A Review on Principles of Molecular Docking and its Application in Drug Discovery

By

Name: Sharif Ezzaz Ahamed
Student ID: 19346068

A thesis submitted to the School of Pharmacy in partial fulfillment of the requirements for the degree of Bachelor of Pharmacy (Hons.)

School of Pharmacy
Brac University
March 2024

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing my degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material that has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

Sharif Ezzaz Ahamed
Student ID: 19346068

Approval

The thesis titled "A Review on Principles of Molecular Docking and its Application in Drug Discovery" submitted by Sharif Ezzaz Ahamed (19346068) in Summer, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy (Hons.).

Examining Committee:

Supervised By:

Dr. Humair Bin Md Omer
Assistant Professor
School of Pharmacy
Brac University

Approved By:

Program Director:

Professor Dr. Hasina Yasmin
Program Director and Assistant Dean
School of Pharmacy
Brac University

Dean:

Professor Dr. Eva Rahman Kabir
Dean
School of Pharmacy
Brac University

Ethics Statement

The study does not involve any kind of animal or human trial.

Abstract

Molecular docking (MD) is a computational technique that provides structural information and binding affinity of protein-ligand complexes in its thermodynamically favored conformation. This aids researchers in identifying active ligands that form stable complexes upon binding with targeted proteins for high throughput virtual screening and potential biological activities. This article provides features of molecular docking along with its application, accuracy, drawbacks, and significance in drug discovery projects and is expected to fill up a much-needed knowledge gap in this area. MD is divided into 3 types: rigid, flexible, and covalent docking. MD applies search algorithms to analyze feasible ways of attaching with targets and a scoring function is applied to determine binding affinity. Thus, the desired structure of the ligand preferably fitting in the target is obtained. Due to the ability of analyzing ligand-protein molecular processes, MD has significantly aided in drug discovery and selected success stories are presented in this thesis as well.

Keywords: Docking; virtual screening; conformational algorithms; rigid; covalent; flexible docking.

Dedication

I want to dedicate this paper to my parents who never let me feel alone, who always stood by me in my bad days.

Acknowledgement

I want to thank my supervisor Dr. Humair Bin Md Omer for many effective discussions and his feedback.

Table of Contents

List of Tables

List of Figures

List of Acronyms

CSD	Cambridge Structural Database
ACD	Available Chemical Directory
MDDR	MDL Drug Data Report
GT	Genetic Technique
NCI	National Cancer Institute
IC	Incremental Construction
GT	Genetic Technique
MC	Monte Carlo
$\Delta G_{binding}$	Gibbs Binding Free Energy
HB	Hydrogen Bonding
MD	Molecular Docking
RMSD	Root Mean Square Deviation

Chapter 1

1.1 General Introduction to Molecular Docking

Molecular docking (MD) is known as a type of computational modeling through which we can predict about what would be the ligand's favorable binding orientation at the moment of its interaction with a receptor. Ligand does so to result in a stable complex formation (receptor-ligand complex). Not only the binding free energy but also the strength of complexes can be found from the bound molecules' favorable orientation data. This favorable orientation data of bound molecules can be figured out by using molecular docking techniques. Currently, the use of molecular docking favors us to make assumptions about the most possible binding orientation through which drug candidates bind to their target. Drug candidate's targets are usually protein, carbohydrate, and nucleic acids. This also facilitates us by letting us know the tentative binding parameters of drug candidates to biomolecular targets. (Agarwal, Shweta & Mehrotra, Ranjana. (2016). Mini Review_ An Overview of Molecular Docking. JSM Chemistry. 2. 1024.) With the help of this information, we can design new drugs that have greater efficacy and specificity and find active and inactive compounds (hits). For example: a drug is present in the market that can produce its indicated therapeutic effect in the human body. By using molecular docking, we can do structure-based development of the existing drug to make (optimize) it to have better efficacy and more specificity than the previous one or perhaps resolve off-target binding issues.

The main motive behind doing molecular docking is to find out the binding conformation of drug candidates to biomolecular targets that require minimal energy because then the formation of a stable complex of ligand-receptor will occur resulting in quick identification of hits. To apply molecular docking, we need a structural data bank that will help us detect the target of

interest. Secondly, we need to select an appropriate methodology with a view to assessing ligands. In these two cases, tools of molecular docking and methodologies play a vital role. We can make an ascending or descending ranking of potential ligands according to their sufficiency to bind with given target candidates (binding affinity/ scoring data). (Agarwal et al., 2015; Dias & de Azevedo Jr., 2008)

Molecular docking lets us know the feasible poses of the ligand by which it binds in the pocket of the biological target to form a stable complex. Here optimal binding geometry is achieved by the formation of a stable complex. The feasible poses of the ligand are visualized by a sampling method. Molecular docking software along with subsidiary visualization software helps us in this regard. Not only NMR spectroscopy but also X-ray crystallography are the predominant methodologies employed to obtain experimental structural data (atom coordinates) of receptors that are used in the examination and validation of three-dimensional structural information pertaining to biomolecular targets. If the receptor structure is unknown or partially resolved then bioinformatics tools such as homology modeling can enable the estimation of the potential structure of unknown proteins, which exhibit significant sequence similarity to proteins with known structures. In this review article, an alternative methodology for the development of target structure, serving as a starting point for the computational identification of drug candidates with high affinity will also be introduced.

For ligand structures, the Cambridge Structural Database (CSD), the Available Chemical Directory (ACD), the MDL Drug Data Report (MDDR), and the National Cancer Institute Database (NCI) are just some of the databases that contain information on small ligand compounds. During the process of molecular docking, a variety of docked conformers, also known as poses, are created. These conformers are subsequently scored and compared with one another. The acceptance or rejection of a posture is determined by the scoring function of the docking program. In instances of rejection, novel postures are developed and the search

method is once again iterated until it reaches the point of accepting a single pose. In the field of molecular docking, the processes of searching and scoring are intricately interconnected. Nevertheless, the task of prioritizing docked conformers based on their experimental binding affinities and binding free energy appears to be more challenging compared to the process of identifying their binding orientation. In order to address this difficulty, researchers adopt several scoring functions, including consensus scoring. This involves applying multiple scoring algorithms on the same docked posture, with the aim of eliminating false positives. The review article will address some of these challenges in calculating and accurately predicting binding affinity data that researchers are undertaking.

Furthermore, in order to have computational approaches influence target recognition, it's necessary that the approaches are potent and swift. The attainment of a favorable outcome in docking simulations is contingent upon the critical characteristics of speed and precision. The primary goal in the creation of a docking algorithm is to provide a rapid technique that can effectively identify new lead compounds during virtual screening or accurately duplicate experimental conformations for conformation using experimental data. MCDOCK, DOCK, M-ZDOCK, AUTODOCK, Surflex, GOLD, FLEXX, MSDOCK, and ZDOCK are the names of the most used docking programs that are currently state-of-the-art. Every docking application relies on a distinct search technique, such as Incremental Construction (IC), Genetic technique (GT), Monte Carlo (MC), and others. Each entity possesses its own distinct set of parameters and search methodology, as exemplified in the subsequent sections of this review article. The docking program is capable of conducting a search to identify the optimal alignment between two or more molecules. This search takes into consideration various parameters derived from the input coordinates of the receptor and ligand molecules. These parameters include geometric complementarity, which accounts for factors such as atomic van der Waals radius and charge, as well as the flexibility of the receptor or ligand structure. Additionally, interatomic

interactions, such as hydrogen bonds and hydrophobic contacts, are also considered during the docking process. As a result, docking applications provide the anticipated orientations (poses) of a ligand within the binding site of the target. Typically, the process of posing yields several potential conformations. In order to optimize and rank data obtained from the docking technique, scoring functions are utilized to measure intermolecular binding affinity or binding free energy. (Dias & de Azevedo Jr., 2008; Lamb & Jorgensen, 1997; Seeliger & De Groot, 2010). These functions aim to get the optimal orientation. We can understand which docking method is more potent and swifter by assessing the efficacy of various docking methods. This can be done by docking-based virtual screening and comparing with experimental binding data if available. The protocols of virtual screening facilitate finding active candidates from a database containing a huge number of molecules. This is why, extensive efforts have been undertaken to advance and implement effective scoring and docking systems. Details of which have been discussed in the subsequent chapters.

1.2 Objective

Molecular docking has played a significant role in the field of drug discovery and design. MD is benefiting us in discovering improved as well as novel potent and selective drugs. The purpose of this review paper is to provide a clear understanding of the basic physical principles of molecular docking, its underlying theories, its advantages its limitations, and how it is playing a significant role in drug discovery projects worldwide.

1.3 Methodology

For this review paper, a vast amount of research papers related to the topic were collected from PubMed and Google Scholar. In order to write this review article, cross-citations were used and the focus was mainly on journal articles that were heavily cited as well as leading papers in the field. The use of cross-citation made the author able to gather information in more detail

as well. The author has used the American Psychological Association 7th edition (APA) citation style to cite all the references.

1.4 Physical Principles Behind Calculations of Binding Affinity

Empirical scoring functions provide an estimation of the ligand binding affinities. This function does so in accordance with van der Waals interaction, hydrophobic interaction, intermolecular hydrogen bonds, entropy, deformation effect, and others. Intermolecular hydrogen bonds, van der Waals contacts, and shape and charge complementarity among ligand and protein are the factors that determine the particularity of ligand for a protein. For this reason, the evaluation of intermolecular hydrogen bonding interactions via crystallographic data holds an important role in defining the particular affinity of ligands towards proteins. The prevailing approach in contemporary empirical scoring functions is to employ a model that allows for the decomposition of binding affinity into distinct components that represent the many factors influencing the binding process. Empirical scoring functions can be formulated by applying previously stated theory. We can use the empirical scoring functions to calculate the Gibbs free energy of binding ($\Delta G_{binding}$). When the summation of interactions is multiplied by the weighting coefficients (cj), we will get $\Delta G_{binding}$. (Huang et al., 2006)

Eq. 1: $\Delta G_{binding} = c0 + c1.f1(x, y, z) + c_2 .f_2 (x, y, z) + ...+ c_N .f_N (x, y, z)$ (de Azevedo Jr. & Dias, 2008)

Atoms in the ligand and protein play an important role in determining each term in the above formula. Based on the cartesian coordinates of these atoms, the value of each term can be calculated. c0 is a regression constant, f's are functions that account for intermolecular interactions, such as intermolecular hydrogen bonds, electrostatic interactions (charge-charge, charge-dipole, charge induced dipole), deformation, hydrophobic effect, and others that may be included. We need to get values for terms of the scoring function, that's why we use a

training set to get those values. After that, a method known as multivariate regression analysis is utilized in order to locate the best possible match between the predicted and experimental protein-binding affinities. The physical laws that govern the creation of binary complexes can be applied to explain empirical scoring functions but there may be loss of a significant portion of the intermolecular interaction. (de Azevedo Jr. & Dias, 2008)

Despite encountering several challenges in comprehending the crucial structural characteristics governing binding affinity, the majority of existing experimental evidence suggests that employing additive functions to model protein-ligand interactions might be a promising strategy for constructing empirical scoring functions. The aforementioned functions can be represented in terms of Gibbs free energy of binding, as demonstrated in the PEARLS method, or in terms of pK_d values, as employed in the XSCORE approach. The estimation of binding analysis for protein-ligand complexes may be conducted using atomic coordinates (x, y, z). This estimation involves calculating the total of interactions, which is then multiplied by weighting coefficients (cj). This can be shown by the following equation-

Eq. 2: $$pK_d = c_0 + \sum_{j=1}^{N} c_j f_j(x,y,z)$$ (de Azevedo Jr. & Dias, 2008)

Here, c_0 can be defined as a regression constant, intermolecular interactions are indicated by f_j, and pK_d can be denoted by -log K_d. The use of empirical scoring functions, which break down the binding free energy into individual terms, poses an inherent issue from a physical standpoint, as this decomposition is not permissible. The free energy of binding is a state function, while its individual terms do not possess this property. In addition, additive approaches are limited in their ability to capture nuanced cooperative phenomena. Notwithstanding these challenges, the computational modeling of biomolecular systems has significant promise in terms of its ability to provide predictions and provide valuable insights

that can inform the process of molecular design. Some efficient empirical scoring function programs are described further below:

XSCORE

It is the empirical functioning program that is applied in most cases. The reason behind this is that it is more efficient than other scoring functions. When we calculate three empirical scoring functions, we can get SCORE. There are five parameters contained in each of those functions. The ligand's atomic coordination data and the protein's atomic coordination data are used to measure those five parameters. (de Azevedo Jr. & Dias, 2008)

Hydrogen Bonding

The atoms that are engaged in hydrogen bonding have been categorized with a view to evaluating the parameter of hydrogen bonding. The metal ions involved in the protein binding site and atoms of nitrogen and oxygen that are connected in bonding with hydrogen atoms are called Hydrogen bond donors. On the other hand, when sp^2 or sp hybridized atoms of nitrogen and oxygen contain lone pairs, they are recognized as Hydrogen bond Acceptors. Generally, we take nitrogen atoms and oxygen as acceptors/donors. It means that they have the ability to work as either one of them (H bond donor or H bond acceptor). The formula that can be used to measure H bonding is-

Eq. 3: $$HB_{ij} = f(d_{ij}) f(\theta_{1,ij}) f(\theta_{2,ij})$$ (de Azevedo Jr. & Dias, 2008)

In this equation i represents atoms of protein and j is the symbol for ligand atoms. d is called the distance function. The distance between i and j is the distance function. There is an angle among the Donor root, Acceptor atoms, and Donor atoms that we would take for angular

functions f($\theta 1$) and f($\theta 2$). The angle among Donor, Acceptor, and Acceptor root atoms is represented as angle $\theta 2$. Interatomic distances and angles attribute the higher contribution to the Hydrogen bond measured by some characteristics of fuzzy functions, these functions are represented by f(d), f(θ 1), and f(θ 1) in the equation.

One of the most effective ways to show the equality between the long-range attractive dispersion force and short-range attractive dispersion force is using the Lennard-Jones equation. This equation can be also used to calculate Van der Waals interaction (VDW). The equation of Lennard Jones is-

Eq. 4: $$VDW_{ij} = \sum_{i}^{ligand} \sum_{j}^{protein} \left[\left(\frac{d_{ij,0}}{d_{ij}} \right)^{8} - 2 \times \left(\frac{d_{ij,0}}{d_{ij}} \right)^{4} \right]$$ (de Azevedo Jr. & Dias, 2008)

Here i is the representative of protein atoms and j is the representative of ligand atoms, the gap among atoms of i and j is represented as d. The addition of van der Waals radii of protein atoms(i) and ligand atoms(j) is denoted as d0 in this equation. (de Azevedo Jr. & Dias, 2008)

Deformation Effect

The deformation effect is represented as RT. There is a formula in which ligand flexibility is measured depending on the number of bonds which are rotatable. The effect of deformation can be measured by applying the formula-

Eq. 5: $$RT = \sum_{i}^{ligand} RT_i$$ (de Azevedo Jr. & Dias, 2008)

Here the value of RT*i* depends on the number of rotatable bonds in atom *i*. When there is no rotable bond in atom *i*, it is considered that RT*i* = 0. When there is only one rotatable bond in atom *i*, the value of RT*i* is considered 0.5. When there are two rotatable bonds in atom *i,* the

value of RT*i* is considered 1.0 ("Computational Biochemistry and Biophysics," 2001; de Azevedo Jr. & Dias, 2008)

Hydrophobic Effect (HS, HC, and HM)

Despite having 5 common parameters among XSCORE's 3 scoring functions, there are some parameters that are different among them. The hydrophobic effect is one of them.

Eq. 6: $$HS = \sum_{i}^{ligand} SAS_i$$ (de Azevedo Jr. & Dias, 2008)

This is the formula which is abide by pKd(1). pKd(1) is the first empirical scoring function. In this formula, protein atoms (atom i) solvent accessible area is denoted by SAS.

Assessment of the hydrophobic effect for pKd(2) is done by using a different formula compared to pKd(1).

Eq. 7: $$HC = \sum_{i}^{ligand\ protein} \sum_{j} f\left(d_{ij}\right)$$ (de Azevedo Jr. & Dias, 2008)

This is the formula that is used to measure the hydrophobic effect for the Second empirical scoring function. At the time of the formation of van der Waals force by the two hydrophobic atoms, the hydrophobic interaction's tenacity becomes most. f (d) is defined as the distance function in that formula which gives us a view of the hydrophobic interaction's power.

SCORE has accepted an equational formula that can be applied to measure the 3rd empirical scoring function which is stated below:

Eq. 8: $$HM = \sum_{i}^{ligand} \log P_i \times HM_i$$ (de Azevedo Jr. & Dias, 2008)

At the time, protein hydrophobic atoms were able to bind atoms of ligand i, and it is enclosed by 6A, it is considered that the value of HMi is 1. The hydrophobic scale is denoted as Log P.

A formula that can be used to evaluate the terminal three empirical scoring functions is given below:

$$pK_{d(1,2,3)} = C_0 + C_1 \times VDW + C_2 \times HB + C_3 \times RT + C_4 \times HE$$

Eq. 9: (de Azevedo Jr. & Dias, 2008)

Here the hydrophobic effect is denoted by HE. It is the same as HS which is used in the 1st scoring function, same as HC which is used in the 2nd scoring function, and likely to HM which is applied in the 3rd scoring function. C represents the coefficients that are gained by analysis of linear regression.

DrugScore

http://pc1664.pharmazie.unimarburg.de/drugscore/index.php is the online address from where we can use DrugScore. There are reactions where atoms of ligand and protein do nonbonding interactions among them. These interactions of nonbonding can be efficiently measured by this scoring function. $\Delta W_{I,J}$ denotes pseudo energy. With the help of pair potentials which are distance-dependent and atom type that have been got arithmetically, we can find out $\Delta W_{I,J}$. ReLiBase system plays an important role here that helps to achieve arithmetically found atom-type and distance-dependent pair potential.

Distance Dependent Potentials

It is denoted by $\Delta W_{I,J}(r)$. It can be calculated by applying this formula-

Eq. 10: $$\Delta W_{ij}(r) = -\ln \frac{g_{ij}(r)}{g(r)}$$ (de Azevedo Jr. & Dias, 2008)

Type i and j atom's normalized radical pair distribution is represented by *gij* (*r*). It can be calculated by Drugscore. Drugscore implements the following formula to calculate it-

Eq. 11: $$g_{ij}(r) = \frac{N_{ij}(r)/4\pi r^2}{\sum_r \left(N_{ij}(r)/4\pi r^2\right)}$$ (de Azevedo Jr. & Dias, 2008)

Here the considered distance ranges are between 1 and 6 Å and it is represented by r.

When any two of atoms are far away in the range of r and r+dr, we can use the formula of g(r) to calculate the normalized mean radical pair distribution-

Eq. 12: $$g(r) = \frac{\sum_i \sum_j g_{ij}(r)}{i \times j}$$ (de Azevedo Jr. & Dias, 2008)

Solvent-Accessible Surface-Dependent Potentials

The calculation method of g(r) is different from the calculation formula of solvent accessible surface dependent potentials.

Eq. 13: $$g_i(SAS) = \frac{N_i(SAS)}{\sum_{SAS} N_i(SAS)}$$ $$g_i(SAS_0) = \frac{N_i(SAS_0)}{\sum_{SAS_0} N_i(SAS_0)}$$ (de Azevedo Jr. & Dias, 2008)

The chance of getting a type i atom remaining in the complexed state and having the characteristics of uncovered solvent accessible SAS surface can be measured by the formula of gi(SAS). There is another formula that is applied to calculate the possibility of a similar atom having the exact characteristics present in the condition other than the complex state.

We can again use DrugScore with a view to measuring the value of binding score among two types of atoms. The equation that Drugscore applies to do that-

$$\Delta W_{I,J} = \sum_{i \in I} \sum_{j \in J} \Delta W_{i,j}(r)$$

Eq. 14: (de Azevedo Jr. & Dias, 2008)

This is the way of finding out the involvement of every type i atom and type j atom.

PEARLS

This scoring function is found at the "http://ang.cz3.nus.edu.sg/cgi-bin/prog/rune.pl" web address. Drugscore can be also found online but the difference is that PEARLS is based on the force field scoring function but drugscore is based on the knowledge scoring function. Significant details about hydrogen bonds, solvation, electrostatic, van der Waals, and entropy that occur because of the intermolecular interactions can be known by PEARLS. Moreover, it gives us the facility of knowing interactions that are intramolecular.

Hydrogen Bonding

PEARLS uses Morse Potential to make assumptions about hydrogen bonding. By using the gap between the hydrogen bond donor and hydrogen bond acceptor, Morse Potential determines hydrogen bonding. The formula that can be used to predict hydrogen bonding is given below-

Eq. 15: $$V_{H-bonds} = \sum_{H-bonds}\left[V_0\left(1-e^{-a(r-r0)}\right)^2 - V_0\right]$$ (de Azevedo Jr. & Dias, 2008)

Here, the interatomic gap among the hydrogen bond donor and hydrogen bond acceptor is denoted by r, the others are the Morse Potential constants.

Van der Waals Interaction

Finding out the free energy of van der Waals interaction by PEARLS is very easy. In order to do so, PEARLS uses a molecular mechanics field of force that is known as AMBER.

Eq. 16: $$V_{VdW} = \sum_{VdW}\left[A_{ij} / r_{ij}^{12} - B_{ij} / r_{ij}^{6}\right]$$ (de Azevedo Jr. & Dias, 2008)

This is the formula to calculate Van Der Waals interaction. In this formula, the Field of Force (AMBER) plays an important role by detecting the values of A_{ij} and B_{ij}. In order to represent the gap among atoms of i and j, r_{ij} is in the formula.

Electrostatic Interaction

Sanderson charge points out the partial changes and it helps us to measure electrostatic interaction. The formula of evaluating electrostatic interaction is given below-

Eq. 17: $$V_{elect} = \sum_{elect}\left[q_i q_j / \varepsilon_r r_{ij}\right]$$ (de Azevedo Jr. & Dias, 2008)

Here, the gap among two atoms is denoted by r_{ij}, and atoms of i and j give partial charges in the field of force (AMBER) is represented by q_i and q_j. A dielectric constant is needed to calculate electrostatic interaction which is denoted by ε_r.

Metal Ligand Bonding

PEARLS is also effective in finding out metal ligand bonding through the formula of metal ligand bonding energy.

Solvation

When molecular binding occurs, a change in the free energy of solvation is seen. This can be evaluated by applying the atomic solvation parameters equation given by Eisenberg.

Eq. 18: $$\Delta G_{solv} = \sum_{atoms(i)} \Delta\sigma_i A_i$$ (de Azevedo Jr. & Dias, 2008)

Here, according to Lee and Richard the surface area that is accessible by the i atom is denoted by A_i and $\Delta\sigma$ represents the parameter of atomic solvation.

Entropy

There is an empirical equation that can be applied in order to evaluate entropy-

Eq. 19: $$\Delta G_{entropy} = 0.59N$$ (de Azevedo Jr. & Dias, 2008)

Here, the sum of bonds that are rotable is denoted by N.

POLSCORE

25 empirical scoring functions that have formulas made of coefficients and variables are used to design the POLSCORE program.

Hydrogen Bond

POLSCORE determines the hydrogen bonding by evaluating f(d) and f(θ) functions. There are two conditions for these functions. Only when both of the conditions are met, the result of these functions is 1. (de Azevedo Jr. & Dias, 2008) On the other hand, when only one condition is

met, the result is zero. The formula POLSCORE uses to determine intermolecular hydrogen bonds is given below:

$$HB = \sum_{i=1}^{N} f_i(d) f_i(\blacklozenge)$$

Eq. 20: (de Azevedo Jr. & Dias, 2008)

Contact Surface

Contact surface is denoted by A. The formula that can be used to measure the changed contact surface is given below:

$$A = \frac{AI^2}{(A2 - AI)}$$

Eq. 21: (de Azevedo Jr. & Dias, 2008)

Contact surface area and ligand total contact area is respectably denoted by A1 and A2. There is a question arises here which is why we are also considering the (A2-A1) value, it is because only taking the A1 value would result in bigger ligands with superior weight in the function. Taking the value of (A2-A1) to calculate the contact surface would solve this problem. (Becker, O. M., et al. 2001)

Chapter 2

2.1 Types of Molecular Docking

Molecular docking is mainly divided into 3 categories: rigid docking, flexible docking, and covalent docking.

Rigid Docking: It is usually done when the compounds are not flexible. In this method, one of the compounds which is present in the three-dimensional space is readjusted. It is done to achieve optimal fit to different substances according to the scoring system's parameters. One of the advantages of this docking type is that we can develop the configuration of the ligand even without the target binding activity. (Dnyandev et al., 2021; Raval & Ganatra, 2022) By using this docking method, we can find out the conformation of two molecules where the least amount of energy is required (Figure: 1)

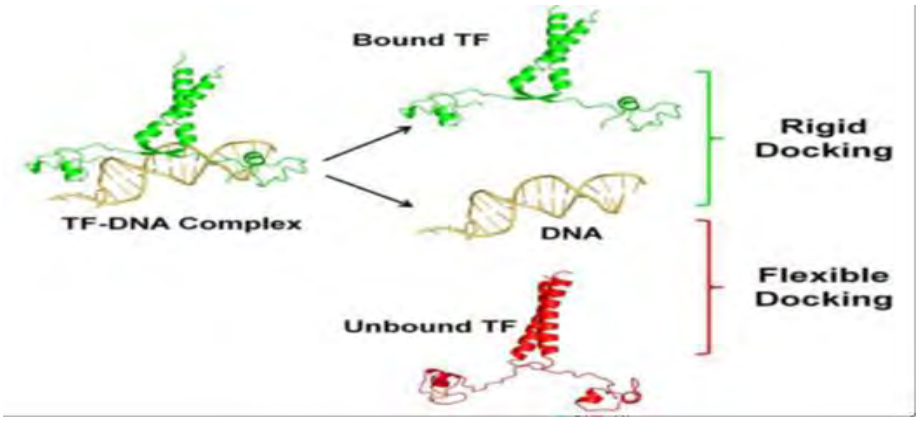


Figure 1: Rigid and flexible docking. (Dnyandev et al., 2021)

Flexible Docking: This type of docking is done when the compounds are flexible. Considering transformation, the flexibility of molecules is generally measured so that we can get the conformations of the ligand-receptor complex. (Dnyandev et al., 2021; Raval & Ganatra, 2022) It is shown in the following figure 2:

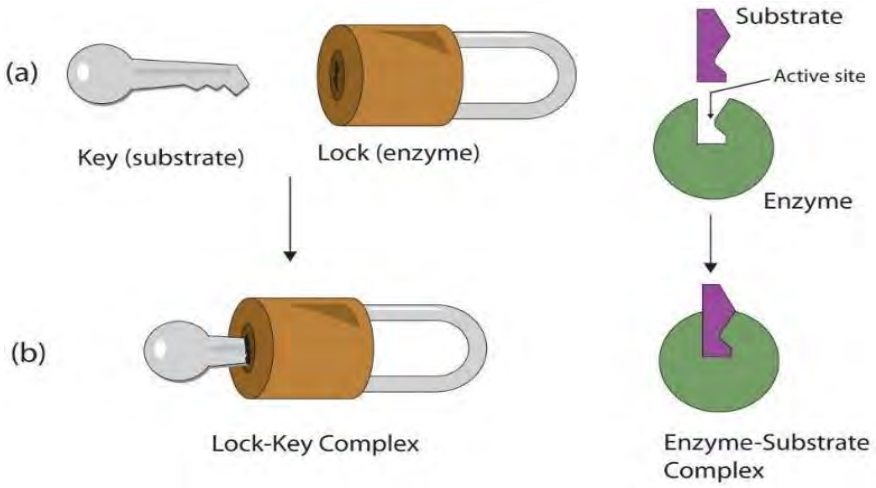


Figure 2: Flexible docking. (Raval & Ganatra, 2022)

Covalent Docking: With a view to doing inhibitor's covalent docking to the proteins that are targeted, we have been able to find many ways. Nevertheless, it is seen in most cases that software used for covalent docking is only able to determine the energy of binding which is among the ligand (electrophilic) and receptor (nucleophilic).

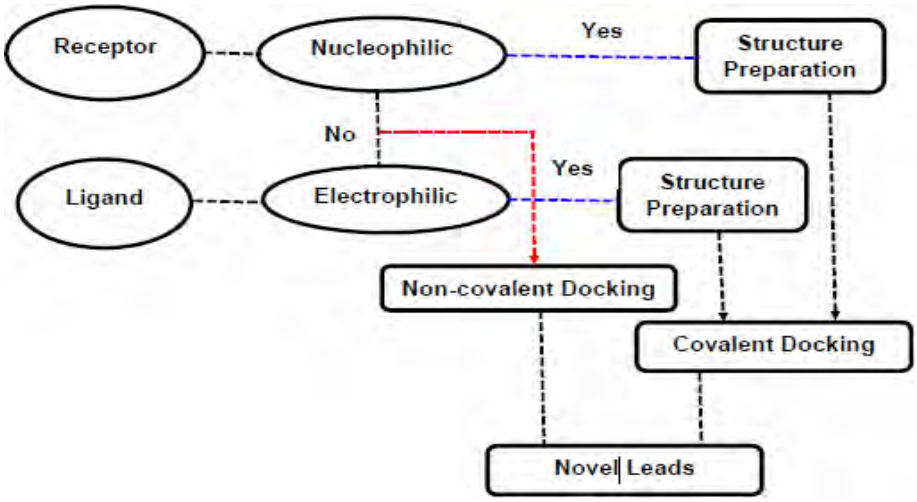


Figure 3: Flowchart of covalent docking. (Dnyandev et al., 2021; Raval & Ganatra, 2022)

Among all the approaches of covalent docking, one is the link atom. As the name suggests, it determines a link atom in protein as well as in ligand. This approach works in a way that both the link atom of protein and ligand have equal steric volume (figure: 3). This is a pre-requisite

of covalent bonding and by doing so covalent bonding is achieved. We can use Gold Software in order to link atoms. There are some other software that use more than one method so that covalently docking to receptors can be achievable. It uses two methods named modification of the flexible side chain and grid-based method. A single side chain is made by the extension of protein and ligand that is bound covalently in the modification of the flexible side chain method. This method considers it as the receptor's portion. The other method is based on the Gaussian biasing function. This approach of covalent docking is not considered ideal since it fails to ease the trouble that arises in screening purposes' up-scaling process. This can be solved if we use ligand files which are automatically prepared. Though this program is able to solve the problem, it has only measured reactions. The basic principle of covalent binding is that initially, non-covalent interaction occurs among the ligand and protein in such a posture that is good for the reaction. Lastly, a reaction occurs among the substances that result in the formation of covalent binding. The Covalent Dock exactly follows this basic principle of covalent docking. (Kumalo et al., 2015a) The resulting covalent linkage in CovalentDock can be determined through the following equation-

$$E \begin{cases} D\left(e^{-2\alpha(r-r_0)} - 2e^{-\alpha(r-r_0)}\right) - T\Delta S + C & r \le r_m \\ 0, & r > r_m \end{cases}$$

Eq. 22: (Kumalo et al., 2015b)

In this equation, the breadth of the bond is denoted by r, and the symmetrical breadth of the bond is represented by r_0. The highest breadth of bond that can be achieved beyond disjunction is denoted by r_m, the structural entropy determined by Gaussian is represented by ΔS_{est} while the constant named correcting empirical is denoted by C. (Kumalo et al., 2015a)

2.2 Methodologies for Docking Calculations

The job of molecular docking is to find out the interaction among two molecules (biological macromolecule and ligand). The purpose behind doing so is to detect the conformation of ligands that will require the least binding energy to bind with the target. The energy needed for interaction is turned into a score of docking by the functions of scoring. Apparatus of molecular docking such as Rasmol, and Pymol enable us to see the ligands attached to the target in 3D pose for visualization. The processes that are usually involved in docking are-

Selection of Target: In order to do molecular docking, initially the target has to be detected. For doing so, it's also essential to determine the sites of the target where the ligand can attach. Docking allows us to visualize the structure of the target in 3D pose and thus we are also able to detect where the ligand can bind. Several processes such as hydrogen atom inclusion, and solvent evacuation are needed with a view to making a target.

Making Assumption About Active Site: To make this assumption, we need a server called CASTp. With the help of this server, we can determine the active sites in our desired target. This server has the ability to detect the sites in the conformation of protein where ligands can bind. It also gives information about the volume as well as the area of the binding pockets in the target. Molecules such as water or heteroatoms are needed to be eliminated from the active site of the target where ligands can bind. (Paramashivam et al., 2015)

Structural Cleaning and Energy Minimization of Receptor: The purpose of doing structural cleaning is so that disappeared atoms can be entered into the disappeared residues. The way of getting target proteins Protein Data Base is RCB. Studio of discovery is used to view the structure of the protein. With the help of 14SB which is an energy refinement force field, a prototypical building is made and at the same time, root mean square restraint 0.3 Å is applied.

This results in the energy minimization of receptors. A toolkit named Molecular Modeling Toolkit is utilized to do the routine minimization. (Pettersen et al., 2004)

Ligand preparation and energy minimization: Chem Draw professional can be applied to formulate the conformation of ligands in 2D pose. In order to do the energy minimization of ligand, we need PDB files. These files can be achieved by using the studio of Biovia Discovery. Before choosing the ligand, a ligand archive is made and it only contains the data that are known. It's essential to follow Lipinski's rule of 5 to make a perfect choice of ligand. (Khan T. et al., 2018)

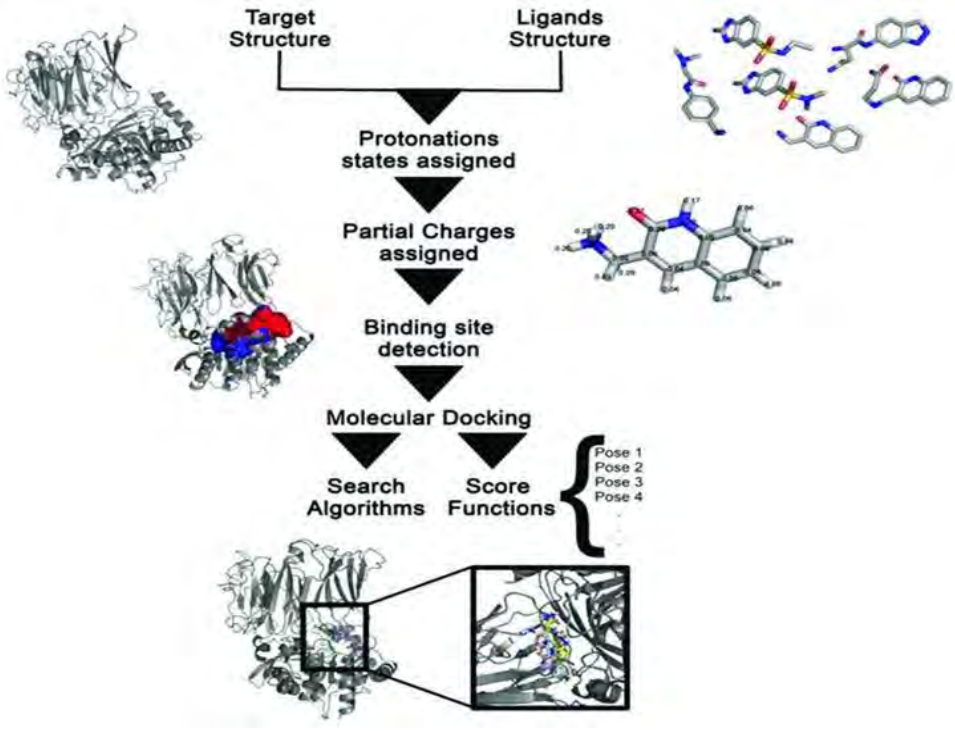


Figure 4: Way of doing calculations of MD. (Torres et al., 2019)

In order to do the calculation of docking, the first step is to get the conformation of ligands as well as our target in three dimensions (Figure: 4). Secondly, situations of protonation and non-integer charge value are allotted. Thirdly, the active site of binding in the target is found. Lastly, posing and scoring are both applied with a view to resulting in the feasible ligand-protein complexes. (Torres et al., 2019)

Chapter 3

3.1 Applications of Molecular Docking in High Throughput Virtual Screening

When our motive is to find biologically effective substances from vast libraries containing molecules in opposition to a specific target of interest, we can apply a high throughput screening process to quickly do that. Docking of selected ligands to the target is considered the main footprint in virtual screening. Motive of molecular docking is to accelerate the formation of the noncovalent complex from the binding of ligand to its target since it is seen in most cases that when minute molecules bind with their target, it results in covalent bonding. (Liao et al., 2013) If we think about the lock and key model, it's the perfect way to visually present molecular docking. In this, the ligand works as a key which is made to perfectly adjust into the lock. This process has been presented in the following figure 5:

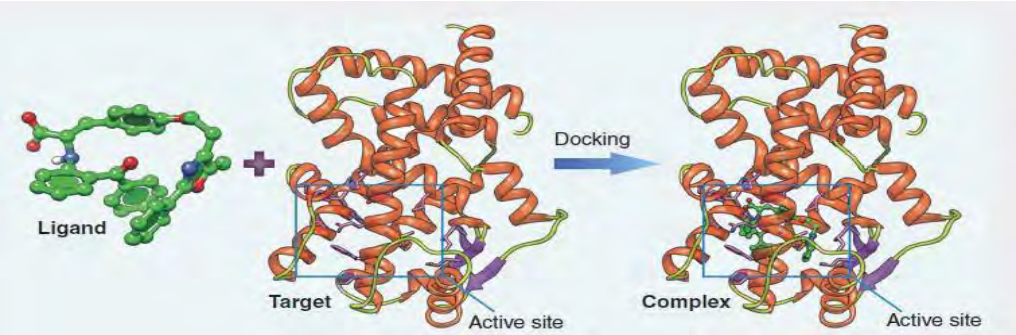


Figure 5: High Throughput Virtual Screening Process. (Liao et al., 2013)

By using molecular docking, we can get a thorough knowledge about the binding mode. This mode shows about the ligand's location and conformation compared to the target. Molecular docking also has the ability to check whether the ligand suits into the target or not. If the ligand suits into the target, molecular docking can also show how perfect the interaction is among the ligand and the binding site of the target.

3.2 Posing Compounds

Flexibility of ligand generates the conformation of ligand that is needed for docking, that's various different docking softwares usually make the ligand flexible. (Liao et al., 2013)There are three methods by which we can make ligands flexible-

Systematic methods: In this method, a library of possible configurations has been made for one and all ligands. Here gradual formation of ligand occurs from the binding site's portions.

Stochastic methods: Not only generic algorithms but also Monte Carlo simulations are based on this method. In this method, initially, a series of modifications are made in the conformation of the ligand and then it is checked which altered ligand has more possibility to fit into the binding site of the target. (Liao et al., 2013)

Stimulation methods: Molecular dynamics and energy minimization are based on this method. Because of two severe drawbacks of this method, this method is not the first. One of them is that it is a costly method and the other is one- it easily falls in relative minima. However, this method is effective in rectifying ligand conformation that resulted from another technique.

Docking program	Developer	Search algorithm	Scoring function	Free for academia?
AutoDock	Scripps Research Institute (CA, USA)	GA, MC	FF	Yes
DOCK	University of California, San Francisco (USA)	IC, EM	FF	Yes
FlexX	BioSolveIT (Sankt Augustin, Germany)	IC	E	No
FRED	OpenEye Scientific Software (NM, USA)	ES	KB	Yes
Glide	Schrödinger (NY, USA)	ES, EM, MC	E	No
GOLD	Cambridge Crystallographic Data Centre (UK)	GA	E, KB	No
ICM	Molsoft (CA, USA)	MC	E	No
Surflex	Tripos (MO, USA)	IC	E	No

E: Empirical; EM: Energy minimization; ES: Exhaustive search; FF: Force field based; GA: Genetic algorithm; IC: Incremental construction; KB: Knowledge based; MC: Monte Carlo simulation.

Table 1: The most widely used docking programs in structure-based drug design. (Lazarova, n.d.)

3.3 Scoring Compounds

It's very important to check whether the generated docking conformation fits well in the target's bind site or not. In order to do so, we can use scoring functions. The advantage of using it is that it does not only measure whether the ligand's docked conformation is perfect or not, but also it measures how well the ligand's binding occurred to the binding site of the target. Scoring functions can be divided into three types-

Physics-based scoring functions: This method calculates the energy that occurs among ligands and targets when they interact. This method does so on the basis of molecular fields of mechanics force.

Empirical scoring functions: This function provides us information about how much surface area is buried, how many hydrogen bonds, and how many rotatable bonds occur in the binding of ligands to protein. There is a series of facts based on which this function provides this information.

Knowledge-based scoring functions: This method relies on the potentials that are distance dependent and occur among the atoms of ligands to their target binding site atoms.

Though the ligand score shows its affinity towards the target protein, results obtained hypothetically do not fully match with this score. There are some methods that can be used to enhance the scoring function's interpretation. One of the methods can be applying not less than two scoring functions jointly with a view to measuring the ligand's affinity for binding to the target. The invention of a new scoring function that is able to calculate the solvation and configuration energies more precisely. (Liao et al., 2013)

3.4 Validation

With the change of the target proteins, it's necessary to switch docking programs and scoring functions in order to increase the efficiency of the docking program and function. There are some parameters that we need to select while using any docking system such as how many shapes of every compound would be stored, the compound's minimum amount of score that is needed to be taken into account, and others.

Before doing molecular docking, it's good to do validation of that docking program so that we can measure how well the program is able to perform. It can be done by taking known ligands and compounds taking indiscriminately. So, when molecular docking would be applied here, it will provide a list of compounds based on most suitability with those ligands (known). Since the ligands are known, we already have the data and now we can cross examine this data with the program's given data. That's how we would be able to measure the efficacy of the molecular docking program. (Liao et al., 2013; Ripphausen et al., 2010)

Postdocking Compound Selection

After doing docking of the compound, we will analyze their scores of binding, and on the basis of it, we will choose compounds. There are some important parameters that need to be checked in all compounds such as conformation, how well it interacts with the target's binding site, and also the size and chemical structure. Mostly it is seen that ligands having high scores lack essential interactions with the binding site of the target. It is assumed that choosing compounds which has the greatest distinct in not only their structural configuration but also interactions of binding to targets can be easily identified by data mining techniques. Here is an example of applying molecular docking in virtual screening.

3.5 Virtual Screening of HIV-1 Protease by AUTODOCK

The binding of ligand to the integrase of HIV-1 can be analyzed by a molecular docking software named AutoDock 4.0. In order to do so, a vast amount of data on compounds is needed. Autodock uses the NCL library to get those data. AutoDock takes into consideration elongated PDBQ files of ligands, elongated PDBQS files of the receptor, files needed for 3D search space, Docking Parameter files containing bibliotheca, file names of Grid Parameter, a ligand that to be employed, search algorithm that to be applied, algorithm specifications, number of dockings that to be executed, scoring function of energy. Figure:7 displays the HIV-1 Protease and the inhibitor Indinavir's molecular docking while the figure:6 visualizes the grid map of the search space. (Lazarova, n.d.)

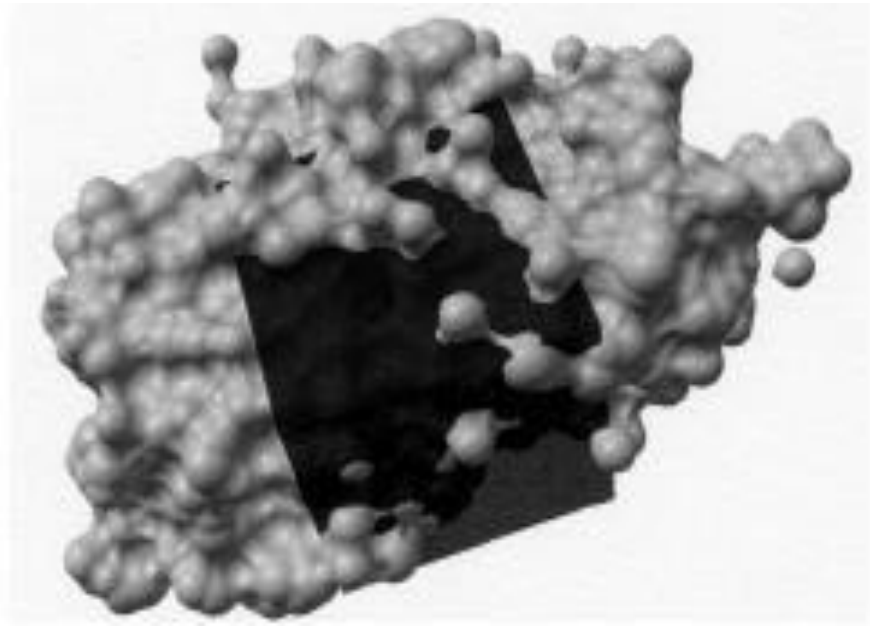


Figure 6: Grid map of the search space. (Lazarova, n.d.)

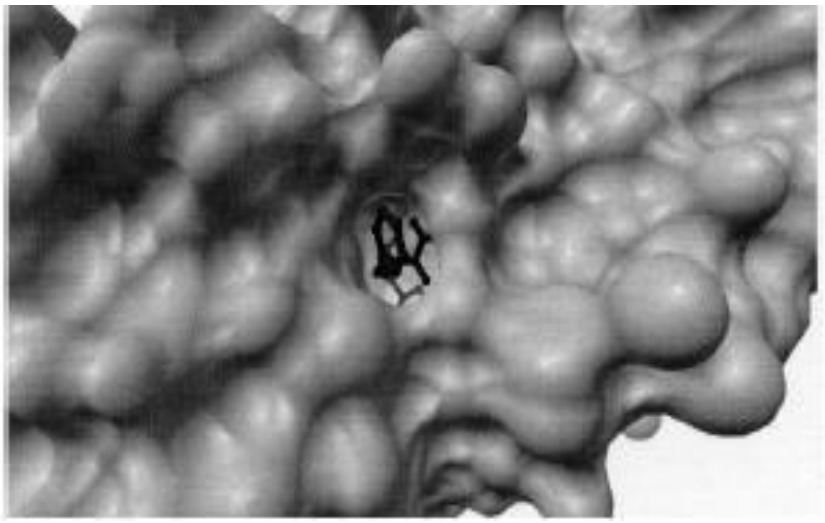


Figure 7: Docking of HIV-1 Protease and the inhibitor Indinavir. (Lazarova, n.d.)

3.6 Conformational Space Algorithm Search for Ligands

An algorithm for ligands can be found by conformational search. Conformational search has various approaches that are used to gradually change the specifications of the ligand structure. By this way, freedom degrees of translation, and rotation of ligands can be modified step by step. Docking is done with a view to achieving precise modeling of the structure as well as forecasting the exact activity. Here search algorithms play an important role by providing the reasonable configuration of the complex. Search algorithms for docking are not the same in every case, it is divided into two types named flexible docking and rigid body docking. The main difference between these two types is that they have incompatible search algorithms. (Yadava, 2018)

The docking method named Rigid-body takes into account only the necessary geometric correspondents. The result of this method has a restriction on its being inch-perfect because this method does not consider ligand and receptor flexibility. The advantage of using this method is that it can result in sites of ligands that will attach to proteins. This result is very helpful in getting crystallographic structures. In order to distinguish between the structures we need Root mean square deviation (RMSD) among the corresponding atoms. Search algorithm

of docking stimulation gives the RMSD. When the value of RMSD is lower than 1.5 Å, it indicates that the output of docking stimulation is perfect. This docking method is the quickest way to get the very beginning stage of the minute molecule's database screening done. DOCK, an algorithm of rigid body docking, is capable of determining which molecules have a greater degree of structure corresponding to the site of attaching. For doing so, this program named DOCK monitors the receptor's molecular superficial part and obtains the negative images of the pocket where the ligand will attach. The coinciding spheres of different radii are gained and it makes the negative site. The technique of checking whether the set is fitted or not is analyzing the space between the central points of the sphere and the space among the ligand atoms. The set is considered as matched only when the space among the central points of the sphere is the same as the space among the ligand atoms (shown in figure:8). In a matched set, just a fit of the square is needed to align the ligand with the central point of the sphere. The algorithm of this docking technique also provides us the facility to examine whether there is a steric conflict among the receptor and ligand. The way of solving an objectionable orientation is converting it into an acceptable orientation. (Ferreira et al., 2015; Yadava, 2018) It is done by realigning the ligand in no more than the lowest fit of the square and at a time acceptable orientation is gained. Then it's possible to successfully put the docked molecule in the target's active site (figure:9)

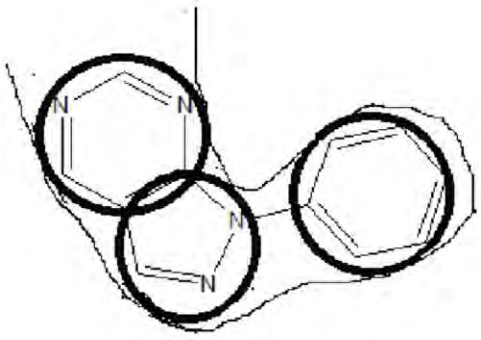


Figure 8: Atoms and central points of sphere centers are matched. (Yadava, 2018)

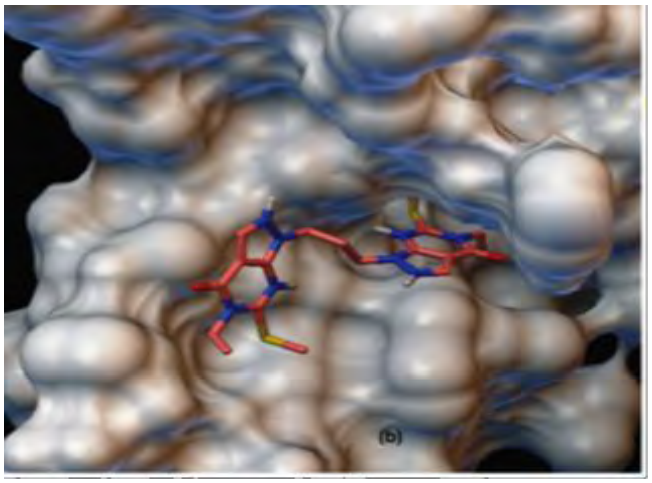


Figure 9: Docked molecule is successfully able to be inside the enzyme's active site. (Yadava, 2018)

Rigid body docking cannot do the lead optimization and accurate refining because it takes not more than six degrees of rotational and transitional freedom. In order to overcome this limitation, flexible docking is applied since it takes into account not only the rotational and transitional freedom degrees but also the freedom degrees of the ligand's shape as well as the receptor's shape. Other docking methods are not able to provide the facility of lead optimization and accurate refining because those methods do not take into account the receptor's conformational space. In spite of having similar techniques in algorithms, the same search algorithms cannot be used for ligand and protein flexibility. Search algorithms for ligand flexibility are categorized in three types-

3.7 Systemic Search Algorithms

The systemic search algorithm makes it possible to result in each and every degree of freedom within a molecule by doing a little change in the structure of the ligand which ultimately causes the alteration in the ligand's shape. This algorithm does so by spinning the angles as well as the bonds. The systemic search algorithm takes all the shapes of molecules that can be feasible. Though this is an advantage of this system, it also creates issues like combinatorial explosion.

Configurational space's energy landscape is examined by this method and after doing the analysis it gives the way where the binding mode requires the least possible energy. Despite having the benefit of showing all the areas of configuration, it fails to meet the global minimum. If we do a continuous search of the various parts situated in the energy landscape from the very beginning, it is much more possible to solve the above issue. Name of some systemic search algorithms are Distance Geometry, Conformational search, etc. (Kier, 1997; Yadava, 2018)

3.8 Random Search Methods

This method does the irregular variations in the ligands or in just one ligand. Among them, the variations that are beneficial are only considered. The advantage of this method is that it can produce all the molecular configurations as well as the landscape of energy in vast amounts. This method can bypass getting the last way as the local energy minimum rather it pumps up the chance of getting the global energy minimum as the last solution. The only drawback of this method is that this method is costly since it induces a vast area of landscape of energy. Name of some random search methods are Tabu search, Genetic algorithm, Monte Carlo simulation, etc. (Yadava, 2018)

3.9 Simulation Methods

Molecular dynamics simulation is the widely used stimulation method in docking. It is based on the second law of Newton. The way of determining the force is to compute the potential energy whenever a change occurs in an atom's position. In this case, the difference between the potential energy in the previous position and the new position of the atom is determined. This force of atoms and their masses are placed in the 2^{nd} law of motion formula provided by

Newton. Integration is done to get the location where the atoms will remain within a short period of time. This simulation method cannot find ligands having global minimum energy. It is because this stimulation method is not able to overcome the obstacles related to high energy. In order to overcome this problem, stimulated annealing is applied. There are some molecular simulations that are used to detect the events of a drug attaching to its specific target, these are called long MD simulations. Only the techniques that minimize energy are not used generally since they result in ligands with local minimum energy. The way of overcoming this drawback is by using those techniques in addition to other search methods. (Hicks & Gulick, 2009; Yadava, 2018)

Chapter 4

Importance of Ligand Flexibility in Molecular Docking

It is the proteins that carry the information of the genome within them. The role of proteins is to do the jobs that are necessary in the human body as well as involved in pathological and physiological processes. However, when the interaction between protein and other molecules occurs, it tends to change the role of protein. That's why, it's very essential to be aware of the possible interactions between proteins and particular molecules in order to design a drug. The advantage of knowing the possible interactions between proteins and particular molecules is that it will help us to design drug molecules that will bind to the specific protein of interest. At first, the shape of a protein is found out and then according to that shape drug is made that will exactly bind to the protein of interest. To do so, a very effective way is to do molecular docking.

There can be various requisitions and purposes behind doing molecular docking, based on that molecular docking has been designed on two elementary apparatuses. One of them is sampling and the other one is scoring. When the specific ligand binds to the protein, there are freedom degrees in the configuration of that binding. Sampling examines that and tells us what is the mode of attaching of specific ligand to the protein. It's also necessary to know the robustness of the mode of attaching. Though the function of sampling and scoring in docking is quite different, their proper functioning is very important with a view to getting accurate results from molecular docking.

In only the initial phases of the rigid docking method, there is no need to detect the flexibility of ligands. But afterward, without detecting the flexibility of ligand, it's quite impossible to do docking. That's why, the detection of ligand flexibility is important. In order to detect it, there are several configurational sampling techniques. Systemic search and stochastic search are the usually used techniques among the different configurational sampling techniques.

For example, from the study of dihydrofolate reductase (DHFR), it is seen in NMR that there are two configurational positions of protein when it remains as unattached and the configurational position of protein alters when a ligand binds to the protein. In order to identify the ligand, we need to know the flexibility of the protein. The two characteristics of proteins are that they are charismatic and they also have flexibility. The protein which is attached by a ligand has a different conformational shape than the protein's conformational shape which is not attached by a ligand. The result of docking is greatly influenced by the flexibility of the protein. (Antunes et al., 2015; Hicks & Gulick, 2009). The docking method that will be used is chosen upon the flexibility of the protein. The concept of docking which can identify biomolecules can be used to categorize the docking methods. Since a significant change occurs in the result of docking because of the flexibility of the ligand, detecting the exact flexibility of the ligand is important to get accurate results in molecular docking. (Antunes et al., 2015; Cavasotto & Abagyan, 2004) The attaching of ligand in the stretchy twist of the protein and formation of the complex of protein-ligand is shown below (figure:10)-

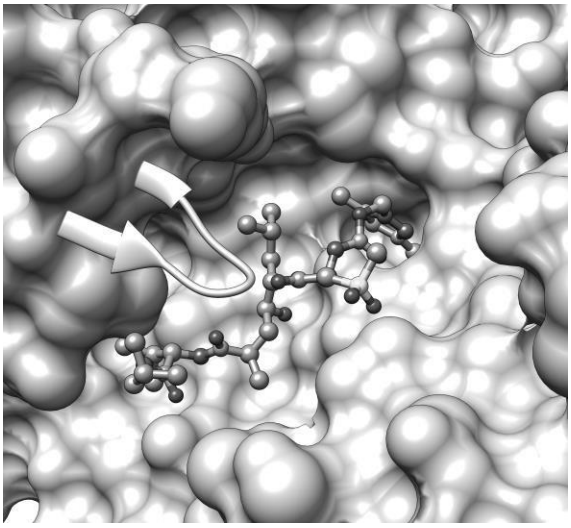


Figure 10: Complex of protein-ligand. Ligand attaches to the stretchy twist of the protein.(Antunes et al., 2015)

Chapter 5

Some Success Stories About Leads or Drugs Obtained from Molecular Docking Studies or Are in Clinical Trials

By using molecular docking, kinase inhibitors are designed which target P13K**α** to treat non-small cell lung cancer.

At first, the protein was prepared. The process of choosing protein is done by using Uniprot. But while choosing a protein, we need to check whether the protein matches with required perquisition such as the existence of a co-crystal ligand, the protein needs to have the same organism as the protein in Homo sapiens and there should be alteration in the gene of the PIK3CA gene. Energy minimization of protein is done because minimization of protein results in the force of inter atom insignificant. That ultimately produces the structure of protein which is best for docking. Then the process for generating a receptor grid is done to identify the site of protein where ligands can bind. The requirement of docking protocol is fulfilled with the help of the composite of protein and co-crystal ligand. In order to match the requirement, the co-crystal ligand is docked in such a way that it would have root mean square deviation that is exactly required to match the requirement of docking (figure:11). Energy minimization of the co-crystal ligand is also required to fulfill the docking requirement.

Secondly, the ligand is prepared. To do so, a protein kinase library is required. Database of ChemDiv is used to get the desired compounds of the protein kinase library. After that, the compound's energy optimization through introduced in Maestro. An application named LigPrep is applied to get 3D conformations. PubChem was adopted to choose a drug that would act as a PI3K pan-inhibitor and a force field named OPLS3e was applied to do energy minimization of ligand.

The third step was to do structure-based virtual screening. In the beginning, 36324 molecules were screened. From it, the finest 1000 molecules were nominated. Again, with a view to doing docking 30 finest molecules were taken from the previous 1000 molecules. Now how well the binding with the desired protein as well as the score of docking are examined to select the best 10 molecules. At the same time, the outcome of these 10 best molecules is analyzed in contrast to the result of the selected drug named Copanlisib.

In the fourth step, molecular mechanics generalized Born surface area is figured out. This value helps to detect what would be the ligand's affinity when it will attach to the receptor. Then drug-likeness is checked. For doing so, a module in Maestro named QikProp is used to examine the ADMET properties of the best 10 hit molecules as well as the standard drug that was selected from PubChem (table 2) Then according to Lipinski's rule of five, drug-likeness is checked.

With a view to making an assumption about flexible ligand docking's impact on the conformation of the protein, the protocol of induced fit docking protocol is applied. (Halder et al., 2022)

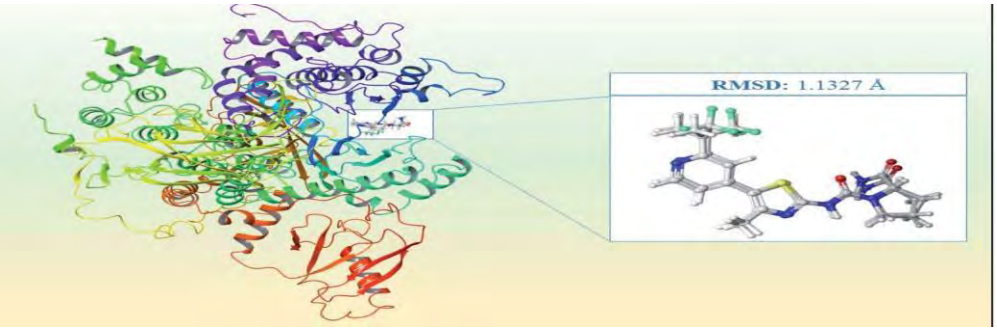
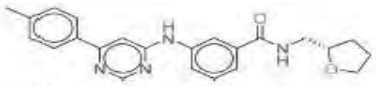
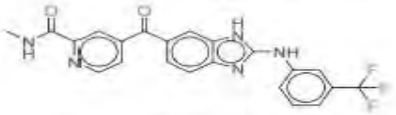
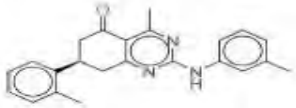
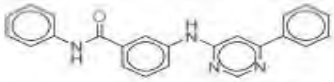
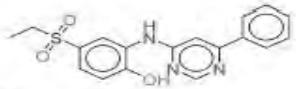
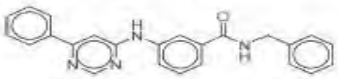
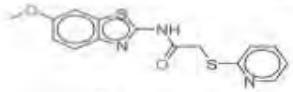
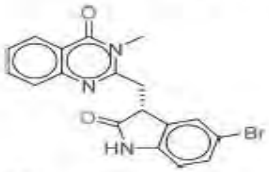
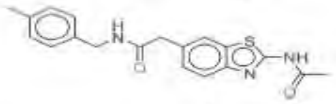
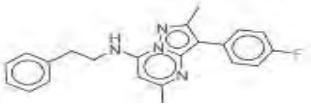
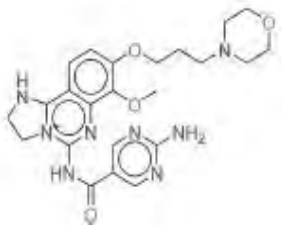


Figure 11: Co-crystal ligand resulted from docking and the co-crystal ligand having RMSD value of 1.1327 °A is laid over each other in order to do the study of validation. (Halder et al., 2022)

Table 2: Best 10 hits from docking along with their conformation and affinity of binding (Halder et al., 2022)

Compound	Structure	Docking score (kcal mol^{-1})	MMGBSA ΔG (kcal mol^{-1})
6943		−11.973	−62.97
34100		−11.312	−55.18
31140		−11.079	−49.53
12500		−11.060	−60.91
14178		−10.822	−53.09
7165		−10.927	−62.46
438		−10.846	−59.11
6450		−10.830	−55.45
19885		−10.823	−52.15
16021		−10.522	−61.09

.

Compound	Structure	Docking score (kcal mol^{-1})	MMGBSA ΔG (kcal mol^{-1})
Copanlisib		−3.941	−42.58

\

A molecular dynamics simulation was done then. The purpose of doing this was to check the perceived best two leads' flexibility. The standard drug Copanlisib, without bonding interaction among the ligand and protein, is contrasted to do the molecular dynamics stimulation. Lastly, synthetic accessibility analysis is done to find out the best two leads score of synthetic accessibility and then it is differentiated with the standard drug Copanlisib. (Halder et al., 2022) That's the method of designing kinase inhibitors that target P13Kα to treat non-small cell lung cancer by molecular docking and it's shown in the figure:11-

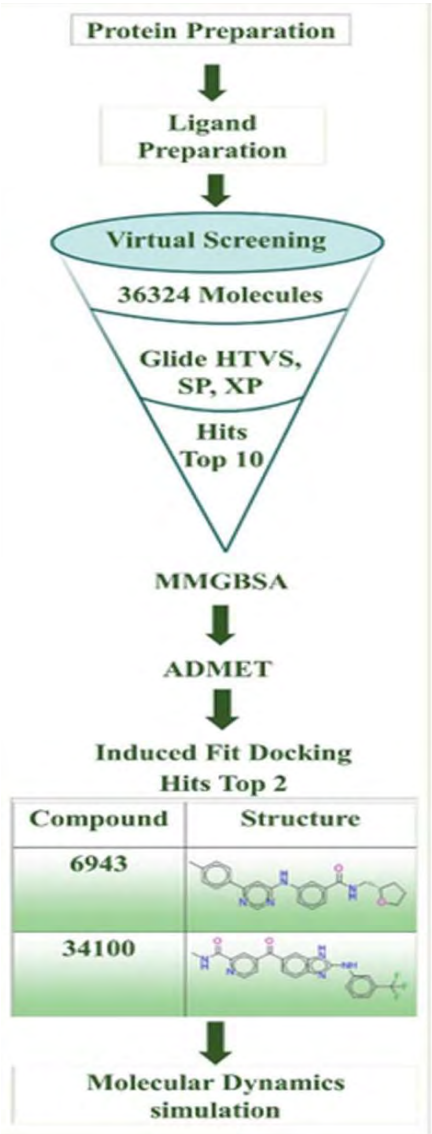


Figure 12: Methods of designing kinase inhibitors that target P13K**α** to treat non-small cell lung cancer by molecular docking presented in the flowchart. (Halder et al., 2022)

5.1 Anti-tubercular Agents Obtained from In-silico Molecular Docking

At first, 35 molecules of *N*-ethyl imidazole pyridine-3-carboxamide are taken. Then docking stimulation of these molecules with DNA gyrase of the target is done. The purpose behind doing that is to examine the attaching affinity of *N*-ethyl imidazole pyridine-3-carboxamide to the target. At first, ChemDraw Ultra Software is used in order to get the conformational structure of those 35 molecules. Then optimization is needed, for doing so a software called Spartan 14 is utilized. At the same time, a virtual screening software named PyPx is applied in order to do the operation of docking. Since the 35 No. molecule has the affinity of attaching (-7.2 kcal/mol) which is greater than the attaching affinity of others, this molecule is considered as the lead molecule. Slight alternation in the conformation of this molecule is made and it results in D1, D2, D3, and D4 named four newborn molecules (figure:13). After comparing the ability to form H bond among M35, isoniazid drug, and D4, it is seen that D4 has the greater ability to form H bonds. D4 complex's interaction of docking is shown in Figure 14. M35 and all four newborn molecules have lower toxicity, good oral bioavailability, and greater absorption. These molecules have the ability to be lead that can be used to design drugs. These drugs would be very effective in the treatment of tuberculosis where multidrug resistance has occurred. (Abdullahi & Adeniji, 2020)

Chemical structure of M35, Isoniazid, and newborn four molecules (D1, D2, D3, D4) with their respective affinity of binding in kcal/mol and D4 complex's interaction of docking are shown in the following page (figure:13 and figure:14).

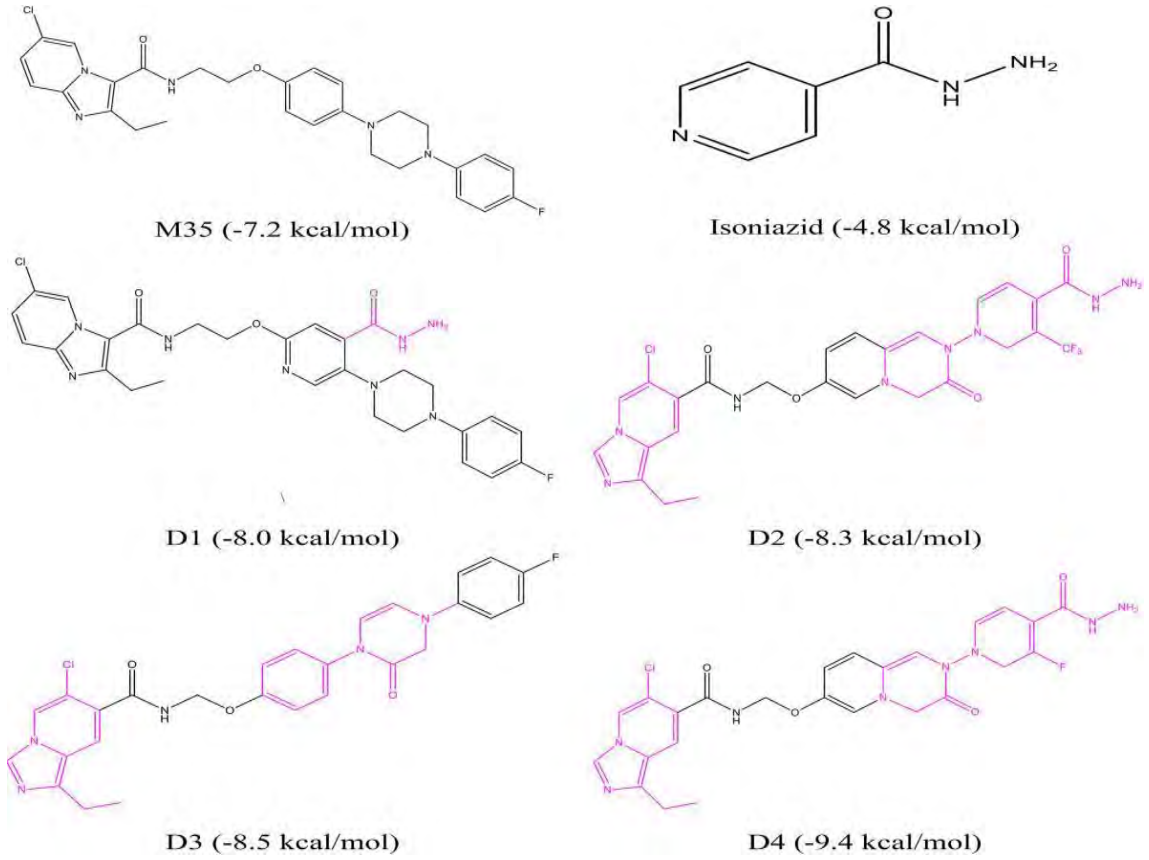


Figure 13: Chemical structure of M35, Isoniazid, and newborn four molecules (D1, D2, D3, D4) with their respective affinity of binding in kcal/mol. (Abdullahi & Adeniji, 2020)

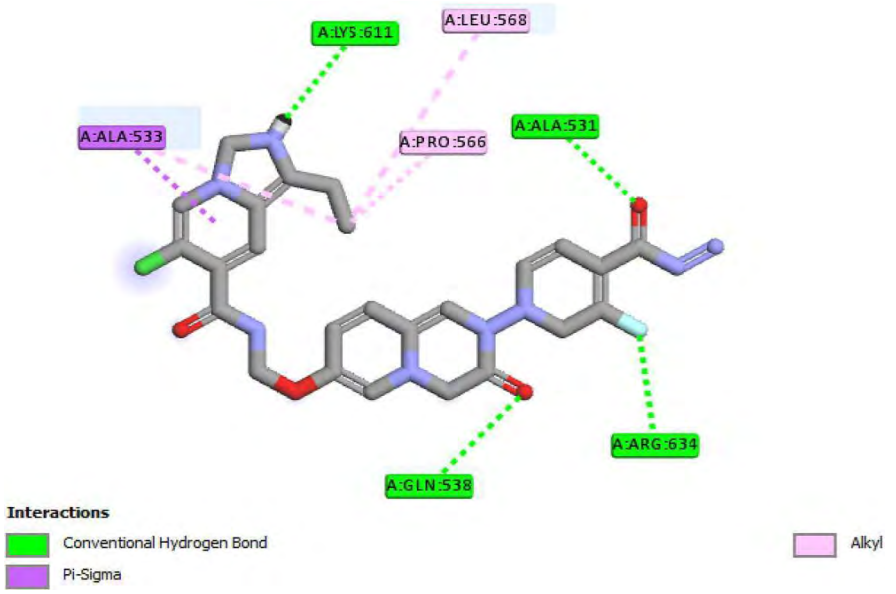


Figure 14: D4 complex's interaction of docking. (Abdullahi & Adeniji, 2020)

Chapter 6

6.1 Discussion

Basically, the pose of the ligand which contains the shape as well as the location of the ligand is detected by molecular docking when the ligand binds to the targeted protein. To find the particular pose of ligand that will result stable ligand-protein complex upon binding to the target, molecular docking uses search algorithms and scoring functions. By applying the search algorithms, molecular docking shows the feasible orientation and shape of the ligand and then scoring functions of molecular docking detect which pose of the ligand would able to form a stable ligand-protein complex. If we can carry out molecular docking correctly, we will get the geometry of the ligand in the attaching site of the target. Molecular docking plays a significant role in designing the structure of a drug. For designing a drug, it's essential to find the molecule that can bind to the target. Molecular docking takes into account the affinity of binding, hydrogen bonding, specificity of binding, van der Waals interactions, complementarity of shape, and flexibility of structure properties to result in the possible leads for drug design. It also examines how well the ligands interact with the target's binding site, ligand's conformation, and also the ligand's size and chemical structure to result in the lead compounds. Mostly it is seen that ligands having high scores lack essential interactions with the binding site of the target. Choosing compounds that have not only the greatest distinct in their structural configuration but also interactions of binding to targets increases the chance of getting the best drug candidate from molecular docking.

We can use docking to assume the pose of the ligand and improve it so that the ligand perfectly fits in the active site of our target. It has become possible because of the tools of molecular docking that help us to calculate every significant ligand's modes of binding in the active site of the target. I think we can use these pieces of information in generating analogs that have

more specificity along with improved efficacy. So, I think molecular docking can be helpful in the optimization of lead.

Due to the progress made in the area of genome theories and unleashing the structures of protein, the use of structure-based docking has become an ideal choice to identify lead compounds in drug discovery. Although it is seen a noticeable improvement in database docking, still we cannot think of database docking not more than a technique of screening. That's why, those who want to reach a decisive projection about the ligands that will be used as inhibitors might be unsuccessful. But docking can be advantageous for us in lead synthesis when the situation is commendatory. In that situation, rates of hits can be lifted as well and what will be the structure of hits while attaching to the target can be known briefly by the use of docking screens. Mainly screens of docking provide us with three types of data. The characteristics of molecules in the first type are those containing the attributes that enable them to attach to various targets. These kinds of molecules are considered in the lists of hits. The molecules that are not drug-like compounds and can react with proteins are in the second type of molecules. Usually, these molecules are eliminated by filters. Molecules having the size of 50-400 nm and can accumulate in solution are the third type. These types of molecules are taken in the lists of hits.

Sometimes scoring functions cannot able to provide the energies of binding correctly, that's the prime drawback of docking. It's because making accurate assumptions about the effect of solvation and the alteration in entropy is very difficult. Scoring functions do not take into account a few intermolecular interactions that are noteworthy in calculating binding energies. It can be shown by an example- despite the interactions between guanidine-arginine are important to consider in predicting the affinity of binding among protein-ligand, the scoring function does not consider this intermolecular interaction. Moreover, molecular docking fails in correctly tackling molecules of water present in the pocket of binding. It's very important to

solve this drawback of molecular docking. The reason is that unless we are sure about the accurate spot of hydrogen, we won't be so sure that the molecules of water are not in the middle of the ligand and target since there is no available theory currently that lets us know about the effect of ligands displacing the molecules of water and how this can affect the bonding network of hydrogen as well. Furthermore, the problem arises when a rigid receptor is used in doing molecular docking. Protein can acquire various shapes when ligand attaches to it. However, only a single shape of a protein is detected when docking is done by using a rigid receptor. This becomes the reason for getting invalid negatives. Lastly, docking fails to show activity range in opposition to the proteins that are not targeted.

In the future, the problems that are decreasing the efficacy of molecular docking in providing accurate results in some cases should be resolved. In my opinion, if we can deal with the role of receptor flexibility, solvent effects, and entropic effects, the issue of getting wrong results from MD in some cases will be solved. Though molecular docking has some drawbacks, still I think molecular docking has the potential to be recognized as an efficient tool in the discovery of drugs. As an example of another success story, Park et. al designed a more improved Aurora Kinase A inhibitor by using molecular docking. The group of Park et. al applied the genetic algorithm to do sampling. At the same time, they did tiny moderation in the energy of dehydration. They used a docking program named AutoDock to do that. Thus, they were successful in designing more improved Aurora Kinase A inhibitors. Discovered Aurora Kinase A inhibitor's molecular structure is shown in the figure:15-

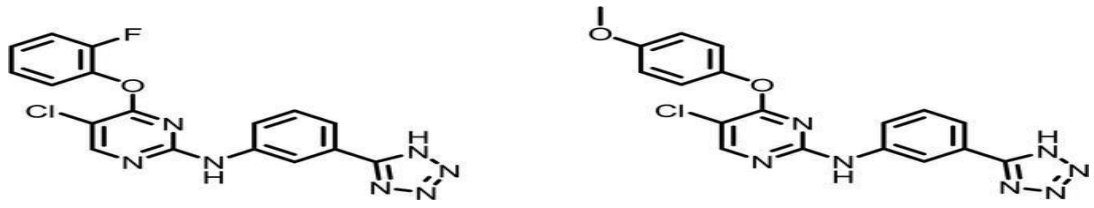


Figure 15: Park et. al discovered Aurora Kinase A inhibitor's molecular structure with IC_{50} 12 and 43 PM subsequently. (Sethi et al., 2019)

Dadashpour et al. discovered COX-2 inhibitors by using molecular docking. This research group used a docking program named AutoDock to construct a lead. They examined all the significant intermolecular interactions and binding energy was accurately calculated. Their designed COX-2 inhibitor's structure is shown below (figure:16)-

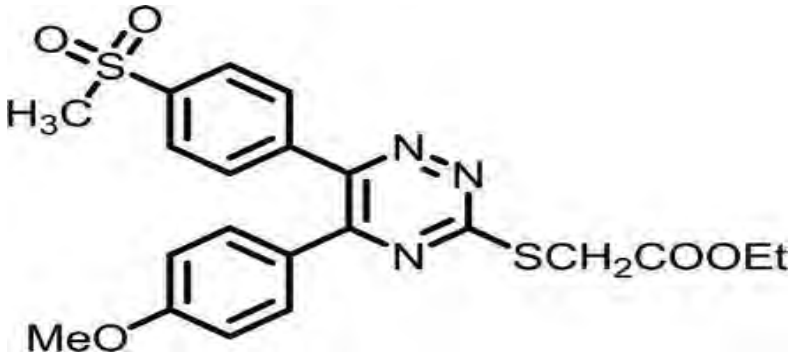


Figure 16: Dadashpour et al. discovered COX-2 inhibitor's structure. (Sethi et al., 2019)

Undoubtedly, it can be said that there are a vast number of reports which show how efficient molecular docking is in pharmaceutical drug discovery and there is no way of declining the utility and potency of molecular docking in drug design.

6.2 Significance and Impact of The Review Article

Researchers interested in molecular docking may have a clear introduction to the field of molecular docking from this review article. This thesis will also help them in finding feasible ways to find physical principles behind the binding affinity of the ligand to the targeted protein. Finding the accurate physical principles behind binding affinity is very important because it helps us select the conformation of the ligand that will result in the stable ligand-protein complex upon binding the ligand to the target. The readers will be aware of the challenges in selecting the correct physical principle behind binding affinity. By reading this thesis, they will be also able to overcome those challenges. Readers will be informed about the various types of molecular docking from this review article. Moreover, readers will get a general idea about

which type of molecular docking should be used on which occasion. I believe the paper will play an important source of information in finding biologically effective substances from vast libraries. The readers will be acknowledged of how to apply a high throughput screening process to quickly find biologically effective substances from vast libraries containing molecules in opposition to specific targets of interest. Various kinds of ligands are docked with the target by the use of molecular docking and then a score of docking shows which one has the highest biological effectivity. Furthermore, this article covers the area of ligand flexibility's importance in getting accurate results from docking. Researchers who want to know the role of molecular docking in drug discovery can get information from this paper. This paper will also enlighten the readers about the successful use of molecular docking in discovered drugs such as kinase inhibitors, anti-tubercular agents, and COX-2 inhibitors. The research group used search algorithms to find out the favorable conformation of the ligand and applied scoring functions to examine the binding affinity of that ligand toward the target. At the same time, all the intermolecular interactions were taken into account. In the future, enthusiastic persons who want to discover drugs by using molecular docking will highly benefit from this paper in my humble opinion.

Chapter 7

Conclusion

Molecular docking is applied to find out the favorable binding orientation for the ligand when it attaches to the binding target. Favorable binding orientation facilitates the stable complex formation between the ligand and the target molecule. Another important aspect of molecule docking is that the tentative binding parameters of drug candidates to their biomolecular target can be known by using molecular docking. Based on the type of molecule, molecular docking is divided into three different categories. If the molecule is not flexible, a molecular docking type named rigid docking is used. On the other hand, the flexible nature of the molecule will facilitate us to do flexible docking. Among rigid and flexible docking, the motive of doing rigid docking is to achieve optimal fit to different substances according to the scoring system's parameters. But in flexible docking, we take into account the transformation and thus we achieve the conformations of the ligand-receptor complex. Another docking method, covalent binding is done to determine a link atom in protein as well as in ligand. This approach works in a way that both the link atom of protein and ligand have equal steric volume. Molecular docking enlightens us about the thorough knowledge of binding mode. The advantage of this binding knowledge is that it exactly tells us about the ligand's location and conformation compared to the target. Molecular docking also has the ability to check whether the ligand suits the target or not. If the ligand suits the target, molecular docking can also show how perfect the interaction is between the ligand and the binding site of the target. The accuracy of molecular docking's result is vastly dependent on the flexibility of the ligand. However, when the interaction between protein and ligand occurs, it tends to change the role of protein. That's why, it's very essential to be aware of the possible interactions between proteins and particular molecules to design a drug. The advantage of knowing the possible interactions between

proteins and particular molecules is that it will help us to design drug molecules that will bind to the specific protein of interest. Molecular docking plays playing important role in drug discovery. It easily finds biologically effective substances from vast libraries containing molecules in opposition to specific targets of interest. Among the important drug discoveries by docking, some of them are- kinase inhibitors targeting P13Kα which is prescribed to heal non-small cell lung cancer, anti-tubercular agents that have been designed to treat tuberculosis, novel RARα agonists that are used to heal serious dermatological disorders.

References

Abdullahi, M., & Adeniji, S. E. (2020). In-silico Molecular Docking and ADME/Pharmacokinetic Prediction Studies of Some Novel Carboxamide Derivatives as Anti-tubercular Agents. *Chemistry Africa*, *3*(4), 989–1000. https://doi.org/10.1007/S42250-020-00162-3

Agarwal, S., Chadha, D., & Mehrotra, R. (2015). Molecular modeling and spectroscopic studies of semustine binding with DNA and its comparison with lomustine-DNA adduct formation. *Journal of Biomolecular Structure & Dynamics*, *33*(8), 1653–1668. https://doi.org/10.1080/07391102.2014.968874

Agarwal, Shweta & Mehrotra, Ranjana. (2016). Mini Review_ An Overview of Molecular Docking. *JSM Chemistry,* *2*(1024). *h*ttps://www.researchgate.net/publication/303897563

Antunes, D. A., Devaurs, D., & Kavraki, L. E. (2015). Understanding the challenges of protein flexibility in drug design. *Expert Opinion on Drug Discovery*, *10*(12), 1301–1313. https://doi.org/10.1517/17460441.2015.1094458

Becker, O. M., MacKerell Jr., A. D., Roux, B., & Watanabe, M. (2001). Computational Biochemistry and Biophysics (O. M. Becker, A. D. MacKerell Jr., B. Roux, & M. Watanabe, Eds.). CRC Press. https://doi.org/10.1201/9780203903827

Cavasotto, C. N., & Abagyan, R. A. (2004). Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *Journal of Molecular Biology*, *337*(1), 209–225. https://doi.org/10.1016/j.jmb.2004.01.003

Chao, J., Dai, Y., Verpoorte, R., Lam, W., Cheng, Y. C., Pao, L. H., Zhang, W., & Chen, S. (2017). Major achievements of evidence-based traditional Chinese medicine in treating major diseases. *Biochemical Pharmacology*, *139*, 94–104. https://doi.org/10.1016/J.BCP.2017.06.123

de Azevedo Jr., W., & Dias, R. (2008). Computational methods for calculation of ligand-binding affinity. *Current Drug Targets*, *9*(12), 1031–1039. https://doi.org/10.2174/138945008786949405

Dias, R., & de Azevedo Jr., W. (2008). Molecular docking algorithms. *Current Drug Targets*, *9*(12), 1040–1047. https://doi.org/10.2174/138945008786949432

Dnyandev, K. M., Babasaheb, G. V., Chandrashekhar, K. V., Chandrakant, M. A., & Vasant, O. K. (2021). A Review on Molecular Docking. *International Research Journal of Pure and Applied Chemistry*, 60–68. https://doi.org/10.9734/IRJPAC/2021/V22I330396

Ferreira, L. G., Dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular Docking and Structure-Based Drug Design Strategies. *Molecules*, *20*(7), 13384. https://doi.org/10.3390/MOLECULES200713384

Guedes, I. A., de Magalhães, C. S., & Dardenne, L. E. (2014). Receptor-ligand molecular docking. *Biophysical Reviews*, *6*(1), 75–87. https://doi.org/10.1007/S12551-013-0130-2

Halder, D., Das, S., Aiswarya, R., & Jeyaprakash, R. S. (2022). Molecular docking and dynamics based approach for the identification of kinase inhibitors targeting PI3Kα against non-small cell lung cancer: a computational study. *RSC Advances*, *12*(33), 21452. https://doi.org/10.1039/D2RA03451D

Hicks, C., & Gulick, R. M. (2009). Raltegravir: the first HIV type 1 integrase inhibitor. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *48*(7), 931–939. https://doi.org/10.1086/597290

Huang, N., Kalyanaraman, C., Bernacki, K., & Jacobson, M. P. (2006). Molecular mechanics methods for predicting protein–ligand binding. *Physical Chemistry Chemical Physics*, *8*(44), 5166–5177. https://doi.org/10.1039/B608269F

Khan T, Lawrence AJ, Azad I, Raza S, Khan AR (2018) Molecular Docking Simulation with Special Reference to Flexible Docking Approach, *JSM Chem, 6*(1): 1053. https://doi.org/10.47739/2334-1831/1053

Kier, L. B. (1997). Molecular Modeling: Principles and Applications By Andrew Leach. Longman, Edinburgh. 1996. xvi + 595 pp. ISBN 0-582-23933-8. $35.00. *Journal of Medicinal Chemistry*, *40*(18), 2969–2969. https://doi.org/10.1021/JM970383K

Kumalo, H. M., Bhakat, S., & Soliman, M. E. S. (2015a). Theory and Applications of Covalent Docking in Drug Discovery: Merits and Pitfalls. *Molecules*, *20*(2), 1984. https://doi.org/10.3390/MOLECULES20021984

Kumalo, H. M., Bhakat, S., & Soliman, M. E. S. (2015b). Theory and applications of covalent docking in drug discovery: Merits and pitfalls. In *Molecules* (Vol. 20, Issue 2, pp. 1984–2000). MDPI AG. https://doi.org/10.3390/molecules20021984

Lamb, M. L., & Jorgensen, W. L. (1997). Computational approaches to molecular recognition. *Current Opinion in Chemical Biology*, *1*(4), 449–457. https://doi.org/10.1016/S1367-5931(97)80038-5

Lazarova, M. (n.d.). *Virtual Screening-Models, Methods, and Software Systems*. https://www.researchgate.net/publication/264849817

Liao, C., Peach, M. L., Yao, R., & Nicklaus, M. C. (2013). Molecular docking and structure-based virtual screening. *In Silico Drug Discovery and Design*, 6–20. https://doi.org/10.4155/EBO.13.181

Paramashivam, S. K., Elayaperumal, K., Natarajan, B., Ramamoorthy, M., Balasubramanian, S., & Dhiraviam, K. (2015). In silico pharmacokinetic and molecular docking studies of small molecules derived from Indigofera aspalathoides Vahl targeting receptor tyrosine kinases. *Bioinformation*, *11*(2), 73–84. https://doi.org/10.6026/97320630011073

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/JCC.20084

Raval, K., & Ganatra, T. (2022). Basics, types, and applications of molecular docking: A review. *IP International Journal of Comprehensive and Advanced Pharmacology*, *7*(1), 12–16. https://doi.org/10.18231/J.IJCAAP.2022.003

Ripphausen, P., Nisius, B., Peltason, L., & Bajorath, J. (2010). Quo Vadis, virtual screening? A comprehensive survey of prospective applications. *Journal of Medicinal Chemistry*, *53*(24), 8461–8467. https://doi.org/10.1021/JM101020Z/ASSET/JM101020Z.FP.PNG_V03

Seeliger, D., & De Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of Computer-Aided Molecular Design*, *24*(5), 417–422. https://doi.org/10.1007/S10822-010-9352-6

Sethi, A., Joshi, K., Sasikala, K., Alvala, M., Sethi, A., Joshi, K., Sasikala, K., & Alvala, M. (2019). Molecular Docking in Modern Drug Discovery: Principles and Recent Applications. *Drug Discovery and Development - New Advances*. https://doi.org/10.5772/INTECHOPEN.85991

Torres, P. H. M., Sodero, A. C. R., Jofily, P., & Silva-Jr, F. P. (2019). Key topics in molecular docking for drug design. In *International Journal of Molecular Sciences* (Vol. 20, Issue 18). MDPI AG. https://doi.org/10.3390/ijms20184574

Yadava, U. (2018). Search algorithms and scoring methods in protein-ligand docking. *Endocrinology & Metabolism International Journal*, *6*(6). https://doi.org/10.15406/EMIJ.2018.06.0021