Computational Methods in Drug Discovery and Development

Sadettin Y. Ugurlu

Department of Computer Science, University of Birmingham, the UK

s.yavuz.ugurlu@gmail.com

Abstract

The rapid advancements in computational methods have revolutionized drug discovery and development. These methods, ranging from molecular modelling to machine learning algorithms, have drastically increased in number and sophistication. However, a comprehensive understanding of these diverse approaches is essential for researchers aiming to make significant contributions to this evolving field. This review aims to provide a detailed overview of the most prominent computational methods currently used in drug discovery. It will analyze their underlying principles, discuss their applications, and highlight their potential for future advancements in the field. Through this examination, we aim to equip researchers with the necessary insights to navigate and contribute to the rapidly expanding landscape of computational drug discovery.

Keywords: Drug discovery; drug development; computational methods; molecular docking; molecular simulation

1 Introduction

Traditional methodologies for drug discovery can be classified according to the availability of target and ligand structures (Figure 2). The conventional drug categorisation discovery methodologies encompass four primary groups ([3]): (i) Library design, (ii) Structure-based design, (iii) Ligand-based design, (iv) De Novo Design (Figure 1). In addition to traditional classification, it is possible to introduce a novel category known as the quantum mechanical simulations and chemoinformatics approach, which can be considered novel classes.

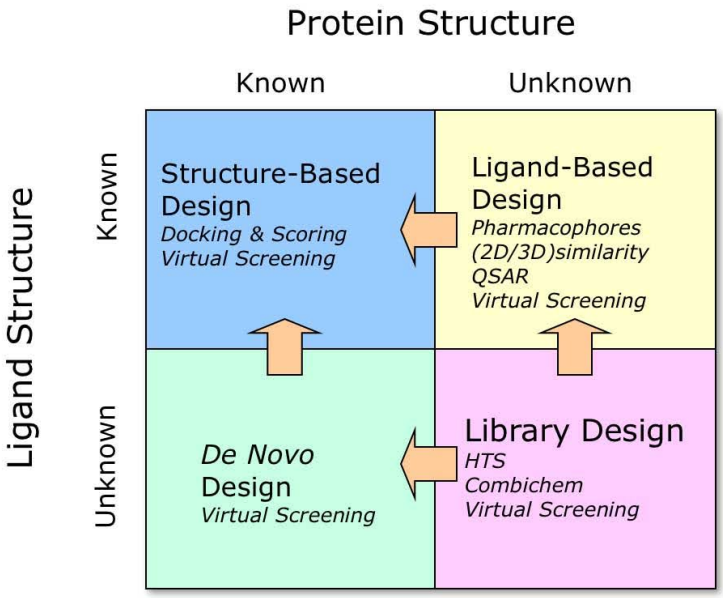

Figure 1: Overview of Conventional Approaches for Drug Discovery

The overview shows that target and ligand structures define the categorisation of conventional approaches for drug discovery. Structural-based drug discovery (SBDD) is a technique for developing small molecules using the three-dimensional conformation of the target protein. By contrast, ligand-based drug discovery (LBDD) mostly focuses on structural and chemical analysis of known ligands. While de novo design is carried out to produce new chemical entities, library design involves screening chemical libraries. Every tactic has special advantages and is quite important for the medication development process.

Commonly categorised based on the presence of target and ligand structures, traditional approaches to drug discovery each provide unique advantages and challenges. First, structure-based drug discovery, or SBDD, is the method of developing novel medications using knowledge of the three-dimensional form of the target protein to produce small molecules that can precisely bind to particular areas on the surface of the protein and alter its action. Often utilised to identify potential therapeutic candidates in the structure-based medicinal design process are techniques such as virtual screening, molecular dynamics simulations ([151]) and molecular docking ([108, 109, 166, 144, 219, 190]). Second, ligand-based drug discovery (LBDD) is more concerned with investigating well-known ligands' chemical and structural characteristics that firmly bind to the target protein. By analysing ligand similarities and differences, LBDD techniques—such as pharmacophore-based virtual screening and quantitative structure-activity relationship (QSAR) modelling ([187]), can forecast novel compounds with similar biological properties. Third, using either experimental or computational approaches, library design is essentially about identifying molecules with specific pharmacological characteristics among vast collections of chemicals. The last, de novo design ([103, 226]) aims, in essence, to create new chemical entities not seen in the natural world before (Figure 1). Besides these four conventional approaches, cheminformatics, as the last group, uses computer methods to organise, analyse, and predict chemical data and attributes to identify drug candidates and optimise their efficacy and safety. It streamlines medication design by integrating chemistry and biology to uncover new medicinal molecules faster and more accurately.

To provide a comprehensive overview of conventional approaches in drug discovery, the literature has been investigated under six sections: (i) Ligand-Based Drug Discovery,

focusing on techniques that rely on known ligands to find new drugs; (ii) Structure-Based Drug Discovery, which delves into methods utilizing the 3D structure of target proteins; (iii) Ligand-Based Drug Discovery, focusing on approaches that rely on known ligands to find new drugs; (iv) De Novo Drug Discovery, exploring strategies to design new drugs from scratch; (v) Quantum Mechanical Simulations, which forecast atomic-level molecular behaviour, revealing electronic structures, reaction mechanisms, and binding interactions, and (vi) Cheminformatics Approaches for Drug Discovery, highlighting computational techniques to analyze chemical data.

1.1 Library design for drug screening

Library design for drug screening is one of the key steps in drug discovery (Figure 1 and 2). Library design is the most time-consuming process in drug discovery since there is no target or ligand at the beginning of the drug discovery. The most logical way to define a target library is since the possible target number is significantly lower than possible ligands and drug candidates. A library with a target-focused approach refers to a compilation of chemicals that have been intentionally created or constructed to target a protein or protein family specifically. The rationale behind screening such a library is based on the notion that a reduced number of compounds is required to identify hit compounds. Moreover, it is commonly observed that there is a higher rate of successful hits when comparing the screening of diverse sets. Additionally, the hit clusters resulting from a successful focused library screening campaign typically display transparent structure-activity relationships, which aid in the subsequent analysis and investigation of these hits ([80]).

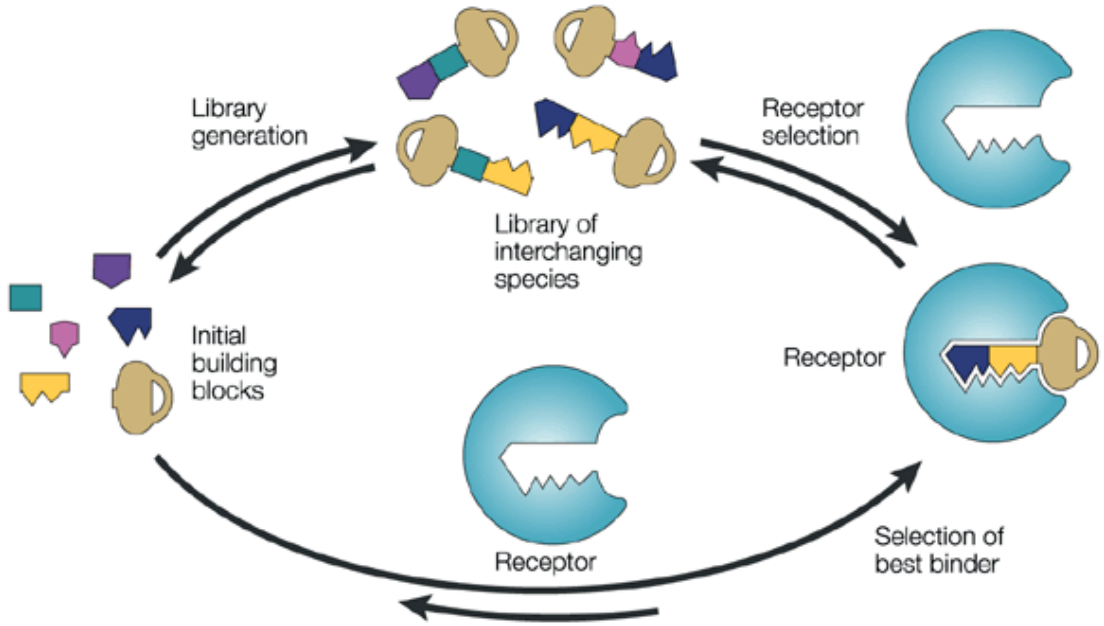

Figure 2: The schematic representation of library design

The library design procedure for drug and target screening encompasses many crucial elements to guarantee the inclusion of a wide-ranging and all-encompassing assortment of probable therapeutic candidates. The process commences with the choice of primary constituents, which are diminutive, structurally diverse molecules employed as the fundamental basis for the collection. Subsequently, these fundamental components are merged in different manners during the phase of library formation, resulting in a wide range of chemical compounds. Subsequently, the library undergoes global propagation and the establishment of a collection of diverse species, wherein various molecular variations are methodically generated and organised. Subsequently, these species undergo testing to determine their capacity to attach to the designated receptor, thereby identifying potential candidates that show promise for subsequent advancement. The technique is recursive, where the successful binding species guide the selection of new building blocks, thus continuously improving and enlarging the library for succeeding screening cycles ([149]).

Target-oriented libraries usually have a single core or scaffold with one or more attachment points, usually two or three. Different substituents or side chains are added to get the desired molecules. If all conceivable combinations were considered, a scaffold that is diversified at two or three attachment locations of diversity would provide a library consisting of numerous chemicals. Generally, a subset of these compounds is often selected for synthesis, ranging from 100 to 500. The selection is made in order to effectively investigate the design hypothesis and ensure adherence to drug-like features with the help of systematic exploration ([80]).

The systematic exploration of the chemical space and the identification of prospective therapeutic candidates are facilitated by constructing a library for drug screening, a critical element of drug development ([80]). To enhance the probability of identifying active matches, choosing molecules that demonstrate a diverse array of structural characteristics is imperative. Furthermore, the design approach frequently employs computational techniques to predict the pharmacokinetic and pharmacodynamic properties of the medications, thereby enhancing the efficacy of the screening process. Consequently, these libraries can be implemented in various drug discovery methodologies, including structure-based, ligand-based, de novo drug development, and cheminformatics.

1.2 Structure-based drug discovery

The utilisation of three-dimensional structures of biological targets, such as proteins or nucleic acids, in the computational drug discovery approach known as structure-based design (SBDD), enables the formulation of novel therapies with a high degree of specificity and affinity. To put it differently, SBDD is crucial in contemporary drug development since it utilises molecular knowledge about target-ligand interactions to inform the logical creation of small molecules or biologics. SBDD allows for the identification of crucial chemical interactions and the optimisation of compound structures to improve binding affinity and selectivity by comprehending the spatial arrangement of atoms within the target binding site. This methodology encompasses a diverse array of methods, such as molecular docking, virtual screening, fragment-based design, and molecular dynamics simulations. The primary objective is to leverage structural data in order to accelerate the process of drug exploration and advance the development of safer and more efficacious therapies for a multitude of diseases ([16]), such as using Molecular docking.

1.2.1 Molecular Docking in Drug Discovery

Molecular Docking (MD) is one of the most common methods to investigate drug-target correlation (Figure 3). Regrettably, conventional and ML-based docking methods have been plagued by a significant false-positive rate, leading to limited effectiveness ([2, 222]). ML models trained using the outcomes of molecular docking programs can effectively decrease the occurrence of false positives in MD and ML-based docking ([190]). Therefore, high false positives reduce the performance of MD. Initially, comprehending the concept of molecular docking is the primary prerequisite for constructing a proficient machine-learning model on molecular docking software.

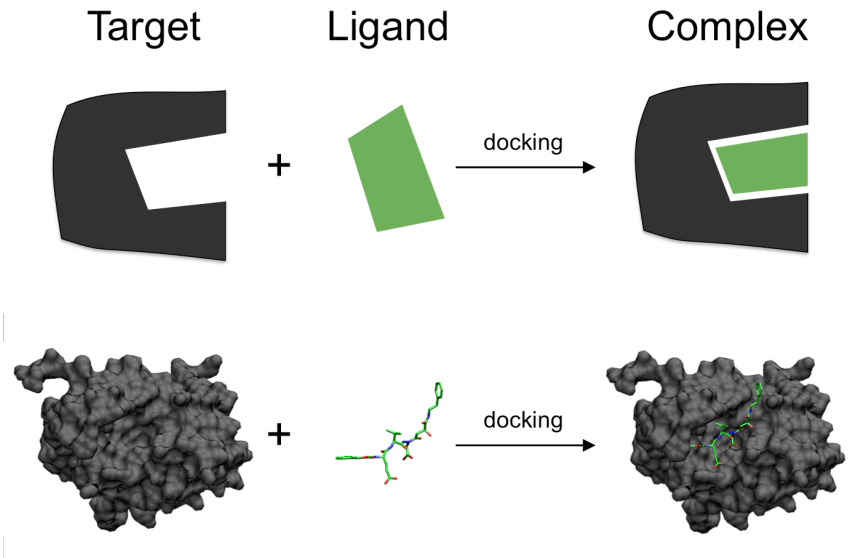

Figure 3: A basic component of molecular docking

The graphic shows the molecule bonding process. Fundamental to molecular docking is the computer prediction of the binding mechanism and affinity of small molecules (ligands) within the active site of a target protein. This method helps to find possible therapeutic options by assessing the degree of interaction and complementarity between the ligand and the protein target ([148]).

The first use of molecular docking in drug discovery was in the early 1980s ([29]), with a simplified function based on "hard sphere repulsions" and "hydrogen bonding" ([2]). The research on docking has enhanced its streamlined functionality by considering different variables in the scoring function besides "hard sphere repulsions" and "hydrogen bonding". The enhanced functionalities augmented the precision of docking and gradually introduced innovative phases. For example, the enhanced functionalities include data on the binding strength and the molecules' shape. Consequently, the efficiency of MD has progressively increased due to the implementation of new features, including enhanced functionalities.

Two steps define docking primarily: (i) prediction of the binding site and (ii) prediction of a ligand conformation and binding affinity ([125]). Unfortunately, even with developments in molecular docking methods, accurate docking cannot be guaranteed. Consequently, the success percentage of docking ranges from 0% to 92.66% ([27]). Therefore, establishing successful docking—which directly affects our machine—learning model's efficacy—depends on understanding docking classifications and selecting among the current approaches. Thus, building strong and very effective models depends on understanding the mechanism of molecule docking. Therefore, the fundamentals of molecular docking are discussed in the following seven sections, from the Molecular Mechanism of Docking to the Classification of Docking by Search Space.

Molecular Mechanism of Docking In molecular docking simulations, evaluating the quality of contacts between ligands and receptors depends on scoring functions, so the molecular mechanism of docking mainly consists of their usage. By evaluating several elements, including intermolecular forces, steric conflicts, hydrogen bonding, and electrostatic interactions, scoring systems in docking algorithms evaluate and rank possible binding positions. Forecasting the binding affinity between a ligand and a receptor is one of these purposes; this is crucial for discovering potential drug candidates. Standard scoring systems are empirical, which uses pre-defined criteria, and physics-based, which uses computational models derived from basic physical principles. The dependability of docking predictions depends much on the precision of scoring systems, affecting structural biology and drug discovery research's decision-making. Maximising docking protocols and improving the accuracy and efficiency of molecular docking simulations depend on a knowledge of the complexity of scoring systems.

The scoring functions in molecular docking programs are essential in computational drug discovery and the research of protein-ligand interactions ([100]). It is a mathematical model employed to evaluate and prioritise the strength of the interaction between a tiny chemical (a ligand) and a target protein receptor. The scoring function assesses the potency of the ligand-receptor interaction, forecasting the probability of a favourable binding position. This forecast is crucial for the identification of possible therapeutic candidates or the comprehension of protein-ligand interaction mechanisms ([100, 58]). A comprehensive scoring function considers multiple aspects, including van der Waals contacts, electrostatic interactions, hydrogen bonding, solvation effects, and entropy variations. The significance of this rests in its capacity to effectively sift through extensive collections of chemical compounds, prioritising those with the strongest binding affinity for subsequent experimental confirmation. An accurately calibrated scoring function can significantly expedite drug development by directing medicinal chemists towards molecules

with the highest therapeutic potential. This minimises the time and resources required for synthesising and testing candidate compounds ([203]).

A scoring function is used to estimate the binding affinity of a tiny molecule, which is a crucial component of docking software. A scoring function typically consists of three main subclusters: (i) physical force field-based, (ii) empirical, and (iii) knowledge-based scoring functions ([105]) (Figure 4).

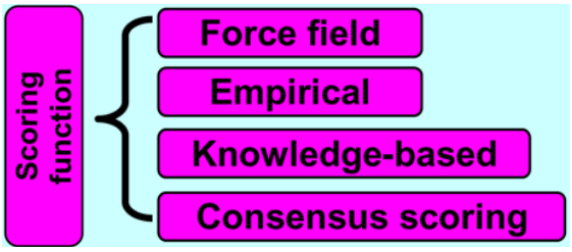

Figure 4: The scoring function classification of molecular docking programs

The diagram depicts three distinct categories of scoring functions that are frequently employed in molecular docking investigations: (i) scoring functions based on the force field, (ii) scoring functions based on empirical, (iii) scoring functions based on knowledge and (iv) consensus scoring function. The scoring functions utilised in structure-based drug discovery employ unique approaches to assess the binding affinity between a ligand and its target protein. This contributes to the systematic development of possible therapeutic candidates.

Scoring functions that utilise physical force fields (or force fields) are employed to analyse molecular interactions (Figure 4). The approach integrates molecular dynamics (MD), binding affinity, and free energy perturbation (FEP) methods. Medusa Score, for example, is one of the physical force field-based approaches. The research demonstrated that the Medusa Score success rate is around 82% ([223]). The success rate is better than various standard scoring functions, including DrugScore, F-Score, LigScore, ChemScore, PLP, LUDI, PMF, X-Score, G-Score, D-Score, and AutoDock. When the scoring method was hybridised with DrugScore, it became 85% ([223]). However, the drawbacks of techniques are speed and sampling limitations ([223]).

Empirical scoring functions aim to calculate binding free energy by leveraging chemical interactions, such as hydrogen bonds ([67]). In essence, binding energy determination depends on the molecular interactions. Molecular interaction variables include Van der Waals, dipole-dipole interactions, London dispersion forces, and hydrogen bonds. Some examples of docking programs that utilise empirical scoring functions include DOCK 4.0 ([54]) and AutoDock ([125, 173]). Molecular docking programs using empirical scoring function examples have already demonstrated their efficacy in the field ([125, 173, 189]). Hence, empirical scoring functions are the most auspicious methodologies.

The other scoring method is knowledge-based scoring functions, which use statistical analysis of protein complex structures. These functions model uncommon atoms, such as sulphur-aromatic. They also work on the statistical analysis of the ligand-target 3D complex structure. For example, Bleep, DrugScore, PMF, and SMoG are the most common knowledge-based scoring functions ([75, 66, 197]). Knowledge-based scoring functions have demonstrated satisfactory performance in molecular docking programs.

Molecular docking programs that employ consensus scoring functions integrate the outcomes of various scoring methods to enhance the precision and dependability of fore-

casting ligand-receptor interactions ([185]). For example, CompScore utilises a consensus scoring function in docking ([142]). The other example is CoBDock, which benefits not only molecular docking scoring function but also cavity detection tools to build a consensus approach ([190]). The programs utilise consensus scoring to effectively balance the merits and drawbacks of individual scoring systems, thereby improving the overall accuracy of predictions ([21]). Consensus scoring offers a significant benefit by effectively decreasing the occurrence of incorrect positive results and enhancing the reliability of forecasts regarding binding affinity. Nevertheless, the drawback is that it frequently necessitates additional computer resources and time, as it involves many scoring calculations that must be done and combined. In addition, the intricacy of including many scoring algorithms might occasionally result in incongruous outcomes if the consensus approach is not optimised ([185, 99]). Consequently, our machine learning model underwent training using several scoring function outcomes in order to enhance its performance. Our ML model's method enhances molecular docking accuracy by mitigating false positives.

Bound vs unbound molecular docking A protein's conformation is categorised into bound (complex) and unbound (one outside of a complex) structures (Figure 5). The bound docking separates a complex and then redocks parts of the complex to build the original complex. While bound docking is essential for developing new docking programs, it does not hold much value in biology. When an unbound docking program predicts a new interaction between a ligand and target (where the ligand and target are not already bonded), it enhances our understanding and becomes highly beneficial.

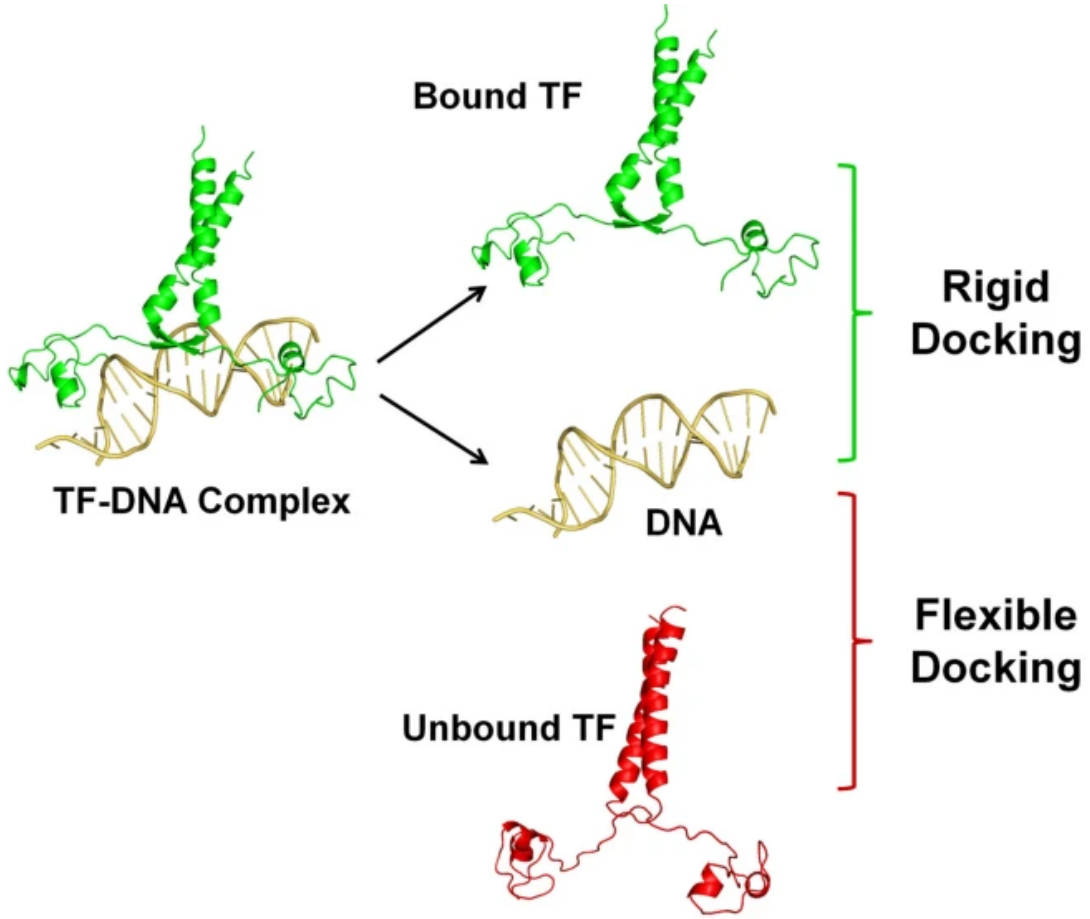

Figure 5: The representation of bound and unbound input for rigid and flexible docking

Bound (in green) and unbound (in red) TF-DNA docking test case construction proceeds to assess docking performance methodically. The bound complex is broken into two binding components: TF and DNA. These elements then are employed for docking, sometimes known as "bound docking," which usually leads to better outcomes on validation sets because of their pre-existing interface compatibility. This method might not, however, fairly depict real-world conditions in which the bound conformation is not always accessible. "Unbound docking" is exploited to overcome this restriction, whereby molecular docking uses the unbound TF, as shown in red. Under this situation, the unbound TF lacks a pre-formed interface fit for complexing with DNA. Hence, flexibility is needed to enhance docking performance. This adaptability enables conformational changes, raising the possibility of effective docking in useful contexts. Comparative bound and unbound docking approaches help one understand the need for structural adaptation to reach precise TF-DNA interactions ([90]).

Bound docking software cannot be significantly successful for an unknown compound because of limited performance in real-life cases. On the other hand, unbound docking provides vital information about unknown ligand-target complexes ([192]). Therefore, unbound docking is called "real-life docking" ([48]) because of the impact on research. Regrettably, the progress of unbound docking approaches has been hindered due to a lack of understanding of binding parameters. Hence, our machine learning techniques and pipelines have been optimised for unbound docking, making them the superior choice for new drug discovery and development.

Template-Based (Homology) Docking Using known protein structures (templates), template-based (homology) Docking is a computer method indispensable in structural biology and drug development that predicts the 3D structure of a target protein and performs molecular binding. This method depends on the idea that proteins with similar sequences usually show identical shapes and activity. Two main phases comprise the process: first, using a 3D model of the target protein derived from sequence comparison for template-based molecular docking, whereby possible ligands are assessed against the projected structure to identify potential drug candidates([38]).

Predicting 3D model for target Approximately 6% of the protein correlations in the human interactome, predicted to be researched experimentally, have been examined ([182]). The scarcity of three-dimensional target models poses a significant obstacle in structural-based drug discovery and development. As a result, various techniques have been created to anticipate three-dimensional target models, one of which is Template-based modelling (TBM). TBM predicts a protein model structure by using the known structures (Figure 6). Several TBMs exist in the literature, including MODELLER ([206]), SWISS-MODEL ([205]), and FoldX ([14]).

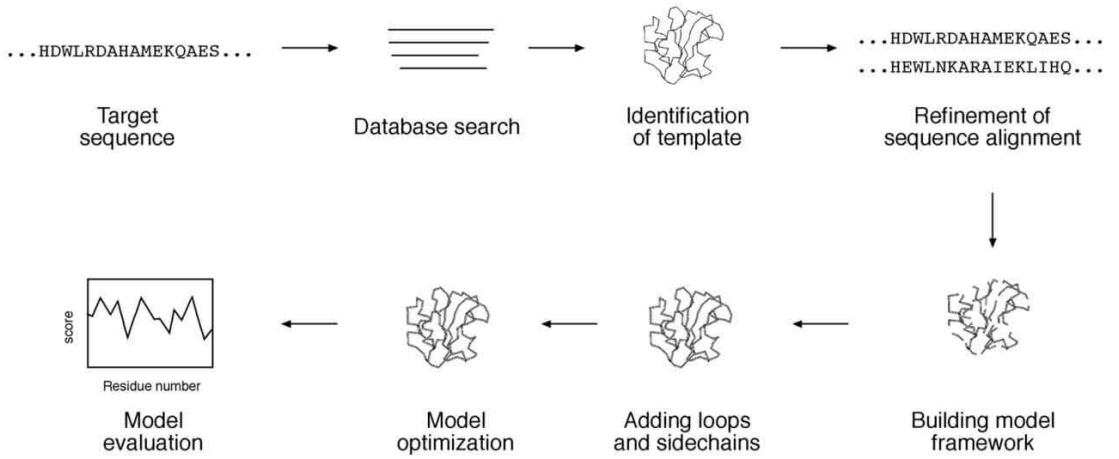

Figure 6: The figure illustrates the procedure for constructing a homology model based on a protein sequence.

The process entails aligning the sequence with homologous proteins with a known structure, selecting templates depending on the alignment quality, constructing a model using comparative modelling approaches, and refining the model to enhance its structural accuracy. Homology modelling allows for anticipating protein structures in three dimensions, helping develop structure-based drug design and other molecular investigations.

Homology models, depicted in Figure 6, are crucial for investigating targets, as wet-lab procedures have only been employed to a limited extent for examining less significant targets. As a result, other techniques, such as ab initio 3D structure prediction approaches besides TBM, have been enhanced over the past ten years. Another way to predict 3D structures is through ab initio investigations. Ab initio predicts the 3D structure of protein "from scratch" using physical principles. Examples of ab initio structure prediction programs are I-Tasser, Raptor-x, Robetta, and PSIPRED ([172]).

The main limitation of TBM is the lowered sequence similarity with known proteins, which significantly influences the prediction accuracy. Low sequence similarity compromises the structural model's dependability, thereby leading to mistakes ([160]). Conversely, ab initio methods—which rely not on current protein structures—can provide answers in these contexts. Still, these methods are computationally demanding and labour-intensive; exact results often require significant resources. This double issue emphasises the need for discoveries in both TBM and ab initio methods to improve the dependability and efficiency of protein structure prediction ([135]). However, TBMs are still practical to perform template-based molecular docking in drug discovery.

Template-based molecular docking Sometimes known as template-based or homology modelling, template-based docking is a computer technique used in molecular docking to predict the three-dimensional arrangement of a protein-ligand complex by using the established structure of a comparable protein-ligand complex ([65]). This method is predicated on the idea that proteins with similar sequences or structures often bind similarly to ligands. Matching the sequence or structure of the target protein with that of the template protein models the structure of the target protein in template-based docking. To create the missing or variable elements, one then uses computational methods, including side-chain prediction or loop modelling. Molecular docking techniques are applied to anticipate the binding shape and affinity of ligands within the binding site of the target protein once a model of the protein is generated ([216, 146, 65]).

When the experimental structure of the target protein is not easily obtainable or accessible, template-based docking is quite beneficial not only to understand the structure but also to use it in structure-based drug discovery and development. It substantially helps to identify new medications and provides essential new perspectives on the interactions between proteins and ligands. Still, it is imperative to confirm the accuracy and reliability of the expected models by rigorous computational analyses and experimental validation ([216, 146]). Understanding the classification of molecular docking can be beneficial for minimising the need for experimental validation.

Classification of Docking by Molecule Type Different types of molecules used in the docking process help to classify molecular docking, a fundamental computational tool used in structural biology and drug development. This classification distinguishes among several docking situations. Each is meant to address specific research hypotheses and objectives. Three varieties of molecular docking models are known to exist: small molecule-protein ([184]), peptide-protein ([234]), and protein-protein ([139]). Mostly in terms of the scoring systems, they have many parallels. The scoring system determines the strength of the contact between a target and a molecule. The three molecular docking techniques differ mainly in the dimensions of the molecules and the size of the search area.

Small molecule-protein docking Small molecule-protein docking is an essential computational technique in structural biology and drug development. It aims to ascertain the binding modes and affinities inside the binding site of small compounds or ligands, thereby guiding their binding to a target protein([184, 56]) (Figure 7). Examples of small molecule docking programs are AutoDock ([49]), BetaDock ([89]), PLANTS ([55, 91]), and GalaxyDock3 ([219]). Also, rational drug design depends on this method since it

provides essential knowledge on the molecular interactions between ligands and proteins. The data about interactions enhance the binding properties of potential drugs and helps to identify them.

Small molecule-protein docking systems also scan the conformational space of ligands and proteins using different scoring systems and search strategies in order to forecast energetically favourable binding locations ([211, 56, 229, 84]). By exposing the fundamental architecture of protein-ligand interactions, small molecule-protein docking helps to generate more selective and successful treatments. This helps advance new therapies for many diseases and accelerates drug discovery ([84]).

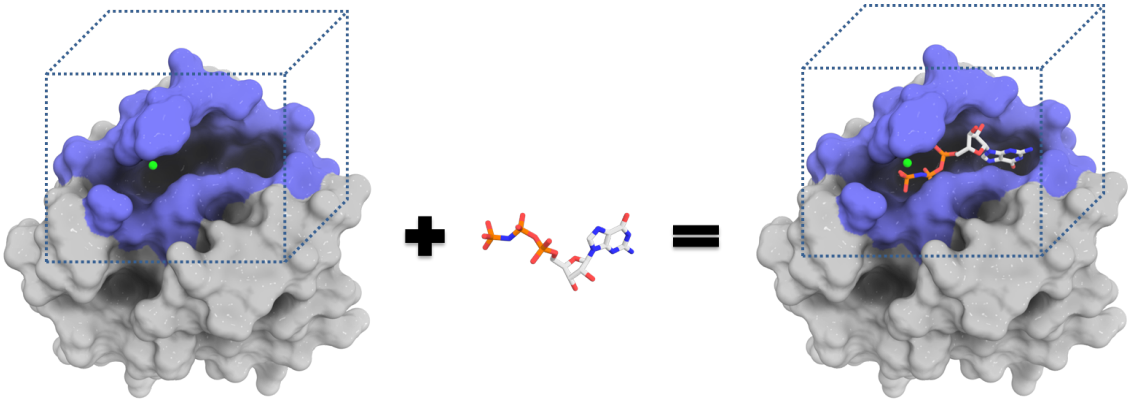

Figure 7: The figure depicts the process of small-molecule (ligand) docking into a protein target.

Molecular docking techniques computationally predict the binding mode and affinity of small molecules within the active site of the protein. The figure illustrates the exploration of ligand conformational space, docking pose generation, and scoring to identify potential drug candidates for further optimization in structure-based drug design studies ([165]).

The conformation of a ligand is one of the significant values to evaluate docking results, such as RMSD ([125]). RMSD calculates the average distance between the atoms of stacked proteins or ligands and assesses the similarity between the reference structure and the expected docked location. A known experimental structure is often used to determine the accuracy of docking predictions using docked conformation. Once the conformation of ligands approaches the natural structure, the RMSD of small ligands is close to zero.

Peptide-protein docking Peptides have vital roles in many biological processes, including cellular communication, control of enzymes, and modification of immune responses ([119]). Also, targets for drug development are peptide-protein complexes since small peptides either act as inhibitors or modulators of protein activity. Furthermore, peptides derived from proteins can be the basis for developing peptide-based treatments such as peptide mimics or vaccinations ([121]). Therefore, it is essential to understand the binding topologies and strengths of peptide-protein complexes.

Molecular docking offers a vital tool for estimating the binding topologies and strengths of peptide-protein complexes. Investigating the interactions between peptides and proteins computationally is accomplished by docking. This technique guarantees the prediction of the strength of the binding, finds the particular sites where these interactions occur, and helps to identify the relevant residues. Understanding peptide-protein interactions in

biological systems ([119, 229, 121]) and developing peptide-based drugs depend on this knowledge. Finding the operational processes and possible therapeutic applications for peptides and proteins depends on understanding their interactions ([119]). Therefore, programs including pepATTRACT ([40]), FlexPepDock ([111]), HADDOCK2 ([193]) and PEP-SiteFinder ([162]) have been utilised to comprehend the binding topologies and strengths of peptide-protein complexes.

A comprehensive comprehension of the binding topologies and strengths of peptide-protein complexes is necessary to elucidate their functional functions and facilitate the development of therapies based on peptides. The process of peptide-protein docking generally consists of two primary stages using molecular docking ([234]): (1) the creation of peptide conformations and (2) the anticipation of their interaction with the protein target. The initial stage involves the utilisation of diverse conformational sampling methodologies, such as Monte Carlo simulations or molecular dynamics simulations, to investigate the conformational space of the peptide ([150]). Docking algorithms are employed in the second stage to forecast the most favourable binding position and strength of the peptide within the binding site of the protein target. The algorithms frequently employ scoring functions to assess the compatibility between the peptide and protein, as well as to choose the binding mode that is most energetically advantageous (Figure 8).

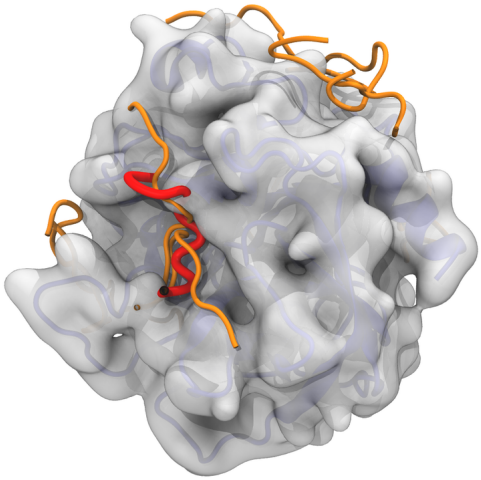

Figure 8: The top 10 peptide poses on the target protein (in grey).
The orange peptides and the protein receptor shown in white are strikingly shown in the visualiser. These highest-ranking models were chosen according to their docking scores, which reflect their possible binding affinity and stability. The complex interactions between the peptides and the protein receptor are stressed by emphasising the important binding sites and potential structural changes. This graphic provides a complete overview of the docking results, therefore supporting additional research and understanding of the interactions between the protein and peptide ([31]).

Figure 8) shows the top 10 poses derived from peptide-protein docking simulations, therefore illustrating the several orientations and likely binding modalities of the peptides inside the binding region of the protein receptor. These studies help identify meaningful interactions and structural elements and help create novel therapeutic drugs for particular protein interfaces. Analysing several docking positions allows one to assess the

binding strength and project biologically relevant interactions. This mechanism helps to better appreciate how peptides support protein activities and their possible applications in biomedical research.

Protein-protein dockings A computer method used to predict the three-dimensional shape of a complex resulting from the interaction of two or more proteins is protein-protein docking (Figure 9). Many biological functions, including enzyme activity and cellular signalling, depend critically on a knowledge of these relationships. Known examples of docking programs for protein-protein binding are HDOCK ([215]), MEGA DOCK ([174]), and ZDOCK ([144]) to investigate these interactions. Simulating the binding interaction between proteins using the protein docking program helps one to find the best orientation and position at which the two proteins bind. In the drug development framework, the given knowledge is quite valuable as it allows the creation of molecules that specifically target protein-protein interactions and inhibit pathogenic pathways. Due to the complex design of protein-protein interactions and the broad spectrum of possible binding methods, protein docking remains a challenging task, even with significant advancement. Still, ongoing studies help to improve the accuracy and efficiency of docking methods, therefore transforming them into a powerful tool for understanding the intricate terrain of protein interactions.

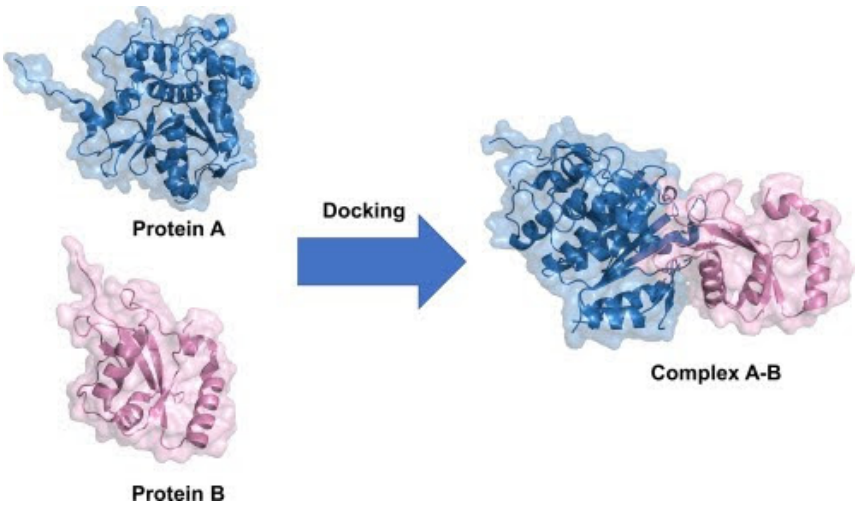

Figure 9: The figure shows an example of protein-protein docking, in which two protein molecules connect to form a complex.

A protein-protein docking approach predicts two proteins' most advantageous binding modes and affinity. The figure shows the study of conformational space, the generation of docking poses, and the scoring methods used to ascertain the binding configuration most energetically favourable among the proteins engaged in the interaction ([178]).

Classification of Docking by Flexibility Molecular docking is a prevalent computational method in structural biology and drug development. It is used to forecast the binding interactions of molecules, such as proteins and ligands. The flexibility of molecules, specifically proteins, is vital in influencing their ability to bind and selectivity. Three primary methodologies are typically utilised in molecular docking research to accommodate protein flexibility: ([96]): (i) rigid docking, (ii) semi-flexible docking, and (iii) flexible

docking (Figure 10). Each methodology presents unique benefits and constraints, and the method selection relies on the research goals and attributes of the studied biological system.

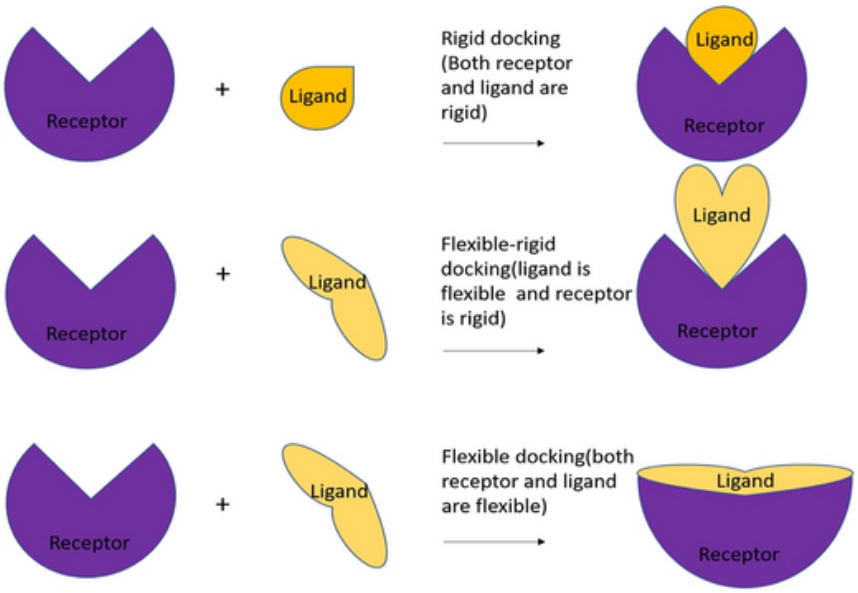

Figure 10: Three different protein docking techniques—rigid docking, flexible-rigid (semi-flexible) docking, and flexible docking—are shown in the diagram.

Every method forecasts the binding interactions among protein molecules using different approaches. While semi-flexible docking enables limited flexibility in some areas, rigid docking requires the absence of any changes in the shape of the protein structures. Conversely, flexible docking considers significant conformational changes in proteins and ligands during binding. Understanding the differences among these approaches will help one decide which is best for studying protein-ligand interactions ([126]).

Rigid docking Rigid docking is a computational approach utilised in structural biology and drug development to predict molecule binding interactions. Here are several examples of grid docking programs that have been employed in drug discovery and development, including MS-DOCK ([167]), pyDock ([28]), and RDOCK ([101]). Such rigid docking programs assume that both the ligand and receptor molecules have constant and unchanged shapes during the docking process ([4]). The technique helps determine the binding modes and affinities of molecular complexes. Rigid docking reduces the computational complexity by disregarding any changes in the shape or structure of the ligand or receptor when they bind together. It allows for a quick examination of the binding possibilities. Rigid docking methods utilise several algorithms and scoring functions to systematically explore energetically favourable binding positions, hence aiding in detecting potential interactions between ligands and receptors.

Although rigid docking may oversimplify the dynamic nature of molecular interactions, it continues to be a valuable tool for virtual screening, lead optimisation, and structure-based drug design initiatives. Rigid docking is essential to the drug discovery process because of its computational efficiency and capability to handle massive datasets. It allows researchers to choose potential therapeutic candidates for further experimental validation and optimisation ([167, 28, 4, 101]).

Semi-Flexible Docking (Flexible-rigid docking) Semi-flexible docking (Figure 10) is a computational method that combines the features of rigid and completely flexible docking approaches. It aims to balance computational efficiency with the ability to account for ligand flexibility during the docking process. Numerous molecular docking programmes, such as DiffBind ([236]) and CANDOCK ([59]), have been documented in the literature and can be utilised to explore ligand-receptor interactions. Semi-flexible docking, such as DiffBind ([236]), involves keeping the receptor structure fixed while allowing the ligand to undergo limited conformational flexibility. This flexibility enables the ligand to make structural alterations to match the binding site better. The semi-flexible docking approach recognises the significance of considering the flexibility of ligands in accurately forecasting binding modes and affinities, especially in situations where ligands can take on many conformations when binding to the receptor ([236]). Semi-flexible docking methods utilise algorithms and scoring functions that can effectively explore the flexibility of ligands while quickly sampling the space for binding. Semi-flexible docking is vital to enhance the reliability of virtual screening and drug design studies by effectively modelling ligand-receptor interactions while considering computing cost and accuracy ([183, 59]).

Flexible Docking Flexible docking (Figure 10) is an advanced computational method used in molecular docking to consider the flexibility of both the ligand and receptor while performing docking ([152]). Flexible docking methods accommodate conformational changes in both the ligand and receptor, unlike rigid docking methods that assume constant conformations for both molecules. Flexible docking makes predictions more accurate and better than rigid and semi-flexible docking because it adds complete flexibility to the docking process ([158]). Therefore, there are plenty of flexible docking programs in the literature, such as CABS-dock ([94]), ATTRACT ([41]), DREAM++ ([120]) and SwarmDock ([188]). As a result, they provide a more thorough understanding of the landscape of interactions between ligands and receptors; therefore, it is a helpful tool in drug discovery, virtual screening, and structure-based drug design efforts ([158]).

Classification of Docking by Input Number Classification of Docking by Input Number involves categorizing docking methods based on the number of input molecules or targets involved in the process. This classification helps in understanding the scope and application of different docking approaches. There are three main groups under this classification: (i) Reverse (Inverse) Docking, which involves screening a single ligand against multiple protein targets to identify potential binding sites and off-target effects ([87]); (ii) Virtual Screening, where an extensive library of ligands is screened against a single protein target to identify potential drug candidates ([37]); and (iii) Cross-Docking, which involves docking multiple ligands against multiple protein targets to explore a wide range of possible interactions and binding affinities. Each group offers unique insights and advantages, making them valuable tools in computational drug discovery ([106]).

Reverse(inverse) docking Reverse docking techniques utilise advanced algorithms and scoring functions to assess the binding affinity between the ligand and different protein targets ([214]) (Figure 11). Reverse docking allows for ranking candidate targets based on their projected interaction strength. It is a method that involves methodically analysing protein structures to identify potential biological targets for small compounds.

Therefore, reverse docking is a method that differs from typical docking approaches as it prioritises the prediction of protein-ligand interactions. Instead of guessing how a ligand will interact with a protein, reverse docking looks through a library of protein structures to see which ones might interact with a specific ligand ([214]).

Reverse docking techniques are very beneficial in drug discovery, as they can assist in identifying targets, predicting off-target effects, and understanding the polypharmacology of small compounds ([96, 72]). The off-target bindings may be an option to design polypharmacological drugs, or they cause side effects. Distinguishing between two possibilities is critical to saving funds and time. For instance, Pfizer designed sunitinib, which is cardiotoxic. Off-target bindings on AMP-activated protein kinase (AMPK) families and the ribosomal S6 kinase (RSK) are the reasons for cardiotoxicity ([60]). The compound wasted significant time and funds of the pharmaceutical company ([60]). Therefore, reverse docking is promising to decide whether off-target binding is a reason for side-effect or a polypharmacology opportunity.

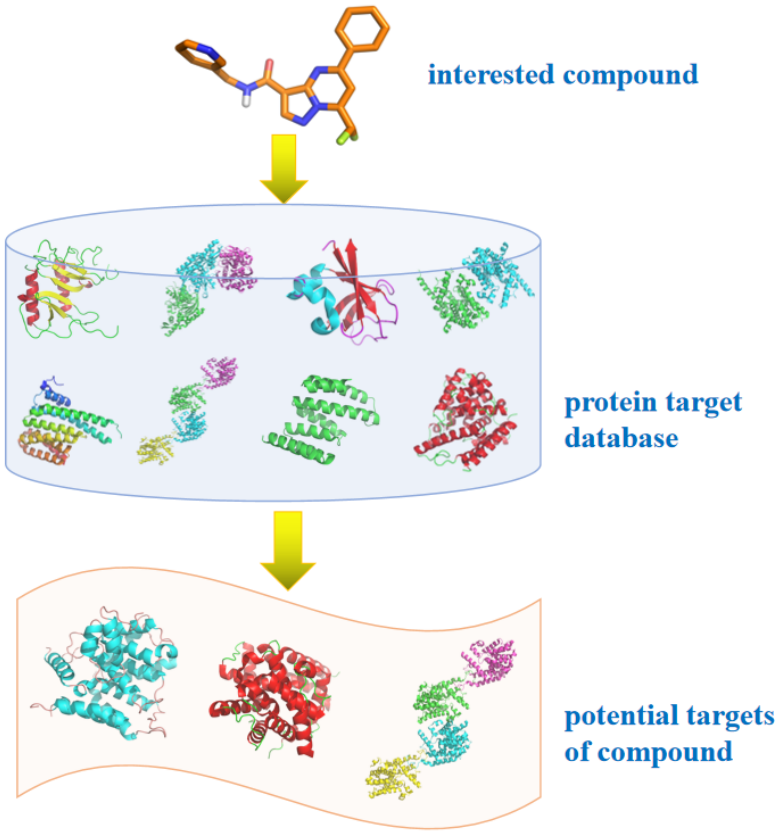

Figure 11: The representation of reverse docking for a small compound into target database

The concept of reverse docking, a computer process used in drug discovery to identify possible protein targets for a given small molecule or ligand, is illustrated here. Reverse docking looks at the interactions between a ligand and a set of protein structures to identify likely binding partners, unlike traditional docking techniques that predict the binding mode of a ligand inside a specified protein target. In drug discovery research, the approach described has great relevance for target identification, lead optimisation, and drug repurposing ([23]).

Virtual Screening Virtual screening (Figure 12), alternatively referred to as computational screening, is a robust computer methodology employed in the field of drug exploration to expeditiously assess extensive collections of chemical compounds and ascertain prospective drug contenders that exhibit a high probability of binding to a specific target protein of interest ([156]). The procedure involves docking several ligands, commonly of small size, into a target protein's binding site and then evaluating their interactions to determine the relative importance of compounds exhibiting the most significant binding affinity. Virtual screening is of utmost importance in the initial phases of drug development since it accelerates the identification of lead compounds with favourable pharmacological characteristics, including potency, selectivity, and drug-likeness ([37, 199, 177]).

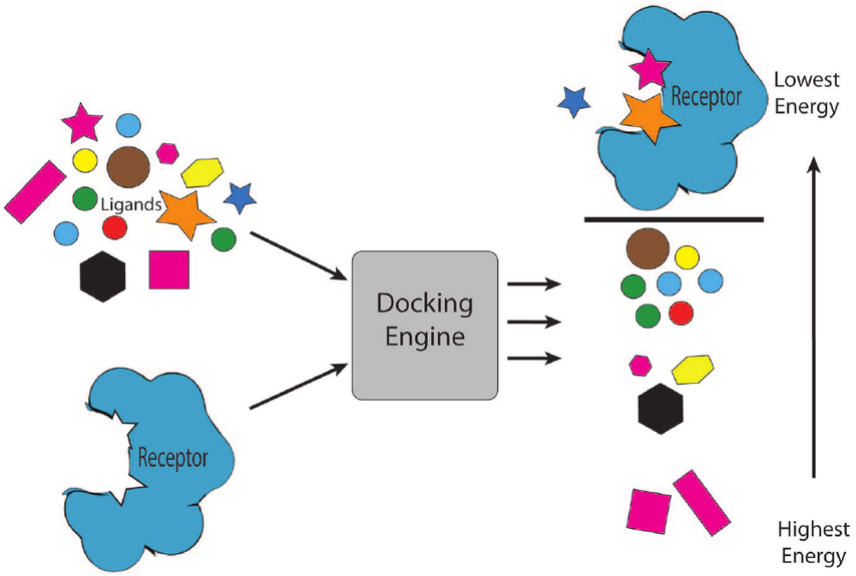

Figure 12: The schematic representation of virtual screening

The figure shows virtual screening, a computer technique used in drug research to precisely arrange multiple ligands into the binding area of a target protein. Virtual screening is a method that helps to quickly evaluate large chemical libraries in search of potential drug candidates with substantial pharmacological action and binding affinity. Since it speeds up the discovery of possible leads and improves the effectiveness of drug development pipelines ([77]), this method is essential in the first phase of drug research.

Cross-docking Cross-docking is a complex computational technique for simultaneously binding several ligands into several target protein configurations ([106]) (Figure 13). Therefore, it provides an essential understanding of the selectivity and specificity of interactions between ligands and proteins. For example, it can be helpful to determine off-target binding, which indicates side effects. However, Cross-docking has a disadvantage in that, particularly for large-scale datasets, the considerable processing resources required to dock multiple ligands into several targets concurrently are a burden. Furthermore, cross-docking may have trouble with the precision of scoring systems and the complexity of ligand-protein interactions, which may lead to erroneous positive or negative forecasts of binding affinities. Furthermore, it is limited to applying cross-docking outcomes to different protein families and structural modifications. Thus, careful analysis and result validation are much more critical. Cross-docking remains a valuable technique for examining the interactions between ligands and targets and for spotting new treatment

candidates with diverse pharmacological profiles, even if there are challenges ([170]).

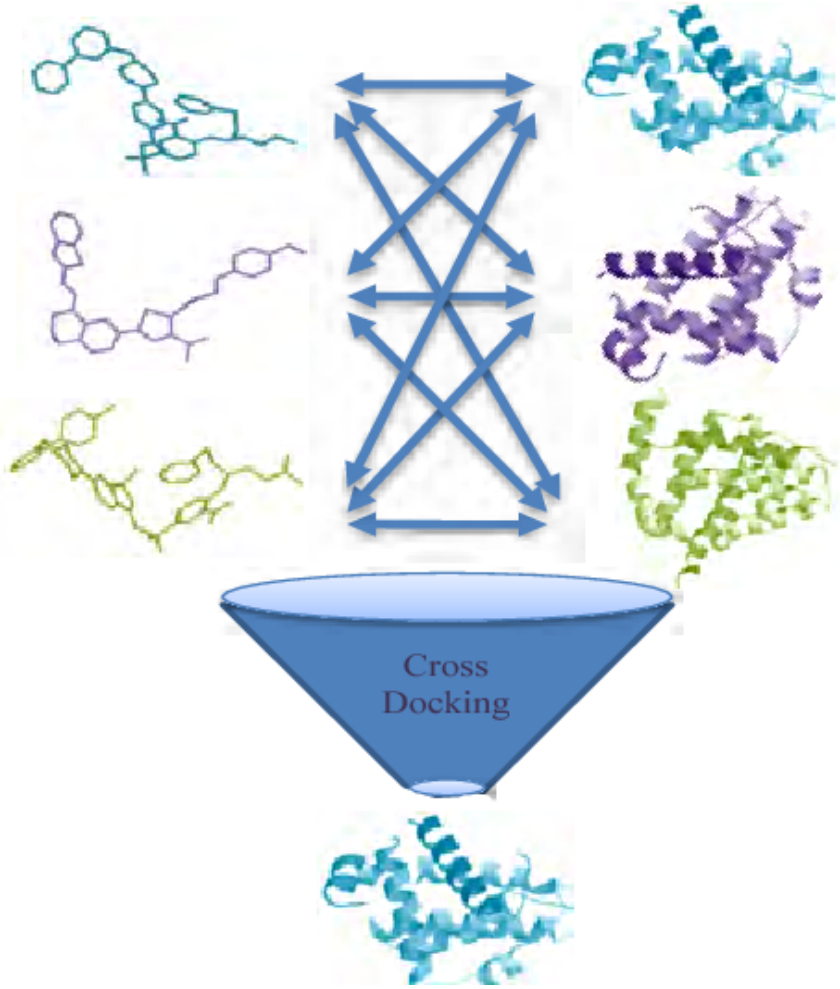

Figure 13: The overview of cross-docking for multiple ligands into multiple targets
A computational method applied in structural biology and drug development to dock many ligands into a binding region of a target protein. Cross-docking is unlike conventional docking, which concentrates on a single ligand-target complex in that it allows the evaluation of ligand binding modes and interactions across several ligand-target combinations by docking a varied range of ligands into a single protein structure. This method improves the development of structure-based drug design techniques by helping to comprehend protein-ligand recognition patterns and pointing up shared binding motifs ([79]).

Classification of Docking by Search Space The classification of docking by search space is the grouping of docking techniques depending on the extent of the search area taken into account during the docking procedure. Understanding the attention and computing needs of several docking techniques depends on this classification. This classification has two main categories: (i) Local Docking, which limits the search area to a particular region or binding site on the protein, and (ii) Global Docking, which looks over the whole surface of the protein to find possible binding sites and binding poses ([218]) (Figure 14).

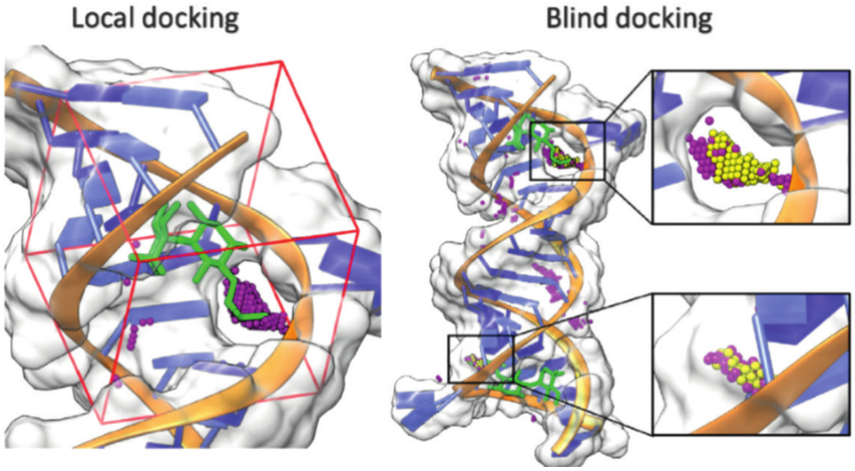

Figure 14: The representation of local and blind (global) docking

Local and global docking simulations of the aminoglycoside antibiotic Gentamicin (shown in green) with bacterial ribosome's 16S rRNA A-site. RLDOCK ([181]) predicts binding locations in the image in pink and yellow. The red cup shows that local docking concentrates especially on a limited area. Therefore, it optimises the search for possible binding sites inside a particular target molecule. Global docking, on the other hand, searches the whole protein surface, looking for several likely binding sites where the ligand might engage. This all-encompassing strategy lets one broadly investigate binding options around the target structure. Understanding molecular recognition and creating effective antibiotics depends on knowledge of the different intensities and orientations of ligand interactions ([97]).

Local docking: Local docking requires a binding site and search space from a user-defined one (Figure 14). There are two main approaches to defining a location for local docking: (i) experimental ligand binding sites and (ii) theoretical predictions. (i) Experimental techniques capture the location of small natural molecules on targets as a binding site. Small natural molecule binding sites are called ligand binding sites (LBSs) ([230]). Most natural LBSs are located on the surface of a protein because of the high affinity obtained by large interfaces. By utilising the coordinates of LBSs, a molecular docking programme can be employed to identify the potential positions of ligands on the coordinates. (ii) Also, theoretical approaches have been developed to identify potential binding regions. For example, Deep-learning cavity finders are the most effective method ([230]), but they suffer from interoperability and extended training time. Recently, the quantum algorithm increased the predictive power of machine learning in a short time ([164]). The research provided Polar+, the first biological modelling, and it was tested on quantum computers ([164]). However, it has significantly higher training costs than classical machine learning approaches ([230]).

Regrettably, the prediction methods used by LBSs are inadequate for fully resolving the issue of detecting LBSs due to factors such as protein flexibility, the limited efficacy of computational approaches, the intricate nature of molecular interactions, and the difficulties in accounting for solvent effects ([76]). Also, cryptic sites become clear when proteins are in a complex (bounded form). There are some studies to determine LBSs successfully. For example, molecular dynamics simulation is a popular method to assess LBSs since it analyses the physical movements of atoms and molecules. Also, machine learning or deep learning integrated with molecular dynamics is promising ([230]). Finally, although combining computational predictions and experimental data is currently

the best solution ([44, 171, 186, 70]), performing global (blind) docking is another option to overcome the limitation of identification of binding sites.

Global(blind) docking Global docking—also known as blind docking—involves the thorough study of the whole surface of the protein to identify likely binding sites and project the ligand binding mechanisms without first understanding the exact location of the binding site ([180, 159]) (Figure 14). Unlike local docking, global docking does not necessitate prior knowledge of specific binding cavities, enabling an impartial evaluation of the binding affinity between the target and ligand ([43]). Global docking comprehensive technique facilitates the identification of previously unnoticed binding sites that more targeted methods may disregard. Hence, global docking is especially advantageous during the initial phases of drug development since it facilitates the creation of novel pharmaceuticals by offering a comprehensive perspective of potential interaction sites throughout the complete target protein. Global docking can potentially uncover previously undiscovered binding sites, which can then be used to create more potent and groundbreaking medicinal medicines. The most cited global docking programs are ZDOCK ([25]), FlexX ([92]), GOLD ([196]) and MEGA DOCK 4.0 ([136]).

Global docking offers several advantages in molecular docking by exploring all potential binding sites on a target protein. This comprehensive approach ensures that no potential binding region is overlooked, providing a complete understanding of possible ligand interactions. One significant advantage is its utility in predicting side effects, as it examines every cavity on the target protein, identifying off-target binding sites that might lead to adverse impacts. The therapeutic effect or side-effect of a ligand depends on where and how it binds to a target ([74]). Any cavity on a target may be a reason for side effects. Therefore, cavities should be considered to predict side effects ([161]). These requirements make global docking more suitable to investigate side effects. A unique consensus-global docking method can destroy the limitations of global dockings, such as high false-positive and low accuracy ([220]). Despite the advantages of global docking programs, they have been plagued by lower performance than local docking methods.

A global docking program has been suffering from a lack of critical location features for binding. Binding location helps local docking focus on the correct location, while global docking should define that position first before increasing performance in ligand pose ([35]). Therefore, global docking's performance is lower than that of local docking. As a result, hybrid molecular docking has been published to improve global docking performance ([73]).

Hybrid molecular docking combines the strengths of both global and local docking approaches. It initially employs global docking to explore potential binding sites across the target surface. Then, it refines the search using local docking techniques to focus on the most promising regions, enhancing the accuracy and efficiency of the docking process ([73]). For example, the hybrid global docking example is LigDockCSA ([175]), which combines conformational space annealing (CSA) with AutoDock's energy function. It has an 84.7% success rate, compared to 80.5% for GOLD and 81.7% for AutoDock. Also, the success rate of LigDockCSA becomes 89.4% with the help of conformational entropy ([175]). The examples indicate that hybrid molecular docking provides more accurate results.

1.2.2 Molecular Dynamics Simulations in Drug Discovery

The classical molecular dynamics (MD) methodology is a computationally taxing technique enabling quantitative study of molecular events. Classical all-atom MD is a modelling method that precisely simulates all atoms in a given system, including the solvent. Considering interatomic forces, it uses classical bonded and nonbonded potentials (Figure 15). Its better performance has resulted in significant developments and has been efficiently applied to handle conformational changes, folding binding penetration, and many other problems ([107]).

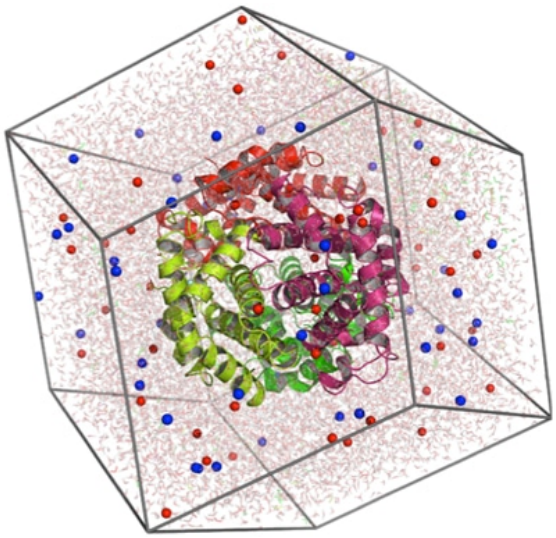

Figure 15: The schematic representation of molecular dynamic simulation
Interactions between proteins and substrates within a molecular dynamic simulated period have dynamic character. The trajectory clarifies important contact sites and conformational changes, therefore providing insightful analysis of the molecular-level stability and binding mechanisms ([107]).

MD has faced two main challenges: first, the computation of interatomic potential tables, sometimes known as force fields, has historically been a laborious process requiring excellent refinement; second, it is computationally demanding despite reasonable efforts and developments in expediting molecular dynamics codes ([124, 45]). To overcome these challenges, machine learning (ML) techniques in MD simulations have been enhanced in terms of their value and efficiency in drug development ([18]). Machine learning methods can analyse large amounts of simulation data to identify trends and project molecular behaviours. This so accelerates the process of spotting possible drug candidates with promise. ML-driven MD simulations offer a potent mix of accuracy and efficiency by improving force fields, anticipating binding affinities, and maximising sample efficiency. MD simulations and ML streamline the drug development process and allow logical synthesis of more specific drugs ([18, 163]).

1.2.3 Binding Site Identification in Drug Discovery

Medications' effects are manifested by their interactions with distinct binding sites on target proteins. These binding sites can be categorised into groups according to their respective mechanisms and locations. The binding sites can be classified into three primary

groups: (i) orthosteric, (ii) allosteric, and (iii) cryptic binding sites ([191]).

Orthosteric binding site: Orthosteric drugs bind to a protein's active site, competing with the natural substrate or ligand (Figure 16. Their effects are exerted by outcompeting the native substrate and obstructing the active site when they possess a strong affinity for the site. Most drugs available in the market are traditionally orthosteric ([210, 141]). Also, the orthosteric active sites within a protein family exhibit a high degree of conservation, implying that a drug designed to target the active site of one protein can also interact with the active sites of other proteins belonging to the same family ([115]).

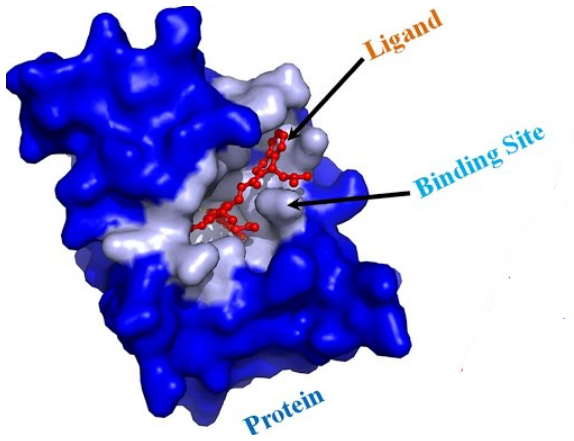

Figure 16: The representation of Orthosteric binding site on the target protein
Interactions between proteins and substrates within a molecular dynamic simulated period have dynamic character. The trajectory clarifies important contact sites and conformational changes, providing insightful analysis of molecular-level stability and binding mechanisms.

Although extensively employed, orthosteric binding sites and pharmaceuticals also have specific drawbacks in drug design and therapeutic uses ([50]). A notable constraint is the possibility of off-target effects caused by the extensive similarity of active sites throughout protein families ([213]). This can result in unintentional interactions with proteins that have similar structures, leading to adverse effects and diminishing the selectivity of the medicine. Furthermore, orthosteric medications frequently compete with endogenous ligands or substrates for binding, which might restrict their effectiveness in specific physiological situations or disease states characterised by fluctuating substrate concentrations ([50, 213, 194]). Also, the total suppression of protein function by orthosteric medications may not always be preferable, as it can interfere with regular cellular processes that depend on regulated enzyme activity ([34]). The significance of taking into account alternative drug design techniques, such as allosteric modulation, is emphasised by these aspects. These strategies aim to obtain more accurate and specific therapeutic results while reducing the possible disadvantages associated with orthosteric binding.

Allosteric binding site Often called allosteric control, allostery is a fundamental biological occurrence relevant to signal transduction pathways, metabolic activities, and genomic transcription ([20, 51]). A localised variation in conformation at the active site results from the fast change in the conformational ensemble balance at an allosteric site ([82, 129]). Potential disturbances cover the interplay between localised chemical

changes ([36, 63]) and small molecules/ions. Thus, allostery is the primary way to regulate the function of biological macromolecules (Figure 17).

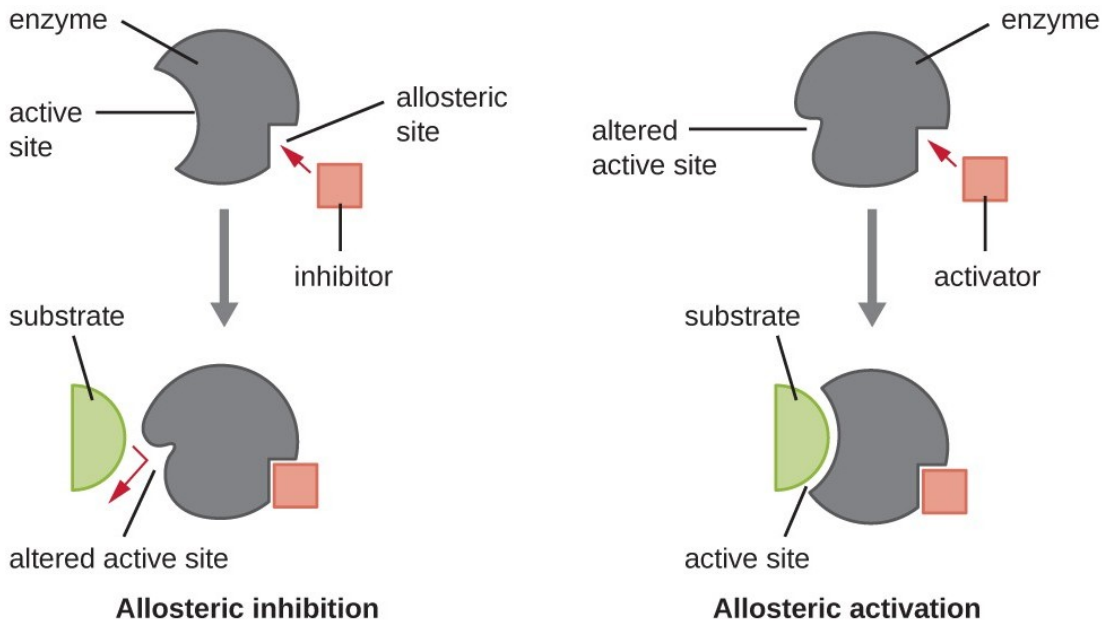

Figure 17: The representation of how allosteric activation and deactivation work.
Controlling protein activity is done by attaching parts of the protein not in the active site, the "orthosteric" site. The figure shows allosteric inhibition, which happens when a ligand binds to an allosteric site and causes a conformational change on the protein's orthosteric side to inhibit binding. In contrast, allosteric activation (right) occurs when a ligand attaches to an allosteric site, rearranging the protein's orthosteric site structure that enhances its activity. Allosteric regulatory mechanisms are of utmost importance in the context of cellular signalling and the regulation of enzymes.

Knowing allostery can give critical new perspectives for the progress of allosteric drug discovery and development ([133, 30]). Among the essential roles allostery plays in many biological processes are those of enzyme catalysis, signal transmission, and gene regulation. Allostery is the phenomenon wherein activity occurs at a distance when a disruption at one point inside a macromolecule causes functional changes at another. Several processes can lead to the modulation of protein activity by allosteric mechanisms: effector-binding interactions involving small molecules, liquids, DNA/RNA, or proteins; covalent modifications including phosphorylation; and photoabsorption ([118, 20, 82, 129]).

Allosteric pharmaceuticals exhibit binding affinities or catalytic efficiency of biological macromolecules using a perturbation signal propagation but at a place distinct from the active site. Allosteric medications have various advantages compared to orthosteric drugs ([30, 134]). Based on sequence conservation analysis, it has been observed that allosteric sites exhibit a lower degree of conservation compared to orthosteric sites ([217, 117]). The lower degree of conservation of allostery enables allosteric modulators to effectively target specific subtypes within receptor families, leading to enhanced selectivity and reduced occurrence of adverse effects compared to orthosteric drugs ([22]). Also, allosteric medicines can regulate protein activity without directly competing with natural ligands, decreasing the probability of adverse effects related to unintended interactions ([132]). They offer more refined regulation of protein activity, enabling partial activation or inhi-

bition instead of complete blockade ([143]). Partial activation or inhibition can be advantageous for preserving regular cellular processes. Furthermore, allosteric pharmaceuticals do not impede the interactions between substrates and proteins; a maximum limit exists to allosteric regulation ([143]). In addition, Allosteric pharmaceuticals are beneficial due to two primary factors: firstly, they can provide a less disruptive method to modulate the activity of a pathway, and secondly, they are more likely to have a reduced incidence of adverse effects ([42, 68, 86]). The other advantage of allostery is that the utilisation of techniques that combine allosteric modulators with orthosteric medications can offer advantages due to the issue of drug resistance, which arises from mutations in the protein target that surpass the inhibitory effects of both orthosteric and allosteric pharmaceuticals ([127, 68, 86, 46, 224]).

Using allosteric modulators could help orthosteric treatments become even more effective. GNF-2 is one instance of an allosteric modulator; it shows binding affinity to T315I human Bcr-Abl's myristate-binding sites. On the mutant Bcr-Abl protein, GNF-2 and the substrate-competitive inhibitor imatinib show synergistic inhibitory effects ([227]). As such, the co-administration of these two drugs offers a possible approach to overcoming drug resistance in patients with chronic myelogenous leukaemia (CML). Also, the US FDA has so approved several allosteric medicines. For example, developed by Genzyme ([131, 113]), plerixafor is one example of an allosteric blocker of the C-X-C chemokine receptor type 4 (CXCR4) that helps haematopoietic stem cells (HSCs) be mobilised. The debate mentioned above on the benefits of allostery and the proof of successful allosteric drugs underlines the great possibilities of allostery. It is crucial to recognise its constraints, including the unknown positions of allosteric binding sites on target molecules, to overcome the restrictions of allostery.

Allosteric pharmaceuticals have various restrictions, even if they offer some encouraging benefits. The critical difficulty is that, for most pharmaceutical targets, the exact areas of allosteric activity are yet unknown ([112]). This ambiguity about the allosteric areas makes designing and developing medications that can attach to these locations efficiently challenging. Moreover, several obstacles hinder the identification of allosteric modulators, including restricted binding strengths and the usually unknown structural properties of putative tiny allosteric compounds ([102, 202]). Furthermore, allosteric sites show less conservation than orthosteric sites, which causes differences in the therapeutic efficacy of several protein targets ([209]). Furthermore, the inherent adaptability of allosteric sites complicates the search for medications, which may only show themselves under particular structural states of the protein ([198]). These constraints hinder allosteric pharmaceutical development and call for more studies to grasp better and use allosteric pathways for therapeutic uses.

Cryptic binding site Cryptic binding sites are hidden or transient regions that are not evident when the protein is inactive or not bound ([78]). However, these cryptic sites either arise or become accessible when a ligand hooks to the protein or when its form changes (Figure 18). These cryptic sites depend on particular conditions or the presence of specialist ligands for their visibility, so they are often invisible using typical structural research techniques such as X-ray crystallography or NMR spectroscopy ([157]). Since they offer fresh drug discovery and development opportunities, especially for targets that have been difficult to control using conventional orthosteric or allosteric sites, identifying

and understanding hidden binding sites is vital. Thus, various compotation strategies have been designed to investigate cryptic binding sites and understand their mechanisms.

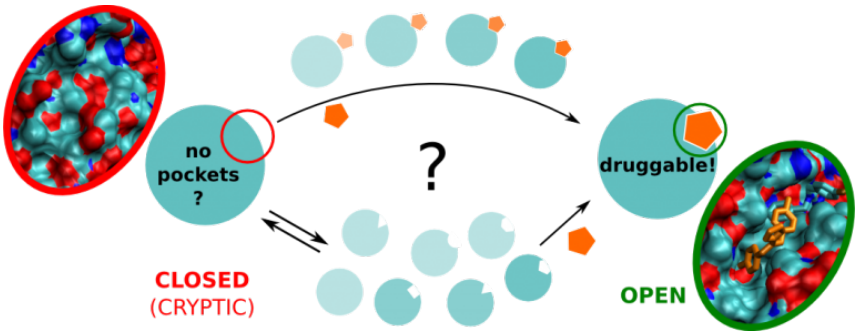

Figure 18: Initially lacking a pocket structure until the ligand binds, the concept of a cryptic binding site exposes the hidden binding site.

Often concealed within proteins, cryptic binding sites become accessible for ligand binding via conformational changes brought about by ligand binding or protein-protein interactions. Understanding and focusing on mysterious binding sites offer interesting chances for investigating drugs and applying therapeutic actions. These cases show how the MD technique is the accepted method for locating hidden places([137]).

Various computational strategies have been employed to detect cryptic or ”transient” locations, considering protein dynamics ([5, 95]). For example, Markov state models on molecular dynamics simulations detect cryptic sites that effectively reveal the hidden locations of two β-lactamases ([61]). In their study, Gao et al. successfully produced bound conformations in lengthy microsecond molecular dynamics (MD) simulations employing unbound initial structures for a mere 8 out of the 39 systems under investigation ([61]). Also, Oleinkovas et al. did not identify hidden locations for three systems using microsecond-length molecular dynamics simulations. As a result, they devised a method to improve sampling by utilising scaled Hamiltonians to sample water interfaces based on replica exchange molecular dynamics ([137]). Moreover, Cimermancic et al. ([32]) uncovered a set of proteins with cryptic sites for their web server, Cryptosite, which predicts binding sites. The term ”cryptic” was used to describe a site that could not be identified using FPocket ([98]) when utilising the unbound structure.

Drug discovery depends on identifying cryptic binding sites, yet traditional computational and experimental approaches are somewhat limited. The always-shifting properties of cryptic sites, which usually go undetectable in the frozen protein structures obtained by crystallography or cryo-electron microscopy, provide a significant challenge. Furthermore, complicating the identification process is the natural flexibility of proteins ([6]) and the limited resolution of experimental instruments. Potential approaches to effectively overcome these limitations and find cryptic binding locations ([233]) come from machine learning (ML). Using large databases of protein structures and binding interactions, ML models could forecast hidden sites that are not readily apparent with conventional methods. Combining molecular dynamics simulations with machine learning techniques allows one to precisely find hidden spots on proteins by recording their transient shapes ([204]). Furthermore, machine learning can help to analyse large amounts of experimental data by identifying relationships and traits that would point to the presence of latent binding sites, therefore accelerating the process of developing drugs.

1.3 Ligand-based drug discovery approaches

Ligand-based drug discovery strategies are fundamental in contemporary pharmaceutical research. They concentrate on comprehending and enhancing the chemical characteristics of drug molecules to attain specific therapeutic outcomes. These approaches utilise ideas based on molecular interactions and physical features of ligands, which are small ligands that preferentially attach to biological targets like proteins or nucleic acids. Standard methodologies include Lipinski Rule of Five ([104]), LogP ([93]), Biopharmaceutics Classification System ([17]), and In-vitro in-vivo correlation (IVIVC) ([114]). While these methodologies are essential in drug discovery and development, the ”key” and ”lock” ideas have drastically impacted new tools and approaches ([33]).

The idea of ”key” and ”lock” in drug discovery is that “similar ligands bind similar targets”, so molecular similarity is one of the target identification methods ([33]). Similarity methods require a representation method for compounds, such as the Simplified Molecular Input Line Entry System (SMILES). SMILES is the most common method to represent and compare the compounds in 1D ([208, 207]). It converts a compound into a string, using symbols such as C, c, N, O for atoms and =, # for bonds (www.daylight.com/dayhtml/doc/theory/theory.smiles.html). SMILES are available in quantity structure-activity (QSAR), virtual screening, and toxicity prediction. An example of a similarity search algorithm is the fingerprint Similarity Search Algorithm (MuSSeL), which can provide IC50 or Ki values for ligands ([228]). Finally, other compound similarity methods exist in the literature, such as 2D-based compound similarity kernels (Figure 19).

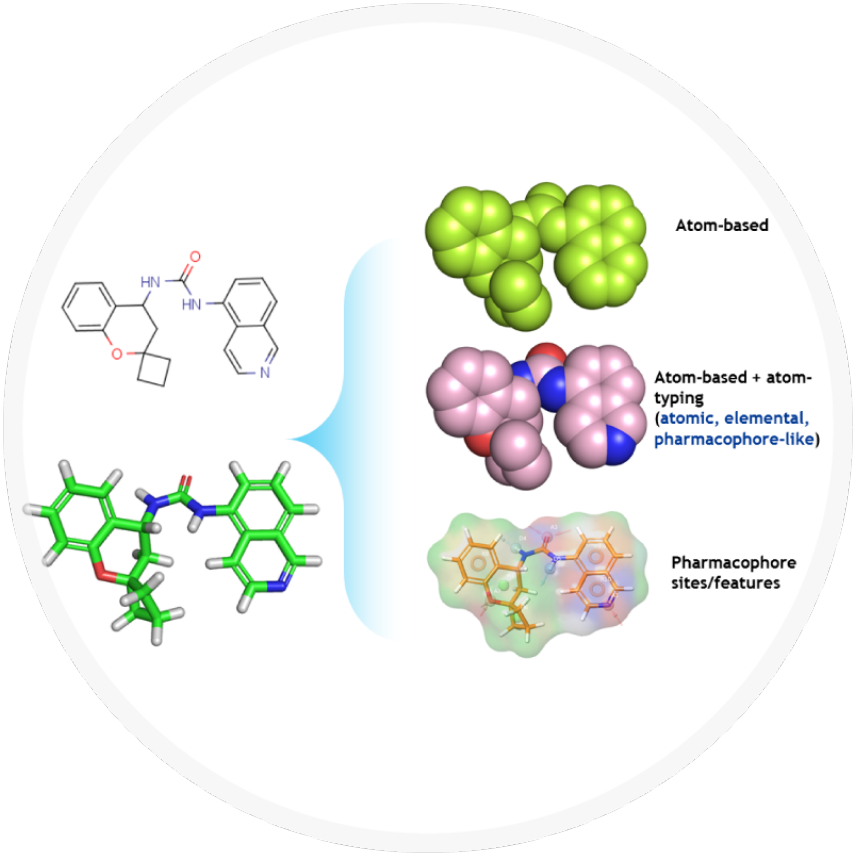

Figure 19: The representation of ligand-based drug discovery approaches

Using data from known ligands that attach to target proteins to identify or synthesise new compounds with equivalent functionality, ligand-based drug discovery is necessary in developing new medicines. Atom-based modelling is one method used in this approach whereby one may understand the interaction of the ligand with the target by analysing its spatial arrangement of atoms. By focussing on the precise arrangement of every atom, researchers may estimate likely binding affinities and create more potent molecules. Moreover, atom-based + atom type modelling considers the spatial organisation as well as the particular atom types—hydrogen, carbon, or nitrogen. This method helps one to understand more fully how different atomic interactions support the intensity and specificity of binding. The major functional groups of a ligand causing its biological effect are investigated using pharmacophore sites. This work finds and models these groups on a more abstract level. The pharmacophore sites identify key features such as hydrophobic regions and hydrogen bond donors or acceptors. These properties enable the synthesis of new ligands capable of strong interaction with the target protein. These methods let ligand-based drug development effectively speed up the identification and improvement of strong therapeutic prospects.

Generally, 2D-based compound similarity kernels, such as SIMCOMP ([138]), are preferred to predict drug-target prediction. Here are some examples of 2D-based compound similarity techniques to indicate their success. One of the 2D-based compound similarities is TargetHunter, a web-based tool ([201]). TargetHunter was trained on ChEMBL data, and PubChem bioassay was utilised as test data ([201]). Compared to 2D and 1D representation, SMILES-based similarity may be computationally more efficient than 2D-based approaches ([138]). Consequently, the ligand-based drug discovery approach can be more successful with other techniques, such as De Novo Drug Discovery.

1.4 De Novo Drug discovery

The concept of de novo drug design (DNDD) pertains to creating new chemical entities that adhere to a predetermined set of limitations through computational growth algorithms ([168]). The term ”de novo” denotes the process of creating new molecular entities without the need for a starting template, as it involves starting from scratch ([47]). De Novo drug design can be classified into four main groups: (i) structure-based, (ii) atom-based, (iii) ligand-based, and (iv) fragment-based. (Figure 20) Also, the next frontiers for machine-learning-enabled de novo drug creation, as a new group, include future directions such as toxicogenomics integration and vaccine development opportunities.

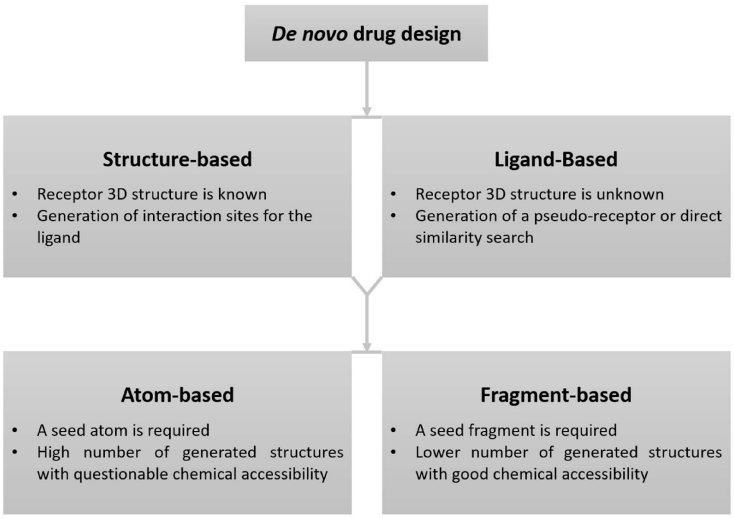

Figure 20: Classificataion of De novo drug design methods

The de novo drug-design process calls for several cutting-edge technologies, each of which uniquely helps to produce new medicinal molecules. Structure-based drug design uses the complex 3D structure of the target protein to produce molecules that exactly suit its active site, hence improving binding interactions for best efficacy. Second, ligand-based drug design uses information from known ligands interacting with the target to generate new molecules with similar or improved potency. This approach often uses computer models to predict how changes to the ligand can increase binding affinity and specificity. To guarantee the best interaction with the binding site of the target protein, atom-based drug design gives spatial configuration and atom composition top priority. This degree of precision helps to maximise the molecular interactions, therefore producing the best possible therapeutic effect. Fragment-based drug design involves the identification of small chemical fragments attaching to different parts of the target protein. These then are chemically linked or amplified to create a strong and targeted pharmacological molecule. Combining these four techniques allows de novo drug design to effectively generate novel compounds with a high probability of therapeutic efficacy ([130]).

De novo drug design offers several benefits, such as the ability to explore a broader range of chemical possibilities, the creation of compounds that represent innovative intellectual property, the possibility of developing new and enhanced therapies, and the efficient development of drug candidates in terms of cost and time. One of the primary obstacles encountered in de novo drug design is the synthetic inaccessibility of the molecular structures produced ([69]). Although de novo drug design benefits, it has limitations encompass several desired properties or chemical characteristics, such as a predetermined range of solubility, toxicity below a certain threshold, and the inclusion of specified chemical groups in the structure ([39]). Fortunately, machine learning applications in De Novo

Drug discovery have the potential to overcome limitations such as computational intensity and limited performance ([122]).

The section provides supporting terms to explain the terms and increase understanding of the research. The supporting literature review is divided into three sections: (i) How do drugs work based on binding site classifications? (ii) Quantum Mechanical Simulations in Drug Discovery and (iii) Cheminformatics Approaches for Drug Discovery

1.5 Quantum Mechanical Simulations

Quantum mechanics operates on the domain of electrons and nuclei, disregarding the influence of chemical bonds. Solving the Schrödinger equation offers a valuable means of understanding systems at the atomic level ([13]). The equation's answer interprets the spatial arrangement of electrons and their respective energy levels. Furthermore, it offers insight into molecule structure, chemical bonding, and molecular interactions ([7]). Nevertheless, the Schrödinger equation can only be solved for the Hydrogen atom. Therefore, approximations of the equation's outcomes are used for the remaining atoms (Figure 21).

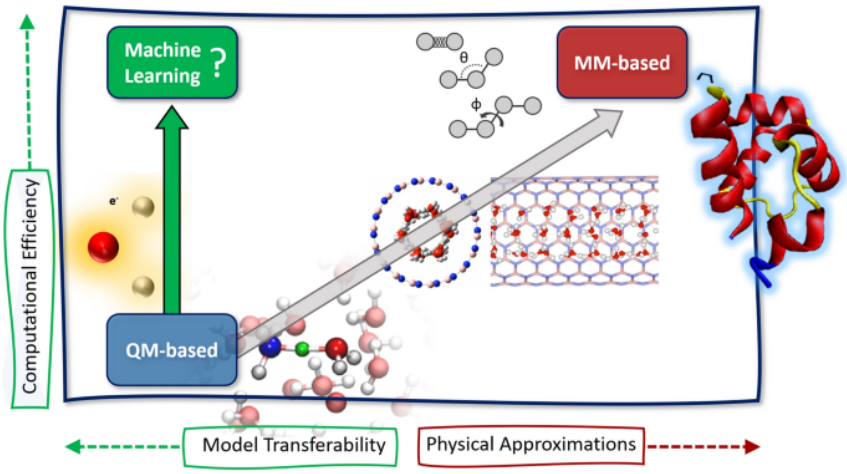

Figure 21: Based on the description of the system, two types of atomistic simulation techniques can be distinguished: quantum mechanical (QM) computations depending on the electronic structure or molecular mechanics (MM) procedures using predefined functional forms.

Their more considerable computing cost limits QM-based simulations to smaller systems. While more efficient, MM-based methods sometimes derive from experimental data and depend on various approximations. QM-based machine learning aims to improve the efficacy of QM techniques while keeping their capacity to be applied to various scenarios, precisely anticipate outcomes, and adequately explain complex bonding patterns, including the formation and breaking of chemical bonds ([128]).

Density functional theory is a computer tool for determining the ideal molecule arrangement, vibrational frequencies, free energy shift during a chemical process, and dipole moments (DFT ([11])). Furthermore, DFT is quite important in determining the affinities of protein-ligand interaction, a fundamental feature in the discipline of drug development ([52]). By providing in-depth knowledge of the electronic structure of molecules, density functional theory (DFT) allows exact predictions of the interactions between possible drug candidates and their target proteins ([88]). DFT's properties make

it a vital tool for the logical development of new drugs since they help to find exciting compounds and improve their binding capacity. This computational approach increases the efficiency and output of the drug development process, hence producing more strong and targeted drugs ([53]).

Quantum Mechanics (QM) approaches show promise but have encountered limitations such as computational power constraints, the absence of atoms and residues on proteins, and inadequate entropic methods. Rather than imposing restrictions on quantum mechanics (QM), QM possesses significant predictive capabilities in binding free energy ([19]). Machine learning techniques in the context of quantum mechanisms can yield distinctive attributes for drug design and development by overcoming the limitations of conventional QM ([128]).

1.6 Cheminformatics Approaches for Drug Discovery

Cheminformatics methods use computational and informational tools to solve chemical problems and enhance the discovery process of new drugs. Combining data from chemistry, biology, and pharmacology, cheminformatics helps to handle, examine, and present large datasets efficiently ([24]). Accelerating the identification of potential pharmacological candidates, improving their features, and predicting their performance in biological systems depend on this multidisciplinary field ([123]). Among the various advantages cheminformatics provides include the ability to examine large chemical databases rapidly, lower the cost and length of experimental procedures, and improve target identification and lead optimisation accuracy. Ultimately, these approaches enable drug research and development procedures' success and efficiency, generating fresh and creative treatments. Three chemogenomic techniques include (i) machine learning-based, (ii) graph methods and (iii) network models approaches ([212]).

1.6.1 Machine learning-based methods in Cheminformatics

Machine learning techniques in cheminformatics transform the drug development process by utilising sophisticated algorithms to analyse intricate chemical and biological data ([110]). These techniques utilise patterns and correlations in data to forecast the characteristics and behaviours of possible drug candidates, expediting the process of identifying and refining new therapeutic substances. The significance of machine learning in drug discovery is its capacity to manage extensive information, reveal concealed insights, and enhance the precision of predictions in contrast to conventional methods. The benefits encompass improved efficacy in analysing extensive chemical libraries, the capability to simulate complex biological interactions, and the possibility to decrease expenses and durations linked to medication development ([110]).

With the help of ML techniques in cheminformatics, several successful cheminformatics studies have been reported in the scientific literature ([155]). The preferable ML model is a supervised model used to study DTIs. For example, the PaDEL descriptor utilised the 1287-dimensional target descriptor and the 1024-dimensional drug descriptor from these datasets to predict DTIs ([200]). The standard classification models used in DTI research are random forest, random walk with restart, support vector machines (SVM), and decision trees ([200]). In another example, Yu et al. designed a method to indicate drug-target interactions from heterogeneous biological data using Random Forest

and SVM ([225]). Also, several machine learning models have been built on a structure-activity relationship (SAR) and structure-property relationships (SPR) ([232, 221]). An instance of the SAR model application is TargetNet ([221]). TargetNet, containing 623 SAR models, is a web service working with Naıve bayes based multi-target SAR models to predict DTIs ([221]). The last example of the QSAR model is that Bender et al. ([12]) benefit from the Bayesian-based method to build QSAR models. Finally, deep-learning algorithms in chemoinformatics also promise to identify targets for a compound ([110]).

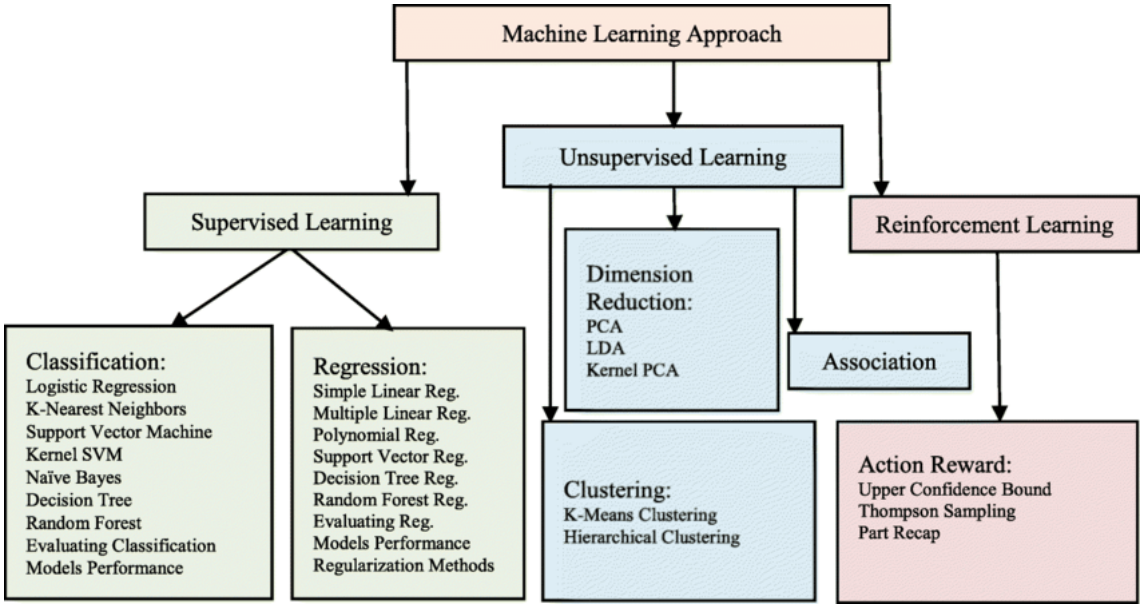

Figure 22: Various machine learning techniques applied in the field of drug discovery are shown in this diagram.

The approaches comprise unsupervised learning, which exposes hidden patterns and relationships in the data without predefined labels; supervised learning, in which models are trained on labelled datasets to forecast drug-target interactions; and reinforcement learning, in which algorithms acquire optimal strategies for drug design by experimentation and improvement. Furthermore, underlined in the picture is the use of deep learning techniques, including neural networks, to depict complex, non-linear relationships and improve the prediction accuracy. Together, these machine-learning techniques increase the accuracy and efficiency of identifying potential therapeutic candidates ([1]).

Deep learning in cheminformatics is an advanced method that utilises artificial neural networks with numerous layers to analyse and simulate intricate chemical data ([110]). Deep learning approaches can improve drug discovery by allowing more precise forecasts of drug-target interactions, molecular characteristics, and potential adverse effects ([15, 62, 231]). For example, the chemogenomics neural network (CN) is the formulation of chemogenomics with deep learning. The deep learning CN approach is superior to novel shallow methods ([145]). In addition, a deep-learning model has been designed to predict retrosynthetic pathways ([169]). Also, Feng et al. ([57]) proposed a Deep-Belief Network (DBN) to foresee DTIs, and DBN has 8420-dimensional Protein Sequence Composition (PSC) of target proteins and 6144-dimensional Extended-Connectivity Fingerprints (ECFP) of drugs ([57]). The last example is that Rayhan et al. ([153]) designed FRnet-Encode to distinguish 4096 features. FRnet-Encode is constructed on a deep convolutional neural network ([153]). These accomplished researches indicate that the impact

of the deep-learning algorithm on hit identification will increase over time.

Deep-learning models in cheminformatics have substantial difficulties in accurately identifying targets because of their intricate nature and the constraints in analysing extensive training datasets ([154]). Deep-learning models may encounter problems identifying meaningful patterns within large datasets, resulting in biases in target selection rather than generating new insights ([154]). To overcome these limitations, a potential solution is to develop integrated models that merge ligand and target data to construct complete machine learning frameworks ([154]). Aligned with this approach, our methodology creates a resilient machine-learning model by fusing molecular docking techniques and sophisticated chemogenomic models.

1.6.2 Graph-based Method in Cheminformatics

In cheminformatics, graph-based methods use graph representations to show molecule structures and interactions, offering a flexible means of understanding and predicting chemical properties and behaviour ([154]). These techniques use graph representations to explain molecules using atoms as nodes and bonds as edges. This allows for relational as well as structural elements. This approach is significant in drug development since it can effectively control complex chemical structures and their interactions, surpassing more traditional techniques. Graph theory and algorithms let researchers rapidly examine molecular fingerprints, predict biological activity, and optimise lead compounds ([179]). Therefore, graph-based methods are crucial for the advancement of computational chemistry as well as for the quick identification of new therapeutic compounds (Figure 23.

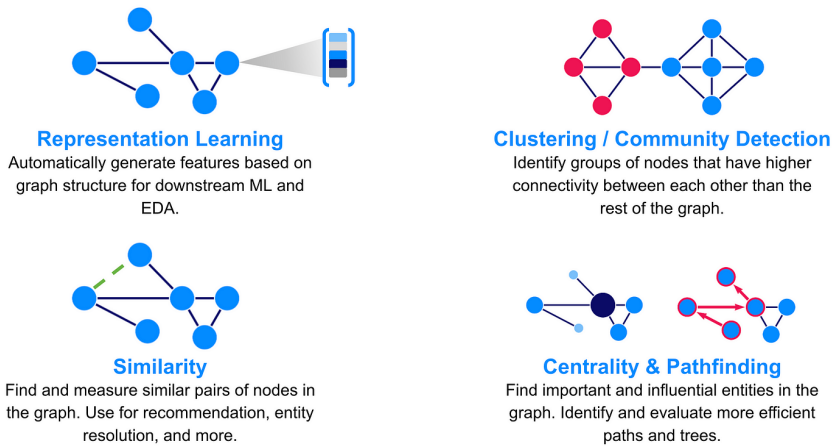

Figure 23: Using graph structures to show complex connections and interactions in data, this figure summarises graph-based machine learning methods.

Representation learning techniques capture intricate traits and patterns, helping nodes and edges in a graph undergo metamorphosis. Comparable measurements evaluate the degree of similarity between nodes or subgraphs, therefore facilitating the identification of objects with comparable architectures. By grouping nodes with like characteristics, clustering methods help to detect communities and trends. Crucially for identifying significant nodes or paths in biological networks, centrality and pathfinding algorithms assess the value of nodes and select the optimal paths within the graph. These graph-based approaches in many disciplines, including drug development and protein interaction studies, help researchers find latent insights and make well-informed decisions jointly ([154]).

Graph-based methods in cheminformatics encompass diverse applications, such as

molecular fingerprinting, molecular similarity assessment, and predictive modelling of biological activities ([110]). These methods leverage graph representations to capture intricate structural details and relational data within molecular structures, offering powerful tools for drug discovery and computational chemistry. Here are examples of graph embedding methods ([64]) based on knowledge graphs that boost DTI prediction performance with the help of ML or DL-based models constructed on low-dimensional feature representation. The graph-based method uses correlations between correlation drug-drug, target-target, and similar matrices, such as DASPfind ([10]). DASPfind orders correlations based on their path scores to determine the top 1%. The DASPfind approach is superior to most network-based models ([10]). Also, DTINET ([116]) uses graph embedding approaches and matrix factorisation to foresee novel DTIs from a heterogeneous graph. DTINET integrates several types of correlation knowledge, such as protein-protein interaction, drug-drug similarity, drug-disease association, drug-drug interactions, protein-protein similarities, drug-side effect associations, and protein-disease association ([26]). The DTINET protocol is used to build a full heterogeneous graph and then learn a low-dimensional feature using matrix factorisation ([116]). These approaches make DTINET outperform others; however, DTINET cannot predict the interaction of new compounds or targets ([116]). Although they still have drawbacks, the example studies indicate that graph-based methods are competitive strategies to identify DTIs.

1.6.3 Network-based Models in Cheminformatics

Network-based cheminformatics models represent molecular structures, interactions, and biological data as networks or graphs by using network science ideas ([147]). Network-based cheminformatics models offer a methodical technique to investigate and grasp complex interactions within biological and chemical systems. Network-based models provide essential insights into network pharmacology, interactions between medications and their targets, and the operation of molecules. Network-based models are significant in fitting several data kinds—including chemical structures, biological pathways, and protein-protein interactions—into a coherent framework ([147, 85]). This integration helps to investigate network properties, identify critical molecular players, and project new therapeutic targets or cooperative drug combinations. Using linked data benefits finding emergent properties, improving knowledge of pharmacological activities at a systems level, and creating logical drug design methods emphasising network-level interactions ([85, 176]). Improving our understanding of complex biological systems and accelerating drug discovery depends critically on network-based models.

Network pharmacology models are still the bottleneck of modern drug discovery, especially target identification ([85, 200]). Network pharmacology is to study the mechanism of a drug candidate at a metabolic level ([71]). It needs network analysis, bioinformatics, and integration of multiple knowledge sources ([140]). Several databases are employed in network-based methods, including Gene Ontology (GO) ([8]) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) ([81]). The databases have information about a drug–target–pathway network, which is essential for network pharmacology. For example, Yamanishi et al. extracted data from KEGG BRITE, BRENDA, SuperTarget, and DrugBank databases ([138]).

Although network-based models have achieved significant breakthroughs, their application is still restricted by the complex intricacies of human metabolism ([83]). The com-

plexity originates from the extensive interconnection of biochemical events and regulatory mechanisms that govern metabolic pathways ([195]). Existing models frequently encounter difficulties in comprehensively capturing the dynamic interactions and metabolic fluxes within this intricate system, which presents obstacles in precisely forecasting drug metabolism, toxicity, and efficacy ([235]). Continuous progress in data integration, modelling approaches, and computational resources is necessary to overcome these restrictions and attain more extensive and dependable forecasts in drug development and personalised medicine.

2 Future direction

Developing state-of-the-art artificial intelligence and machine learning algorithms has the potential to enhance the precision and effectiveness of structural-based drug development. By incorporating these models with detailed protein structures, the accuracy of predicting protein-ligand interactions can be improved, which expedites the discovery of promising pharmaceutical candidates. Moreover, these models aid in predicting alterations in protein structure and their impact on the strength of molecular interactions. Consequently, the current accuracy of computational methods can be improved.

Another future direction is utilising deep learning techniques, such as geometric deep learning ([9]), to analyse complex ligand-binding data and generate prediction models to create novel medications. Deep learning enhances virtual screening by identifying novel ligand binding patterns and improving chemical libraries based on known ligands. This approach may aid in discovering compounds that exhibit reduced off-target effects and enhanced efficacy. However, deep learning can mitigate the interpretability of the model, making understanding how the model works harder ([190]). Nevertheless, deep learning has critical potential to improve the performance of currently available computational methods.

Hybrid methods are promising to enhance the performance of the current method without losing interoperability ([190]). For example, a conventional molecular docking program, Vina ([49]), can be executed to produce ligand poses. Then, an ML model can only order the outputs to improve the overall performance of molecular docking. As a result, such a method improves the performance without losing interpretability.

While computer power and ML techniques are drastically improving, more accurate but computationally intense methods, such as Density Functional Theory simulation ([11]), will quickly provide higher performance and dominate computational drug discovery and development methods.

3 Conclusion

Integrating sophisticated computer techniques has fundamentally changed the terrain of drug discovery and development. From molecular modelling and structure-based approaches to ligand-based strategies and creative de novo design techniques, these computational tools have greatly improved our capacity to find and create new therapeutic medicines. Constant improvement and integration of these techniques promise to propel more discoveries as the area develops.

This review clarifies the fundamental ideas and uses of several computational techniques, giving a whole picture of their contributions to drug development. Future developments have great promise from high-resolution structural data, advanced algorithms, and developing technologies, including artificial intelligence. However, the intricacy and variety of these approaches call for sophisticated knowledge and ongoing adaptation to match the fast developments in the area.

Overcoming obstacles and opening new possibilities will depend critically on developing more accurate predictive models, integrating multi-dimensional biological data, and optimising computational procedures. Staying current with these developments and using the insights offered in this review can help researchers shape the direction of drug discovery and development, therefore hastening the introduction of fresh and potent treatments to meet unmet medical needs.

References

[1] Abhishek and Neeru Jindal. Copy move and splicing forgery detection using deep convolution neural network, and semantic segmentation. *Multimedia Tools and Applications*, 80(3):3571–3599, 2021.

[2] Laeeq Ahmed, Hiba Alogheli, Staffan Arvidsson McShane, Jonathan Alvarsson, Arvid Berg, Anders Larsson, Wesley Schaal, Erwin Laure, and Ola Spjuth. Predicting target profiles with confidence as a service using docking scores. *Journal of Cheminformatics*, 12:1–11, 2020.

[3] Wafa Mohamed Al Madhagi. Importance and application of computational studies in finding new active quinazoline derivatives. In *Recent Advances on Quinazoline*. IntechOpen, 2023.

[4] Hiba Alogheli, Gustav Olanders, Wesley Schaal, Peter Brandt, and Anders Karlén. Docking of macrocycles: comparing rigid and flexible docking in glide. *Journal of chemical information and modeling*, 57(2):190–202, 2017.

[5] Rommie E Amaro. Will the real cryptic pocket please stand out? *Biophysical Journal*, 116(5):753–754, 2019.

[6] Dinler A Antunes, Didier Devaurs, and Lydia E Kavraki. Understanding the challenges of protein flexibility in drug design. *Expert opinion on drug discovery*, 10(12):1301–1313, 2015.

[7] Olayide A Arodola and Mahmoud ES Soliman. Quantum mechanics implementation in drug-design workflows: does it really help? *Drug design, development and therapy*, pages 2551–2564, 2017.

[8] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000.

[9] Kenneth Atz, Francesca Grisoni, and Gisbert Schneider. Geometric deep learning on molecular representations. *Nature Machine Intelligence*, 3(12):1023–1032, 2021.

[10] Wail Ba-Alawi, Othman Soufan, Magbubah Essack, Panos Kalnis, and Vladimir B Bajic. Daspfind: new efficient method to predict drug–target interactions. *Journal of cheminformatics*, 8:1–9, 2016.

[11] Libero J Bartolotti and Ken Flurchick. An introduction to density functional theory. *Reviews in computational chemistry*, pages 187–216, 1996.

[12] Andreas Bender, Josef Scheiber, Meir Glick, John W Davies, Kamal Azzaoui, Jacques Hamon, Laszlo Urban, Steven Whitebread, and Jeremy L Jenkins. Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem: Chemistry Enabling Drug Discovery*, 2(6):861–873, 2007.

[13] Feliks Aleksandrovich Berezin and Mikhail Shubin. *The Schrödinger Equation*, volume 66. Springer Science & Business Media, 2012.

[14] Oliver Buß, Jens Rudat, and Katrin Ochsenreither. Foldx as protein engineering tool: better than random based approaches? *Computational and structural biotechnology journal*, 16:25–33, 2018.

[15] Alexander Button, Daniel Merk, Jan A Hiss, and Gisbert Schneider. Automated de novo molecular design by hybrid machine intelligence and rule-driven chemical synthesis. *Nature machine intelligence*, 1(7):307–315, 2019.

[16] Dong-Sheng Cao, Zhen-Ke Deng, Min-Feng Zhu, Zhi-Jiang Yao, Jie Dong, and Rui-Gang Zhao. Ensemble partial least squares regression for descriptor selection, outlier detection, applicability domain assessment, and ensemble modeling in qsar/qspr modeling. *Journal of Chemometrics*, 31(11):e2922, 2017.

[17] J-M Cardot, A Garcia Arieta, P Paixao, I Tasevska, and B Davit. Implementing the biopharmaceutics classification system in drug development: reconciling similarities, differences, and shared challenges in the ema and us-fda-recommended approaches. *The AAPS journal*, 18:1039–1046, 2016.

[18] Paula Carracedo-Reboredo, Jose Liñares-Blanco, Nereida Rodríguez-Fernández, Francisco Cedrón, Francisco J Novoa, Adrian Carballal, Victor Maojo, Alejandro Pazos, and Carlos Fernandez-Lozano. A review on machine learning approaches and trends in drug discovery. *Computational and structural biotechnology journal*, 19:4538–4558, 2021.

[19] Claudio N Cavasotto, Natalia S Adler, and Maria G Aucar. Quantum chemical approaches in structure-based virtual screening and lead optimization. *Frontiers in chemistry*, 6:188, 2018.

[20] Jean-Pierre Changeux. The concept of allosteric modulation: an overview. *Drug Discovery Today: Technologies*, 10(2):e223–e228, 2013.

[21] Paul S Charifson, Joseph J Corkery, Mark A Murcko, and W Patrick Walters. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *Journal of medicinal chemistry*, 42(25):5100–5109, 1999.

[22] Alexios Chatzigoulas and Zoe Cournia. Rational design of allosteric modulators: Challenges and successes. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 11(6):e1529, 2021.

[23] Fangling Chen, Zhuoya Wang, Chaoyi Wang, Qingliang Xu, Jiazhen Liang, Ximing Xu, Jinbo Yang, Changyun Wang, Tao Jiang, and Rilei Yu. Application of reverse docking for target prediction of marine compounds with anti-tumor activity. *Journal of Molecular Graphics and Modelling*, 77:372–377, 2017.

[24] Hongming Chen, Thierry Kogej, and Ola Engkvist. Cheminformatics in drug discovery, an industrial perspective. *Molecular Informatics*, 37(9-10):1800041, 2018.

[25] Rong Chen, Li Li, and Zhiping Weng. Zdock: an initial-stage protein-docking algorithm. *Proteins: Structure, Function, and Bioinformatics*, 52(1):80–87, 2003.

[26] Ruolan Chen, Xiangrong Liu, Shuting Jin, Jiawei Lin, and Juan Liu. Machine learning for drug-target interaction prediction. *Molecules*, 23(9):2208, 2018.

[27] Yu-Chian Chen. Beware of docking! *Trends in pharmacological sciences*, 36(2):78–95, 2015.

[28] Tammy Man-Kuang Cheng, Tom L Blundell, and Juan Fernandez-Recio. pydock: Electrostatics and desolvation for effective scoring of rigid-body protein–protein docking. *Proteins: Structure, Function, and Bioinformatics*, 68(2):503–515, 2007.

[29] Gaurav Chopra and Ram Samudrala. Exploring polypharmacology in drug discovery and repurposing using the cando platform. *Current pharmaceutical design*, 22(21):3109–3123, 2016.

[30] Arthur Christopoulos. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nature reviews Drug discovery*, 1(3):198–210, 2002.

[31] Maciej Pawel Ciemny, Mateusz Kurcinski, Andrzej Kolinski, and Sebastian Kmiecik. Towards protein-protein docking with significant structural changes using cabs-dock. *arXiv preprint arXiv:1605.09266*, 2016.

[32] Peter Cimermancic, Patrick Weinkam, T Justin Rettenmaier, Leon Bichmann, Daniel A Keedy, Rahel A Woldeyes, Dina Schneidman-Duhovny, Omar N Demerdash, Julie C Mitchell, James A Wells, et al. Cryptosite: expanding the druggable proteome by characterization and prediction of cryptic binding sites. *Journal of molecular biology*, 428(4):709–719, 2016.

[33] Natanya Civjan. *Chemical biology: approaches to drug discovery and development to targeting disease*. John Wiley & Sons, 2012.

[34] Robert A Copeland. *Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists*. John Wiley & Sons, 2013.

[35] Jason B Cross, David C Thompson, Brajesh K Rai, J Christian Baber, Kristi Yi Fan, Yongbo Hu, and Christine Humblet. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *Journal of chemical information and modeling*, 49(6):1455–1474, 2009.

[36] Peter Csermely, Robin Palotai, and Ruth Nussinov. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in biochemical sciences*, 35(10):539–546, 2010.

[37] Sheisi FL da Silva Rocha, Carolina G Olanda, Harold H Fokoue, and Carlos MR Sant'Anna. Virtual screening techniques in drug discovery: review and recent applications. *Current topics in medicinal chemistry*, 19(19):1751–1767, 2019.

[38] Pankaj R Daga, Ronak Y Patel, and Robert J Doerksen. Template-based protein modeling: recent methodological advances. *Current topics in medicinal chemistry*, 10(1):84–94, 2010.

[39] Andrew M Davis, Simon J Teague, and Gerard J Kleywegt. Application and limitations of x-ray crystallographic data in structure-based ligand and drug design. *Angewandte Chemie International Edition*, 42(24):2718–2736, 2003.

[40] Sjoerd J de Vries, Julien Rey, Christina EM Schindler, Martin Zacharias, and Pierre Tuffery. The pepattract web server for blind, large-scale peptide–protein docking. *Nucleic Acids Research*, 45(W1):W361–W364, 2017.

[41] Sjoerd J de Vries, Christina EM Schindler, Isaure Chauvot de Beauchêne, and Martin Zacharias. A web interface for easy flexible protein-protein docking with attract. *Biophysical journal*, 108(3):462–465, 2015.

[42] Gregory J Digby, P Jeffrey Conn, and Craig W Lindsley. Orthosteric-and allosteric-induced ligand-directed trafficking at gpcrs. *Current opinion in drug discovery & development*, 13(5):587, 2010.

[43] David J Diller and Christophe LMJ Verlinde. A critical evaluation of several global optimization algorithms for the purpose of molecular docking. *Journal of computational chemistry*, 20(16):1740–1751, 1999.

[44] Joseph A DiMasi, Henry G Grabowski, and Ronald W Hansen. Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics*, 47:20–33, 2016.

[45] Stefan Doerr, Maciej Majewski, Adrià Pérez, Andreas Kramer, Cecilia Clementi, Frank Noe, Toni Giorgino, and Gianni De Fabritiis. Torchmd: A deep learning framework for molecular simulations. *Journal of chemical theory and computation*, 17(4):2355–2363, 2021.

[46] Ryan JO Dowling, Ivan Topisirovic, Bruno D Fonseca, and Nahum Sonenberg. Dissecting the role of mtor: lessons from mtor inhibitors. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1804(3):433–439, 2010.

[47] Oranit Dror, Alexandra Shulman-Peleg, Ruth Nussinov, and Haim J Wolfson. Predicting molecular interactions in silico: I. a guide to pharmacophore identification and its applications to drug design. *Current medicinal chemistry*, 11(1):71–90, 2004.

[48] Dina Duhovny, Ruth Nussinov, and Haim J Wolfson. Efficient unbound docking of rigid molecules. In *Algorithms in Bioinformatics: Second International Workshop, WABI 2002 Rome, Italy, September 17–21, 2002 Proceedings 2*, pages 185–200. Springer, 2002.

[49] Jerome Eberhardt, Diogo Santos-Martins, Andreas F Tillack, and Stefano Forli. Autodock vina 1.2. 0: New docking methods, expanded force field, and python bindings. *Journal of chemical information and modeling*, 61(8):3891–3898, 2021.

[50] Christiane Ehrt, Tobias Brinkjost, and Oliver Koch. Impact of binding site comparisons on medicinal chemistry and rational molecular design. *Journal of medicinal chemistry*, 59(9):4121–4151, 2016.

[51] David Eisenberg, Edward M Marcotte, Ioannis Xenarios, and Todd O Yeates. Protein function in the post-genomic era. *Nature*, 405(6788):823–826, 2000.

[52] Murtala A Ejalonibu, Ahmed A Elrashedy, Monsurat M Lawal, Mahmoud E Soliman, Sphelele C Sosibo, Hezekiel M Kumalo, and Ndumiso N Mhlongo. Dual targeting approach for mycobacterium tuberculosis drug discovery: Insights from dft calculations and molecular dynamics simulations. *Structural Chemistry*, 31:557–571, 2020.

[53] Murtala A Ejalonibu, Segun A Ogundare, Ahmed A Elrashedy, Morufat A Ejalonibu, Monsurat M Lawal, Ndumiso N Mhlongo, and Hezekiel M Kumalo. Drug discovery for mycobacterium tuberculosis using structure-based computer-aided drug design approach. *International Journal of Molecular Sciences*, 22(24):13259, 2021.

[54] Todd JA Ewing, Shingo Makino, A Geoffrey Skillman, and Irwin D Kuntz. Dock 4.0: search strategies for automated molecular docking of flexible molecule databases. *Journal of computer-aided molecular design*, 15:411–428, 2001.

[55] Thomas Eckart Exner, Oliver Korb, and Tim Ten Brink. New and improved features of the docking software plants. *Chemistry Central Journal*, 3(1):1–1, 2009.

[56] Federico Falchi, Fabiana Caporuscio, and Maurizio Recanatini. Structure-based design of small-molecule protein–protein interaction modulators: the story so far. *Future medicinal chemistry*, 6(3):343–357, 2014.

[57] Qingyuan Feng, Evgenia Dueva, Artem Cherkasov, and Martin Ester. Padme: A deep learning-based framework for drug-target interaction prediction. *arXiv preprint arXiv:1807.09741*, 2018.

[58] Philippe Ferrara, Holger Gohlke, Daniel J Price, Gerhard Klebe, and Charles L Brooks. Assessing scoring functions for protein- ligand interactions. *Journal of medicinal chemistry*, 47(12):3032–3047, 2004.

[59] Jonathan Fine, Janez Konc, Ram Samudrala, and Gaurav Chopra. Candock: Chemical atomic network-based hierarchical flexible docking algorithm using generalized statistical potentials. *Journal of chemical information and modeling*, 60(3):1509–1527, 2020.

[60] Thomas Force and Kyle L Kolaja. Cardiotoxicity of kinase inhibitors: the prediction and translation of preclinical models to clinical outcomes. *Nature reviews Drug discovery*, 10(2):111–126, 2011.

[61] Cen Gao, Jeremy Desaphy, and Michal Vieth. Are induced fit protein conformational changes caused by ligand-binding predictable? a molecular dynamics investigation. *Journal of computational chemistry*, 38(15):1229–1237, 2017.

[62] Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.

[63] Nina M Goodey and Stephen J Benkovic. Allosteric regulation and catalysis emerge via a common route. *Nature chemical biology*, 4(8):474–482, 2008.

[64] Palash Goyal and Emilio Ferrara. Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, 151:78–94, 2018.

[65] Marianne A Grant. Protein structure prediction in structure-based ligand design and virtual screening. *Combinatorial chemistry & high throughput screening*, 12(10):940–960, 2009.

[66] Bartosz A Grzybowski, Alexey V Ishchenko, Jun Shimada, and Eugene I Shakhnovich. From knowledge-based potentials to combinatorial lead design in silico. *Accounts of chemical research*, 35(5):261–269, 2002.

[67] Isabella A Guedes, Felipe SS Pereira, and Laurent E Dardenne. Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges. *Frontiers in pharmacology*, 9:1089, 2018.

[68] Alexis S Hammond, Alice L Rodriguez, Steven D Townsend, Colleen M Niswender, Karen J Gregory, Craig W Lindsley, and P Jeffrey Conn. Discovery of a novel chemical class of mglu5 allosteric ligands with distinct modes of pharmacology. *ACS chemical neuroscience*, 1(10):702–716, 2010.

[69] Markus Hartenfeller and Gisbert Schneider. De novo drug design. *Chemoinformatics and computational chemical biology*, pages 299–323, 2011.

[70] Stefan Henrich, Outi MH Salo-Ahen, Bingding Huang, Friedrich F Rippmann, Gabriele Cruciani, and Rebecca C Wade. Computational approaches to identifying and characterizing protein binding sites for ligand design. *Journal of Molecular Recognition: An Interdisciplinary Journal*, 23(2):209–219, 2010.

[71] Andrew L Hopkins. Network pharmacology: the next paradigm in drug discovery. *Nature chemical biology*, 4(11):682–690, 2008.

[72] Kun-Yi Hsin, Samik Ghosh, and Hiroaki Kitano. Combining machine learning systems and multiple docking simulation packages to improve docking prediction reliability for network pharmacology. *PloS one*, 8(12):e83922, 2013.

[73] Sheng-You Huang, Min Li, Jianxin Wang, and Yi Pan. Hybriddock: a hybrid protein–ligand docking protocol integrating protein-and ligand-based approaches. *Journal of Chemical Information and Modeling*, 56(6):1078–1087, 2016.

[74] Georgios Iakovou. *Simulating molecular docking with haptics.* PhD thesis, University of East Anglia, Norwich, UK, 2015.

[75] Alexey V Ishchenko and Eugene I Shakhnovich. Small molecule growth 2001 (smog2001): An improved knowledge-based scoring function for protein- ligand interactions. *Journal of medicinal chemistry*, 45(13):2770–2780, 2002.

[76] Md Ashraful Islam. Atomlbs: An atom based convolutional neural network for druggable ligand binding site prediction. Master's thesis, The University of Texas Rio Grande Valley, 2022.

[77] Reed B Jacob, Tim Andersen, and Owen M McDougal. Accessible high-throughput virtual screening molecular docking software for students and educators. *PLoS computational biology*, 8(5):e1002499, 2012.

[78] Ursula Jakob, Richard Kriwacki, and Vladimir N Uversky. Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chemical reviews*, 114(13):6779–6805, 2014.

[79] Mohammad Hasan Jamei, Mehdi Khoshneviszadeh, Najmeh Edraki, Maryam Firoozi, Zahra Haghighijoo, Rmin Miri, and Amirhossein Sakhtaman. Cross docking study directed toward virtual screening and molecular docking study of phenanthrene 1, 2, 4-triazine derivatives as novel bcl-2 inhibitors. *Trends in Pharmaceutical Sciences*, 2(4):253–258, 2016.

[80] C John Harris, Richard D Hill, David W Sheppard, Martin J Slater, and Pieter FW Stouten. The design and application of target-focused compound libraries. *Combinatorial chemistry & high throughput screening*, 14(6):521–531, 2011.

[81] Minoru Kanehisa. The kegg database. In *'In silico'simulation of biological processes: Novartis Foundation Symposium 247*, volume 247, pages 91–103. Wiley Online Library, 2002.

[82] Gozde Kar, Ozlem Keskin, Attila Gursoy, and Ruth Nussinov. Allostery and population shift in drug discovery. *Current opinion in pharmacology*, 10(6):715–722, 2010.

[83] Supratik Kar and Jerzy Leszczynski. Recent advances of computational modeling for predicting drug metabolism: a perspective. *Current Drug Metabolism*, 18(12):1106–1122, 2017.

[84] Kristian W Kaufmann and Jens Meiler. Using rosettaligand for small molecule docking into comparative models. *PloS one*, 7(12):e50769, 2012.

[85] Aman Chandra Kaushik, Aamir Mehmood, Dong-Qing Wei, Sadia Nawab, Shakti Sahi, and Ajay Kumar. *Cheminformatics and bioinformatics at the interface with systems biology: bridging chemistry and medicine*, volume 24. Royal Society of Chemistry, 2023.

[86] Terry Kenakin and Arthur Christopoulos. Analytical pharmacology: the impact of numbers on pharmacology. *Trends in pharmacological sciences*, 32(4):189–196, 2011.

[87] Prashant S Kharkar, Sona Warrier, and Ram S Gaud. Reverse docking: a powerful tool for drug repositioning and drug rescue. *Future medicinal chemistry*, 6(3):333–342, 2014.

[88] Samima Khatun, Rinki Bhagat, Sk Abdul Amin, Tarun Jha, and Shovanlal Gayen. Density functional theory (dft) studies in hdac-based chemotherapeutics: Current findings, case studies and future perspectives. *Computers in Biology and Medicine*, page 108468, 2024.

[89] Deok-Soo Kim, Chong-Min Kim, Chung-In Won, Jae-Kwan Kim, Joonghyun Ryu, Youngsong Cho, Changhee Lee, and Jong Bhak. Betadock: shape-priority docking method based on beta-complex. *Journal of Biomolecular Structure and Dynamics*, 29(1):219–242, 2011.

[90] RyangGuk Kim, Rosario I Corona, Bo Hong, and Jun-tao Guo. Benchmarks for flexible and rigid transcription factor-dna docking. *BMC structural biology*, 11:1–10, 2011.

[91] Oliver Korb, Thomas Stutzle, and Thomas E Exner. Empirical scoring functions for advanced protein- ligand docking with plants. *Journal of chemical information and modeling*, 49(1):84–96, 2009.

[92] Bernd Kramer, Matthias Rarey, and Thomas Lengauer. Evaluation of the flexx incremental construction algorithm for protein–ligand docking. *Proteins: Structure, Function, and Bioinformatics*, 37(2):228–241, 1999.

[93] Jacek Kujawski, Hanna Popielarska, Anna Myka, Beata Drabińska, and Marek K Bernard. The log p parameter as a molecular descriptor in the computer-aided drug design–an overview. *Computational Methods in Science and Technology*, 18(2):81–88, 2012.

[94] Mateusz Kurcinski, Michal Jamroz, Maciej Blaszczyk, Andrzej Kolinski, and Sebastian Kmiecik. Cabs-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic acids research*, 43(W1):W419–W424, 2015.

[95] Antonija Kuzmanic, Gregory R Bowman, Jordi Juarez-Jimenez, Julien Michel, and Francesco L Gervasio. Investigating cryptic binding sites by molecular dynamics simulations. *Accounts of chemical research*, 53(3):654–661, 2020.

[96] Margherita Lapillo, Tiziano Tuccinardi, Adriano Martinelli, Marco Macchia, Antonio Giordano, and Giulio Poli. Extensive reliability evaluation of docking-based target-fishing strategies. *International journal of molecular sciences*, 20(5):1023, 2019.

[97] Vy TT Le, Tu HT Nguyen, and Phuc-Chau Do. Global ligand-protein docking tools: Comparation and case study. 2024.

[98] Vincent Le Guilloux, Peter Schmidtke, and Pierre Tuffery. Fpocket: an open source platform for ligand pocket detection. *BMC bioinformatics*, 10:1–11, 2009.

[99] Dong-Dong Li, Xiang-Feng Meng, Qiang Wang, Pan Yu, Lin-Guo Zhao, Zheng-Ping Zhang, Zhen-Zhong Wang, and Wei Xiao. Consensus scoring model for the molecular docking study of mtor kinase inhibitor. *Journal of Molecular Graphics and Modelling*, 79:81–87, 2018.

[100] Jin Li, Ailing Fu, and Le Zhang. An overview of scoring functions used for protein–ligand interactions in molecular docking. *Interdisciplinary Sciences: Computational Life Sciences*, 11:320–328, 2019.

[101] Li Li, Rong Chen, and Zhiping Weng. Rdock: refinement of rigid-body protein docking predictions. *Proteins: Structure, Function, and Bioinformatics*, 53(3):693–707, 2003.

[102] Xiaobai Li, Yingyi Chen, Shaoyong Lu, Zhimin Huang, Xinyi Liu, Qi Wang, Ting Shi, and Jian Zhang. Toward an understanding of the sequence and structural basis of allosteric proteins. *Journal of Molecular Graphics and Modelling*, 40:30–39, 2013.

[103] Yibo Li, Liangren Zhang, and Zhenming Liu. Multi-objective de novo drug design with conditional graph generative model. *Journal of cheminformatics*, 10:1–24, 2018.

[104] Christopher A Lipinski, Franco Lombardo, Beryl W Dominy, and Paul J Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 23(1-3):3–25, 1997.

[105] Jie Liu and Renxiao Wang. Classification of current scoring functions. *Journal of chemical information and modeling*, 55(3):475–482, 2015.

[106] Kai Liu and Hironori Kokubo. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations: a cross-docking study. *Journal of chemical information and modeling*, 57(10):2514–2522, 2017.

[107] Xuewei Liu, Danfeng Shi, Shuangyan Zhou, Hongli Liu, Huanxiang Liu, and Xiaojun Yao. Molecular dynamics simulations and novel drug discovery. *Expert opinion on drug discovery*, 13(1):23–37, 2018.

[108] Yang Liu, Maximilian Grimm, Wen-tao Dai, Mu-chun Hou, Zhi-Xiong Xiao, and Yang Cao. Cb-dock: A web server for cavity detection-guided protein–ligand blind docking. *Acta Pharmacologica Sinica*, 41(1):138–144, 2020.

[109] Yang Liu, Xiaocong Yang, Jianhong Gan, Shuang Chen, Zhi-Xiong Xiao, and Yang Cao. Cb-dock2: Improved protein–ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Research*, 50(W1):W159–W164, 2022.

[110] Yu-Chen Lo, Stefano E Rensi, Wen Torng, and Russ B Altman. Machine learning in chemoinformatics and drug discovery. *Drug discovery today*, 23(8):1538–1546, 2018.

[111] Nir London, Barak Raveh, Eyal Cohen, Guy Fathi, and Ora Schueler-Furman. Rosetta flexpepdock web server—high resolution modeling of peptide–protein interactions. *Nucleic acids research*, 39(suppl_2):W249–W253, 2011.

[112] Shaoyong Lu, Wenkang Huang, and Jian Zhang. Recent computational advances in the identification of allosteric sites in proteins. *Drug discovery today*, 19(10):1595–1600, 2014.

[113] Shaoyong Lu, Shuai Li, and Jian Zhang. Harnessing allostery: a novel approach to drug discovery. *Medicinal research reviews*, 34(6):1242–1285, 2014.

[114] Ying Lu, Sungwon Kim, and Kinam Park. In vitro–in vivo correlation: Perspectives on model development. *International journal of pharmaceutics*, 418(1):142–148, 2011.

[115] R Frederick Ludlow, Marcel L Verdonk, Harpreet K Saini, Ian J Tickle, and Harren Jhoti. Detection of secondary binding sites in proteins using fragment screening. *Proceedings of the National Academy of Sciences*, 112(52):15910–15915, 2015.

[116] Yunan Luo, Xinbin Zhao, Jingtian Zhou, Jinglin Yang, Yanqing Zhang, Wenhua Kuang, Jian Peng, Ligong Chen, and Jianyang Zeng. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature communications*, 8(1):573, 2017.

[117] Buyong Ma, Tal Elkayam, Haim Wolfson, and Ruth Nussinov. Protein–protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proceedings of the National Academy of Sciences*, 100(10):5772–5777, 2003.

[118] Xiaomin Ma, Hu Meng, and Luhua Lai. Motions of allosteric and orthosteric ligand-binding sites in proteins are highly correlated. *Journal of Chemical Information and Modeling*, 56(9):1725–1733, 2016.

[119] Rucha Mahadik, Paul Kiptoo, Tom Tolbert, and Teruna J Siahaan. Immune modulation by antigenic peptides and antigenic peptide conjugates for treatment of multiple sclerosis. *Medical research archives*, 10(5), 2022.

[120] Shingo Makino, Todd JA Ewing, and Irwin D Kuntz. Dream++: flexible docking program for virtual combinatorial libraries. *Journal of computer-aided molecular design*, 13:513–532, 1999.

[121] Ryan J Malonis, Jonathan R Lai, and Olivia Vergnolle. Peptide-based vaccines: current progress and future challenges. *Chemical reviews*, 120(6):3210–3229, 2019.

[122] Dominic D Martinelli. Generative machine learning for de novo drug discovery: A systematic review. *Computers in Biology and Medicine*, 145:105403, 2022.

[123] Karina Martinez-Mayorga, Abraham Madariaga-Mazon, José L Medina-Franco, and Gerald Maggiora. The impact of chemoinformatics on drug discovery in the pharmaceutical industry. *Expert opinion on drug discovery*, 15(3):293–306, 2020.

[124] Gerard Martinez-Rosell, Toni Giorgino, Matt J Harvey, and Gianni de Fabritiis. Drug discovery and molecular dynamics: methods, applications and perspective beyond the second timescale. *Current topics in medicinal chemistry*, 17(23):2617–2625, 2017.

[125] Xuan-Yu Meng, Hong-Xing Zhang, Mihaly Mezei, and Meng Cui. Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design*, 7(2):146–157, 2011.

[126] Madhuchhanda Mohanty and Priti S Mohanty. Molecular docking in organic, inorganic, and hybrid systems: a tutorial review. *Monatshefte für Chemie-Chemical Monthly*, 154(7):683–707, 2023.

[127] Klaus Mohr, Christian Tränkle, Evi Kostenis, Elisabetta Barocelli, Marco De Amici, and Ulrike Holzgrabe. Rational design of dualsteric gpcr ligands: quests and promise. *British journal of pharmacology*, 159(5):997–1008, 2010.

[128] Tobias Morawietz and Nongnuch Artrith. Machine learning-accelerated quantum mechanics-based atomistic simulations for industrial applications. *Journal of Computer-Aided Molecular Design*, 35(4):557–586, 2021.

[129] Hesam N Motlagh, James O Wrabl, Jing Li, and Vincent J Hilser. The ensemble nature of allostery. *Nature*, 508(7496):331–339, 2014.

[130] Varnavas D Mouchlis, Antreas Afantitis, Angela Serra, Michele Fratello, Anastasios G Papadiamantis, Vassilis Aidinis, Iseult Lynch, Dario Greco, and Georgia Melagraki. Advances in de novo drug design: from conventional to machine learning methods. *International journal of molecular sciences*, 22(4):1676, 2021.

[131] Christa E Müller, Anke C Schiedel, and Younis Baqi. Allosteric modulators of rhodopsin-like g protein-coupled receptors: opportunities in drug development. *Pharmacology & therapeutics*, 135(3):292–315, 2012.

[132] Ruth Nussinov and Chung-Jung Tsai. The different ways through which specificity works in orthosteric and allosteric drugs. *Current pharmaceutical design*, 18(9):1311–1316, 2012.

[133] Ruth Nussinov and Chung-Jung Tsai. Allostery in disease and in drug discovery. *Cell*, 153(2):293–305, 2013.

[134] Ruth Nussinov and Chung-Jung Tsai. The design of covalent allosteric drugs. *Annual review of pharmacology and toxicology*, 55(1):249–267, 2015.

[135] Marc Nathan Offman. *Protein structure prediction and refinement*. University of London, University College London (United Kingdom), 2008.

[136] Masahito Ohue, Takehiro Shimoda, Shuji Suzuki, Yuri Matsuzaki, Takashi Ishida, and Yutaka Akiyama. Megadock 4.0: an ultra–high-performance protein–protein docking software for heterogeneous supercomputers. *Bioinformatics*, 30(22):3281–3283, 2014.

[137] Vladimiras Oleinikovas, Giorgio Saladino, Benjamin P Cossins, and Francesco L Gervasio. Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *Journal of the American Chemical Society*, 138(43):14257–14263, 2016.

[138] Hakime Öztürk, Elif Ozkirimli, and Arzucan Özgür. A comparative study of smiles-based compound similarity functions for drug-target interaction prediction. *BMC bioinformatics*, 17:1–11, 2016.

[139] Nataraj S Pagadala, Khajamohiddin Syed, and Jack Tuszynski. Software for molecular docking: a review. *Biophysical reviews*, 9:91–102, 2017.

[140] Musun Park, Sa-Yoon Park, Hae-Jeung Lee, and Chang-Eop Kim. A systems-level analysis of mechanisms of platycodon grandiflorum based on a network pharmacological approach. *Molecules*, 23(11):2841, 2018.

[141] Alessio Peracchi and Andrea Mozzarelli. Exploring and exploiting allostery: Models, evolution, and drug targeting. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1814(8):922–933, 2011.

[142] Yunierkis Perez-Castillo, Stellamaris Sotomayor-Burneo, Karina Jimenes-Vargas, Mario Gonzalez-Rodriguez, Maykel Cruz-Monteagudo, Vinicio Armijos-Jaramillo, M Natália DS Cordeiro, Fernanda Borges, Aminael Sánchez-Rodríguez, and Eduardo Tejera. Compscore: boosting structure-based virtual screening performance by incorporating docking scoring function components into consensus scoring. *Journal of chemical information and modeling*, 59(9):3655–3666, 2019.

[143] Kosmas Alexandros Pervanidis, Giovanni Danilo D'Angelo, Jörn Weisner, Sven Brandherm, and Daniel Rauh. Akt inhibitor advancements: From capivasertib approval to covalent-allosteric promises. *Journal of Medicinal Chemistry*, 67(8):6052–6063, 2024.

[144] Brian G Pierce, Kevin Wiehe, Howook Hwang, Bong-Hyun Kim, Thom Vreven, and Zhiping Weng. Zdock server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics*, 30(12):1771–1773, 2014.

[145] Benoit Playe and Veronique Stoven. Evaluation of deep and shallow learning methods in chemogenomics for the prediction of drugs specificity. *Journal of cheminformatics*, 12(1):11, 2020.

[146] Kathryn A Porter, Israel Desta, Dima Kozakov, and Sandor Vajda. What method to use for protein–protein docking? *Current opinion in structural biology*, 55:1–7, 2019.

[147] Rajani Pydipalli. Network-based approaches in bioinformatics and cheminformatics: Leveraging it for insights. *ABC Journal of Advanced Research*, 7(2):139–150, 2018.

[148] Hojjat Rakhshani, Lhassane Idoumghar, Julien Lepagnot, Mathieu Brévilliers, and Edward Keedwell. Automatic hyperparameter selection in autodock. In *2018 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pages 734–738. IEEE, 2018.

[149] Olof Ramström and Jean-Marie Lehn. Drug discovery by dynamic combinatorial libraries. *Nature Reviews Drug Discovery*, 1(1):26–36, 2002.

[150] L Ramya and N Gautham. Conformational space exploration of met-and leu-enkephalin using the mols method, molecular dynamics, and monte carlo simulation—a comparative study. *Biopolymers*, 97(3):165–176, 2012.

[151] Arjun Rao, Tin M Tunjic, Michael Brunsteiner, Michael Müller, Hosein Fooladi, Chiara Gasbarri, and Noah Weber. Bayesian optimization for ternary complex prediction (botcp). *Artificial Intelligence in the Life Sciences*, 3:100072, 2023.

[152] Matthias Rarey, Bernd Kramer, Thomas Lengauer, and Gerhard Klebe. A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*, 261(3):470–489, 1996.

[153] Farshid Rayhan, Sajid Ahmed, Zaynab Mousavian, Dewan Md Farid, and Swakkhar Shatabda. Frnet-dti: Deep convolutional neural network for drug-target interaction prediction. *Heliyon*, 6(3), 2020.

[154] Daniel Reker, Petra Schneider, Gisbert Schneider, and JB Brown. Active learning for computational chemogenomics. *Future medicinal chemistry*, 9(4):381–402, 2017.

[155] Raquel Rodríguez-Pérez, Filip Miljković, and Jürgen Bajorath. Machine learning in chemoinformatics and medicinal chemistry. *Annual review of biomedical data science*, 5(1):43–65, 2022.

[156] Judith M Rollinger, Hermann Stuppner, and Thierry Langer. Virtual screening for the discovery of bioactive natural products. *Natural compounds as drugs Volume I*, pages 211–249, 2008.

[157] J Rondeau, Gerhard Klebe, and Alberto Podjarny. Ligand binding: the crystallographic approach. *Biophysical approaches determining ligand binding to biomolecular targets: detection, measurement and modelling. modelling*, 1:56–135, 2011.

[158] R Rosenfeld, S Vajda, and C DeLisi. Flexible docking and design. *Annual review of biophysics and biomolecular structure*, 24(1):677–700, 1995.

[159] Christopher D Rosin, R Scott Halliday, William E Hart, and Richard K Belew. A comparison of global and local search methods in drug docking. In *ICGA*, pages 221–229. Citeseer, 1997.

[160] Ashish Runthala and Shibasish Chowdhury. Refined template selection and combination algorithm significantly improves template-based modeling accuracy. *Journal of Bioinformatics and Computational Biology*, 17(02):1950006, 2019.

[161] Kanica Sachdev and Manoj K Gupta. A comprehensive review of computational techniques for the prediction of drug side effects. *Drug Development Research*, 81(6):650–670, 2020.

[162] Adrien Saladin, Julien Rey, Pierre Thévenet, Martin Zacharias, Gautier Moroy, and Pierre Tufféry. Pep-sitefinder: a tool for the blind identification of peptide binding sites on protein surfaces. *Nucleic acids research*, 42(W1):W221–W226, 2014.

[163] Outi MH Salo-Ahen, Ida Alanko, Rajendra Bhadane, Alexandre MJJ Bonvin, Rodrigo Vargas Honorato, Shakhawath Hossain, André H Juffer, Aleksei Kabedev, Maija Lahtela-Kakkonen, Anders Støttrup Larsen, et al. Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes*, 9(1):71, 2020.

[164] Samarth Sandeep, Vaibhav Gupta, and Torin Keenan. Utilizing quantum biological techniques on a quantum processing unit for improved protein binding site determination. *BioRxiv*, pages 2020–03, 2020.

[165] Karina B Santos, Isabella A Guedes, Ana LM Karl, and Laurent E Dardenne. Highly flexible ligand docking: Benchmarking of the dockthor program on the leads-pep protein–peptide data set. *Journal of Chemical Information and Modeling*, 60(2):667–683, 2020.

[166] Diogo Santos-Martins, Stefano Forli, Maria João Ramos, and Arthur J Olson. Autodock4zn: an improved autodock force field for small-molecule docking to zinc metalloproteins. *Journal of chemical information and modeling*, 54(8):2371–2379, 2014.

[167] Nicolas Sauton, David Lagorce, Bruno O Villoutreix, and Maria A Miteva. Ms-dock: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC bioinformatics*, 9:1–12, 2008.

[168] Petra Schneider and Gisbert Schneider. De novo design at the edge of chaos: Miniperspective. *Journal of medicinal chemistry*, 59(9):4077–4086, 2016.

[169] Marwin HS Segler, Mike Preuss, and Mark P Waller. Planning chemical syntheses with deep neural networks and symbolic ai. *Nature*, 555(7698):604–610, 2018.

[170] Lucia Sessa, Luigi Di BIasi, Rosaura Parisi, Simona Concilio, and Stefano Piotto. Receptor flexibility in molecular cross-docking. *PeerJ Preprints*, 4:e2199v1, 2016.

[171] Attila A Seyhan. Lost in translation: the valley of death across preclinical and clinical divide–identification of problems and overcoming obstacles. *Translational Medicine Communications*, 4(1):1–19, 2019.

[172] Bilal Shaker, Myung-Sang Yu, Jingyu Lee, Yongmin Lee, Chanjin Jung, and Dokyun Na. User guide for the discovery of potential drugs via protein structure prediction and ligand docking simulation. *Journal of Microbiology*, 58:235–244, 2020.

[173] Jamal Shamsara. Crossdocker: a tool for performing cross-docking using autodock vina. *SpringerPlus*, 5:1–5, 2016.

[174] Takehiro Shimoda, Takashi Ishida, Shuji Suzuki, Masahito Ohue, and Yutaka Akiyama. Megadock-gpu: acceleration of protein-protein docking calculation on gpus. In *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, pages 883–889, 2013.

[175] Woong-Hee Shin, Lim Heo, Juyong Lee, Junsu Ko, Chaok Seok, and Jooyoung Lee. Ligdockcsa: protein–ligand docking using conformational space annealing. *Journal of computational chemistry*, 32(15):3226–3232, 2011.

[176] Peter K Sorger, Sandra RB Allerheiligen, Darrell R Abernethy, Russ B Altman, Kim LR Brouwer, Andrea Califano, David Z D'Argenio, Ravi Iyengar, William J Jusko, Richard Lalonde, et al. Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms. In *An NIH white paper by the QSP workshop group*, volume 48, pages 1–47. NIH Bethesda Bethesda, 2011.

[177] Cristoph Sotriffer and H Matter. *Virtual screening*. Wiley Online Library, 2011.

[178] Francesca Stanzione, Ilenia Giangreco, and Jason C Cole. Use of molecular docking computational tools in drug discovery. *Progress in medicinal chemistry*, 60:273–343, 2021.

[179] Maciej Staszak, Katarzyna Staszak, Karolina Wieszczycka, Anna Bajek, Krzysztof Roszkowski, and Bartosz Tylkowski. Machine learning in drug design: Use of artificial intelligence to explore the chemical structure–biological activity relationship.

Wiley Interdisciplinary Reviews: Computational Molecular Science, 12(2):e1568, 2022.

[180] Vladimir B Sulimov, Danil C Kutov, and Alexey V Sulimov. Advances in docking. *Current medicinal chemistry*, 26(42):7555–7580, 2019.

[181] Li-Zhen Sun, Yangwei Jiang, Yuanzhe Zhou, and Shi-Jie Chen. Rldock: a new method for predicting rna–ligand interactions. *Journal of chemical theory and computation*, 16(11):7173–7183, 2020.

[182] Andras Szilagyi and Yang Zhang. Template-based structure modeling of protein–protein interactions. *Current opinion in structural biology*, 24:10–23, 2014.

[183] Xuan Tao, Yukun Huang, Chong Wang, Fang Chen, Lingling Yang, Li Ling, Zhenming Che, and Xianggui Chen. Recent developments in molecular docking technology applied in food science: a review. *International Journal of Food Science & Technology*, 55(1):33–45, 2020.

[184] Richard D Taylor, Philip J Jewsbury, and Jonathan W Essex. A review of protein-small molecule docking methods. *Journal of computer-aided molecular design*, 16:151–166, 2002.

[185] Reiji Teramoto and Hiroaki Fukunishi. Supervised consensus scoring for docking and virtual screening. *Journal of chemical information and modeling*, 47(2):526–534, 2007.

[186] Amy Hin Yan Tong, Becky Drees, Giuliano Nardelli, Gary D Bader, Barbara Brannetti, Luisa Castagnoli, Marie Evangelista, Silvia Ferracuti, Bryce Nelson, Serena Paoluzi, et al. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science*, 295(5553):321–324, 2002.

[187] Weida Tong, William J Welsh, Leming Shi, Hong Fang, and Roger Perkins. Structure-activity relationship approaches and applications. *Environmental Toxicology and Chemistry: An International Journal*, 22(8):1680–1695, 2003.

[188] Mieczyslaw Torchala, Iain H Moal, Raphael AG Chaleil, Juan Fernandez-Recio, and Paul A Bates. Swarmdock: a server for flexible protein–protein docking. *Bioinformatics*, 29(6):807–809, 2013.

[189] Oleg Trott and Arthur J Olson. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.

[190] Sadettin Y Ugurlu, David McDonald, Huangshu Lei, Alan M Jones, Shu Li, Henry Y Tong, Mark S Butler, and Shan He. Cobdock: an accurate and practical machine learning-based consensus blind docking method. *Journal of Cheminformatics*, 16(1):5, 2024.

[191] Sandor Vajda, Dmitri Beglov, Amanda E Wakefield, Megan Egbert, and Adrian Whitty. Cryptic binding sites on proteins: definition, detection, and druggability. *Current opinion in chemical biology*, 44:1–8, 2018.

[192] Ilya A Vakser. Protein-protein docking: From interaction to interactome. *Biophysical journal*, 107(8):1785–1793, 2014.

[193] GCP Van Zundert, JPGLM Rodrigues, M Trellet, C Schmitz, PL Kastritis, E Karaca, ASJ Melquiond, Marc van Dijk, SJ De Vries, and AMJJ Bonvin. The haddock2. 2 web server: user-friendly integrative modeling of biomolecular complexes. *Journal of molecular biology*, 428(4):720–725, 2016.

[194] Patrick ML Vanderheyden and Nerdjes Benachour. Influence of the cellular environment on ligand binding kinetics at membrane-bound targets. *Bioorganic & Medicinal Chemistry Letters*, 27(16):3621–3628, 2017.

[195] Goutham N Vemuri and Aristos A Aristidou. Metabolic engineering in the-omics era: elucidating and modulating regulatory networks. *Microbiology and Molecular Biology Reviews*, 69(2):197–216, 2005.

[196] Marcel L Verdonk, Jason C Cole, Michael J Hartshorn, Christopher W Murray, and Richard D Taylor. Improved protein–ligand docking using gold. *Proteins: Structure, Function, and Bioinformatics*, 52(4):609–623, 2003.

[197] Marcel L Verdonk and Wijnand TM Mooij. Knowledge-based methods in structure-based design. In *Computational and Structural Approaches to Drug Discovery*, pages 111–126. 2007.

[198] Jeffrey R Wagner, Christopher T Lee, Jacob D Durrant, Robert D Malmstrom, Victoria A Feher, and Rommie E Amaro. Emerging computational methods for the rational discovery of allosteric drugs. *Chemical reviews*, 116(11):6370–6390, 2016.

[199] W Patrick Walters, Matthew T Stahl, and Mark A Murcko. Virtual screening—an overview. *Drug discovery today*, 3(4):160–178, 1998.

[200] Cheng Wang, Wenyan Wang, Kun Lu, Jun Zhang, Peng Chen, and Bing Wang. Predicting drug-target interactions with electrotopological state fingerprints and amphiphilic pseudo amino acid composition. *International Journal of Molecular Sciences*, 21(16):5694, 2020.

[201] Lirong Wang, Chao Ma, Peter Wipf, Haibin Liu, Weiwei Su, and Xiang-Qun Xie. Targethunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. *The AAPS journal*, 15:395–406, 2013.

[202] Qi Wang, Mingyue Zheng, Zhimin Huang, Xinyi Liu, Huchen Zhou, Yingyi Chen, Ting Shi, and Jian Zhang. Toward understanding the molecular basis for chemical allosteric modulator design. *Journal of Molecular Graphics and Modelling*, 38:324–333, 2012.

[203] Renxiao Wang, Yipin Lu, and Shaomeng Wang. Comparative evaluation of 11 scoring functions for molecular docking. *Journal of medicinal chemistry*, 46(12):2287–2303, 2003.

[204] Michael D Ward. *Combining Computer Simulations and Deep Learning to Understand and Predict Protein Structural Dynamics*. PhD thesis, Washington University in St. Louis, 2022.

[205] Andrew Waterhouse, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T Heer, Tjaart A P de Beer, Christine Rempfer, Lorenza Bordoli, et al. Swiss-model: homology modelling of protein structures and complexes. *Nucleic acids research*, 46(W1):W296–W303, 2018.

[206] Benjamin Webb and Andrej Sali. Comparative protein structure modeling using modeller. *Current protocols in bioinformatics*, 54(1):5–6, 2016.

[207] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.

[208] David Weininger, Arthur Weininger, and Joseph L Weininger. Smiles. 2. algorithm for generation of unique smiles notation. *Journal of chemical information and computer sciences*, 29(2):97–101, 1989.

[209] Cody J Wenthur, Patrick R Gentry, Thomas P Mathews, and Craig W Lindsley. Drugs for allosteric sites on receptors. *Annual review of pharmacology and toxicology*, 54(1):165–184, 2014.

[210] Michael R Wood, Corey R Hopkins, John T Brogan, P Jeffrey Conn, and Craig W Lindsley. "molecular switches" on mglur allosteric ligands that modulate modes of pharmacology. *Biochemistry*, 50(13):2403–2410, 2011.

[211] Qi Wu, Zhenling Peng, Yang Zhang, and Jianyi Yang. Coach-d: improved protein–ligand binding sites prediction with refined ligand-binding poses through molecular docking. *Nucleic acids research*, 46(W1):W438–W442, 2018.

[212] Arthur Wuster and M Madan Babu. Chemogenomics and biotechnology. *Trends in biotechnology*, 26(5):252–258, 2008.

[213] Lei Xie, Li Xie, and Philip E Bourne. Structure-based systems biology for analyzing off-target binding. *Current opinion in structural biology*, 21(2):189–199, 2011.

[214] Xianjin Xu, Marshal Huang, and Xiaoqin Zou. Docking-based inverse virtual screening: methods, applications, and challenges. *Biophysics reports*, 4:1–16, 2018.

[215] Yumeng Yan, Huanyu Tao, Jiahua He, and Sheng-You Huang. The hdock server for integrated protein–protein docking. *Nature protocols*, 15(5):1829–1852, 2020.

[216] Yumeng Yan, Zeyu Wen, Xinxiang Wang, and Sheng-You Huang. Addressing recent docking challenges: A hybrid strategy to integrate template-based and free protein-protein docking. *Proteins: Structure, Function, and Bioinformatics*, 85(3):497–512, 2017.

[217] Jae-Seong Yang, Sang Woo Seo, Sungho Jang, Gyoo Yeol Jung, and Sanguk Kim. Rational engineering of enzyme allosteric regulation through sequence evolution analysis. *PLoS computational biology*, 8(7):e1002612, 2012.

[218] Jianyi Yang, Ambrish Roy, and Yang Zhang. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, 29(20):2588–2595, 2013.

[219] Jinsol Yang, Minkyung Baek, and Chaok Seok. Galaxydock3: Protein–ligand docking that considers the full ligand conformational flexibility. *Journal of Computational Chemistry*, 40(31):2739–2748, 2019.

[220] Su-Qing Yang, Qing Ye, Jun-Jie Ding, Ming-Zhu Yin, Ai-Ping Lu, Xiang Chen, Ting-Jun Hou, and Dong-Sheng Cao. Current advances in ligand-based target prediction. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 11(3):e1504, 2021.

[221] Zhi-Jiang Yao, Jie Dong, Yu-Jing Che, Min-Feng Zhu, Ming Wen, Ning-Ning Wang, Shan Wang, Ai-Ping Lu, and Dong-Sheng Cao. Targetnet: a web service for predicting potential drug–target interaction profiling via multi-target sar models. *Journal of computer-aided molecular design*, 30:413–424, 2016.

[222] Wen-Ling Ye, Chao Shen, Guo-Li Xiong, Jun-Jie Ding, Ai-Ping Lu, Ting-Jun Hou, and Dong-Sheng Cao. Improving docking-based virtual screening ability by integrating multiple energy auxiliary terms from molecular docking scoring. *Journal of Chemical Information and Modeling*, 60(9):4216–4230, 2020.

[223] Shuangye Yin, Lada Biedermannova, Jiri Vondrasek, and Nikolay V Dokholyan. Medusascore: an accurate force field-based scoring function for virtual drug screening. *Journal of chemical information and modeling*, 48(8):1656–1662, 2008.

[224] Calvin K Yip, Kazuyoshi Murata, Thomas Walz, David M Sabatini, and Seong A Kang. Structure of the human mtor complex i and its implications for rapamycin inhibition. *Molecular cell*, 38(5):768–774, 2010.

[225] Hua Yu, Jianxin Chen, Xue Xu, Yan Li, Huihui Zhao, Yupeng Fang, Xiuxiu Li, Wei Zhou, Wei Wang, and Yonghua Wang. A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data. *PloS one*, 7(5):e37608, 2012.

[226] Yaxia Yuan, Jianfeng Pei, and Luhua Lai. Ligbuilder v3: a multi-target de novo drug design approach. *Frontiers in chemistry*, 8:142, 2020.

[227] Jianming Zhang, Francisco J Adrián, Wolfgang Jahnke, Sandra W Cowan-Jacob, Allen G Li, Roxana E Iacob, Taebo Sim, John Powers, Christine Dierks, Fangxian Sun, et al. Targeting bcr–abl by combining allosteric with atp-binding-site inhibitors. *Nature*, 463(7280):501–506, 2010.

[228] Jing Zhang, Huajun Li, Yubo Zhang, Chaoran Zhao, Yizi Zhu, and Mei Han. Uncovering the pharmacological mechanism of stemazole in the treatment of neurodegenerative diseases based on a network pharmacology approach. *International journal of molecular sciences*, 21(2):427, 2020.

[229] Mingzhen Zhang, Jun Zhao, and Jie Zheng. Molecular understanding of a potential functional link between antimicrobial and amyloid peptides. *Soft Matter*, 10(38):7425–7451, 2014.

[230] Jingtian Zhao, Yang Cao, and Le Zhang. Exploring the computational methods for protein-ligand binding site prediction. *Computational and structural biotechnology journal*, 18:417–426, 2020.

[231] Alex Zhavoronkov, Yan A Ivanenkov, Alex Aliper, Mark S Veselov, Vladimir A Aladinskiy, Anastasiya V Aladinskaya, Victor A Terentiev, Daniil A Polykovskiy, Maksim D Kuznetsov, Arip Asadulaev, et al. Deep learning enables rapid identification of potent ddr1 kinase inhibitors. *Nature biotechnology*, 37(9):1038–1040, 2019.

[232] Shuangjia Zheng, Xin Yan, Yuedong Yang, and Jun Xu. Identifying structure–property relationships through smiles syntax analysis with self-attention mechanism. *Journal of chemical information and modeling*, 59(2):914–923, 2019.

[233] Wenjun Zheng. Predicting cryptic ligand binding sites based on normal modes guided conformational sampling. *Proteins: Structure, Function, and Bioinformatics*, 89(4):416–426, 2021.

[234] Pei Zhou, Bowen Jin, Hao Li, and Sheng-You Huang. Hpepdock: a web server for blind peptide–protein docking based on a hierarchical algorithm. *Nucleic acids research*, 46(W1):W443–W450, 2018.

[235] Wei Zhou, Yonghua Wang, Aiping Lu, and Ge Zhang. Systems pharmacology in small molecular drug discovery. *International journal of molecular sciences*, 17(2):246, 2016.

[236] Jintao Zhu, Zhonghui Gu, Jianfeng Pei, and Luhua Lai. Diffbind: A se (3) equivariant network for accurate full-atom semi-flexible protein-ligand docking. *arXiv preprint arXiv:2311.15201*, 2023.