synthesis.
- Incorporate structural modifications to improve binding affinity, selectivity, solubility, and other pharmacological properties. Utilize structure-activity relationship (SAR) analysis and QSAR modelling to guide the optimization process.

G. Experimental Validation:

- Synthesize prioritized ligands for experimental testing in biochemical assays or cellular assays.
- Evaluate the binding affinity, potency, selectivity, and pharmacokinetic properties of the synthesized ligands. Iteratively refine the ligand design based on experimental feedback and computational predictions.

H. Lead Optimization and Development:

- Select lead compounds with favorable pharmacological profiles for further optimization.
- Conduct preclinical studies to assess the safety, efficacy, and pharmacokinetics of lead compounds. Advance promising leads to preclinical and clinical development stages for drug candidate selection and optimization.

- ✓ Throughout the LBDD process, computational techniques play a crucial role in guiding the design and optimization of ligands, accelerating the drug discovery process, and reducing the time and cost associated with experimental screening and synthesis.
- ✓ LB-CADD is particularly useful in situations where the target protein structure is unavailable or difficult to determine, such as in the case of membrane-bound receptors or protein-protein interactions. By focusing on the ligand's properties, LB-CADD enables the design and optimization of small molecule modulators for a wide range of biological targets, making it an invaluable tool in drug discovery and development.

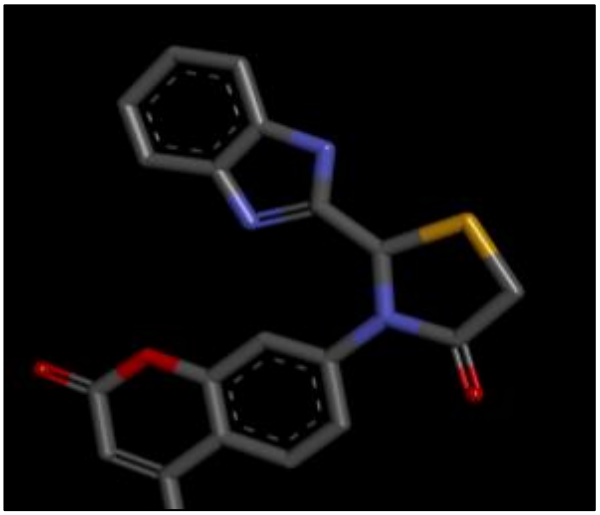

Figure 3.8 3D View of Ligand

3.4 Theoretical View of Computational Chemistry:

- Molecular Docking: Docking is attempted to find best matching between the two molecules one is ligand & protein molecules. Docking is a method which predicts the preferred orientation of one Ligand when bound in an active site to from a stable complex.
- Requirment For Molecular Docking:
  - ✓ Macromolecule (protein)
  - ✓ Ligand
- Principle: To achieve an optimized conformation for both receptor & ligand & relative orientations between protein & ligand such that the free energy of the overall system is minimized successful docking methods search high-dimensional spaces effectively & use a scoring function that correctly ranks candidate docking.
- Types of Interaction:
  - Protein-ligand interaction
  - Protein-protein interaction
  - Protein-DNA interaction

3.4.1 Molecular Docking Theory:

A. The Lock and Key Model:

It was introduced by Fischer in 1894, named as the lock and key model for its biological molecules According to this model, a ligand or a substrate molecule fits within the active site pocket of a macromolecule like a key fitting within a lock.

B. The Induced Fit Model:

This model is suggested by the Daniel Kashan in 1958. The model is more accepted for enzyme substrate complex than the lock and key model.

For proper catalysis, only a proper substrate can have induced the proper alignment of active site.

C. Binding Affinity:

Strength of the binding interaction between a single biomolecule to its ligand.

The example includes, protein-drug binding affinity, protein-inhibitor binding affinity.

- Measured and by the equilibrium dissociation constant (KD).
- Evaluates the strengths of bimolecular interactions.
- The smaller value of K = greater binding affinity of the ligand with its target.
- The larger value of Kp = target molecule and ligand bind weakly.

a. In Drug Discovery Process:

- binding affinity is used as a measure to rank the order hits that are able to bind to the target help design drugs that bind their targets selectively and specifically. Strength of the binding interaction between a single macromolecule to its ligand.
- E.g.: protein-drug affinity, protein-inhibitor binding affinity.
- Binding affinity is used as a measure to rank the order hits that are able to bind to the target help design drug that bind their target selectively and specifically.

D. Scoring Functions:

The scoring function generates a score for each pose. Based on the scores the different poses of ligand molecules are ranked.

In molecular docking process, scoring is used for a quantitative estimation of the pose's quality.

E. Modes of Docking:

- Rigid Docking: it's a method where both ligand and protein molecule are non-flexible state

- Flexible Docking: it's a method where both ligand and protein molecule are flexible state
- Semi Flexible Docking: it's a method where one among the ligand or protein molecule is nonflexible state and other one is

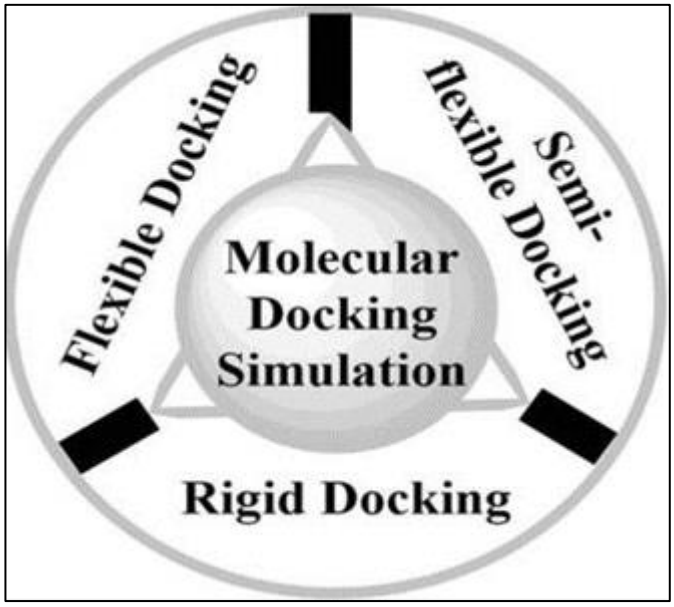

Figure 3.9: Modes of Docking

Legal Free Sources Used for Preparation and Selection of Protein and Ligands:

Molecular Modeling Task	Software/Web Server	Open Source*/Commercial	Website
1. Protein Preparation			
Crystal Structure PDB	RCSB	Open Source	https://www.rcsb.org/
Protein Visualization	UCSF Chimera	Open Source	https://www.cgl.ucsf.edu/chimera/
Homology Modeling	NCBI BLAST	Open Source	https://blast.ncbi.nlm.nih.gov/Blast.cgi
	SWISS-MODEL	Open Source	https://swissmodel.expasy.org/
Assigning Protein Protonation States	UCSF Chimera	Open Source	https://www.cgl.ucsf.edu/chimera/
Protein Energy Minimization	UCSF Chimera	Open Source	https://www.cgl.ucsf.edu/chimera/
2. Ligand Preparation			
Drawing Chemical Structures	Avogadro	Open Source	https://avogadro.cc/
	Chemdraw	Commercial	https://www.perkinelmer.com/category/chemdraw
Downloading Established Ligand Databases	PubChem	Open Source	https://pubchem.ncbi.nlm.nih.gov/

3.5 Conclusion:

- Computational chemistry plays a pivotal role in modern scientific research by providing invaluable insights into molecular structures, properties, and interactions. Through the application of quantum mechanics and molecular modeling techniques, computational chemists are able to simulate and predict various chemical phenomena with remarkable accuracy.
- Throughout this discussion, we have explored the diverse applications of computational chemistry, ranging from drug discovery and materials science to environmental chemistry and catalysis. By leveraging powerful computational algorithms and high-performance computing resources, researchers can efficiently explore vast chemical spaces and design novel compounds with desired properties.
- Moreover, computational chemistry serves as a complementary tool to experimental methods, aiding in the interpretation of experimental data, guiding experimental design, and even predicting experimental outcomes. This synergy between theory and experiment accelerates the pace of scientific discovery and enables the development of innovative solutions to complex chemical challenges.
- Looking ahead, the field of computational chemistry continues to evolve rapidly, driven by advances in algorithms, software development, and hardware capabilities. As computational methods become increasingly sophisticated and accessible, we can anticipate even greater contributions to diverse areas of chemistry and interdisciplinary research.