Republic of Iraq

Ministry of Higher Education and Scientific Research

University of Misan

College of Pharmacy

"The Role of Molecular Docking Computational Tools in Drug Discovery"

A Thesis Submitted to the Council of the College of Pharmacy, University of Misan, as a Partial Fulfillment of the Requirements for the Bachelor's Degree in Pharmacy Sciences.

Prepared by:

Fatima Farhan Abdulhamid

Fatima Ghazi Basi

Fatima Anin Mohammad

Fatima Mitham Taama

Supervised by:

Assist. Prof. Dr. Mohammed Talib Jasim

2025 **Maysan** **1446**

بسم الله الرحمن الرحيم

(يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ)

(سورة المجادلة: 11)

SUPERVISOR CERTIFICATION

I attest that the research project, "The Role of Molecular Docking Computational Tools in Drug Discovery," was completed as a capstone project at the University of Misan's College of Pharmacy under my supervision.

Supervisor signature:

Dedication

-﴿ وَآخِرُ دَعْوَاهُمْ أَنْ الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ﴾ سورة يونس -10

لَمْ تَكُنِ الرِّحْلَةُ قَصِيرَةً وَلَمْ يَكُنِ الطَّرِيقُ مَحْفُوفًا بِالتَّسْهِيلَاتِ، لَكِنِّي فَعَلْتُهَا وَهَا أَنَا قَدْ وَصَلْتُ إِلَى نِهَايَةِ رِحْلَتِي الْجَامِعِيَّةِ بَعْدَ تَعَبٍ وَمَشَقَّةٍ وَبَعْدَ خَمْسِ سِنِينَ فِي سَبِيلِ الْحُلْمِ وَالْعِلْمِ حَمَلْتُ فِي طَيَّاتِهَا أَمَانِيَّ وَمَنَاكِبَ سَعْيِي فَكُلُّهُ أَمَسَّي مِيْعَادُ الْيَوْمِ.

الْحَمْدُ لِلَّهِ الَّذِي يُسِّرَ الْبِدَايَاتِ وَبَلَّغَنَا النِّهَايَاتِ بِفَضْلِهِ وَكَرَمِهِ.

أُهْدِي هَذَا الْعَمَلَ الْمُتَوَاضِعَ لِنَفْسِي أَوَّلًا ثُمَّ لِمَنْ سَعَى مَعِي لِإِتْمَامِ هَذِهِ الْمَسِيرَةِ دُمْتُمْ سَنَدًا لَا عُمْرَ لَهُ لِـ:

مَنْ أَوْصَانِي لِمَا أَنَا عَلَيْهِ الآنَ وَسَهِرُوا اللَّيَالِيَ مِنْ أَجْلِي لِمَنْ قَالَ فِيهِمَا عَزَّ وَجَلَّ: "وَاخْفِضْ لَهُمَا جَنَاحَ الذُّلِّ مِنَ الرَّحْمَةِ وَقُلْ رَبِّ ارْحَمْهُمَا كَمَا رَبَّيَانِي صَغِيرًا" لِنُورِ عَيْنَيَّ وَالِدَيَّ الْعَزِيزَيْنِ.

لِمَنْ جَعَلَ اللَّهُ الْجَنَّةَ تَحْتَ أَقْدَامِهَا، لِمَنْ كَانَتْ دُعَواتُهَا تُحِيطُنِي وَتُسَعِدُنِي فِي كُلِّ وَقْتٍ قُرَّةَ عَيْنِي وَسِرَّ نَجَاحِي وَسَنَدِي -أُمِّي-

لِمَنْ كَلَّلَ الْعَرَقُ جَبِينَهُ، لِمَنْ عَلَّمَنِي أَنْ النَّجَاحَ لَا يَأْتِي إِلَّا بِالصَّبْرِ وَالْإِصْرَارِ لِلنُّورِ الَّذِي أَنَارَ دَرْبِي -أَبِي-

لِضَلْعِي الثَّابِتِ وَأَمَانِ أَيَّامِي إِلَى مَنْ شَدَدْتُ عَضُدِي بِهِمْ فَكَانُوا لِي يَنَابِيعَ أَرْتَوِي مِنْهَا لِخَيْرَةِ أَيَّامِي وَصَفْوَتِهَا -أَخَوَاتِي وَإِخْوَانِي-

لِمَنْ دَعَمُونِي وَكَانُوا شُرَكَائِي فِي هَذِهِ الرِّحْلَةِ.

مَنْ قَالَ "أَنَا لَهَا نَالَهَا"، وَأَنَا لَهَا وَإِنْ أَبَتْ رَغْمًا عَنْهَا أَتَيْتُ بِهَا فَالْحَمْدُ لِلَّهِ شُكْرًا وَحُبًّا وَامْتِنَانًا عَلَى الْبَدْءِ وَالْخِتَامِ.

Acknowledgement

Here it ends, all the nights and days of hard work, here is the time that the present become a memory and the mentorship of **Assist. Prof. Dr. Mohammed Talib Jasim** become an experience for our future, here is the turning point in which we grow, glow and look forward, here we look for to become a figure for the future generation just like the man who taught us, guided us and helped in all the steps of becoming who we are, here is the end of the journey, where we finish this thesis, but our respect and appreciation will always remain for you.

Thank you, Dr. Mohammed Talib

Sincerely

Your trustworthy students

Contents

Abstract

Molecular docking has become an important component of the drug discovery process. Since first being developed in the 1980s, advancements in the power of computer hardware and the increasing number of and ease of access to small molecule and protein structures have contributed to the development of improved methods, making docking more popular in both industrial and academic settings. Over the years, the modalities by which docking is used to assist the different tasks of drug discovery have changed. Although initially developed and used as a standalone method, docking is now mostly employed in combination with other computational approaches within integrated workflows.

Despite its invaluable contribution to the drug discovery process, molecular docking is still far from perfect. In this chapter we will provide an introduction to molecular docking and to the different docking procedures with a focus on several considerations and protocols, including protonation states, active site waters and consensus, that can greatly improve the docking results

Introduction and Literature Review

The discipline of molecular docking has developed over the past thirty years, propelled by the demands of structural molecular biology and structure-based drug discovery. The significant increase in the availability and capability of computers, along with the enhanced accessibility of small chemical and protein databases, has tremendously assisted this process. [1–4]

The docking procedure entails predicting the structure and orientation of a ligand within a certain binding site. Docking investigations often pursue two objectives: precise structural modeling and accurate activity prediction. Nonetheless, the identification of molecular characteristics responsible for particular biological recognition, or the prediction of compound alterations that enhance potency, presents intricate challenges that are frequently difficult to comprehend and, even more so, to model computationally. Considering these problems, docking is typically structured as a multi-stage procedure, with each step adding one or more layers of complexity.[5] The procedure commences with the utilization of docking algorithms that position tiny molecules within the active region. This is inherently hard, as even very basic organic molecules may possess numerous conformational degrees of freedom. Sampling these degrees of freedom must be executed with adequate precision to discern the conformation that most closely aligns with the receptor structure, and must be sufficiently rapid to facilitate the assessment of thousands of molecules in a single docking run. Algorithms are enhanced by SCORING FUNCTIONS intended to forecast biological activity by assessing interactions between substances and prospective targets.

Initial scoring functions assessed compound compatibility by calculations of approximate form and electrostatic complementarity. Basic scoring functions are extensively utilized, particularly in the first phases of docking simulations. Pre-selected conformers are frequently subjected to additional assessment utilizing intricate scoring methodologies that provide a more comprehensive analysis of electrostatic and van der Waals interactions, as well as incorporating certain solvation or entropic effects.[6]

Molecular docking serves numerous functions in drug discovery, encompassing structure–activity analyses, lead optimization, identification of potential leads through virtual screening, formulation of binding hypotheses to enhance predictions for mutagenesis studies, support for x-ray crystallography in aligning substrates and inhibitors with electron density, investigation of chemical mechanisms, and design of combinatorial libraries.[7] High-Throughput Screening (HTS) of extensive chemical libraries against molecular target panels has established itself as the benchmark for identifying biologically active hits. Nonetheless, the substantial expenses associated with the establishment and maintenance of these screening platforms frequently impede their application in drug discovery.[8]

Furthermore, in light of recent advancements in computer technology and the swift proliferation of structural, chemical, and biological data pertaining to an expanding arry of therapeutic targets, it is evident that the utilization of in silico methodologies such as chemoinformatics, molecular modeling, and artificial intelligence (AI) has markedly escalated in recent decades. [9–13]

In silico methodologies currently provide the virtual screening of millions of compounds in a cost-effective manner, hence decreasing the preliminary expenses of hit discovery and enhancing the likelihood of discovering suitable drug candidates. Currently, various molecular modeling strategies exist to aid drug discovery, mostly categorized into structure-based and ligand-based methodologies.[14] Structure-based approaches utilize information obtained from the three-dimensional structure of a target of interest, enabling the ranking of molecular databases based on the structural and electronic complementarity of ligands to a certain target.[15]

Molecular docking is a prominent and effective structure-based in silico approach that predicts interactions between molecules and biological targets.[15] This procedure typically involves initially forecasting the molecular orientation of a ligand within a receptor, followed by assessing their complementarity using a scoring function.[15]

Docking has demonstrated significant efficacy in structure-based drug design (SBDD), leading to its extensive development and enhancement throughout the years. In the past

two decades, over 60 distinct docking technologies have been created in both academic and commercial environments. Numerous studies exist that elucidate the assessment and comparison of various docking programs. [16-19]

Objectives

The main objectives of molecular docking are:

1. Predict Binding Modes: To predict the preferred orientation (pose) of a small molecule (ligand) when it binds to a target macromolecule (usually a protein or DNA).

2. Estimate Binding Affinity: To estimate the strength of the interaction (binding energy or score) between the ligand and the target.

3. Drug Discovery: To identify potential drug candidates by virtually screening large libraries of compounds and selecting those with high binding affinity to a specific biological target.

CADD (COMPUTER-AIDED DRUG DISCOVERY) INVOLVES

a. The application of computational techniques to optimize the drug discovery and development process.

b. Utilize chemical and biological data on ligands and/or targets to identify and enhance innovative pharmaceuticals.

c. Development of in-silico filters to eliminate chemical compounds with undesirable features (suboptimal activity and/or inadequate Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET)) and identify the most promising candidates.

d. Identification of innovative drug targets and acquisition via databases of target protein structures, such as the Protein Data Bank (PDB). CADD (Figure 1) is employed to identify potential medication candidates. Virtual screening is utilized to identify novel drug candidates from diverse chemical scaffolds by examining databases. [20-21]

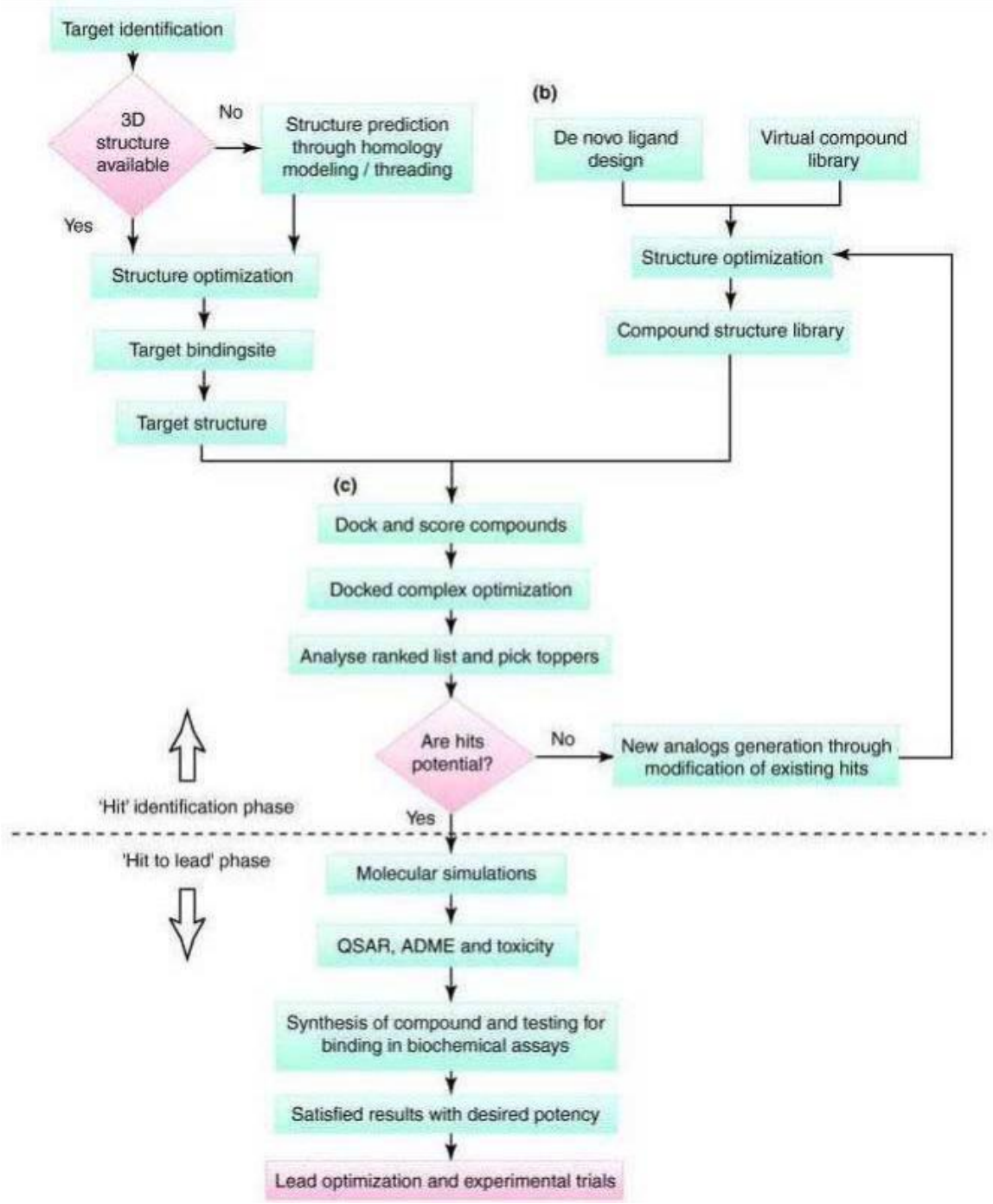

Figure 1 The computer aided drug design and discovery (CADDD) procedure. (Chaudhary, Kamal Kumar, and Nidhi Mishra. 2016)

Diverse forms of docking The primary approach employed for docking is the Lock and Key model.Rigid Docking - Both the receptor and ligand are held stationary during the docking process. Induced fitFlexible docking-involves induced fit docking, wherein both the ligand and the receptor exhibit conformational flexibility. At each rotation, the surface cell occupancy and energy are computed; subsequently, the most optimal position is chosen.[22]

Principal stages involved in the mechanics of molecular docking

Molecular docking is the in-silico examination of intermolecular interactions between two molecules. The protein receptor serves as the macromolecule in this process. The micro molecule is the ligand, which can function as an inhibitor. The docking process comprises the subsequent steps:

Step I – Protein Preparation: The three-dimensional structure of the protein must be obtained from the Protein Data Bank (PDB); then, the recovered structure should undergo pre-processing. This should allow for the evacuation of water molecules from the cavity, stabilising the charges, filling the absent residues, and generating the side chains, in accordance with the supplied specifications.

Step II – Active Site Prediction: Following protein production, the active site of the protein must be predicted. The receptor may contain numerous active sites, but only one of interest should be selected. The removal of water molecules and heteroatoms, if present, is predominantly executed [23-24].

Step III – Ligand Preparation: Ligands can be obtained from several databases, including ZINC and PubChem, or can be illustrated using the ChemSketch tool. The selection of the ligand should adhere to Lipinski's Rule of Five. The Lipinski Rule of Five aids in distinguishing between non-drug-like and drug-like candidates. It offers a significant probability of success or failure for drug-like compounds that adhere to two or more of the established criteria. For the selection of a ligand in accordance with Lipinski's Rule: Fewer than five hydrogen bond donors Fewer than ten hydrogen bond acceptors Molecular mass

under 500 Da High lipophilicity (shown by LogP not exceeding 5) 5 The molar refractivity should range from 40 to 130.

Step IV - Docking: The ligand is positioned against the protein, and the interactions are examined. The scoring function assigns a score based on the selected optimal docked ligand complex.[25]

Approaches of Molecular Docking

Two principal techniques are used for conducting molecular docking. One strategy uses computer simulations to estimate the energy profile for the docked conformer of the ligand target. The second approach employs a technique that assesses the surface complementarity between the ligand and the target [26].

Simulative Methodology

This method involves the separation of ligand and target molecules by a physical distance, after which the ligand is permitted to bind to the groove or pocket of the target molecule following a specified number of movements within its conformational space. The modifications pertain to alterations in the ligand structure, either internally through torsional angle rotations or externally via rigid body transformations, including rotations and translations. Each movement within the conformational constraints of the ligand produces energy, quantified as the "Total Energy of the System." This method is superior to the form complementarity technique as it better accommodates ligand flexibility within molecular modelling tools. An additional advantage of this approach is that it provides a more accurate assessment of the chemical recognition between the ligand and the target molecule. Nonetheless, molecular docking employing this method requires an extended period to evaluate the best docked conformer, as extensive energy landscapes must be computed for each position. Nonetheless, rapid optimisation techniques and grid-based tools have significantly transformed this limitation, enhancing the user-friendliness of computer simulation methods [26,27].

Shape Complementarity Method

This method utilises ligand and target as a collection of surface structure characteristics that enable their molecular docking. To accomplish molecular docking, the molecular surface of the target is characterised in relation to its solvent-accessible surface area, while the ligand's molecular surface is represented through a corresponding surface depiction. The complementarity of two molecular surfaces is assessed through shape matching, which aids in identifying the complementary groove or pocket for ligand docking on the target molecular surface. Specifically, for protein target molecules, hydrophobicity is assessed by evaluating the number of twists in the main-chain atoms. The shape complementarity approach is efficient and reliable, enabling the rapid assessment of hundreds of ligands within seconds to determine their potential binding characteristics on the target molecular surface. [26, 27]

Structure-Based Drug Design (SBDD)

Comprehending the mechanisms governing the recognition and interaction of small-molecule ligands with macromolecules is crucial in pharmaceutical research and development (R & D) [28]. SBDD denotes the methodical application of structural data, such as macromolecular targets or receptors, typically acquired through experimental means or computational homology modelling. The objective is to design ligands with certain electrostatic and stereochemical characteristics to attain elevated receptor binding affinity. The accessibility of three-dimensional macromolecular structures for a thorough examination of the binding site architecture, encompassing clefts, cavities, and sub-pockets. Electrostatic characteristics, including charge dispersion, can be meticulously analysed. Contemporary structure-based drug design (SBDD) techniques facilitate the creation of ligands with essential characteristics for effective regulation of the target receptor [28,29]. The selective regulation of a validated drug target by high-affinity ligands disrupts certain cellular processes, ultimately resulting in the intended pharmacological and therapeutic effects [30]. SBDD is a cyclical process including incremental knowledge gain **(Figure 2)**.

In silico investigations are performed to find possible ligands based on a known target structure. The synthesis of the most promising molecules is conducted subsequent to these molecular modelling approaches [31].

Subsequently, assessments of biological features, including potency, affinity, and efficacy, are conducted utilising several experimental platforms [32]. Upon identification of active molecules, the three-dimensional structure of the ligand-receptor complex can be elucidated. The existing structure facilitates the observation of several intermolecular characteristics that underpin the process of molecular recognition. Structural descriptions of ligand-receptor complexes facilitate the examination of binding conformations, the characterisation of critical intermolecular interactions, the identification of unknown binding sites, mechanistic investigations, and the elucidation of ligand-induced conformational alterations [33]. Upon establishing a ligand-receptor complex, biological activity data are correlated with the structural information [34]. The SBDD process recommences with new steps aimed at integrating molecular changes that may enhance the affinity of novel ligands for the binding site. The adaptability of the target receptor is a critical factor to consider during the modelling phase, as significant conformational alterations may arise upon ligand interaction. Techniques like flexible docking and molecular dynamics are effective in tackling the issue of flexibility [35,36].

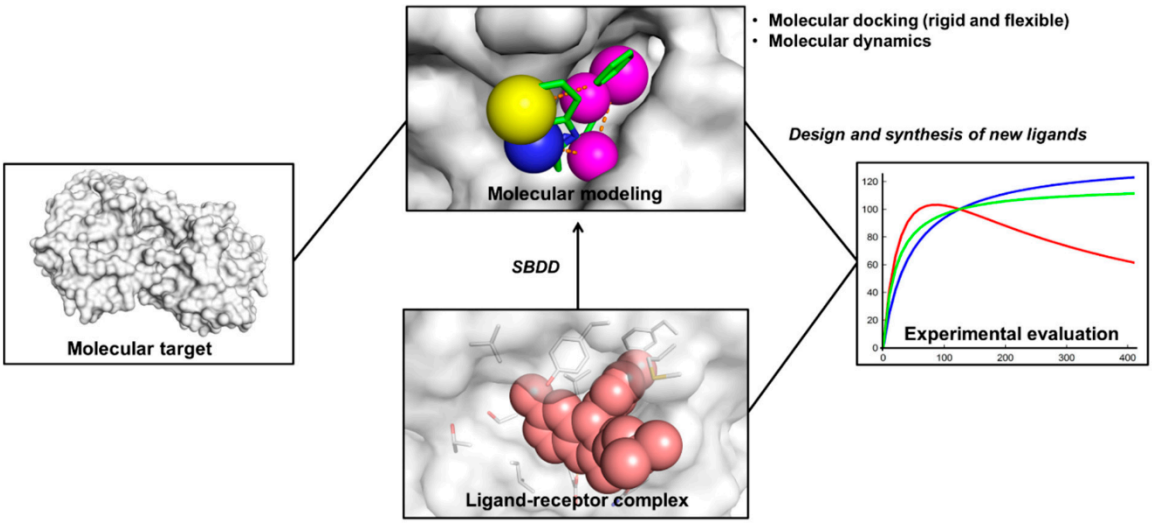

Figure 2: Outline of SBDD. The three-dimensional structure of the molecular target is employed in molecular modeling studies. Promising compounds are synthesized and then experimentally evaluated. Given that bioactive small-molecules are discovered, the structure of a ligand-receptor complex can be obtained. The binding complex is used in molecular modeling studies and novel compounds are designed. (Ferreira, Leonardo G., et al. 2015)

Molecular representations for docking

Molecular representations for docking purposes Evaluating different docking approaches necessitates consideration of the representation of the protein and ligand. The receptor can be represented in three fundamental forms: atomic, surface, and grid [37]. Atomic representation is typically employed solely in conjunction with a potential energy function [38] and frequently only during the final ranking stages due to the computational difficulty associated with analysing pairwise atomic interactions. Surface-based docking programs are generally employed in protein–protein docking, though not exclusively [39,40]. Connolly's initial contributions to molecular surface representations significantly catalysed further research in this domain [41,42]. These strategies seek to align locations on surfaces by reducing the angle between the surfaces of opposing molecules [43]. Consequently, the rigid body approximation remains the norm for numerous protein–protein docking methodologies. Goodford [44] pioneered the utilisation of potential energy grids, which are employed by many docking algorithms for energy computations.

The fundamental concept is to retain data regarding the receptor's energetic contributions at grid points, allowing it to be accessed solely during ligand scoring. Grid points fundamentally store two categories of potentials: electrostatic and van der Waals. (figure 3) depicts a representative grid for the acquisition of electrostatic potentials, while (figure 4) demonstrates the electrostatic potential of a bound inhibitor projected onto its molecular surface.

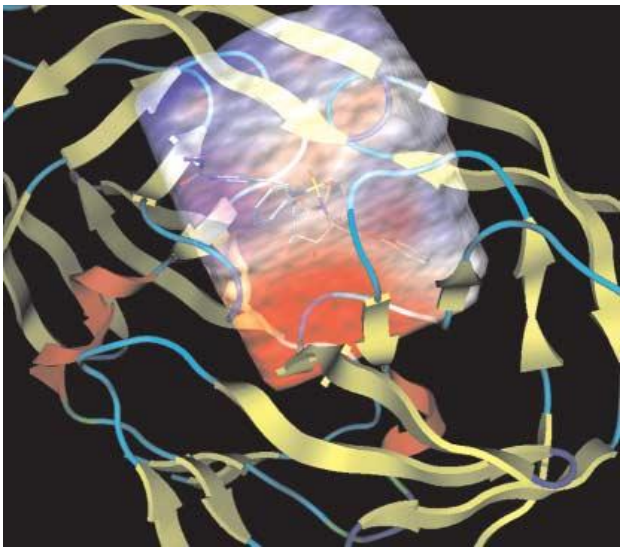

Figure 3a

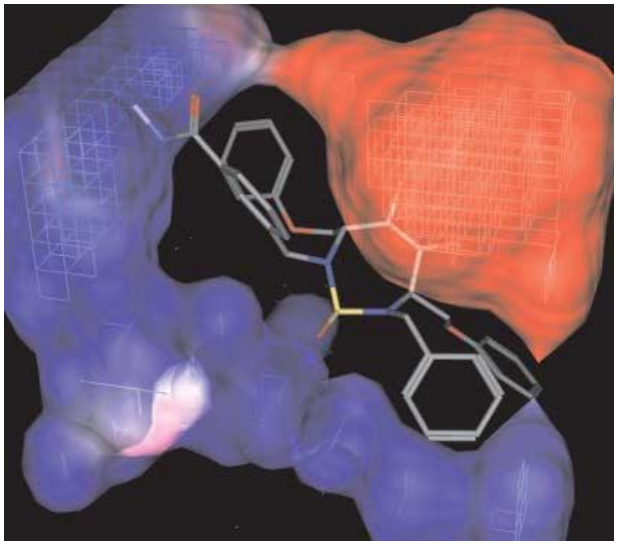

Figure 4b

Figure 3: Grid representations. A surface plot depicting the electrostatic potential of HIV protease (PDB code: 1BVE) surrounding its active site, which includes the binding inhibitor Dmp323, is presented. Red and blue signify regions of negative and positive electrostatic potential, respectively. b | Displays a 'cut-away' electrostatic potential grid of the enzyme surrounding the bound inhibitor (excluded from the computation). (Kitchen, Douglas B., et al. 2004)

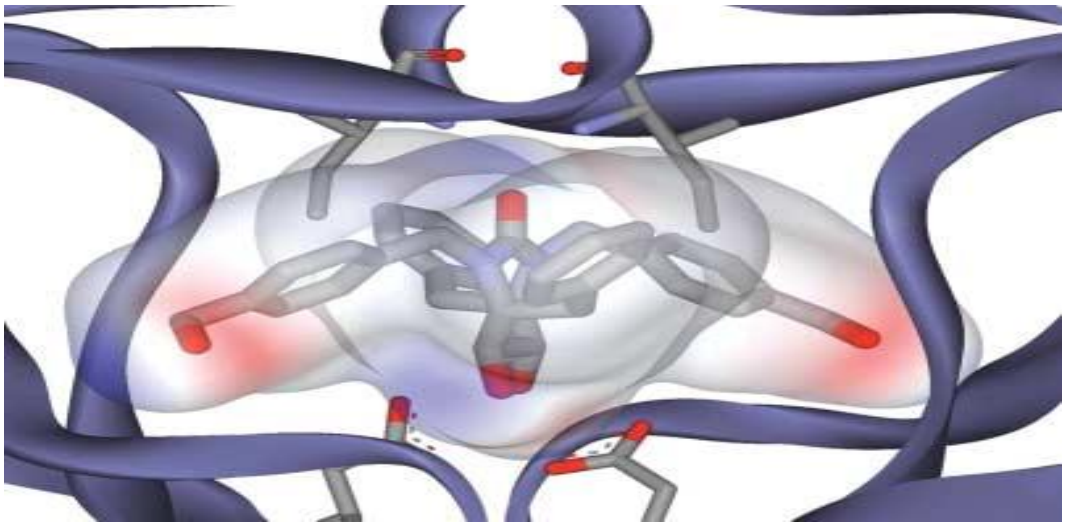

Figure 4: Electrostatic potential of a constrained inhibitor. The inhibitor Dmp323 is depicted in association with HIV protease (PDB code: 1BVE). The electrostatic potential of the symmetrical inhibitor in its binding conformation was delineated on its computed molecular surface. The residues Ile50 and Asp25 from each monomer in HIV protease facilitate the stabilisation of inhibitor binding. (Kitchen, Douglas B., et al. 2004)

1. Protein-protein docking

Protein-protein docking is a burgeoning research area, owing to its capacity for forecasting protein-protein interactions (PPIs) and pinpointing critical residues at the protein-protein interface. While protein-protein docking adheres to the concepts of protein-small molecule docking, the exploration of conformational space in protein-protein docking presents significant challenges. Even for comparatively hard proteins, exploring the rotational-conformational space of mutual orientations that may be sampled by a pair of interacting proteins is challenging. The extensive degrees of freedom render the computational expense of search algorithms significant for protein-protein docking.

Two interacting proteins exhibit considerable differences from ligand-protein binding cavities. Protein-protein binding sites typically have relatively planar surfaces lacking a prominent, well-defined pocket [45]. The prediction of protein associations is further confounded by their inherent flexibility. Proteins are dynamic entities that continually interconvert among conformers with differing energy states, and accurately representing this flexibility remains a difficulty in molecular docking. Notwithstanding the intricacy of the issue, numerous docking techniques are presently accessible for forecasting the configurations of protein-protein complexes. The selection of the approach is contingent upon the characteristics of the docking issue (46). The primary objective is to isolate the two proteins from the complex structure and employ a rigid docking procedure to attempt to recreate a near-native approximation of the complex (bound docking).

Protein-protein docking techniques have significantly advanced in recent years, as evidenced by the outcomes of the Critical Assessment of PRedicted Interactions (CAPRI) [47].

2. Protein-peptide docking

Peptideprotein interaction modeling Peptide therapeutics are increasingly recognized in drug development for their potential to target 'undruggable' intracellular protein-protein interfaces, which are characterized by extensive and relatively featureless surfaces. Peptides can interact with extensive protein surfaces with significant efficacy and selectivity, resulting in reduced off-target side effects and diminished toxicity potential

compared to small molecule medicines [48]. Computational techniques, such as docking, have demonstrated efficacy in the discovery and design of small molecule pharmaceuticals and in peptide therapies. Protein-peptide docking entails computational processes including conformational sampling, structural refinement, and scoring, akin to conventional protein-small molecule docking methodologies. Peptides exhibit greater flexibility than tiny molecules and typically assume many conformations. Consequently, modeling protein-peptide interactions is a formidable and labor-intensive challenge.

Protein-protein docking techniques, like ZDOCK [49] and Hex [50], have been employed to dock peptides onto proteins. In contrast to proteins, peptide molecules exhibit significantly greater flexibility and reduced stability. Protein partners often possess well-defined three-dimensional structures prior to the formation of protein-protein complexes, whereas peptides generally lack such structures. Moreover, peptide-mediated interactions are typically ephemeral and less robust than protein-protein interactions because of the reduced contact between peptides and their protein counterparts.

The substantial rise of peptide-protein structures in the PDB has enabled the advancement of more effective docking and refinement techniques for forecasting peptide-protein interactions. Peptide-docking methodologies can be categorized as template-based docking or template-free docking, according on the volume of requisite input data. Template-based docking approaches, also known as comparative methods, utilize existing structures (templates) as a framework to construct a model of the target complex [51]. These strategies are preferred when the template closely resembles the subject of investigation.

3. Nucleic acid docking

Nucleic acids (NAs) are integral to numerous cellular activities, encompassing cellular reproduction and protein synthesis. Consequently, DNA binders may disrupt the DNA replication process, impacting cell proliferation, or modulate the transcription process, potentially leading to the inhibition of gene expression. Likewise, RNA binders may disrupt the processes of transcription and translation. Consequently, nucleic acids are prospective pharmacological targets for various disorders, particularly in the domains of anticancer, antibacterial, and antiviral therapies [52]. Small molecules engage with nucleic acids by many mechanisms: intercalation, cross-linking, strand cleavage, and reading

molecules. As of this writing, the PDB contains 523 RNA-ligand co-crystallized structures and 730 DNA-ligand co-crystallized structures, with numbers rising annually. This structural data facilitates the examination of molecular interactions between nucleic acids and ligands, while also supporting structure-based computational approaches for the design of nucleic acid-targeting ligands for certain disorders.An efficient method for addressing NA flexibility is to employ ensemble docking, utilising a collection of predefined NA conformations derived from various X-ray crystal structures, NMR models, or normal-mode analysis of molecular dynamics simulations.

NAs-small molecule docking

1.Solvent exposed pockets, high charge density and polarity

2.NAs exhibit induced fit or conformational changes upon small molecule binding

3.Unconventional interactions between NAs and small-molecule binders

4.Strong interactions with water and metal ions.

Theory of docking

The docking of molecules techniques encompass two fundamental processes: predicting the shape of a ligand, often a tiny molecule, along with its position and orientation within the protein binding site (termed a pose), and evaluating the quality of the pose by a scoring function. The experimental binding inversion must be repeatable via sampling and should be prioritized as the highest-ranked among all conformations generated by the scoring function. Another objective of the docking process is to achieve predictably greater ratings for active molecules than for stable inactive compounds. It is challenging to attain this level of precision, as it is typically impacted by outside variables unrelated to the protein. Accurately predicting the ligand's conformation and evaluating its quality are the main objectives of a docking technique.[53]

The sampling process is non-trivial. The conformational space involves many degrees of freedom including the rotation and translation of a molecule relative to another. In practice, it is impossible to explore the search space exhaustively by enumerating all possible

conformations and all possible rotational and translational orientations of a single molecule relative to a protein within a second of elapsed time (a time scale that is realistically needed for virtual screening). Hence, efficient sampling of conformational space is still a challenge in molecular docking.

Early approaches to account for these sampling problems treated both the ligand and the protein as rigid bodies thereby reducing the number of degrees of freedom to just six. Such approaches rely on the shape similarities between the ligand and the protein binding site. A very well cited example of a program using this algorithm is DOCK.[54]

The ligand and the protein binding sites are represented as pharmacophore spheres with different radii in a rigid docking approach. The search algorithm attempts to pair the ligand and protein spheres by comparing the internal distances of all the ligand's spheres and all of the protein binding site's spheres. A least squares fit of the atoms to the sphere centroid is then used to align the ligand within the binding site. When an undesirable orientation occurs (for as when the ligand and the protein binding site conflict), the ligand is reoriented until it achieves an appropriate orientation. The degree of overlap between the ligand and protein pharmacophore spheres is then used to score the orientation.[55]

Since the conformation is linked to protein-ligand interactions, its effectiveness is limited even with its computational efficiency. The ideal binding conformation of a small molecule is a compromise between its best internal geometry and the interactions it forms with the binding site. Rigid docking does not consider this. This is particularly pertinent in predicted docking because to the further complexity arising from the conformational alteration of the ligand from its unbound (isolated) state to its bioactive (bound) state.[56]

Small databases of molecules

For virtual screening, a number of databases of small compounds that resemble leads and drugs are accessible. Even while some of the collections have a lot in common, each database has special qualities that can make it better than the others for a given virtual screening job. Every database has a sufficient number of unique compounds to warrant examining multiple libraries, if at all possible. The 'Lipinski's rule of five', which states

that drug-like compounds should have a molecular weight below 500, lipophilicity (logP) below 5, fewer than five hydrogen bond donors, and fewer than ten hydrogen bond acceptors, allows for the design of many publicly accessible chemical databases to have desirable properties like 'drug-likeness' [57].

Nonetheless, a growing number of substances that violate some of these regulations have been authorized as medications and put on the market over time (for example, 50% of marketed medications do not adhere to the rule of five, and many natural product medications do not) [58]. The range of chemotypes can be severely restricted by rigorous adherence to this guideline, suggesting that it is best to follow it with some leeway [58]. Compounds that violate the rule of five are included in other databases that are accessible because they contain chemical structures from licensed medications or natural items.

Here, we'll give a quick rundown of some of the most widely utilized chemical databases for virtual screening.

1. The ZINC database

The University of California, San Francisco's Department of Pharmaceutical Chemistry created the free database ZINC, which contains information on commercially accessible chemicals [59]. It has an increasing number of ready-to-dock 3D structures from many manufacturers' catalogs, together with features like size, calculated logP, number of rotatable bonds, and pertinent information on protonation and tautomeric states. The primary feature that sets this ZINC apart from other databases is its emphasis on docking and availability, as each molecule in the database also includes vendor and purchasability information.

Over 736 million lead-like compounds (molecular weight less than 400 g/mol, computed logP less than 4, and rotatable bonds less than 7) are included in ZINC20 [60], its most recent edition. Of these, 509 million can be downloaded in 3D and are prepared for docking, along with information.

2.ENAMiNe databases

For screening, ENAMINE offers a variety of commercial collections of compounds. There are already more than 2 million low molecular weight chemical molecules in the screening collection. While the Advanced Collection (>493,000 compounds) is meant for lead discovery, the HTS collection (>2,115,000 compounds) comprises a very broad variety of chemotypes produced from in-house research and partner academic organizations. This collection was created using lead-like characteristics and/or useful pharmacophores like amide, carboxylic, and primary amino groups. Conversely, the Premium collection (>44,500 compounds) includes compounds with the best physicochemical characteristics (low logP, MW, and high Fsp^3). Additionally, Enamine offers a 3D diversity set (50,240 compounds from conformational analysis and shape clustering of the HTS collection), a pharmacologically diverse set (10,240 drug-like compounds clustered by activities from biologically relevant chemical space), and a covalent screening library (10,480 compounds of well-validated covalent binders). Targeted libraries, such as those for the central nervous system (CNS), antibacterial, ion channel, coronavirus, kinase, and lipid G protein-coupled receptor (GPCR) libraries, as well as fragment libraries (e.g. covalent, sp3-rich, Protein-Protein Interactions (PPI), fluorinated and brominated fragments, etc. The REAL database, the largest Enamine database, is a virtual collection of more than 1.36 billion molecules that can be used to find analogues of successful compounds and to discover new hit molecules through extensive virtual screening. Every molecule in the REAL database meets the Veber criterion (TPSA < 140 and rotatable bonds ≤ 10) as well as the rule of five [61]. These different collections' structure data files are updated on a regular basis and are available for direct download from the ENAMINE webpage in MDL SD (.sdf) or MDL ISIS (.db) formats, or upon request. Each molecule's entry in the database includes SMILES and catalogue IDs, as well as key physicochemical parameters (MW, sLogP, HBA, HBD, etc.), structural alerts (PAINS, Brenk, and Eli Lilly medchem rules), chemistry type, and synthesis difficulty ('s', simple chemistry, standard effort,'m', advanced chemistry, higher effort).

3.The NCI open database

The National Cancer Institute's (NCI) Developmental Therapeutics Program created the publicly available NCI Open Database. This database includes a collection of substances that are not protected by confidentiality agreements that the NCI has gathered since 1955 for anticancer testing and since the 1980s for anti-AIDS screening. More than 260,000 compounds from organic synthesis and natural source extracts are presently available for download in.sdf format from the database [62]. A manually curated database of bioactive compounds with characteristics similar to those of drugs is called ChEMBL [63, 64]. More than 1.8 million chemicals and more than 15 million records on their effects on biological systems are available in the ChEMBL database. It includes data on absorption, distribution, metabolism, excretion, and toxicity (ADMET), as well as how tiny molecules interact with their targets and impact cells and the entire body. ChEMBL now includes more information on the clinical development of substances (ChEMBL Drugs). This carefully selected dataset contains both commercially available medications and those that are currently or have been in clinical research. It is annotated with details about the compounds' known therapeutic targets and related indications. A substantial portion of the SAR and the discovery of contemporary medications are covered by the data in ChEMBL, which are taken and curated from the major medicinal chemistry and pharmacology literature. Researchers' deposits and data taken from other public databases are also included in the ChEMBL database. ChEMBL contains 2D structures with computed molecular attributes (e.g., logP, molecular weight, Lipinski parameters) and bioactivity information (e.g., binding constants, pharmacology, and ADMET). The bioactivity information is labeled to indicate connections between published experiments and molecular targets.

The Role of Molecular Docking in Combating the Coronavirus (SARS-CoV-2)

As the COVID-19 pandemic spreads, covalent drugs that form permanent chemical bonds with the virus, enhancing their effectiveness, have gained importance.

Molecular simulations have been used to design covalent inhibitors targeting Mpro, such as ketone-based inhibitors, which have been studied to combat the virus.This has improved strategies to combat the pandemic and reduced the need for lengthy and expensive laboratory experiments.[72]

How molecular simulations are being used in the search for treatments for SARS-CoV-2 (coronavirus)?

By identifying key viral proteins such as Mpro (3CLpro), Spike Protein and RdRp as therapeutic targets, it helped test well-known drugs such as Remdesivir, Lopinavir and Hydroxychloroquine to see how effective they are against the virus.[75]

The role of molecular simulation in the pharmaceutical industry:

1. Many drugs such as Imatinib and Zanamivir were developed using this technology.

2. Reducing costly laboratory experiments:by directing research to compounds most likely to succeed.

3. Predicting the interaction between the drug and biological receptors: By calculating the binding position of molecules to proteins, the mechanism of action of the drug can be understood and modifications can be suggested to improve its compatibility with receptors.

4. Structure-Activity Relationship (SAR) studies: To improve compounds, discover new drugs, predict the position and properties of drug-protein binding and optimize this interaction to increase efficacy and reduce side effects.

5. Exploring drug resistance mechanisms: used to understand how drug resistance develops in bacteria or viruses, helping to design drugs that bypass this resistance.[73-77]

Materials (Databases & Chemical Libraries):

1. ZINC Database: ZINC is a free database of commercially-available compounds developed in the Department of Pharmaceutical Chemistry at the University of California, San Francisco.[78]

It contains a constantly growing number of 3D structures ready-to-dock from catalogues of several vendors with annotated relevant information about protonation and tautomeric states, and properties such as size, calculated logP, number of rotatable bonds, etc. Each molecule in the database also contains purchasability and vendor information, making this ZINC's focus on docking and availability the main distinctive characteristic from other databases.

In its latest version, ZINC20 comprises over 736 million lead-like compounds (molecular weight less than 400 g/mol, calculated logP less than 4 and rotatable bonds less than 7), 509 million of these compounds are avail- able for download in 3D ready for docking, together with information.[72]

2. Drug Bank: is an online database that provides detailed information about FDA-approved drugs as well as experimental drugs undergoing FDA approval. It serves as both a bioinformatics and cheminformatics resource, combining chemical, pharmacological, and pharmaceutical drug data with comprehensive details about drug targets, including sequences, structures, and biological pathways.

All the data in Drug Bank is freely available, non-proprietary, or sourced from non-proprietary origins, ensuring full accessibility and traceability to its original references.

Currently, Drug Bank contains over 13,700 drug entries, including approved small-molecule drugs, biologics (such as proteins, peptides, vaccines, and allergenics), and more than 6,000 experimental drugs still in the discovery phase. Additionally, the database links over 5,000 unique protein sequences, covering drug targets, carriers, transporters, and enzymes, along with information on drug-drug and drug-food interactions.

All chemical structures in DrugBank are available in different formats, including SMILES, sdf, .mol, .pdb, InChI, and InChIKey.[74]

3. The Chemistry European Biochemistry Laboratory (ChEMBL): is a carefully curated database focused on biologically active molecules with drug-like properties. It contains over 1.8 million compounds and over 15 million records detailing their effects on biological systems. The database provides information on how small molecules interact with their targets, their effects on cells and whole organisms, and their key ADMET properties (absorption, distribution, metabolism, excretion, and toxicity).[77]

In addition, ChEMBL provides data on the clinical progress of various compounds through the "ChEMBL Drugs" section, which presents marketed drugs as well as drugs undergoing clinical trials. These entries are accompanied by details about their therapeutic targets and indications.

ChEMBL includes two-dimensional molecular structures with calculated properties, such as logP, molecular weight, and Lipinski criteria, as well as bioactivity data (such as binding constants, pharmacology, and ADMET). All bioactivity data are annotated, linking molecular targets to their published experimental results.[78]

In short, ChEMBL is a powerful tool in drug discovery, providing reliable data to support pharmaceutical research and improve drug development.

4. PDB (Protein Data Bank): Database of optimised existing PDB entries with electron density maps, a description of model changes, and a wealth of model validation data. [77] It is a good starting point for any structural biology project.

All the entries are treated with a consistent protocol that reduces the effects

of differences in age, software, and depositors.[78-79]

PDB is a central archive that stores all experimentally determined protein structures. It is managed by an international organization called wwPDB, which consists of four main members:

RCSB PDB (responsible for maintaining a single, consistent version of the data for all users) PDBe (Protein Data Bank in Europe) PDBj (Protein Data Bank in Japan)

BMRB (Biological Magnetic Resonance Data Bank) RCSB PDB ensures that all users access the same, unaltered data. [80-81]

Software tools used in molecular simulation (Preprocessing & Preparation Tools):

1. ModRefiner: is a tool for improving the quality of protein structures by refining the atomic geometry of the model and reducing structural distortions. The text does not provide any details about this tool or its role in improving the model after it is built.

2. PROPKA: is a computational tool used to predict the pKa values of ionizable amino acids in proteins. These predictions help determine the protonation state of amino acids such as Asp, Glu, Arg, Lys, and His under specific pH conditions.

Assigning the correct protonation state to protein residues is essential for predicting ligand binding modes and affinities, especially in virtual screening, where errors can lead to false positives or missed true binders. Protonation states change dynamically, affecting protein-ligand interactions, so they must align with the bound conformation and experimental pH. Crystal structures and known ligands provide insights into these states, steric clashes, and hydrogen bonding. Aspartic and glutamic acids are usually deprotonated, while arginine and lysine are protonated. Histidine is more complex, as it can adopt different protonation forms, influenced by its environment and sometimes misrepresented due to poor crystal resolution. To ensure accuracy, each histidine in the binding site should be analyzed individually, with hydrogen bonding as a key reference.[82-86]

3. RDKit: This tool is used to create and optimize 3D structures of molecules, facilitating molecular modeling and prediction of chemical reactions.

Several programs exist to generate and optimise the 3D structure of a ligand (e.g. CSD Conformer Generator, Omega, Confab, Confect, RDKit.[87]

4. CSD Conformer Generator: The Cambridge Structural Database (CSD) is a repository for small molecule organic and metal-organic crystal structures. With over 1 million structures from X-ray and neutron diffraction analyses, the CSD includes several subsets, such as the CSD Drug Subset with entries that feature in the approved drug list provided by DrugBank and a CSD COVID-19 subset that includes structures of interest in the fight against COVID-19.

In short, is a tool linked to the CSD (Cambridge Structural Database), a database containing the crystal structures of thousands of organic and metal-organic molecules. This tool is used to generate possible conformations (three-dimensional shapes) of small molecules based on the chemical formula and structural data of the compound. [81-82]

Current challenges

Despite being in a mature stage of development, docking is still far from ideal. With claimed success rates between 70 and 80%, the majority of docking systems can typically predict known binding postures with average accuracies of 1.5 to 2 [83]. The computation of precise binding energies, however, is one of the main drawbacks of molecular docking and is closely tied to all of the approximations made during a docking run (e.g. the treatment of solvent and the flexibility of the macromolecular system). Perhaps the most detrimental flaw in docking is the absence of an appropriate scoring function and searching algorithm that can effectively combine speed and precision. Therefore, the results of a docking experiment should not be interpreted as the final result, but rather as a good starting point or as part of a workflow for a more thorough and accurate analysis, even though it makes a significant contribution to understanding target-ligand interactions in support of drug discovery projects. We will concentrate on some of the areas where the aforementioned restrictions are probably going to have an effect in this section which include:

1-Blind docking: The process of docking a ligand to a protein's whole surface without knowing the target pocket beforehand is known as "blind docking." A far wider search space results from blind docking, which views the entire protein as a site where a ligand might bind. Furthermore, the more binding sites there are, the more difficult blind docking becomes, which significantly restricts its practical application. Blind docking's main challenge is to address the vast search space.

As an alternative to a single blind docking run that covers the entire protein structure, this can be mitigated in one of two ways: either reduce the search complexity and split the docking box into multiple boxes, which would sacrifice the flexibility of some parts of the

ligand, or repeat the search multiple times using different seeds and then combine the results.

The Protein Energy Landscape Exploration (PELE), a Monte Carlo-based method coupled with a protein structure prediction algorithm, represents a more sophisticated method for sampling flexibility in blind docking. There are three primary steps taken. First, side chain sampling (using algorithms), followed by ligand and protein perturbation (using a rotamer library), and finally, minimization and acceptance using the Metropolis acceptance criteria. Despite being computationally expensive, this method is still less expensive than MD simulations [91].

2- Covalent docking: Historically, drug discovery mainly focuses on non-covalent drugs due to potential off-target effects and toxicity issues of irreversible covalent drugs. However, in recent years and with the outbreak of Covid-19, we have witnessed the resurgence of covalent drugs [92-93]. Compared to non-covalent drugs, covalent drugs might have extra advantages, including better efficacy. They also offer a lower patient burden and less drug resis-tance due to lower and less frequent dosing and improved target specificity by careful designs that target specific protein residues [94].Because covalent bond formation, bond breaking, and bond rearrangements are quantum mechanical (QM) phenomena that cannot be sufficiently handled by force fields or empirical approaches typically used for non-covalent protein-ligand interactions, the rational design of covalent ligands continues to face unique challenges [95].

In the past, numerous manual interventions and ad hoc methods have been employed to modify current docking technologies for usage with covalent ligands in order to get around these restrictions [95]. More recently, this has altered. Given the scale of the molecular systems and the quantity of configurations and compounds to take into account, a complete QM treatment is currently impractical in everyday applications, despite the fact that contributions from QM approaches are progressively being incorporated into docking applications. However, quicker and easier modeling techniques could avoid the requirement for QM calculations in covalent docking, and in many situations, the QM treatment of the docking process might not be necessary.

Does anyone know the binding site? Is the reactivity of the targeted amino acid known? Is it known what kind of electrophilic warhead the ligands have? Various scenarios and requirements apply depending on how these questions are answered. In the most straightforward scenario, where the target site is widely recognized.

3-Reverse docking: As the name implies, reverse docking, also known as inverse docking, is the process of docking a group of one or a few ligands against a variety of protein families in order to determine a possible target, their binding affinity, or their polypharmacology profile. Furthermore, RD can be a useful tool for drug rescue, repositioning, and repurposing. It can also help identify targets for medications with mechanisms that are currently unclear and aid in the logical design of less toxic or multitarget medications [96]. Clinically approved medications may therefore be used for conditions other than those for which they were initially intended [97-98]. One well-known example is sildenafil [99], a phosphodiesterase-5 (PDE5) inhibitor that was initially created to treat angina but is now used to treat erectile dysfunction.

Minoxidil, which was first created to treat hypertension but was later repositioned to treat male hair loss, is another example of a successful therapeutic repurposing [100]. Even while RD wasn't used to find these materials, it encouraged the use of computational methods for repurposing[101], which are now being used, for instance, to find cures for infectious disorders like COVID-19 [102]. In recent years, RD has gained increasing attention due to the advancement of computing resources. It has demonstrated some effectiveness in target identification, with anticipated results confirmed by crystallographic investigations and bioassays. These successful examples demonstrate the significance of reverse docking in small molecule protein target prediction.

Limitations

Covalent Docking Limitations:

• The main challenges in designing covalent ligands relate to the fact that the formation, breaking, and rearrangement of covalent bonds are quantum mechanics (QM) phenomena,

and cannot be adequately addressed using approximate forces or traditional methods used in non-covalent docking.

• QM techniques are still impractical for routine application due to the size of molecular systems and the number of configurations and compounds that must be considered, making them only suitable for re-evaluation after the docking process, not during it.

• Blind docking is not applicable to covalent ligands because the binding site and target amino acid must be known in advance.

Reverse Docking Limitations:

• One of the main limitations is the accuracy of currently used scoring functions, as they do not distinguish well between real and fake targets, leading to many false positives.

• Modern methods based on artificial intelligence/machine learning are effective but require massive computational resources and large biological data, which are not always available.

Future Recommendations (Looking Forward)

• Integrating docking with other computational tools such as MD (Molecular Dynamics Simulation) and AI/ML to enhance prediction and improve the accuracy of virtual screening results.

• Optimizing scoring functions to suit different docking scenarios, especially in RD.

• Using cloud computing and GPU clusters to accelerate docking operations, especially for intensive tasks such as blind docking.

References:

1. Hendlich, M. (1998) Databases for protein-ligand complexes. Acta Crystallogr D Biol Crystallogr, 54(Pt 6 Pt 1): 1178–1182.
2. Hu, L., Benson, M.L., Smith, R.D., Lerner, M.G., and Carlson, H.A. (2005) Binding MOAD (Mother Of All Databases). Proteins, 60(3): 333–340.
3. Irwin, J.J. and Shoichet, B.K. (2005) ZINC—a free database of commercially available compounds for virtual screening. J Chem Inf Model, 45(1): 177–182.
4. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000) The Protein Data Bank. Nucleic Acids Res, 28(1): 235–242.
5. Brooijmans, N. & Kuntz, I. D. Molecular recognition and docking algorithms. Annu. Rev. Biophys. Biolmol. Struct. 32, 335–373 (2003)
6. Gohlke, H. & Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew. Chem. Int. Ed. 41, 2644–2676 (2002)
7. Morris GM, Lim-Wilby M. Molecular docking. Molecular modeling of proteins. 2008:365-82.
8. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E.W. Computational methods in drug discovery. Pharmacol. Rev. 2014, 66, 334–395.
9. Song, C.M.; Lim, S.J.; Tong, J.C. Recent advances in computer-aided drug design. Brief. Bioinform. 2009, 10, 579–591.
10. Macalino, S.J.Y.; Gosu, V.; Hong, S.; Choi, S. Role of computer-aided drug design in modern drug discovery. Arch. Pharm. Res. 2015, 38, 1686–1701.
11. D'Agostino, D.; Clematis, A.; Quarati, A.; Cesini, D.; Chiappori, F.; Milanesi, L.; Merelli, I. Cloud Infrastructures for In Silico Drug Discovery: Economic and Practical Aspects. Biomed Res. Int. 2013, 2013, 138012.
12. Jorgensen, W.L. The Many Roles of Computation in Drug Discovery. Science 2004, 303, 1813–1818.
13. Kapetanovic, I.M. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. Chem. Biol. Interact. 2008, 171, 165–176.

14. Pinzi L, Rastelli G. Molecular docking: shifting paradigms in drug discovery. International journal of molecular sciences. 2019 Sep 4;20(18):4331.
15. Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat. Rev. Drug Discov. 2004, 3, 935–949.
16. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. Biophys Rev 2017;9(2):91–102.
17. Wang Z, Sun H, Yao X, Li D, Xu L, Li Y, et al. Comprehensive evaluation of
1. ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. Phys Chem Chem Phys 2016;18 (18):12964–75.
18. Li Y, Shen J, Sun X, Li W, Liu G, Tang Y. Accuracy assessment of protein-based docking programs against RNA targets. J Chem Inf Model 2010;50(6):1134–46.
19. Cross JB, Thompson DC, Rai BK, Baber JC, Fan KY, Hu Y, et al. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. J Chem Inf Model 2009;49(6):1455–74.
20. Pozzan A. Molecular descriptors and methods for ligand based virtual high throughput screening in drug discovery. Curr Pharm Des. 2006; 12: 2099-2110.
21. Green DV. Virtual screening of virtual libraries. Prog Med Chem. 2003; 41: 61- 97.
22. 3. Trosset JY, Scheraga HA. PRODOCK: Software Package for Protein Modeling and Docking. J Comput Chem. 1999; 20: 412-427.
23. McMartin C, Bohacek RS. QXP: Powerful, Rapid Computer Algorithms for Structure-based Drug Design. J Comput Aid Mol Des. 1997; 11: 333-344.
24. Schnecke V, Kuhn LA. Virtual Screening with Solvation and Ligandinduced Complementarity, Perspect. Drug Discov. 2000; 20: 171-190.
25. : Chaudhary KK, Mishra N (2016) A Review on Molecular Docking: Novel Tool for Drug Discovery. JSM Chem 4(3): 1029.
26. B Mukesh, K Rakesh. Molecular docking: a review. IJRAP. 2011; 2: 1746-1751.
27. Lamb ML, Jorgensen WL. Computational approaches to molecular recognition. Curr Opin Chem Biol. 1997; 1: 449-457

28. Blaney, J. A very short history of structure-based design: How did we get here and where do we need to go? J. Comput. Aided Mol. Des. 2012, 26, 13–14.
29. Mandal, S.; Moudgil, M.N.; Mandal, S.K. Rational drug design. Eur. J. Pharmacol. 2009, 625, 90–100.
30. Urwyler, S. Allosteric modulation of family C G-protein-coupled receptors: From molecular insights to therapeutic perspectives. Pharmacol. Rev. 2011, 63, 59–126.
31. Wilson, G.L.; Lill, M.A. Integrating structure-based and ligand-based approaches for computational drug design. Future Med. Chem. 2011, 3, 735–750.
32. Fang, Y. Ligand-receptor interaction platforms and their applications for drug discovery. Expert Opin. Drug Discov. 2012, 7, 969–988.
33. Kahsai, A.W.; Xiao, K.; Rajagopal, S.; Ahn, S.; Shukla, A.K.; Sun, J.; Oas, T.G.; Lefkowitz, R.J. Multiple ligand-specific conformations of the β2- adrenergic receptor. Nat. Chem. Biol. 2011, 7, 692–700.
34. Shoichet, B.K.; Kobilka, B.K. Structure-based drug screening for G-protein-coupled receptors. Trends Pharmacol. Sci. 2012, 33, 268–272.
35. Chandrika, B.R.; Subramanian, J.; Sharma, S.D. Managing protein flexibility in docking and its applications. Drug Discov. Today 2009, 14, 394–400.
36. Durrant, J.D.; McCammon, J.A. Computer-aided drug-discovery techniques that account for receptor flexibility. Curr. Opin. Pharmacol. 2010, 10, 770–774.
37. Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 47, 409–443 (2002).
38. Burnett, R. M. & Taylor, J. S. DARWIN: a program for docking flexible molecules. Proteins 41, 173–191 (2000).
39. Norel, R., Lin, S. L., Wolfson, H. & Nussinov, R. Shape complementarity at protein–protein interfaces. Biopolymers 34, 933–940 (1994).
40. Norel, R., Petrey, D., Wolfson, H. & Nussinov, R. Examination of shape complementarity in docking of unbound proteins. Proteins 35, 403–419 (1999).
41. Connolly, M. L. Analytical molecular surface calculation. J. Appl. Cryst. 16, 548–558 (1983).

42. Connolly, M. Solvent-accessible surface of proteins and nucleic acids. Science 221, 709–713 (1983)..
43. Norel, R., Wolfson, H. & Nussinov, R. Small molecular recognition: solid angles surface representation and shape complementarity. Comb. Chem. High Throughput Screen 2, 177–191 (1999).
44. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J. Med. Chem. 28, 849–857 (1985)
45. Huang S-Y. Search strategies and evaluation in protein-protein docking: principles, advances and challenges. Drug Discov Today 2014;19(8):1081-96.
46. Porter KA, Desta I, Kozakov D, Vajda S. What method to use for protein-protein docking? Curr Opin Struct Biol 2019;55:1-7.
47. Janin J. Henrick K. Moult J. Ten Eyck L, Sternberg MJE, Vajda S, et al. CAPRI: a critical assessment of PRedicted interactions. Proteins 2003:52(1):2-9, [184] Tom N. Peptide therapeutiec tanging the undriigible Fur] Med Chim
48. Tsomaia N. Peptide therapeutics: targeting the undruggable space. Eur J Med Chem 2015:94:459-70.
49. Pierce BG, Wiehe K, Hwang H, Kim B-H, Vreven T, Weng Z. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multi-mers. Bioinformatics 2014;30(12):1771-3.
50. Macindoe G, Mavridis L, Venkatraman V, Devignes M-D, Ritchie DW. HexServer: an FFT-based protein docking server powered by graphics processors. Nucleic Acids Res 2010;38(suppl_2):W445-9.
51. Lee AC-L, Harris JL, Khanna KK, Hong J-H. A comprehensive review on current advances in peptide drug development and design. Int J Mol Sci 2019;20(10):2383.
52. Wang M, Yu Y, Liang C, Lu A, Zhang G. Recent advances in developing small molecules targeting nucleic acid. Int J Mol Sci 2016;17(6):779.
53. Ferreira LG, dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Molecules 2015;20(7):13384–421.
54. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982;161(2):269–88.

55. Kabsch W. A solution for the best rotation to relate two sets of vectors. Acta Crystallogr Sect A 1976;32(5):922–3.
56. Kabsch W. A discussion of the solution for the best rotation to relate two sets of vectors. Acta Crystallogr Sect A 1978;34(5):827–8.
57. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 2001;46(1-3):3-26.
58. Shultz MD. Two decades under the influence of the rule of five and the changing properties of approved oral drugs. J Med Chem 2019;62(4):1701-14.
59. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: a free tool to discover chemistry for biology. J Chem Inf Model 2012;52(7):1757-68.
60. Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, et al. ZINC20-a free Ultralarge-scale chemical database for ligand discovery. J Chem Inf Model 2020;60(12):6065-73.
61. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the Oral bioavailability of drug candidates. J Med Chem 2002;45(12):2615-23.
62. Anon. Downloadable Structure Files of NCI Open Database Compounds. Available from https://cactus.nci.nih.gov/download/nci/.
63. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: an update. Nucleic Acids Res 2014;42(Database issue):D1083-90.
64. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 2012;40(D1): D1100-7.
65. Mohamed K, Yazdanpanah N, Saghazadeh A, Rezaei N. Computational drug discovery and repurposing for the treatment of COVID-19: a systematic review. BioorgChem 2020;106:104490.
66. Jarada TN, Rokne JG, Alhajj R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. J Cheminform 2020;12(1):46.

67. Mohamed K, Yazdanpanah N, Saghazadeh A, Rezaei N. Computational drug discov-ery and repurposing for the treatment of COVID-19: a systematic review. BioorgChem 2020;106:104490.
68. Chen X, Ung CY, Chen Y. Can an in silico drug-target search method be used toprobe potential mechanisms of medicinal plant ingredients? Nat Prod Rep 2003;20 (4):432–44.
69. Eric S, Ke S, Barata T, Solmajer T, Antic Stankovic J, Juranic Z, et al. Target fishing and docking studies of the novel derivatives of aryl-aminopyridines with potential anticancer activity. Bioorg Med Chem 2012;20(17):5220–8.
70. Ji ZL, Wang Y, Yu L, Han LY, Zheng CJ, Chen YZ. In silico search of putative adverse drug reaction related proteins as a potential tool for facilitating drug adverse effect prediction. Toxicol Lett 2006;164(2):104–12.
71. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: a free tool to discover chemistry for biology. J Chem Inf Model 2012;52(7):1757–68.
72. Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, et al. ZINC20—a free Ultralarge-scale chemical database for ligand discovery. J ChemInf Model 2020;60(12):6065–73.
73. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006;34(Database issue):D668–72.
74. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2018;46(D1):D1074–82.
75. Chen X, Lin Y, Liu M, Gilson MK. The binding database: data management and interface design. Bioinformatics 2002;18(1):130–9.
76. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. Binding DB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res 2015;44(D1):D1045–53.
77. Chen X, Liu M, Gilson MK. Binding DB: a web-accessible molecular recognition database. Comb Chem High Throughput Screen 2001;4(8):719–25.

78. Wang R, Fang X, Lu Y, Wang S. The PDBbind database: collection of binding affinities for protein–ligand complexes with known three-dimensional structures. J MedChem 2004;47(12):2977–80.
79. Grudinin S, Popov P, Neveu E, Cheremovskiy G. Predicting binding poses and affinities in the CSAR 2013–2014 docking exercises using the knowledge-basedconvex-PL potential. J Chem Inf Model 2015;56(6):1053–62.
80. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 2013;27(3):221–34.
81. Kalliokoski T, Salo HS, Lahtela-Kakkonen M, Poso A. The effect of ligand-based tautomer and protomer prediction on structure-based virtual screening. J Chem Inf Model 2009;49(12):2742–8.
82. Onufriev AV, Alexov E. Protonation and pK changes in protein-ligand binding. QRev Biophys 2013;46(2):181–209.
83. Li S, Hong M. Protonation, tautomerization, and rotameric structure of histidine: a comprehensive study by magic-angle-spinning solid-state NMR. J Am Chem Soc 2011;133(5):1534–44.
84. Kim MO, Nichols SE, Wang Y, McCammon JA. Effects of histidine protonation and rotameric states on virtual screening of M. tuberculosis RmlC. J Comput Aided MolDes 2013;27(3):235–46.
85. Landrum G, et al. RDKit: Open-source cheminformatics, http://www.rdkit.org/.2018; 2018.
86. Cole JC, Korb O, McCabe P, Read MG, Taylor R. Knowledge-based conformer generation using the Cambridge structural database. J Chem Inf Model 2018;58 (3):615–29.
87. Taylor R, Cole J, Korb O, McCabe P. Knowledge-based libraries for predicting the geometric preferences of drug like molecules. J Chem Inf Model 2014;54(9):2500–14.
88. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. Biophys Rev 2017;9(2):91-102.

89. Grebner C, Iegre J, Ulander J, Edman K, Hogner A, Tyrchan C. Binding mode and induced fit predictions for prospective computational drug design. J Chem Inf Model 2016;56(4):774-87.
90. Singh J, Petter RC, Baillie TA, Whitty A. The resurgence of covalent drugs. Nat Rev Drug Discov 2011;10(4):307-17.
91. Hoffman RL, Kania RS, Brothers MA, Davies JF, Ferre RA, Gajiwala KS, et al. Discovery of ketone-based covalent inhibitors of coronavirus 3CL proteases for the potential therapeutic treatment of COVID-19. J Med Chem 2020:63:12725-47.
92. Li Q, Wang Z, Zheng Q, Liu S. Potential clinical drugs as covalent inhibitors of the priming proteases of the spike protein of SARS-CoV-2. Comput Struct Biotechnol J 2020;18:2200-8
93. Sotriffer C. Docking of covalent ligands: challenges and approaches. Mol Inf 2018;37 (9-10):1800062.
94. Kharkar PS, Warrier S, Gaud RS. Reverse docking: a powerful tool for drug repositioning and drug rescue. Future Med Chem 2014;6(3):333-42.
95. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov 2004;3(8):673-83.
96. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, et al. Predicting new molecular targets for known drugs. Nature 2009;462(7270):175-81.
97. Terrett NK, Bell AS, Brown D, Ellis P. Sildenafil (VIAGRATM), a potent and selective inhibitor of type 5 cGMP phosphodiesterase with utility for the treatment of male erectile dysfunction. Bioorg Med Chem Lett 1996;6(15):1819-24.
98. Zappacosta AR. Reversal of baldness in patient receiving minoxidil for hypertension. N Engl J Med 1980;303:1480-1.
99. Jarada TN, Rokne JG, Alhajj R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. J Cheminform 2020;12(1):46.
100 Mohamed K, Yazdanpanah N, Saghazadeh A, Rezaei N. Computational drug discovery and repurposing for the treatment of COVID-19: a systematic review. Bioorg Chem 2020;106:104490.

101 Chen X, Ung CY, Chen Y. Can an in silico drug-target search method be used to probe potential mechanisms of medicinal plant ingredients? Nat Prod Rep 2003;20 (4):432-44.

102 Erić S, Ke S, Barata T, Solmajer T, Antić Stanković J, Juranić Z, et al. Target fishing and docking studies of the novel derivatives of aryl-aminopyridines with potential anticancer activity. Bioorg Med Chem 2012;20(17):5220-8.

103. Sotriffer C. Docking of covalent ligands: challenges and approaches. Mol Inf 2018;37(9–10):1800062.

104. Fanfrlı ´k J, Brahmkshatriya PS, Řeza´c J, Jı ´lkova ´ A, Horn M, Mares ˇ M, et al. Quantum mechanics-based scoring rationalizes the irreversible inactivation of parasitic Schistosoma mansoni cysteine peptidase by vinyl sulfone inhibitors. J Phys Chem B 2013;117(48):14973–82.