Prioritising Candidate Drugs via Virtual Screening, Molecular Docking, and MD Simulations

Dheeraj Kumar Chaurasia[1,2], Raushan Anjum[2], Ashutosh Shandilya[2], Ashok Kumar Patel[2,3] and B. Jayaram[2,3,4,*]

[1] School of Interdisciplinary Research, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, 110016, India

[2] Supercomputing Facility for Bioinformatics and Computational Biology, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, 110016, India

[3] Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, 110016, India

[4] Department of Chemistry, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, 110016, India

*Author for correspondence

Dheeraj Kumar Chaurasia

Dheeraj Kumar Chaurasia holds a B.Sc. and M.Sc. from the University of Allahabad and has over five years of experience in Computer-Aided Drug Design (CADD) and Molecular Dynamics (MD) simulation-based drug discovery. Dheeraj Kumar Chaurasia works as a Project Scientist at SCFBio, IIT Delhi, and a Ph.D. scholar at the School of Interdisciplinary Sciences, IIT Delhi.

Email: dheeraj110594@gmail.com
Website: https://web.iitd.ac.in/~srz238152/

Raushan Anjum

Raushan Anjum earned her B.Sc. degree in Physics, Chemistry, and Mathematics from I.T. College, Lucknow, India, and M.Sc. degree in Chemistry from IIT Delhi, India. At IIT Delhi, she conducted significant research in drug discovery and computational biology under the guidance of Prof. B. Jayaram. She is currently a Ph.D. student in Chemistry at the University of Arizona, Tucson, USA.

Email: anjum.official2001@gmail.com

Dr. Ashutosh Shandilya

Dr. Ashutosh Shandilya is an expert in Computational Chemistry and Structural Biology, with over 14 years of research experience and 25+ publications with 1000 citations. He earned his Ph.D. from IIT Delhi, focusing on protein-ligand interactions and drug design against Malaria and Alzheimer's. He has held postdoctoral positions and teaching roles, contributing significantly to molecular dynamics simulations and collaborative research projects.

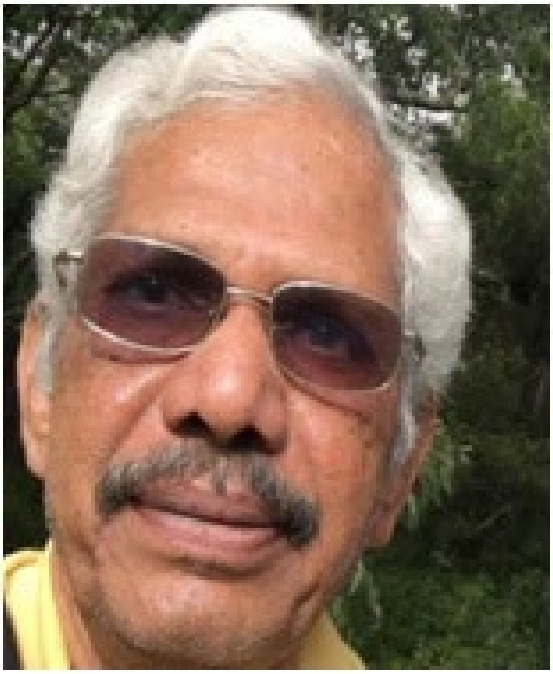
Email: izerokelvin@gmail.com

Prof. Ashok Kumar Patel

Dr. Ashok Kumar Patel, an Associate Professor at Kusuma School of Biological Sciences, IIT Delhi, holds a B.Sc. and M.Sc. in Physics from BHU, Varanasi. His Ph.D. research under Prof. J.V. Medicherla focused on protein structural and biophysical studies. Trained by renowned scientists in the UK, Netherlands, and Finland, he conducted postdoctoral research at Johns Hopkins University. His current research includes studies on viruses like Chikungunya, Dengue, SARS-CoV-2, and cancer.

Email: ashokpatel@bioschool.iitd.ac.in
Website: https://web.iitd.ac.in/~ashokpatel/

Dr. B. Jayaram



Dr. B. Jayaram, a retired professor from IIT Delhi's Chemistry Department, earned his Ph.D. in Chemistry from the City University of New York in 1986 under Prof. David L. Beveridge. He conducted postdoctoral research with Prof. Barry Honig at Columbia University. At IIT Delhi, he established the Supercomputing Facility for Bioinformatics & Computational Biology (SCFBio), advancing in silico solutions through the Gene to Drug suite of software.

Email: bjayaram@chemistry.iitd.ac.in
Website: http://www.scfbio-iitd.res.in/scfbiogroup/biocomputinggroup.htm

Contents

Abbreviation and Symbols

2D: 2-Dimensional
$3CL^{pro}$: 3-chymotrypsin-like protease
3D: 3-Dimensional
AI: Artificial Intelligence
BIMP: Bioactivity of Indian Medicinal Plants
CADD: Computer-Aided Drug Discovery
CG: Conjugate Gradient
Cry-EM: Cryo-Electron Microscopy
DNA: Deoxyribonucleic Acid
EMEA: European Medicines Evaluation Agency
EPR: Electron Paramagnetic Resonance
ER: Estrogen Receptor
ESI: Electrospray Ionization
FBDD: Fragment-Based Drug Design
FDA: Food and Drug Administration
FEP: Free Energy Perturbation
FRET: Förster Resonance Energy Transfer
H-Bond: Hydrogen Bond
HER2: Human Epidermal Growth Factor Receptor 2
HPC: High-Performance Computing
HTS: High-Throughput Screening
LBVS: Ligand-Based Virtual Screening
LC-MS: Liquid Chromatography-Mass Spectrometry
MD: Molecular Dynamics
ML: Machine Learning
MM/GBSA: Molecular Mechanics/Generalized Born Surface Area
MM/PBSA: Molecular Mechanics/Poisson-Boltzmann Surface Area
M^{pro}: Main Protease
MR: Molar Refractivity
NF-κB: Nuclear Factor-kappa B
NMR: Nuclear Magnetic Resonance
NPT: Isothermal-Isobaric Ensemble
NVT: Canonical Ensemble
PBS: Phosphate Buffered Saline
PDB: Protein Data Bank
PME: Particle Mesh Ewald
PR: Progesterone Receptor
QSAR: Quantitative Structure-Activity Relationship
RMSD: Root Mean Square Deviation
RMSF: Root Mean Square Fluctuation
RoG: Radius of Gyration
SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2
SB-CADD: Structure Based-Computer Aided Drug Discovery
SBVS: Structure-Based Virtual Screening

SD: Steepest Descent
SPC/E: Extended Simple Point Charge Model
T. chebula: Terminalia chebula
TI: Thermodynamic Integration
TIP3P: Transferable Intermolecular Potential 3P
TNBC: Triple Negative Breast Cancer
WHO: World Health Organization
°C: Celcius
Å: Angstrom
C: Carbon
g: gram
h: hour
K: Kelvin
kcal: kilo calories
kV: kilo-volt
L: liter
m: mass
mL: milliliter
mm: millimeter
mM: millimolar
mol: mole
N: nitrogen
ns: nanosecond
ps: picosecond
V: volt
z: charge
αC: alpha-helix C
β: beta
κ: kappa
µg: microgram
µm: micro-meter

Abstract
Prioritization of drug candidates is a crucial stage in the drug development process, requiring effective and economical approaches to find potential molecules. This chapter focuses on the transformative role of Computer-Aided Drug Discovery (CADD) techniques, specifically virtual screening, molecular docking, and molecular dynamics (MD) simulations, in streamlining this process. Computational methods are utilized by CADD to predict the interactions between drug candidates and biological targets, thereby improving the selection and optimization of potential therapeutics. Virtual screening enables the rapid evaluation of vast compound libraries, identifying molecules with high binding affinity and specificity. Molecular docking provides detailed insights into the preferred orientation and binding modes of these molecules within target proteins, facilitating rational drug design. MD simulations offer a dynamic perspective on protein-ligand interactions, revealing the stability and conformational changes of biomolecular complexes over time. By using these methods, scientists can effectively traverse the process of discovering new drugs, enhancing the likelihood of success for potential compounds and decreasing the expenses associated with their development. This chapter emphasizes the significance of computational tools for modern drug discovery, demonstrating their influence through specific examples and case studies.

1. Introduction

1.1. Overview and Importance of Computer-Aided Drug Discovery

The process of identifying and developing new drugs is intricate, expensive, and time-consuming. Typically, the process of developing a new drug and getting it approved for sale in the market spans over a decade and requires substantial financial investments, often exceeding billions of dollars. The traditional method [1-3] entails conducting extensive laboratory studies and numerous stages of clinical trials, which frequently experience significant attrition rates. Many potential drug candidates fail during these trials due to their lack of effectiveness or safety issues. In response to these formidable challenges, Computer-Aided Drug Discovery (CADD) [4-6] has emerged as a revolutionary method that leverages computational technologies to improve and simplify the process of discovering new drugs. Through the application of computational models and simulations, researchers are able to predict the interactions between drug candidates and biological targets, thereby enhancing the effectiveness of drugs prior to their synthesis and laboratory testing.

The significance of CADD lies in its ability to transform the early stages of drug discovery [7], where choices made can have profound implications for subsequent development phases. Early identification of promising compounds, facilitated by CADD, can streamline the pipeline, focusing resources on the most viable candidates and thereby enhancing overall productivity. This capability is especially vital at a time when pharmaceutical companies are under increasing pressure to quickly and cost-efficiently introduce novel and efficient treatments to the market. Another advantage of CADD is its potential to significantly reduce the cost and duration of drug development. Drug discovery is traditionally a resource-intensive process, requiring the synthesis and biological evaluation of thousands of compounds to identify a handful of promising candidates. This trial-and-error approach is both costly and time-consuming. CADD addresses these challenges by enabling virtual screening of vast libraries of compounds, allowing researchers to quickly identify those with the highest likelihood of success. It also provides a detailed molecular insight into the interactions between drug candidates and their targets. These insights facilitate the design of molecules with better

efficacy and pharmacokinetic properties, thereby reducing the chances of failure in subsequent phases of development.

1.2. History

Drug discovery can be divided into three key periods. In the nineteenth century, the initial stage of medicinal chemistry mainly relied on the fortuitous discoveries made by chemists. The second stage, commencing in the early twentieth century, was marked by the discovery of new drug structures, leading to significant advancements in antibiotic discovery. During this time, significant advancements were made in several techniques [8-9], including molecular modeling, combinatorial chemistry, and automated high-throughput screening. Additionally, the emergence of recombinant DNA technology allowed for the identification of prospective therapeutic targets. The third period, in the twenty-first century, has been driven by the "Omics" revolution, resulting in a substantial increase in the approval of biopharmaceutical drugs by the FDA and EMEA for therapeutic use.

Thus, the history of CADD goes back to the late 20th century when advancements in computational chemistry and structural biology laid the groundwork for a paradigm shift in drug discovery. The early stages of CADD were marked by the development of molecular modeling techniques, which allowed researchers to visualize and analyze the three-dimensional structures of biological molecules. Key milestones include the pioneering work of Pauling and Corey in the 1950s [10], who proposed the concept of molecular modeling and elucidated the structures of biomolecules. In the 1970s and 1980s, the advent of computational methods such as Molecular Dynamics (MD) simulations [11] and quantum mechanics calculations expanded the scope of CADD, enabling researchers to simulate the behavior of molecules at the atomic level. Concurrently, the emergence of high-performance computing systems accelerated the pace of drug discovery, facilitating the screening of large chemical libraries and the optimization of lead compounds.

The 1990s witnessed significant breakthroughs in structural biology, with the rapid development of X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy techniques. These advancements provided researchers with unprecedented insights into the three-dimensional structures of biological macromolecules, paving the way for structure-based drug design approaches. In the early 21st century, the integration of computational tools with structural biology techniques gave rise to modern CADD methodologies [12]. Molecular docking algorithms [13] emerged as powerful tools for predicting the binding affinity and orientation of small molecule ligands within target proteins, while virtual screening techniques [14] enabled the rapid identification of potential drug candidates from large chemical databases.

1.3. Recent advancements in CADD

In the dynamic landscape of drug discovery, recent years have witnessed remarkable advancements in the realm of CADD [15]. This field, leveraging computational power, machine learning, and deep insights into biological systems, has undergone transformative developments, propelling drug discovery into new frontiers. One of the most significant advancements in CADD stems from the integration of Machine Learning (ML) and Artificial Intelligence (AI) techniques [16]. ML and AI algorithms have revolutionized drug discovery, enabling the development of predictive models for various aspects of drug discovery [17]. These models leverage large datasets of chemical structures and biological activities to predict compound properties. Such algorithms have expedited virtual

screening processes [18], aiding in the identification of potential drug candidates with unprecedented speed and accuracy.

Furthermore, recent years have witnessed a growing emphasis on structure-based drug repurposing [19]. Structure-based CADD techniques, which utilize known protein structures and computational docking methods [20], are increasingly being employed to identify new therapeutic uses for existing drugs. This approach accelerates the drug development process by repurposing approved drugs for new indications [21], thereby reducing costs and timelines associated with traditional drug discovery efforts. This approach also enables researchers to design novel drugs with a deeper understanding of the three-dimensional structure of target proteins, thereby optimizing drug-target interactions.

In parallel, Fragment-Based Drug Design (FBDD) [22] has emerged as a powerful strategy in CADD. By screening libraries of small molecules or fragments, FBDD identifies molecules that bind to target proteins, laying the groundwork for subsequent drug optimization. High-throughput fragment screening techniques combined with SB-CADD approaches enable a rapid identification of fragment hits, which can then be optimized into lead compounds using structure-based design principles. This iterative process has proven to be highly effective in generating novel drug candidates with improved potency and selectivity. Virtual screening, a cornerstone of CADD, has also experienced significant enhancements. Improved scoring functions, innovative ligand-based [23] and structure-based screening methodologies [24], and the integration of machine learning have revolutionized virtual screening processes. These advancements enable researchers to sift through vast compound libraries with unprecedented efficiency, identifying lead compounds for further experimental validation.

Additionally, advancements in cloud computing and high-performance computing (HPC) [25] have democratized CADD access to computational resources. Researchers can now perform large-scale virtual screenings, molecular docking, MD simulations, and other computationally intensive tasks more efficiently and cost-effectively. This has accelerated the pace of drug discovery and enabled researchers to explore a wider range of chemical space in search of novel therapeutics. Recent advancements in CADD have revolutionized the drug discovery process, leading to the identification of novel drug candidates with improved potency, selectivity, and safety profiles. By leveraging ML, AI, structural biology, and computational resources, researchers are poised to continue making significant strides in drug discovery, ultimately improving patient outcomes and addressing unmet medical needs.

1.4. Advantages over Traditional Methods

Traditional high-throughput screening (HTS) [26-27] methods involve testing large numbers of compounds against a biological target to identify active compounds. HTS can be labor-intensive and expensive, often requiring sophisticated equipment and significant amounts of biological material. In contrast, virtual screening can evaluate millions of compounds *in silico* in a fraction of the time and at a fraction of the cost. By prioritizing the most promising candidates for synthesis and experimental testing, CADD can significantly reduce the number of compounds that need to be physically screened.

CADD allows for the exploration of vast chemical spaces, identifying novel compounds that might be overlooked by traditional methods. The chemical space refers to the theoretical space of all possible chemical compounds, which is enormous and largely unexplored. Traditional drug discovery methods are limited by the number of compounds that can be physically synthesized and tested,

whereas CADD can virtually explore millions or even billions of compounds. This capability is particularly important for identifying novel chemical scaffolds [28-29] and designing new molecules with unique properties. By exploring diverse chemical spaces, CADD can identify compounds with improved efficacy, safety, and pharmacokinetic profiles, as well as those that can overcome resistance mechanisms. This approach not only enhances the discovery of new drugs but also provides opportunities for developing innovative therapies for challenging diseases.

For example, in the field of antimicrobial drug discovery, the emergence of antimicrobial-resistant [30] bacteria is a significant public health concern [31-35]. Traditional methods have struggled to keep pace with the rapid evolution of resistance mechanisms. CADD can accelerate the discovery of new antibiotics [36-37] by exploring vast chemical spaces and identifying compounds that target novel bacterial pathways. By expanding the pool of potential drug candidates, CADD can help address the urgent need for new and effective antimicrobial therapies. The impact of CADD on modern drug discovery is profound and multifaceted. By leveraging computational technologies, CADD offers a more efficient, cost-effective, and precise approach to discovering new drugs. It enhances the efficiency of the initial stages of drug discovery, increasing the likelihood of success and reducing the risk of failure in clinical trials [38].

1.5. Objectives of the Chapter

This chapter aims to focus on prioritizing candidate drugs via virtual screening [39-41], molecular docking [42-45], and MD simulations [46-49], which have also been incorporated in the Sanjeevini [50-52] module (http://www.scfbio-iitd.res.in/Sanjeevini/index.php) of SCFBio-CADD pipeline. These methods play crucial roles in expediting the identification and optimization of potential therapeutics by utilizing computational approaches. Virtual screening, in the early stages of drug discovery, emerges as a powerful tool that offers a systematic and efficient means of sifting through vast compound libraries to identify molecules with desired biological activity. The principles of virtual screening, encompassing both structure-based methods, which rely on the three-dimensional structures of target proteins, and ligand-based approaches, which utilize known active compounds as references, will be explored. By examining the algorithms and strategies employed in virtual screening, the utility of this technique in identifying lead compounds with high binding affinity and specificity will be elucidated.

Following the virtual screening, the discussion extends to molecular docking, a linchpin in structure-based drug design. Molecular docking algorithms predict the preferred orientation and binding mode of small molecules within the binding site of target proteins, aiding in the rational design of potential drug candidates. By providing insights into the principles and applications of molecular docking, the utility in predicting ligand-protein interactions and facilitating the identification of promising drug candidates will be elaborated.

Subsequently, the exploration delves into MD simulations, offering a dynamic perspective on protein-ligand interactions. MD simulations simulate the motions of atoms and molecules over a period of time, offering valuable insights into the dynamic behavior and flexibility of biomolecular complexes. The principles and methodologies of MD simulations will be examined to elucidate how these simulations enhance our understanding of protein-ligand binding kinetics, conformational changes, and stability. Quantifying the thermodynamic stability of biomolecular complexes is of paramount importance in understanding ligand binding kinetics and predicting binding affinities accurately. Free

energy calculation methods [53-56], such as Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) [57] and Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) [58], offer valuable insights into the energetics of ligand binding and contribute to the rational design of potent and selective drug candidates. This subsection delves into the theoretical foundations of free energy calculations, exploring the partitioning of free energy into enthalpic and entropic contributions, focusing in particular on developing these quantities based on MD trajectories.

Finally, this chapter will underscore the importance of integrating various CADD techniques, including virtual screening, molecular docking, MD simulations, and free energy analysis, within the drug discovery pipeline. By synergistically combining computational methods, researchers can navigate the drug discovery pipeline with enhanced efficiency and efficacy, from lead identification and optimization to preclinical and clinical development. This section highlights the importance of an integrated approach to CADD, emphasizing the complementary nature of different techniques and their collective impact on accelerating drug discovery timelines and reducing development costs. We illustrate the transformative potential of integrated CADD approaches in modern drug discovery [59] through real-world case studies and examples.

2. Methods in Drug Discovery

Drug discovery is a complex and multi-disciplinary process that involves identifying new candidate medications based on the knowledge of a biological target. Modern drug discovery utilizes various computational and experimental techniques to streamline this process. Among the most critical methods in contemporary drug discovery are Virtual Screening, Molecular Docking, MD Simulations, and Free Energy Analysis. These methods leverage computational power and advanced algorithms to predict and analyze the interactions between potential drug molecules and biological targets. A thorough explanation of these methods, their underlying principles, operational procedures, practical applications, and challenges encountered are presented here.

2.1 Virtual Screening

Virtual Screening is a computational method employed in the field of drug development to detect promising bioactive molecules from a vast collection of chemical structures. The system utilizes computational power to assess and rank chemicals according to their estimated binding capacity to a certain protein. Virtual screening is an essential part of structure-based drug design techniques. It involves using computational models to evaluate how candidate ligands bind to a target protein [60-61]. Usually, the screening process involves placing the ligands at various positions or poses within the three-dimensional structure of the target protein. A scoring function is assessed at each position to estimate the binding-free energy. This information is then used to rank the binding poses and various candidate ligands based on their ability to bind to the target protein. Although suitable docking poses are commonly created, scoring algorithms generally lack the precision required to accurately score poses or ligands [62]. As a result, docking approaches are often enhanced by utilizing MD simulations to calculate more precise binding affinities. Virtual screening of vast quantities of compounds against target protein binding sites has become an essential element of drug discovery procedures. Frequently, this examination is conducted by computationally placing ligands into a specific protein binding site. However, this method has the disadvantage of requiring the evaluation of numerous positions in order to acquire precise estimations of the strength of the bond between the

protein and ligand. There are primarily two types of virtual screening: ligand-based and structure-based.

Ligand-Based Virtual Screening (LBVS): This approach utilizes data on established active chemicals to discover novel ones. Commonly used techniques include quantitative structure-activity relationship (QSAR) models, pharmacophore modeling, and similarity searching [63,64]. The LBVS methodology operates under the assumption that compounds possessing analogous structures will have comparable biological activity.

Structure-Based Virtual Screening (SBVS): This approach is based on the three-dimensional structure of the protein of interest, which is commonly established using X-ray crystallography, NMR spectroscopy, or Cryo-EM more recently. SBVS [65] entails the process of docking a collection of chemicals into the specific binding site of the target protein and making predictions about their binding affinities. Methods such as molecular docking (explained in the sequel) and high-throughput docking are essential components of structure-based virtual screening [66].

2.1.1. Workflow of Virtual Screening

Preparation of Compound Library: A wide range of chemical structures is carefully selected from databases such as BIMP, ZINC [67, 68], PubChem [69], or ChEMBL [70]. These molecules undergo processing to produce 3D conformations with optimized geometries to facilitate the estimation of non-covalent interactions, as well as physicochemical molecular descriptors. The descriptors typically include properties such as molecular weight, the number of hydrogen bond donors and acceptors, the approximate octanol-water partition coefficient log P (logP) [71,72], the molar refractivity (MR) [72], and the Wiener topology index (W) for each ligand. These are precomputed and a library of molecules ready for screening is created.

Target Preparation: The desired protein structure is acquired and processed by delineating the binding site, including hydrogen atoms, and assigning appropriate charge states. A few properties, such as those listed here, of the active site/binding pocket of the target protein are pre-calculated. The molar refractivity and log P values are calculated separately for aromatic and non-aromatic residues. Additionally, the hydrogen bond donors are determined for the backbone amide group PD(Amide-NH) [73]. The amino acid sets that are considered are as follows: positively charged amino acids PD (K+R+HIP), neutral amino groups PD (K+N+Q), heteroaromatic donors PD (W+H), and hydroxyl-containing groups PD (T+S+Y+D+E). The hydrogen bond acceptors are quantified for the backbone amide PA(Amide-O) as well as for the following groups: negatively charged PA (D+E), neutral non-aromatic PA (N+Q+T+S+D-H+E-H), and aromatic acceptors PA (Y+H). Furthermore, the measurement of the protein pocket's volume is referred to as PVol. These are some crucial characteristics of protein structure that must be taken into account during protein target preparation for screening.

Screening Process: During the screening procedure, Molecules are subjected to comprehensive examination using computational techniques such as RASPD+ [73,74] for analysis. In the context of LBVS, compounds are methodically compared to a well-selected library of produced compounds. This comparison evaluates their structural and chemical characteristics in order to discover prospective hits. On the other hand, SBVS uses molecular docking techniques to virtually place compounds into the binding site of the target protein. This helps predict how the compounds will bind

and their strength of interaction, making it easier to select molecules with a higher potential for therapeutic efficacy. These computational approaches are crucial in speeding up the process of discovering new drugs by effectively searching through large collections of chemicals and helping to prioritize which compounds should be tested experimentally and improved further.

Scoring and Ranking: During the scoring and ranking step, compounds are assessed according to their predicted binding affinities or resemblance to known active compounds [75,76]. This procedure depends on a variety of scoring systems specifically designed to estimate binding free energy, a critical measure in forecasting the intensity of the interaction between a chemical and its target protein. The scoring functions mentioned here cover a range of computational methods, including empirical scoring functions and physics-based models. Each of these approaches provides distinct insights into the molecular interactions that influence the binding between ligands and targets. By utilizing these scoring approaches, scientists can methodically rank compounds with the greatest probability of strong and specific binding, accelerating the discovery of potential therapeutic candidates in extensive chemical libraries.

Post-Screening Analysis: After the initial scoring and ranking processes, the top-ranked compounds revealed in the post-screening analysis are subjected to thorough validation using more accurate computational approaches, particularly molecular docking. The purpose of this stage is to improve and confirm the predicted ways in which the selected compounds bind to the target protein's binding site, as well as their strengths of interaction. Molecular docking simulations utilize advanced algorithms [77] to forecast the energetically advantageous positions and structures of ligands inside the protein's active site, yielding significant insights into the probable binding interactions and affinity profiles. By combining molecular docking with experimental data and further computational analyses, scientists can enhance the selection of compounds, prioritize the most promising candidates, and provide guidance for subsequent optimization efforts in the process of drug discovery. This ultimately accelerates the development of effective and specific therapeutic agents.

2.1.2. Applications and Challenges

Virtual screening is a crucial component of contemporary drug discovery, providing notable benefits in terms of cost and time effectiveness by enhancing experimental screening endeavors [78,79]. It aids in the discovery of new chemical frameworks and allows for the improvement of initial molecules, speeding up the process of developing drugs. However, there are still some obstacles that remain in virtual screening approaches. The precision of scoring systems, which play a crucial role in forecasting binding affinities and directing compound prioritization, continues to be a significant concern. Scoring functions that are not accurate can result in false positives or false negatives, which can undermine the reliability of virtual screening results [80,81]. Moreover, the quality of target protein structures employed in molecular docking simulations greatly influences the prediction precision of virtual screening results. Target protein structures with structural flaws or uncertainties might contribute to biases and mistakes in the docking predictions, which can compromise the trustworthiness of the screening process. Furthermore, the ever-changing characteristics of proteins, such as alterations in shape and adaptability, present a significant obstacle in the process of virtual screening [82]. Accurately considering protein flexibility is crucial in docking simulations to estimate ligand binding modes and affinities. However, this process is complicated and requires significant computer resources. To overcome these issues, it is necessary to constantly improve computational

methodologies and closely integrate them with experimental validation techniques. This will improve the reliability and prediction capabilities of virtual screening approaches in drug development efforts.

2.2 Molecular Docking

Molecular docking is a computer technique used to forecast the optimal arrangement of a small molecule (ligand) when it attaches to a target protein (receptor) and to assess the intensity of its interaction. Structure-based drug design relies heavily on this technique. Computer modeling plays a crucial role in modern medicine by helping to identify new lead molecules [83]. One of the main challenges in this field is accurately predicting how candidate ligands bind to target proteins. This involves modeling the structures of protein-ligand complexes, which is an essential step in any CADD application [84-86]. The success of *in-silico* drug design relies primarily on two key factors: firstly, the use of a proficient and resilient molecular docking protocol that can accurately position the candidate molecule in an energetically advantageous conformation relative to its target protein conformation, and secondly, the implementation of an effective scoring function to evaluate the binding affinity of the modeled complex [87]. The docking problem in computational techniques is categorized into two types: rigid body algorithms and flexible algorithms. The rigid body approximation is the initial and fundamental stage of docking, where the flexibility of the protein and ligand is not explicitly taken into account [88]. As a result, the search algorithm quickly explores the optimal position of the ligand in the active site of the receptor within a six-dimensional search space, utilizing the translational and rotational degrees of freedom of the ligand. Some examples of inflexible docking programs [89] are DOCK [90], Yucca [91], and FRED [92]. Flexible docking approaches involve the addition of torsional degrees of freedom to either the ligand alone or to both the receptor and ligand. Several docking techniques have been documented in the past few years. The algorithms are categorized into three groups based on the search criteria: systematic search (including incremental construction, conformational search, and databases), random or stochastic approaches (such as Monte Carlo, evolutionary algorithms, and tabu search), and simulation methods. The incremental construction algorithms initially divide each molecule into a collection of inflexible pieces based on rotatable bonds, then subsequently assemble the fragments gradually around the binding pocket. Several instances of this category are DOCK, FlexX [93], Surflex [94], FLOG [95], and Hammerhead [96]. Stochastic search algorithms such as AutoDock [97], ICM [98], GOLD [99], and MCDOCK [100] are representative examples. These algorithms are based on either genetic algorithms or the Monte Carlo simulated annealing approach. These approaches function by implementing substantial random alterations to the ligand. MD simulation approaches enable the simulation of different components of a protein-ligand system at varying temperatures. In addition to search techniques used in docking exercises, the scoring function is crucial in determining the ranking of different bound states. To accurately determine the optimal conformation of a docked structure, it is crucial to employ both a highly effective search and docking algorithm, as well as a robust scoring system [101]. An excellent search strategy should possess high efficiency in accurately identifying the global minimum. While many docking systems have been tested on different sets of complexes, there is a lack of comprehensive analysis of the projected structures and their binding energetics compared to experimental results. This work aims to investigate two fundamental questions: (1) the effectiveness of the search approach used for stiff protein ligands in predicting the crystal structure and (2) the reliability of the combined docking/scoring process in terms of efficiency.

2.2.1. Workflow of Molecular Docking

Molecular Docking Search Algorithm: The search algorithm is a crucial technique in computational biology for exploring the conformational space of ligands and understanding their possible orientations within the protein's binding site. Researchers utilize advanced methodologies such as genetic algorithms [102], Monte Carlo simulations [103], and systematic search approaches [104] to explore the extensive range of molecular configurations. This enables them to uncover energetically advantageous conformations and binding poses. Genetic algorithms utilize concepts derived from natural selection to progressively enhance ligand conformations, while Monte Carlo simulations randomly explore the conformational and configuration space, guided by thermodynamic considerations. Systematic search methods methodically investigate the parameter space, effectively encompassing potential ligand orientations within the protein binding site. These approaches together facilitate the investigation of intricate molecular interactions, assisting in the development of drugs by anticipating how ligands attach and improving our comprehension of the links between molecular structure and biological activity, which is essential for designing drugs based on reason and logic.

Molecular Docking Scoring Function: The scoring function is a crucial component in computational drug development since it assesses the binding affinity between a ligand and its protein target. These functions are formulated using empirical, knowledge-based, or force-field-based methodologies, with the goal of approximating the binding free energy between the ligand and protein. Empirical scoring functions utilize statistical analyses of experimental data to establish mathematical correlations between molecular characteristics and binding affinities. Knowledge-based scoring systems utilize databases of established protein-ligand complexes to deduce interaction patterns and energetics. Force-field-based scoring functions employ molecular mechanics force fields to compute the intermolecular interactions and energy terms that influence the binding of ligands to proteins. Scoring functions allow for a semi-quantitative assessment of these interactions, enabling the prioritization and ranking of possible ligands. This helps in selecting viable candidates for further experimental validation and drug development.

2.2.2 Applications and Challenges

Molecular docking is a crucial technique in the field of drug development. It encounters many obstacles despite its extensive applicability. Precisely representing protein flexibility is essential for accurately predicting actual binding interactions. This remains a challenge due to the dynamic nature of proteins. Integrating solvent effects into docking simulations introduces additional complexity, as it profoundly affects the energetics of ligand binding and can have a substantial impact on the accuracy of predictions. Moreover, the constraints of scoring algorithms in precisely calculating binding affinities present a difficulty in virtual screening and lead optimization endeavors. Continuing research in the field of molecular docking focuses on boosting the dependability and predictive capability of docking simulations in drug development by improving the accuracy of scoring functions and tackling the challenges posed by protein flexibility and solvent effects. It is crucial to overcome these problems in order to optimize the efficiency of molecular docking and expedite the progress of developing novel medicines.

2.3 MD Simulations

MD simulations are computer techniques employed to investigate the dynamic behavior of atoms and molecules by tracking their physical motions over a period of time. They offer comprehensive analysis of the kinetics and structural alterations of biomolecules, which are essential for comprehending their functionality and interactions. In recent years, the use of MD simulations in molecular biology and drug development has significantly increased. These simulations accurately depict the actions of proteins and other biomolecules at the atomic level and with precise time measurements. The significant advancements in speed, precision, and availability of biomolecular simulation methods, coupled with the widespread availability of experimental structure data, have made biomolecular simulation more attractive to experimentalists. Simulations have been found to be highly useful in understanding the operational mechanisms of proteins and other biomolecules, revealing the underlying structure/mechanisms of diseases, and in the creation and improvement of therapeutically relevant small molecules, peptides, etc. A molecular biologist studying the functionality of a protein or any other biomolecule encounters a comparable obstacle. An atomic-level structure is highly advantageous and usually provides a significant understanding of the biomolecule's functionality. The atoms within a biomolecule exhibit perpetual motion, and the functionality and intermolecular interactions are contingent upon the dynamics of the participating molecules. One desires not only a fixed image but also the capability to observe these biomolecules in motion, manipulate them at the atomic scale, and observe their corresponding reactions. Regrettably, observing the movements of individual atoms and manipulating them in a desired manner poses a significant challenge. A compelling option is to utilize an atomic-level computer simulation to study the pertinent biomolecules.

MD simulations utilize a comprehensive model of interatomic interactions to forecast the movement of each atom within a protein or other molecular system as time progresses [105]. These simulations have the ability to accurately represent a diverse range of significant biomolecular processes, such as conformational change, ligand binding, and protein folding. They provide detailed information about the positions of all the atoms with a temporal resolution of femtoseconds. Significantly, these simulations have the ability to forecast the atomic-level reactions of biomolecules to disturbances like mutation, phosphorylation, protonation, or the introduction or elimination of a ligand. MD simulations are frequently employed alongside diverse experimental structural biology techniques, such as X-ray crystallography, cryo-EM, NMR, electron paramagnetic resonance (EPR), and Förster resonance energy transfer (FRET).

MD simulations are not a recent development. The initial MD simulations of simple gases were conducted in the 1950s by Alder and Wainwright (1957). The initial MD simulation of a protein took place in the 1970s [106]. The fundamental principles that made these simulations possible were acknowledged as a significant contribution to the field of chemistry and were awarded the Nobel Prize in Chemistry in 2013 [107,108]. MD simulations have gained significant popularity and visibility in recent years, especially among experimental molecular biologists. Simulations are increasingly common in experimental structural biology studies, serving the purpose of interpreting experimental findings and providing guidance for further experimental investigations. In the field of neuroscience, for instance, simulations have been used extensively to investigate proteins that play a crucial role in neuronal signaling. These simulations have also been employed to aid in the development of drugs that target the nervous system, uncover the mechanisms behind protein

aggregation in neurodegenerative disorders, and lay the groundwork for the design of more advanced optogenetics tools.

MD simulations utilize classical mechanics, Newtonian dynamics, in particular, to model the behavior of atoms and molecules. In this approach, atoms and molecules are considered as particles that move in response to forces generated by potential energy functions.

Force Fields: Within the realm of MD simulations of proteins, the word "force field" denotes the amalgamation of a mathematical equation and its corresponding parameters, which are employed to articulate the energy of the protein based on its atomic coordinates. Since the early 1980s, there have been ongoing efforts to generate force fields for drug-like molecules due to the growing interest in modeling and simulation in drug discovery. The force fields commonly employed for small molecules in present times are OPLS-All-Atom (OPLS-AA) [109], OPLS3 [110], the CHARMM General force field (CGenFF) [111-113], the General AMBER Force Field (GAFF) [114, 115], Merck Molecular Force Field (MMFF) [116-120], and GROMOS [121-125]. The maintenance and improvement of these force fields have been ongoing, with monthly updates to incorporate new parameters that encompass a broader spectrum of chemical entities. Due to the high potential for errors and the need for extensive experience, algorithms have been created to automatically detect atom types and generate parameters for compounds, eliminating the need for manual assignment. As an illustration, the AnteChamber program was specifically created to produce GAFF and AMBER topologies. Similarly, the CGenFF program, which can be accessed through the ParamChem website [126], was developed to build CHARMM topologies and parameters using CGenFF as a basis. Additional programs for parameter assignment include ATB [127] and PRODRG [128, 129] for GROMOS, as well as MATCH [130] and SwissParam [131] for CHARMM.

2.3.1. Workflow of MD Simulations

System Preparation: The original configuration of the biomolecule is acquired from databases such as the Protein Data Bank and is thoroughly examined for its integrity, with any absent atoms or residues being supplemented as required. Next, the biomolecule is immersed in a container of water/solvent molecules using a suitable water/solvent model (such as TIP3P or SPC/E [132]), and periodic boundary conditions are applied to simulate an endless system. In order to achieve system neutralization, counterions are introduced to offset the charges of the biomolecule, and supplementary ions may be incorporated to mimic the ionic strength found in physiological conditions. The system undergoes energy minimization to remove any unfavorable interactions, followed by equilibration in NVT [133] and NPT [134] ensembles to stabilize temperature and pressure, ensuring that the water and ions are appropriately positioned around the biomolecule.

Energy Minimization: The objective of energy minimization is to eliminate steric conflicts and optimize the geometry of the system so as to achieve a physically realistic and stable configuration prior to dynamics simulations. Steric clashes occur when atoms are in close proximity, resulting in strong repulsive interactions. These conflicts are handled during minimization by altering the locations of the atoms. This procedure depends on a potential energy function determined by a force field (such as AMBER or CHARMM) that includes both bound and non-bonded interactions. Different techniques, such as steepest descent and conjugate gradient, are employed, persisting until a convergence condition, such as a low energy gradient, is achieved. Energy minimization, carried out using software tools such as AMBER [135, 136] or GROMACS [137], leads to a structure that

has lower potential energy and is free from steric conflicts. It also offers an energy profile that can be analyzed.

Equilibration: Equilibration is performed by subjecting the system to carefully regulated temperature and pressure settings in order to attain a state of stability. This procedure consists of two main stages: NVT (constant Number, Volume, Temperature) and NPT (constant Number, Pressure, Temperature) ensembles. During the NVT phase, the system's temperature is incrementally modified to the desired value using a thermostat. This ensures that the biomolecule and the liquid surrounding it reach thermal equilibrium without causing any notable changes in their structure. Subsequently, the NPT phase employs a barostat to regulate the pressure, enabling the system volume to vary and attain the desired pressure, hence ensuring the stability of the system's density. This step guarantees that both temperature and pressure are balanced, resulting in a system that replicates true biological conditions. The equilibration process entails monitoring and verifying the stability of different parameters, including temperature, pressure, density, and potential energy. Typically, these stages are implemented using tools such as GROMACS, AMBER, and CHARMM. Specific input files are utilized to define the thermostat, barostat, target temperature, and pressure. Effective equilibration yields a biomolecular system that is stable and prepared for productive MD simulations, with properly balanced solvent and ion distributions surrounding the biomolecule.

Production Run: The production run is the stage in which the real MD simulation takes place, usually lasting from nanoseconds to microseconds, in order to examine the temporal progression of the system. In this phase, the system is simulated under constant circumstances, typically in the NPT ensemble, to ensure realistic environmental conditions are maintained. The selection of the time step is crucial, often ranging from 1 to 2 femtoseconds, in order to precisely capture the movements of atoms while maintaining numerical stability. The simulation produces trajectories that document the locations, velocities, and forces exerted on every atom during the duration. These trajectories are employed to examine the dynamic characteristics of the biomolecule, encompassing conformational alterations, interactions with solvents and ions, and reaction routes. Advanced techniques like replica exchange or enhanced sampling approaches might be utilized to examine the energy landscape of the system completely. The production run requires significant computing resources and is usually performed using high-performance MD software packages such as GROMACS, AMBER, or CHARMM. The results of the manufacturing run, which include trajectory files, energy profiles, and structural snapshots, offer useful data for the subsequent analysis and interpretation of the biomolecular processes being studied.

Analysis: When analyzing trajectory data, researchers extract many important observations to gain a comprehensive understanding of the dynamic behavior of molecular systems. Conformational modifications are carefully examined, uncovering alterations in molecular structure that are essential for comprehending functionality. An analysis is conducted to unravel the complicated mechanisms involved in molecular recognition and the creation of complexes by studying binding interactions. The dynamic features of a system can be assessed using metrics such as root-mean-square deviation (RMSD), which provides information on the system's stability and fluctuations over time. Additionally, root-mean-square fluctuation (RMSF) can be used to understand the flexibility of individual residues. In addition, the examination of hydrogen bonds offers a comprehensive view of the intermolecular interactions that are crucial for the stability and functioning of molecules.

Collectively, these studies provide a thorough perspective on the structural changes and interactions within the system being studied.

2.3.2 Free Energy Analysis

Free Energy Analysis is a crucial technique in molecular biophysics that allows for the precise assessment of the energy distribution involved in important molecular events such as ligand binding or conformational changes. By calculating the changes in free energy, this method provides a deep understanding of the thermodynamic principles that control these occurrences, therefore enabling the logical development of candidate drugs. Free Energy Analysis is a valuable tool for studying the stability of protein-ligand complexes and the dynamics of biomolecular conformations. It provides important information about the energy involved in molecular interactions, which can be used to optimize drug candidates and predict binding affinities. This methodological framework is a crucial tool in the collection of computational tools used for drug discovery. It enables researchers to make well-informed judgments and speed up the creation of new treatments. Molecular biophysics employs advanced methodologies, including Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA), Molecular Mechanics Generalized Born Surface Area (MMGBSA), and MMBAPPL+ [138] approaches, to accurately measure the alterations in free energy that occur during molecular processes. These methodologies offer a thorough comprehension of the contributions made by various energy components, such as molecular mechanics, solvation effects, and entropy changes. Moreover, Free Energy Perturbation (FEP) [139] and Thermodynamic Integration (TI) [140] methods provide advanced approaches for calculating the differences in free energy between two states by systematically converting one state into the other. The FEP and TI methods, although compute-intensive, play a crucial role in rational drug design and biomolecular engineering by systematically perturbing molecular systems and accurately calculating binding affinities, stability constants, and conformational preferences through the quantification of associated free energy changes. Together, these approaches are essential tools for understanding the thermodynamic principles that control molecular interactions and for driving the creation of new molecules that have improved effectiveness and selectivity.

3. Integrated Workflow

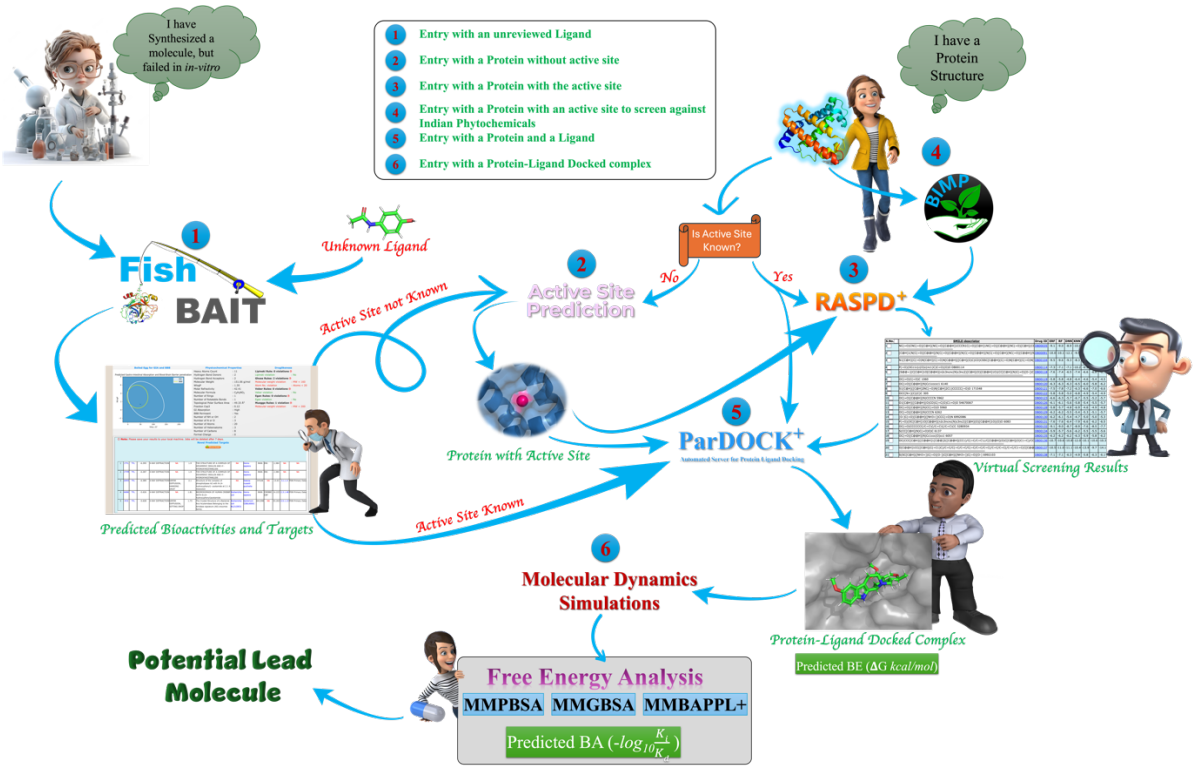


Fig. 1: Workflow for Ligand Bioactivity Prediction, Target Identification, Active Site Prediction, Virtual Screening, Molecular Docking, and MD Simulations

When working with an unknown ligand, a structured workflow is essential to systematically analyze its bioactivity, identify potential protein targets, predict active sites, and perform molecular modeling to propose potential lead molecules. The process begins with the ligand itself, an unknown compound whose properties and interactions need to be elucidated. The first step involves utilizing the FishBAIT web tool (http://www.scfbio-iitd.res.in/fishbait/), a tool designed specifically to "fish" bioactivity information and predict potential targets for the ligand. FishBAIT analyzes the chemical structure of the ligand and cross-references it with known bioactivity data to provide insights into its possible biological functions and target proteins. This initial bioactivity prediction is crucial as it sets the stage for subsequent steps.

Upon obtaining potential protein targets from FishBAIT, the next step involves assessing whether these targets have known active sites. If the predicted protein target has a documented active site, the workflow advances to molecular docking using ParDOCK+ (http://www.scfbio-iitd.res.in/pardock+/). ParDOCK+ is a powerful tool that facilitates the docking of the ligand into the active site of the target protein, predicting how the ligand binds and interacts at the molecular level. This step is critical for understanding the binding affinity and orientation of the ligand within the active site, providing insights into its potential efficacy as a drug or therapeutic agent.

However, if the active site of the predicted protein target is not known, the workflow requires the use of active site prediction software (AADS; http://www.scfbio-iitd.res.in/AADS/). This software

analyzes the protein structure to identify potential active sites where the ligand might bind. Once the active site is predicted, molecular docking using ParDOCK+ can proceed as previously described, ensuring that the ligand's interactions with the protein are thoroughly explored.

For users who have a protein structure with a known active site, the workflow incorporates virtual screening using RASPD+ (http://www.scfbio-iitd.res.in/raspd+/). Virtual screening is a computational technique that involves screening a library of compounds against a protein target to identify potential hits. RASPD+ enhances this process by efficiently screening large libraries and identifying compounds that exhibit high binding affinity to the target protein. In addition to RASPD+, the BIMP database (https://scfbio.iitd.ac.in/bimp/) can be utilized for screening. The BIMP database contains a vast collection of bioactive molecules, and screening against this database increases the chances of identifying promising compounds.

If the user has a protein structure but lacks active site information, the workflow necessitates active site prediction before proceeding to virtual screening. Active site prediction software is employed to determine potential binding sites on the protein, and once these sites are identified, virtual screening using RASPD+ and the BIMP database can be conducted. This ensures that the ligand can be effectively screened against potential binding sites on the protein even without prior knowledge of active sites.

Following the virtual screening, selected hits are subjected to molecular docking using ParDOCK+. This step refines the virtual screening results by providing detailed insights into the binding interactions of the ligand with the protein target. The docked complexes are then analyzed further through MD simulations. MD simulations are computational techniques that simulate the physical movements of atoms and molecules over time. By performing MD simulations on the docked complexes, researchers can assess the stability and dynamics of the ligand-protein interactions, gaining a deeper understanding of how the ligand behaves within the binding site.

To complement MD simulations, free energy analysis is performed using MMPBSA, MMGBSA, and MMBAPPL+ methods. These techniques allow the calculation of the free energy changes associated with ligand binding, providing quantitative insights into the strength and stability of the interactions. MMPBSA and MMGBSA are widely used for their ability to decompose the free energy contributions into various components, such as electrostatic, van der Waals, and solvation energies. MMBAPPL+ offers additional validation by incorporating alternative approaches and ensuring robustness in the free energy calculations.

Throughout this workflow, the integration of various computational tools and techniques allows for a comprehensive analysis of the ligand's bioactivity, target interactions, and binding properties. By starting with bioactivity prediction and target identification, the workflow ensures that the ligand's potential functions and targets are thoroughly explored. Active site prediction and molecular docking provide detailed insights into the binding interactions, while virtual screening expands the scope of potential hits. MD simulations and free energy analysis further refine these findings, offering a dynamic view of ligand-protein interactions and quantifying their stability.

Ultimately, the goal of this workflow is to identify and propose potential lead molecules based on the combined results of bioactivity predictions, molecular docking, virtual screening, MD simulations,

and free energy analysis. A lead molecule is a compound that demonstrates promising bioactivity and binding affinity, making it a candidate for further development and optimization in drug discovery. By following this structured approach, researchers can systematically narrow down the vast chemical space of potential compounds, identifying those with the highest potential for therapeutic applications.

In conclusion, this comprehensive workflow leverages advanced computational tools and techniques to analyze unknown ligands, predict their bioactivity, identify protein targets, and assess their binding interactions. From bioactivity prediction with FishBAIT to molecular docking with ParDOCK+, virtual screening with RASPD+ and the BIMP database, and detailed analysis through MD simulations and free energy calculations, each step builds upon the previous ones to provide a thorough understanding of the ligand's potential as a lead molecule. This systematic approach not only streamlines the drug discovery process but also enhances the accuracy and efficiency of identifying promising therapeutic candidates. Some commercial software offers alternatives to the above modules with nearly similar functionalities.

To navigate the complex workflow of analyzing an unknown ligand and its potential interactions with protein targets, various entry points have been designed, each with specific minimum requirements, to ensure smooth progression through the workflow. The first entry point involves the ligand itself, particularly an unreviewed ligand entering the FishBAIT software. For this step, the essential requirement is the chemical structure of the ligand, which must be provided in a format such as SMILES, InChI, or a structure file like SDF or MOL. FishBAIT utilizes this structural information to predict the bioactivity of the compound and identify potential protein targets by referencing extensive bioactivity databases. This step is crucial as it sets the foundation for all subsequent analyses if the target is unknown and only a bioactive compound is known against a particular disease.

The second entry point is for scenarios where there is a protein target without known active site information. Here, the primary requirement is the three-dimensional structure of the protein, usually provided in the PDB (Protein Data Bank) [141] file format. Active site prediction software, such as AADS [142], CASTp [143], or Fpocket [144], is then used to analyze the protein structure and predict potential active sites where ligands might bind. This prediction is essential for guiding further molecular docking studies.

In cases where the protein has a known active site, the third entry point involves virtual screening using RASPD+ software. The minimum requirements for this step include the protein structure with defined active site coordinates and a library of compounds to screen against. RASPD+ is utilized to identify compounds with high binding affinity to the protein's active site. This step is particularly important for narrowing down potential lead compounds from a large library.

The fourth entry point is for proteins with known active sites that need to be screened against the Indian Phytochemicals database BIMP or other molecular databases such as Zinc, etc. Similar to the RASPD+ screening, this step requires the protein structure with active site information and access to the BIMP, which contains a curated collection of bioactive molecules from Indian medicinal plants. This specialized screening can uncover natural compounds with therapeutic potential. The methodology can also cover Zinc and other small molecule databases that have been pre-prepared for virtual screening using RASPD+.

The fifth entry point involves a protein and a ligand prepared for molecular docking using ParDOCK+ software. The minimum requirements here include both the protein structure (with or without active site information, if previously predicted) and the ligand structure. ParDOCK+ performs detailed molecular docking to predict the binding orientation and affinity of the ligand within the protein's active site. This step refines the understanding of ligand-protein interactions.

Finally, the sixth entry point is for the protein-ligand docked complex, ready for MD simulations. The requirements for this stage include the docked complex structure, typically in PDB format, along with parameters for running MD simulations, such as force field information and simulation conditions (temperature, solvent model, etc.). MD simulations provide dynamic insights into the stability and behavior of the ligand-protein complex over time. Post-simulation, free energy analysis using methods like MMPBSA, MMGBSA, or MMBAPPL+ can be performed to quantify the binding affinity and validate the interactions observed during docking and MD simulations.

In summary, each entry point in this workflow has specific prerequisites that ensure a seamless transition from one step to the next, starting from ligand bioactivity prediction or specification of a protein target to detailed molecular docking and dynamic simulations. This structured approach not only facilitates the efficient identification of potential lead molecules but also integrates a comprehensive analysis of their interactions and stability, ultimately aiding in the drug discovery process. Most phytochemicals obtained from plants, including phenolic compounds and flavonoids, have been demonstrated to have a positive impact on health. Plant-based pharmaceuticals are gaining popularity due to their natural composition, non-addictive properties, easy biodegradability, minimal side effects, and cost-effectiveness in both developing and developed countries [145, 146]. Modern dietary patterns influence the attainment of a nutritionally balanced diet. There is a growing disparity in nutrition, and therefore, the concept of normal living has been altered. Within this framework, dietary supplements and herbal remedies are prevalent as supplementary or alternative products for individuals. Various diseases are currently associated with "oxidative stress," which arises when there is an imbalance between the production of pro-oxidants and their elimination.

4. Some Case Studies

India boasts a long and vibrant history of medicinal knowledge. Traditional systems like Ayurveda and Siddha have flourished for centuries, recognizing the immense potential of plants for healing. Natural plants are regarded as nature's gifts, each with unique properties that promote health and well-being. Ancient Ayurvedic texts, which date back thousands of years, document the use of numerous plants for treating various ailments. Nature’s pharmacy in India, such as Tulsi (Holy basil), Turmeric, Neem, Ginger, Ashwagandha, and Haritaki, to name a few, boasts a vast array of medicinal plants.

Medicinal plants are currently receiving significant interest due to their distinct qualities as abundant reservoirs of medicinal bioactive compounds, which have the potential to facilitate the development of novel pharmaceuticals. According to the WHO reports, 80% of individuals in underdeveloped nations depend on traditional medicine as their main source of healthcare [147]. Ancient Indian medical systems like Ayurveda, Siddha, and Unani promote the utilization of medicinal plants for the treatment of diseases [148]. Ayurveda and Unani hold the belief that medicinal herbs have the potential to decrease the likelihood or diminish the danger of cardiovascular illness, as well as other conditions such as rheumatoid arthritis, lung disorders, cataracts, Parkinson's disease, and improve liver function [149]. Plants produce a wide variety of biologically active substances and serve as an

important reservoir of medicines [150]. Phytochemicals are broadly classified based on their chemical structures and biological activities. The main categories include:

i. Alkaloids: These nitrogen-containing compounds are primarily found in plants like coffee, tea, and certain vegetables. Alkaloids have a diverse range of pharmacological effects, including pain relief, stimulation, and anti-malarial properties. Examples of alkaloids include Morphine (analgesic), quinine (antimalarial), and caffeine (stimulant) etc.

ii. Flavonoids: Flavonoids are Polyphenolic compounds that are found in a wide variety of fruits, vegetables, and beverages like tea and wine. They are known for their antioxidant properties. They help reduce inflammation and may lower the risk of chronic diseases such as heart disease and cancer. A few examples include quercetin, catechins, and anthocyanins.

iii. Terpenoids: They are the largest class of phytochemicals, including essential oils and carotenoids. Beta-carotene, lycopene, and lutein are well-known carotenoids that are linked to eye health and may reduce the risk of certain cancers. Limonene (anticancer), menthol (analgesic), and beta-carotene (antioxidant) are examples of Terpenoids.

iv. Phenolic Acids: Compounds with potent antioxidant activity and help in reducing inflammation and preventing cellular damage. They are found in foods like coffee, fruits, and vegetables. Examples include ferulic acid, gallic acid, and salicylic acid.

v. Glycosides: Molecules consisting of a sugar and a non-sugar component with diverse therapeutic effects. Examples include digoxin (cardiotonic) and sennosides (laxatives).

vi. Saponins: Saponins are present in beans and legumes. Compounds with soap-like properties, known for their immune-boosting and anticancer effects, such as saponins, have been shown to lower cholesterol levels as well. A few examples include ginsenosides and diosgenin.

vii. Glucosinolates: Glucosinolates are sulfur-containing phytochemicals found in cruciferous vegetables like broccoli, cabbage, and Brussels sprouts. These phytochemicals have shown cancer-preventive properties.

4.1. Prioritising Phytochemicals in Haritaki

Medicinal plants such as Haritaki are increasingly attracting attention due to their unique properties as a rich source of bioactive compounds, which hold potential for the development of new medicines. *Terminalia chebula* Retzius*,* often known as Haritaki, is extensively grown in South East Asia, especially India. Haritaki is a term that holds significant significance. It has healing properties and is said to cure all ailments [151]. The tree is a perennial member of the *Terminalia* genus and is widely recognized as a rejuvenating plant. This tree is native to the forests of Northern India, particularly in regions with little precipitation, such as Uttar Pradesh and Bengal. It can also be found in Tamil Nadu, Karnataka, and Southern Maharashtra.

Harad is the Hindi name for this herb, but it is also called Haritaki in Sanskrit. It holds a prominent place in Ayurveda, being well-recognized and used. In Tibet, it is referred to as the "King of Medicine," and its remarkable healing properties have positioned it as the highest-ranking item on the Ayurvedic Materia Medica list [152]. Haritaki is available in seven distinct variations: Vijayan,

Boodhana, Rogini, Abhyan, Amrutha, Boothagi, and Sethagi. These types are recognized in Siddha literature based on their geographical distribution. In addition to salt, the fruit of haritaki possesses five distinct tastes: a pungent outer peel, a sour ridge, an astringent seed, a bitter stem, and a sweet endosperm [153]. It is commonly known as an epicenter of medicinal activity and a significant ethnomedicinal plant in human society. Since ancient times, it has been utilized as a medicinal remedy. It contains a variety of beneficial compounds such as phytochemicals, tannins, flavonoids, sterols, amino acids, fructose, resin, and oils. However, it is especially abundant in tannins (32%–34%), which give it a bitter and astringent taste. This is the main reason why it is not widely accepted by consumers [154].

Consuming Haritaki can not only fulfill people's nutritional requirements but also have a significant role in the production of nutraceuticals. Consequently, the food and flavor sectors are seeking innovative food constituents to use in the production of dietary supplements. Haritaki usage can also help meet nutritional requirements and prevent other degenerative diseases, including cancer, neurological disorders, cardiovascular disorders, and aging. Compounds found in the plant provide biological functions and can exhibit chemically significant impacts on human systems, leading to reduced negative consequences [155]. In Thai traditional medicine, the fruit of the haritaki is utilized for its laxative, carminative, astringent, expectorant, and tonic properties. Fever, cough, diarrhea, gastroenteritis, skin problems, candidiasis, urinary tract infection, and wound infections are all prevalent ailments that Tamil Nadu tribes commonly treat as a kind of traditional medicine [156]. It is used to decelerate the process of aging and enhance longevity and immune system function [157].

Various studies have delved into its chemical constituents, revealing the presence of bioactive compounds such as tannins, gallic acid, chebulagic acid, and chebulinic acid. *T. chebula* exhibits a spectrum of therapeutic properties, encompassing antibacterial [158,159], antifungal [160-162], antiamoebic and immunomodulatory [163,164], antiplasmodial [165], molluscicidal [166], anthelmintic [167], antiviral [168-170], antimutagenic and anticarcinogenic [171,172], antidiabetic [173], antiulcerogenic [174], and radioprotective activities [175]. Its widespread adoption among diverse ethnic groups, ‘vaidyas’, ‘hakims’, and Ayurvedic practitioners underscores its esteemed status as a popular remedy for various ailments, as corroborated by comprehensive literature surveys on *T. chebula* [176].

LC-MS and Mass spectra analysis of Aqueous extract of Haritaki: Two grams of plant (*T. chebula*) were dried, crushed, and mixed in 10 mL Phosphate Buffered Saline (PBS). The mixture was incubated overnight at 50 °C to extract the active ingredients. The suspension was centrifuged at 4000 g for 10 minutes. The resulting supernatant was collected and dried at 37 °C overnight. The dried plant extract powder was dissolved in PBS at 10 mg/mL. The extracts were filter sterilized, and aliquots were made and stored at -20 °C till further use. Freshly prepared dilutions were used in each experiment. The Haritaki extract was analyzed using a QTOF-MS instrument (Waters Xevo G2 QTof, Waters, Milford, MA, USA) in electrospray ionization (ESI) mode. The instrument had a mass resolution of 20,000 and was controlled by MassLynx 4.1 software. A chromatographic separation was performed using a Waters Acquity UPLC BEH C18 column with dimensions of 2.1 mm inner diameter and 100 mm length, packed with 1.7 µm particles. The separation was carried out at a temperature of 35°C. An autosampler was used to inject the sample. The mass spectrometer was calibrated using a solution of sodium formate with a concentration of 0.5 mM. Leucine enkephalin,

at a concentration of 2 µg/mL and with a mass-to-charge ratio (m/z) of 554.2615 in negative mode, was employed as a lock spray at a flow rate of 10 µL/min. The collision energy was 6 V. The source parameters were set as follows: capillary voltage at 2.5 kV, sampling cone voltage at 30 V, extraction cone voltage at 4 V, source temperature at 150 °C, desolvation temperature at 500 °C, gas flow at 1000 L/h, and cone gas flow at 50 L/h. The mass spectrum analysis was conducted utilizing the Waters Xevo G2 QTof instrument. The column employed is the AQUITY UPLC BEH C18 with a diameter of 1 mm and a length of 100 mm.

All the extracted bioactive compounds from Haritaki were subjected to virtual screening, molecular docking, and MD simulations **(Fig. 2)** to analyze their anticancer and antiviral activities.

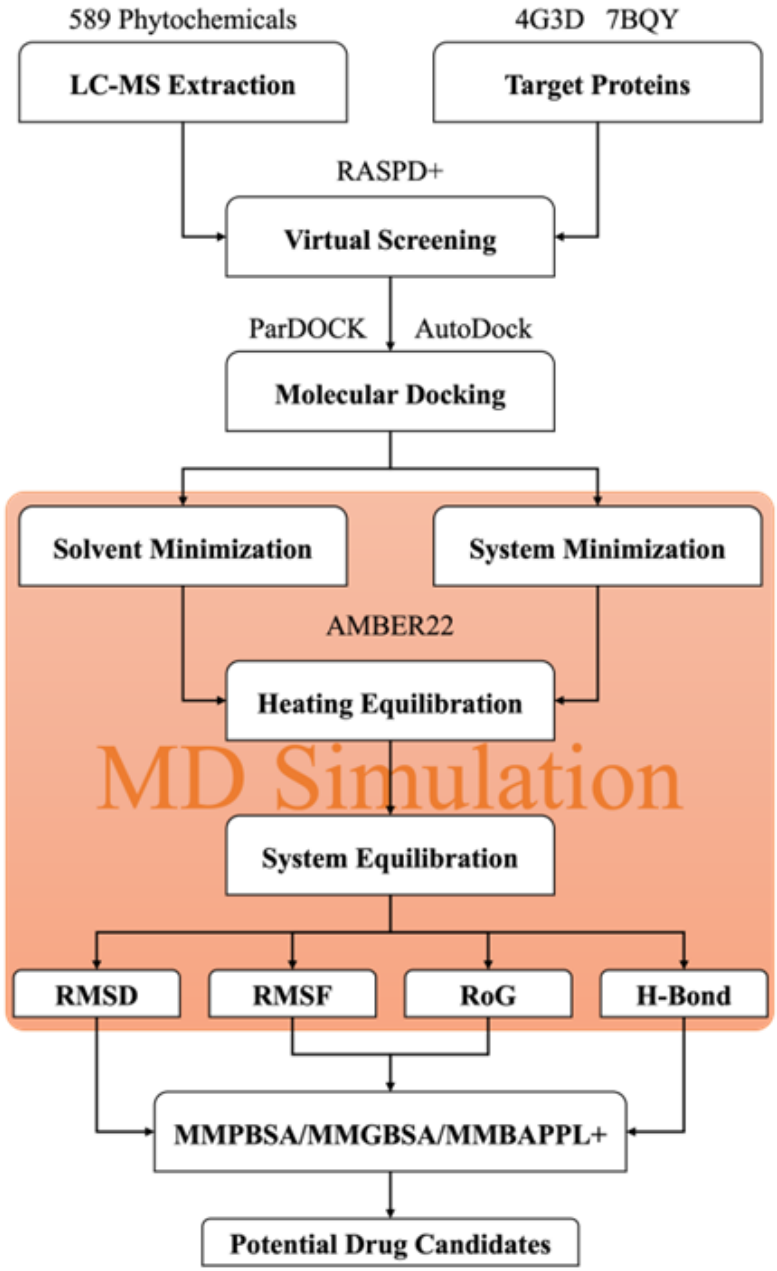


Fig. 2: A workflow for the identification of promising candidates as anticancer and antiviral agents.

4.1.1. Case Study 01: As Anticancer Agents against TNBC

This case study focuses on the prioritization of phytochemicals found in the Haritaki plant as anticancer agents against Triple Negative Breast Cancer (TNBC) using the workflow given in **Fig. 2**. TNBC, a particularly aggressive form of breast cancer lacking estrogen, progesterone, and HER2 receptors, presents unique challenges in diagnosis and treatment. TNBC [177] represents a subtype of breast cancer characterized by the absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) expression, limiting targeted therapeutic options. Consequently, TNBC patients often face aggressive disease progression and poorer prognoses compared to other breast cancer subtypes. In recent years, there has been growing interest in exploring alternative therapeutic targets to improve outcomes for TNBC patients. Among these, nuclear factor-kappa B (NF-κB) signaling has emerged as a promising candidate due to its pivotal role in regulating various cellular processes, including inflammation, proliferation, and apoptosis, all of which are implicated in TNBC pathogenesis. Dysregulated NF-κB signaling has been associated with TNBC aggressiveness, metastasis, and resistance to conventional therapies. Therefore, targeting NF-κB signaling pathways holds significant potential for the development of novel and effective therapeutic strategies against TNBC, addressing an unmet clinical need and providing hope for improved patient outcomes. Its aggressive nature and limited targeted therapy options underscore the urgent need for innovative research and improved healthcare strategies.

The missing residues of NF-κB protein (PDB: 4G3D [178]) were modeled using BhageerathH+ [179-181] via homology modeling by giving the sequence of the protein (UniProt ID: Q99558) [182] as an input. The kinase domain of NF-κB exhibits the characteristic bilobed structure commonly found in protein kinases, comprising an N-terminal region primarily consisting of β-sheets and a C-terminal region composed of helices, which possess several distinctive attributes. Preceding the canonical kinase domain, a conserved N-terminal segment (residues 334–365) forms a helix-strand-helix motif. The initial helix (residues 334–342) extends outward from the core structure, facilitating interactions with adjacent molecules through contacts with the activation loop and C-terminal lobe. The adjacent strand (residues 345–350) and subsequent helix (residues 351–363) closely associate with the N-lobe of the kinase domain, with the strand contributing to a six-strand β sheet and the helix positioned against the catalytically significant "αC" helix. Additionally, the conserved residues at the C-terminus (657–675) of the minimal kinase domain adopt an extended conformation, packing against the C-lobe of the protein. Its active site **(Fig. 3)** is also predicted using AADS active site prediction software.

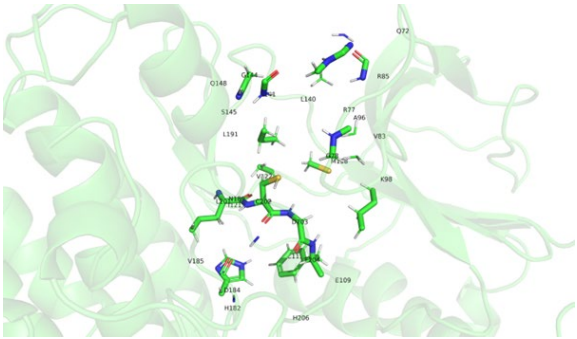


Fig. 3: Active site of NF-κB showing the active site residues

Similarly, target proteins were also prepared by adding missing hydrogens, followed by the assignment of the appropriate ionization states for the side chains of both proteins. All the 589

constituents were virtually screened against the target using RASPD+. Docking studies given in **Table 1** of the top 8 compounds obtained from virtual screening were performed utilizing ParDOCK+ docking software, a module of Sanjeevini, and AutoDock4.

Table 1. Binding energies of ligands NF-κB (PDB: 4G3D) predicted by ParDOCK+ and AutoDock4 software after docking studies. (Example Link for PubChem: https://pubchem.ncbi.nlm.nih.gov/compound/238205; Example Link for BIMP: http://scfbio-iitd.res.in/bimp/compound.php?compound=BIMP068258)

PDB ID	PubChem ID	BIMP ID	ParDOCK	AutoDock
4G3D	38222	BIMP058124	-10.2	-6.5
4G3D	238205	BIMP068258	-11.28	-7.9
4G3D	253793	BIMP070895	-10.26	-4.6
4G3D	315440	BIMP093550	-9.33	-7.5
4G3D	345138	BIMP062086	-7.7	-5.7
4G3D	383130	BIMP051307	-10.86	-5.9
4G3D	427204	BIMP091539	-11.96	-4.5
4G3D	135398646	BIMP043473	-6.71	-5.7

MD simulations were performed on all the 8 protein-ligand docked complexes for up to 100 nanoseconds. The Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), and Radius of Gyration (RoG) were analyzed to ensure the integrity, stability, and compactness of the protein-ligand complex. Out of 8 protein-ligand complexes, only one ligand showed favorable results. The hydrogen bond analysis (**Fig. 4**) of the 4G3D_238205 complex showed that there is at least one constant hydrogen present between the ligand and main chain of Leu141 residue. Furthermore, a free energy analysis (**Table 3**) was conducted to calculate the binding affinity of the ligand toward the protein over time. The results demonstrated a strong and stable interaction between the ligand and the protein, indicating potential efficacy as an anticancer agent.

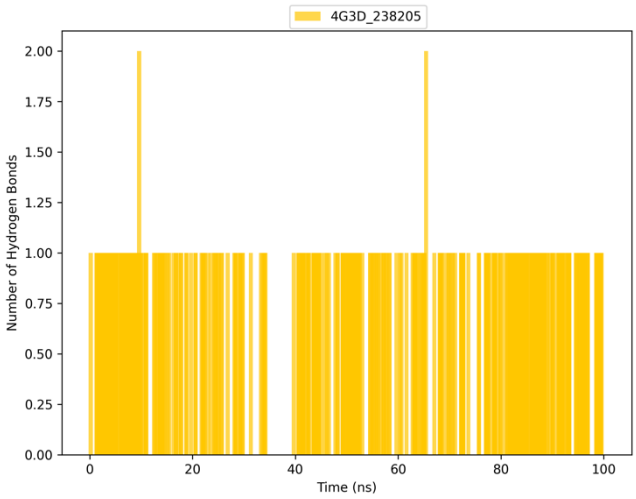


Fig. 4: Hydrogen Bond Analysis of 4G3D_238205 complex

To get more insights into the interaction between the protein and the ligand, 2D representations of the binding pocket of the docked complexes and 3D representations of the binding pocket of the complex before and after the MD simulations are shown in **Fig. 5A** and **5B**, respectively.

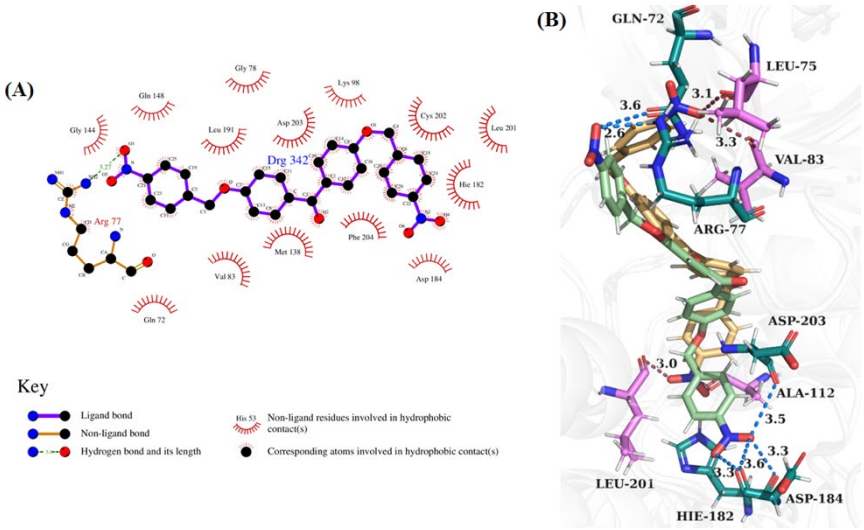


Fig. 5: (A) 2D representation of the binding pocket of the 4G3D_238205 docked complex, and (B) Superimposed 3D representation of the binding pocket of the 4G3D_238205 complex before and after the MD simulations. The 3D interaction diagrams emphasize the change in orientation and stability of the molecules in the binding pockets. Before the MD simulations, the colors used in protein-ligand complexes are: ligand-pale green, protein residues-green, and hydrogen bonding dashes-blue. After the MD simulations, the colors used in protein-ligand complexes are: ligand-yellow, protein residues-pink, and hydrogen bond dashes-maroon.

4.1.2 Case Study 02: As Antiviral Agents against SARS-CoV-2

This case study focuses on the prioritization of phytochemicals found in the Haritaki plant as anticancer agents against SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) using the workflow given in **Fig. 2**. The impact of SARS-CoV-2 on individuals and societies has been profound. SARS, a highly contagious respiratory illness, caused widespread fear and disruption when it emerged in 2020. Its rapid spread highlighted the vulnerabilities of our interconnected world, triggering stringent public health measures and economic repercussions. In one of previous studies, it has been demonstrated that Haritaki inhibits SARS-CoV-2 main protease [183]. SARS-CoV-2 [184, 185], the virus responsible for the COVID-19 [186] pandemic, employs various proteins crucial for its replication and infection. Among these, the main protease (M^{pro}), also known as $3CL^{pro}$ (3-chymotrypsin-like protease) [187], stands out as a prime target for therapeutic intervention. M^{pro} plays a pivotal role in processing viral polyproteins, essential for viral replication, by cleaving them into functional proteins necessary for viral assembly and maturation. Due to its indispensable role in viral replication and the absence of close homologs in humans, M^{pro} presents an attractive target for the development of antiviral drugs. Numerous studies have focused on the design and screening of small molecule inhibitors, peptides, and natural compounds to inhibit M^{pro} activity, aiming to disrupt viral replication and mitigate COVID-19 severity. Targeting M^{pro} holds promise for the development of novel therapeutics against COVID-19 and other coronaviruses, emphasizing the importance of continued research into understanding its structure, function, and inhibition mechanisms.

The M^{pro} (PDB: 7BQY [188]) is a homodimeric protein consisting of two protomers, each composed of three domains (Domains I, II, and III). Domains I and II, spanning residues 8–101 and 102–184, respectively, are structured with six antiparallel β-barrels. Domain III (residues 201–303) forms an antiparallel globular cluster of five α-helices and is connected to domain II via an elongated loop region (residues 185–200). Within the crevice situated between domains I and II lies a catalytic dyad comprised of Cys and His residues, believed to be crucial for proteolytic activity, along with N-terminus residues 1 to 7. The active site of the M^{pro} target protein is given in **Fig. 6**.

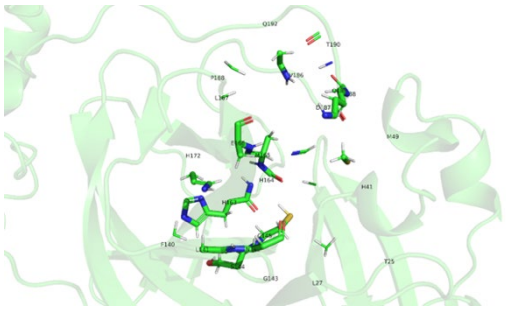


Fig. 6. The active site of main protease M^{pro} showing the active site residues

All the 589 constituents were virtually screened against the target using RASPD+. Docking studies (**Table 2**) of the top 8 compounds obtained from virtual screening were performed utilizing ParDOCK+ and AutoDock4. MD simulations were performed on all the 8 protein-ligand docked complexes for up to 100 nanoseconds. RMSD, RMSF, and RoG ensure the integrity, stability, and compactness of the protein-ligand complex. Out of 8 protein-ligand complexes, four ligands showed favorable results.

Table 2. Binding energies of ligands against main protease M^{pro} (PDB:7BQY) predicted by ParDOCK and AutoDock4 software after docking studies. (Example Link for PubChem: https://pubchem.ncbi.nlm.nih.gov/compound/238205; Example Link for BIMP: http://scfbio-iitd.res.in/bimp/compound.php?compound=BIMP068258)

PDB ID	PubChem ID	BIMP ID	ParDOCK	AutoDock
7BQY	8899	BIMP084541	-7.89	-4.9
7BQY	38222	BIMP058124	-7.43	-6.0
7BQY	238205	BIMP068258	-10.72	-7.7
7BQY	267400	BIMP044029	-7.87	-7.7
7BQY	354330	BIMP084543	-7.98	-6.8
7BQY	427204	BIMP091539	-11.2	-8.6
7BQY	4185717	BIMP084937	-8.23	-8.2
7BQY	442793	BIMP075455	-7.36	-5.7

The hydrogen bond analysis of all 4 protein-ligand complexes showing favorable results is shown in **Fig. 7**. The 7BQY_238205 (**Fig. 7A**) complex forms a maximum of 4 hydrogen bonds, and at least 2 hydrogen bonds are consistent with main chains of Gly146 and Val148 amino acids. 7BQY_267400 (**Fig. 7B**) complex also forms a consistent hydrogen bond with the Hie41 residue. On the other hand, 7BQY_354330 (**Fig. 7C**) and 7BQY_4185717 (**Fig. 7D**) complexes form the maximum number of

hydrogen bonds among all the complexes. In the 7BQY_354330 complex, Thr190, Arg188, Gln192, Gln189, and Glu166 amino acids majorly participate in hydrogen bonding. Similarly, in the 7BQY_4185717 complex, Gln192, Glu166, Thr190, Arg188, and Hie41 amino acids form hydrogen bonds with the ligand.

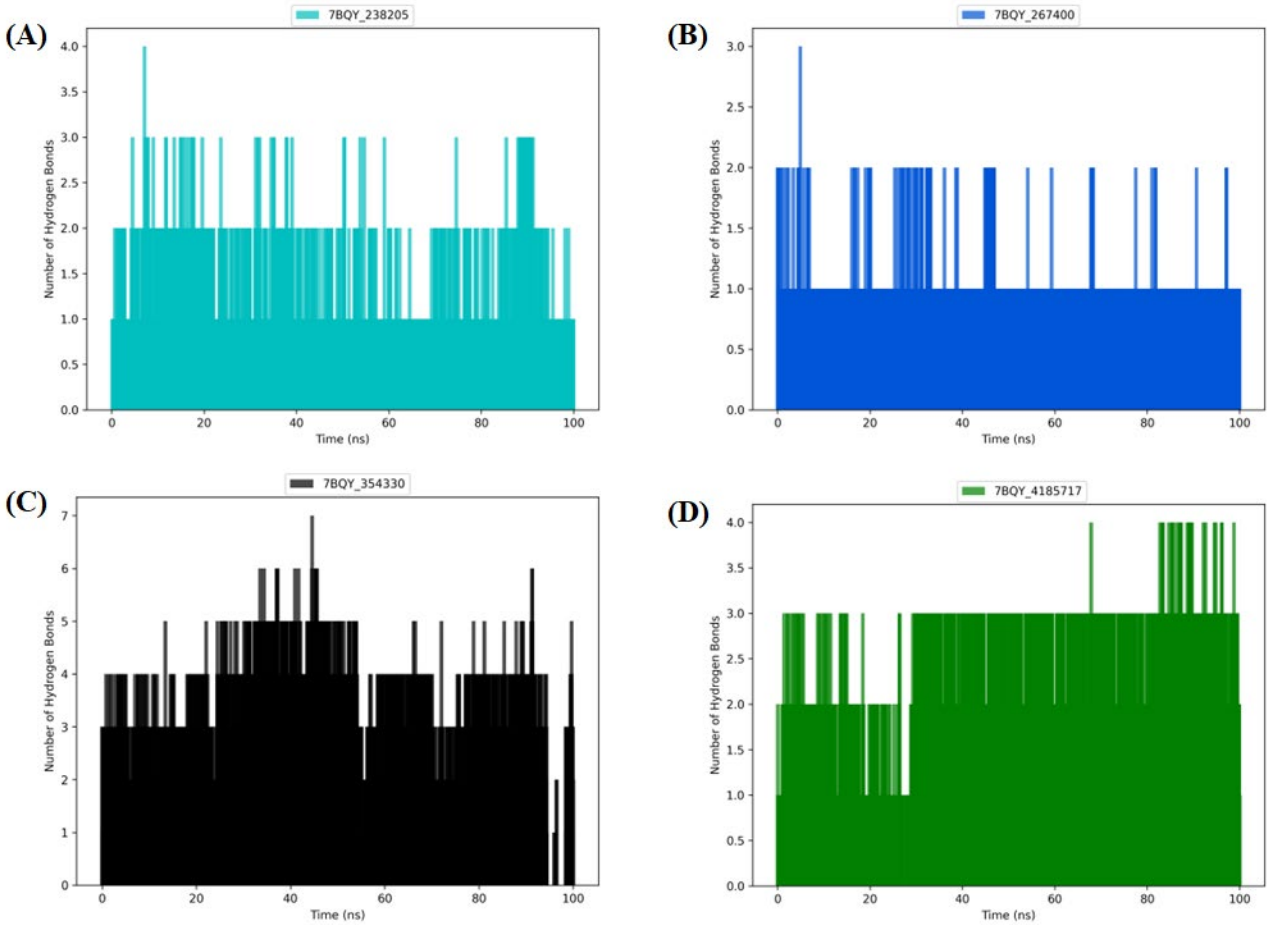
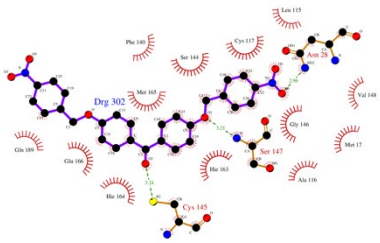
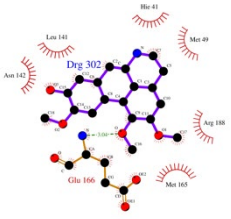


Fig. 7: Hydrogen Bond Analysis of (A) 7BQY_238205, (B) 7BQY_267400, (C) 7BQY_354330 and (D) 7BQY_4185717 complexes.

To gain deeper insights into the interaction between the protein and the ligand, 2D representations of the binding pocket of the docked complexes are shown in **Fig. 8**, and 3D representations of the binding pocket before and after the MD simulations of 100 ns are shown in **Fig. 9**.

(a) 7BQY_238205	(b) 7BQY_267400
	
(c) 7BQY_354330	**(d) 7BQY_4185717**

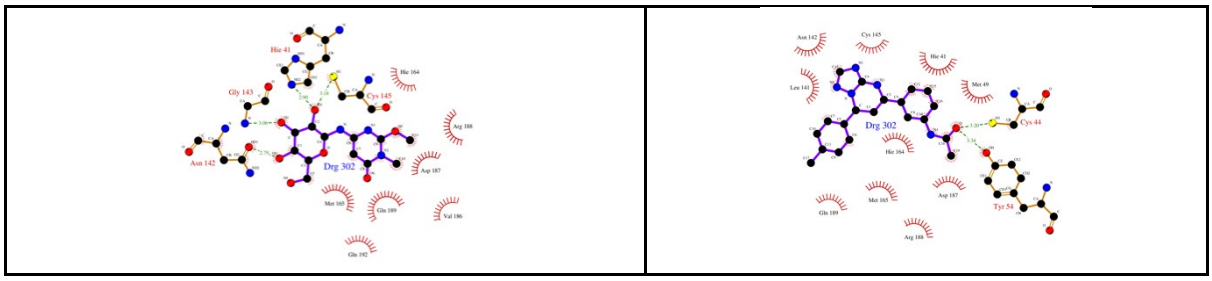


Fig. 8: 2D Representations of docked protein-ligand complexes of 7BQY_238205, 7BQY_267400, 7BQY_354330, and 7BQY_4185717 complexes.

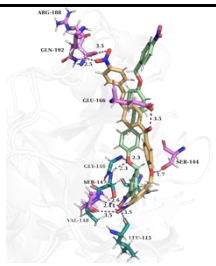
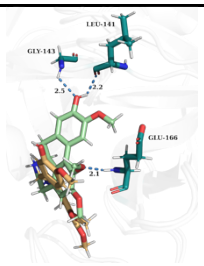
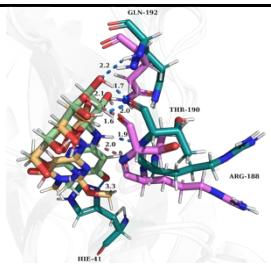
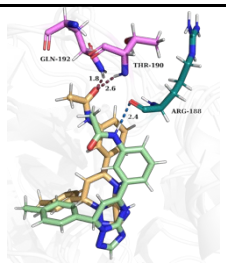
(a) 7BQY_238205	(b) 7BQY_267400
	
(c) 7BQY_354330	**(d) 7BQY_4185717**
	

Fig. 9: Superimposed images of the structures of the protein-ligand complexes before and after the MD simulations. (a) 7BQY_238205, (b) 7BQY_267400, (c) 7BQY_354330 and (d) 7BQY_4185717 complexes. The 3D interaction diagrams emphasize the change in orientation and stability of the molecules in the binding pockets. Before the MD simulations, the colors used in protein-ligand complexes are: ligand-pale green, protein residues-green, and hydrogen bonding dashes-blue; After the MD simulations, the colors used in protein-ligand complexes are: ligand-yellow, protein residues-pink, hydrogen bond dashes-maroon.

Table 3 presents a comprehensive analysis of the binding affinities of various ligands with specific protein structures, utilizing different computational methods to predict the binding free energies. Lower binding free energy values generally indicate stronger interactions, implying better inhibitory potential. In the case of the protein with PDB ID 7BQY, the MMGBSA method predicts a binding free energy of -49.51 kcal/mol, which is indicative of a strong interaction. The MMPBSA and MMBAPPL+ methods predict binding free energies of -11.19 kcal/mol and -10.85 kcal/mol, respectively. Comparing another ligand with PubChem ID 267400 and BIMP ID BIMP044029 binding to the protein 7BQY, the MMGBSA method predicts a binding free energy of -40.99 kcal/mol. The MMPBSA and MMBAPPL+ methods show binding free energies of -6.11 kcal/mol and -9.23 kcal/mol, respectively. For the ligand with PubChem ID 354330 and BIMP ID BIMP084543 binding to the protein 7BQY, the predicted binding free energies are -28.26 kcal/mol (MMGBSA), -5.35 kcal/mol (MMPBSA), and -8.06 kcal/mol (MMBAPPL+). Lastly, the ligand with

PubChem ID 4185717 and BIMP ID BIMP084937 binding to the protein 7BQY shows binding free energies of -44.65 kcal/mol (MMGBSA), -11.51 kcal/mol (MMPBSA), and -10.21 kcal/mol (MMBAPPL+). All these values suggest moderate to strong interaction, reinforcing the ligand's potential as a good inhibitor for the 7BQY protein.

Table 3: Binding free energies of the protein-ligand complexes predicted by MMGBSA, MMPBSA, and MMBAPPL+.

PDB IDs	PubChem IDs	Predicted Binding Free Energies		
		MMGBSA (kcal/mol)	MMPBSA (kcal/mol)	MMBAPPL+ (kcal/mol)
4G3D	238205	-60.2	-2.5	-11.0
7BQY	238205	-49.5	-11.1	-10.8
7BQY	267400	-40.9	-6.1	-9.2
7BQY	354330	-28.2	-5.3	-8.1
7BQY	4185717	-44.6	-11.5	-10.2

4.2 Case Study 03: ZINC Compounds against Hepatitis B Virus as Antivirals

Chronic hepatitis B virus (HBV) infection remains a significant global health challenge [189], affecting approximately 257 million people worldwide and leading to severe liver diseases such as cirrhosis [190] and hepatocellular carcinoma [191]. Current treatment regimens, primarily based on nucleoside and nucleotide analogs like tenofovir and entecavir, aim to suppress HBV replication [192]. However, these therapies often require lifelong administration, can lead to the development of drug-resistant HBV strains, and rarely achieve the complete loss of hepatitis B surface antigen (HBsAg), a crucial therapeutic endpoint. Therefore, there is a pressing need for alternative therapeutic strategies that can effectively target HBsAg to achieve better clinical outcomes.

A case study [193] aimed to identify and evaluate small molecule inhibitors that can bind to HBsAg with high affinity and effectively inhibit its production and the secretion of hepatitis B virions, including strains resistant to existing treatments. To achieve this, a combination of computational virtual screening, molecular docking, and molecular dynamics simulations was employed to sift through a vast library of compounds, seeking those with high binding affinity for HBsAg. The ZINC database, which contains millions of commercially available compounds, was used as the source for the initial screening. This comprehensive approach allowed for the efficient evaluation of the potential anti-HBV activity of numerous compounds.

The study began with the virtual screening of one million molecules from the ZINC database using advanced computational techniques. This process involved several steps. First, virtual screening rapidly evaluated the binding potential of a vast number of compounds against HBsAg. Virtual screening is a crucial initial step that helps narrow down the vast library to a manageable number of promising candidates. Next, the top candidates from the virtual screening were subjected to molecular docking simulations to predict their binding modes and affinities. This step involved placing the molecules in the binding site of HBsAg and assessing their interactions. The docking scores provided insights into how well each compound could potentially inhibit HBsAg.

Table 4: Docking Results by ParDOCK, SwissDock, and AutoDock with the selected ZINC compound with HBsAg.

ZINC ID	IUPAC NAME	ParDOCK	SwissDock	AutoDock
ZINC20451377	(2E)-3-(4-Methoxyphenyl)-1-[4-(3-{[4-(3pyridinylmethyl)-1-piperazinyl]methyl}phenoxy)1-piperidinyl]-2-propen-1-one	-11.1	-8.1	-8.3

To further refine the selections, molecular dynamics simulations were performed. This allowed for the study of the stability of the compound-HBsAg complexes over time, ensuring that the binding was not only strong but also stable. The molecular dynamics simulations provided dynamic insights into the interaction between HBsAg and the potential inhibitors, helping understand the stability and efficacy of the binding in a more realistic biological context.

Table 5. Predicted binding free energies for Molecule 5 against HBsAg calculated using MMBAPPL and AMBER (for average values during MD simulations).

ZINC ID	MMBAPPL Score (kcal/mol)	MMGBSA (kcal/mol)	MMPBSA (kcal/mol)
ZINC20451377	-8.19	-50.01	-16.99

Following the computational phase, a subset of compounds was selected for experimental validation. The selected compounds underwent a series of tests to determine their cytotoxicity profiles and anti-HBV activities. A widely accepted HBV cell culture model was used for these experiments, providing a reliable platform to evaluate the efficacy and safety of the compounds. The cytotoxicity tests were crucial to ensure that the compounds were not harmful to the host cells, while the anti-HBV activity assays measured the compounds' ability to inhibit HBsAg production and HBV virion secretion.

One compound, in particular, stood out during these tests: ZINC20451377. This small molecule demonstrated a high binding affinity for HBsAg, with a KD of 65.3 nM as determined by Surface Plasmon Resonance (SPR) spectroscopy. SPR is a powerful technique that allows real-time measurement of the interaction between molecules, providing precise binding affinity data. The high affinity indicated that ZINC20451377 was very effective at binding to HBsAg, a crucial first step in inhibiting its function.

In vitro studies revealed that ZINC20451377 effectively inhibited HBsAg production and hepatitis B virion secretion at low micromolar concentrations (10 μM). These findings were significant because they demonstrated that ZINC20451377 could effectively reduce the levels of HBsAg and the release of infectious virions, which are critical for the spread and persistence of HBV infection. Moreover, ZINC20451377 was also efficacious against an HBV quadruple mutant (CYEI mutant) resistant to tenofovir, one of the most commonly used antiviral drugs for HBV. This resistance often complicates treatment, making it essential to find new therapies that can overcome it.

The identification of ZINC20451377 highlights a significant advancement in HBV therapeutic research. Its high affinity for HBsAg and efficacy in inhibiting viral production, even in drug-resistant

strains, suggest it could be a potent alternative to existing treatments. The use of computational screening followed by rigorous in vitro validation proved to be an effective strategy for discovering new potential HBV inhibitors.

In conclusion, the study identified a novel small molecule inhibitor, ZINC20451377, which exhibits strong anti-HBV properties. Its ability to bind HBsAg with high affinity and inhibit virus production and secretion, including in tenofovir-resistant strains, underscores its potential as a therapeutic agent. Further preclinical and clinical testing is warranted to establish its efficacy and safety profile fully. Future directions include comprehensive preclinical studies to assess the pharmacokinetics, pharmacodynamics, and toxicity profile of ZINC20451377, as well as mechanistic studies to investigate the molecular mechanisms underlying the inhibition of HBsAg production and HBV secretion. Additionally, initiating clinical trials will be crucial to evaluate the therapeutic potential and safety of ZINC20451377 in patients with chronic HBV infection. Exploring the potential of ZINC20451377 in combination with existing HBV therapies could also enhance efficacy and reduce the risk of resistance development. This case study demonstrates the successful application of computational and experimental approaches in identifying a promising new therapeutic candidate for the treatment of chronic HBV infection.

5. Challenges

Despite the significant advancements in CADD, prioritizing candidate drugs through virtual screening, molecular docking, and MD simulations remains fraught with several challenges [194]. These challenges impact the efficiency and accuracy of identifying viable drug candidates and can hinder the drug discovery process.

i. Accuracy of Computational Predictions: One of the foremost challenges is the accuracy of computational predictions. Virtual screening and molecular docking rely on algorithms that predict the binding affinity of small molecules to target proteins. However, these predictions can be compromised by inaccuracies in the scoring functions used to evaluate binding interactions. Scoring functions often fail to capture the full complexity of molecular interactions, leading to false positives (inactive compounds predicted as active) and false negatives (active compounds overlooked).

ii. Quality of Structural Data: The quality of the structural data used in these simulations is critical. Inaccurate or incomplete structural information about the target protein can significantly impact the outcome of virtual screening and docking studies. Experimental methods such as X-ray crystallography and NMR spectroscopy, which provide these structures, are not always able to capture all possible conformations of a protein, particularly its dynamic states. This limitation can lead to incomplete or misleading representations of the binding site.

iii. Computational Costs and Resources: The computational cost associated with high-throughput virtual screening, extensive molecular docking, and detailed MD simulations is substantial. These processes require significant computational power and time, especially when dealing with large libraries of compounds or conducting long-term simulations. Small research labs with limited access to high-performance computing resources may find these requirements prohibitive.

iv. Integration of Protein Flexibility: Proteins are inherently flexible, undergoing conformational changes that can significantly affect ligand binding. Traditional docking methods often assume a static protein structure, which does not account for this flexibility. While MD simulations can provide

insights into protein dynamics, integrating these dynamic aspects into docking and screening workflows remains a complex task. Accurately modeling and predicting how these conformational changes impact binding interactions is an ongoing challenge.

v. Correlation Between *in silico* and *in vitro* Results: There is often a gap between *in silico* predictions and *in vitro* or *in vivo* experimental results. Computational models may not always accurately replicate the biological environment, leading to discrepancies between predicted and actual biological activity. This lack of correlation necessitates extensive experimental validation of computational hits, which can be time-consuming and resource-intensive.

vi. Handling Large Data Volumes: Virtual screening, docking, and MD simulations generate vast amounts of data. Efficiently managing, analyzing, and interpreting this data to prioritize the most promising drug candidates is a significant challenge. Advanced data analytics and machine learning techniques are increasingly employed to tackle this issue, but these approaches also require expertise and robust computational infrastructure.

vii. Validation of Computational Methods: Continuous validation and benchmarking of computational methods against experimental data are essential to maintain their reliability. However, the iterative process of refining algorithms and validating them with experimental data is resource-intensive and requires close collaboration between computational and experimental researchers.

Addressing these challenges necessitates a multifaceted approach involving advancements in computational algorithms, improved integration of experimental and *in silico* data, enhanced computational resources, and robust validation frameworks. By overcoming these hurdles, the drug discovery process can become more efficient, accurate, and ultimately successful in identifying and prioritizing candidate drugs.

6. Future Perspectives

The future of drug repurposing [195] through virtual screening, docking, and simulations is set to transform pharmaceutical research and development, leveraging advanced computational methods to uncover new therapeutic uses for existing drugs. These techniques will play an increasingly important role in identifying repurposing opportunities addressing the need for efficient and cost-effective drug development. As computational power and algorithm efficiency continue to improve, high-throughput virtual screening will enable the rapid evaluation of extensive libraries of approved drugs against various targets. This capability is particularly valuable for identifying and repurposing candidates for emerging and neglected diseases where traditional drug development timelines are too slow to meet urgent public health needs.

MD simulations will also undergo significant advancements, providing deeper insights into drug-target interactions. As simulation algorithms and force fields become more sophisticated, they will enable more accurate modeling of complex biological systems. These improvements will help researchers better understand the dynamic behavior of drugs within different biological contexts, revealing potential repurposing opportunities that static models might miss. Enhanced MD simulations will elucidate the mechanisms by which drugs interact with their targets, facilitating the optimization of existing compounds for new therapeutic applications.

In the context of personalized medicine [196, 197], the discovery of new molecules and drug repurposing will increasingly intersect with individualized treatment approaches. By integrating patient-specific genomic and proteomic data into virtual screening and docking workflows, researchers can identify existing drugs that are most likely to be effective for individual patients. This personalized approach will optimize therapeutic outcomes and minimize adverse effects, offering more targeted and patient-centric treatment options. The convergence of drug repurposing and personalized medicine holds great promise for improving healthcare by delivering tailored therapies that are both effective and safe.

Collaborative databases and open science initiatives will play a crucial role in the future of drug discovery. Shared repositories of high-quality molecular and clinical data will enable researchers worldwide to contribute to and benefit from collective efforts. Open access to computational tools and datasets will democratize the drug discovery process, fostering innovation and accelerating the identification of new therapeutic uses for existing drugs. Collaborative platforms will facilitate the pooling of expertise and resources, driving forward the field and ultimately enhancing global health outcomes.

In parallel, the discovery of drugs from natural products [198] is set to be reinvigorated by modern computational techniques. Natural products derived from diverse biological sources, such as plants, microorganisms, and marine organisms, offer a rich chemical space that is often underexplored by synthetic libraries. Virtual screening, docking, and simulations will streamline the identification and optimization of bioactive compounds from these natural sources. As high-throughput virtual screening becomes more advanced, researchers will be able to efficiently explore vast libraries of natural products, rapidly identifying those with potential therapeutic applications.

Structural elucidation and modeling of natural products, which have traditionally posed challenges due to their complexity, will be greatly enhanced by improved molecular docking and simulation techniques. These advancements will enable more accurate modeling of the intricate structures of natural products, facilitating the identification of their interactions with biological targets. By leveraging the unique structural features and bioactivities of natural products, researchers can discover new drug candidates with novel mechanisms of action, offering innovative solutions to pressing medical challenges.

The seamless integration of virtual screening, docking, and simulations [199, 200] into new molecule and natural product discovery workflows will also foster the development of multi-target drugs. Complex diseases, such as cancer and neurodegenerative disorders, often involve multiple biological pathways. Identifying compounds that can modulate several targets simultaneously can enhance therapeutic efficacy and reduce the likelihood of resistance. This multi-target approach, supported by advanced computational techniques, will open new avenues for treating multifactorial diseases more effectively.

7. Conclusion

The adoption of CADD techniques has revolutionized the drug development process, providing a robust, efficient, and cost-effective framework for prioritizing drug candidates. By leveraging virtual screening, molecular docking, and MD simulations, researchers can predict and evaluate the interactions between drug candidates and biological targets with high precision. Virtual screening

expedites the identification of promising molecules from extensive compound libraries, while molecular docking elucidates their preferred binding orientations and modes within target proteins. MD simulations further contribute by offering a dynamic understanding of the stability and conformational adaptability of protein-ligand complexes over time in a near-physiological milieu. Collectively, these computational methods enhance the rational design and optimization of therapeutics, ultimately increasing the success rates and reducing the financial burden of drug discovery. This chapter underscores the transformative impact of CADD on modern drug discovery, illustrating its efficacy through specific examples and case studies that highlight its vital role in advancing medical science.

8. Acknowledgments

Funding from the Department of Biotechnology, Govt of India and National Supercomputing Mission administered by DST, DEITY, Govt of India & CDAC to SCFBio, IIT Delhi, is gratefully acknowledged.

9. References

1. Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *British journal of pharmacology*, *162*(6), 1239–1249.

2. Pirintsos, S., Panagiotopoulos, A., Bariotakis, M., Daskalakis, V., Lionis, C., Sourvinos, G., Karakasiliotis, I., Kampa, M., & Castanas, E. (2022). From Traditional Ethnopharmacology to Modern Natural Drug Discovery: A Methodology Discussion and Specific Examples. *Molecules (Basel, Switzerland)*, *27*(13), 4060.

3. Choudhury, C., Arul Murugan, N., & Priyakumar, U. D. (2022). Structure-based drug repurposing: Traditional and advanced AI/ML-aided methods. *Drug discovery today*, *27*(7), 1847–1861.

4. Hopfinger A. J. (1985). Computer-assisted drug design. *Journal of medicinal chemistry*, *28*(9), 1133–1139.

5. Marshall G. R. (1987). Computer-aided drug design. *Annual review of pharmacology and toxicology*, *27*, 193–213.

6. Muegge, I., Bergner, A., & Kriegl, J. M. (2017). Computer-aided drug design at Boehringer Ingelheim. *Journal of computer-aided molecular design*, *31*(3), 275–285.

7. Eder, J., & Herrling, P. L. (2016). Trends in Modern Drug Discovery. *Handbook of experimental pharmacology*, *232*, 3–22.

8. Pina, A. S., Hussain, A., & Roque, A. C. (2009). An historical overview of drug discovery. *Methods in molecular biology (Clifton, N.J.)*, *572*, 3–12.

9. Drews J. (2000). Drug discovery: a historical perspective. *Science (New York, N.Y.)*, *287*(5460), 1960–1964.

10. Schlick T. (2010). Biomolecular Structure and Modeling: Historical Perspective. *Molecular Modeling and Simulation: An Interdisciplinary Guide: An Interdisciplinary Guide*, *21*, 1–40.

11. Field M. J. (2015). Technical advances in molecular simulation since the 1980s. *Archives of biochemistry and biophysics*, *582*, 3–9.

12. Dar, K. B., Bhat, A. H., Amin, S., Hamid, R., Anees, S., Anjum, S., Reshi, B. A., Zargar, M. A., Masood, A., & Ganie, S. A. (2018). Modern Computational Strategies for Designing Drugs to Curb Human Diseases: A Prospect. *Current topics in medicinal chemistry*, *18*(31), 2702–2719.

13. Dias, R., & de Azevedo, W. F., Jr (2008). Molecular docking algorithms. *Current drug targets*, *9*(12), 1040–1047.

14. da Silva Rocha, S. F. L., Olanda, C. G., Fokoue, H. H., & Sant'Anna, C. M. R. (2019). Virtual Screening Techniques in Drug Discovery: Review and Recent Applications.

15. Song, C. M., Lim, S. J., & Tong, J. C. (2009). Recent advances in computer-aided drug design. *Briefings in bioinformatics*, *10*(5), 579–591.

16. Vemula, D., Jayasurya, P., Sushmitha, V., Kumar, Y. N., & Bhandari, V. (2023). CADD, AI and ML in drug discovery: A comprehensive review. *European journal of pharmaceutical sciences: official journal of the European Federation for Pharmaceutical Sciences*, *181*, 106324.

17. Selvaraj, C., Chandra, I., & Singh, S. K. (2022). Artificial intelligence and machine learning approaches for drug design: challenges and opportunities for the pharmaceutical industries. *Molecular diversity*, *26*(3), 1893–1913.

18. Singh, S., Kumar, R., Payra, S., & Singh, S. K. (2023). Artificial Intelligence and Machine Learning in Pharmacological Research: Bridging the Gap Between Data and Drug Discovery. *Cureus*, *15*(8), e44359.

19. Adasme, M. F., Parisi, D., Sveshnikova, A., & Schroeder, M. (2021). Structure-based drug repositioning: Potential and limits. *Seminars in cancer biology*, *68*, 192–198.

20. Lengauer, T., & Rarey, M. (1996). Computational methods for biomolecular docking. *Current opinion in structural biology*, *6*(3), 402–406.

21. Sahragardjoonegani, B., Beall, R. F., Kesselheim, A. S., & Hollis, A. (2021). Repurposing existing drugs for new uses: a cohort study of the frequency of FDA-granted new indication exclusivities since 1997. *Journal of pharmaceutical policy and practice*, *14*(1), 3.

22. Erlanson D. A. (2012). Introduction to fragment-based drug discovery. *Topics in current chemistry*, *317*, 1–32.

23. Stahura, F. L., & Bajorath, J. (2005). New methodologies for ligand-based virtual screening. *Current pharmaceutical design*, *11*(9), 1189–1202.

24. Li, Q., & Shah, S. (2017). Structure-Based Virtual Screening. *Methods in molecular biology (Clifton, N.J.)*, *1558*, 111–124.

25. Wang, J., Yin, J., Nguyen, M. H., Wang, J., & Xu, W. (2024). Editorial: Big scientific data analytics on HPC and cloud. *Frontiers in big data*, *7*, 1353988.

26. Ozaki, H., Imaizumi, Y., Oishi, K., & Kohama, K. (2001). *Nihon yakurigaku zasshi. Folia pharmacologica Japonica*, *118*(3), 187–196.

27. Szymański, P., Markowicz, M., & Mikiciuk-Olasik, E. (2012). Adaptation of high-throughput screening in drug discovery-toxicological screening tests. *International journal of molecular sciences*, *13*(1), 427–452.

28. Giordano, D., Biancaniello, C., Argenio, M. A., & Facchiano, A. (2022). Drug Design by Pharmacophore and Virtual Screening Approach. *Pharmaceuticals (Basel, Switzerland)*, *15*(5), 646.

29. Seidel, T., Wieder, O., Garon, A., & Langer, T. (2020). Applications of the Pharmacophore Concept in Natural Product inspired Drug Design. *Molecular informatics*, *39*(11), e2000059.

30. Tang, K. W. K., Millar, B. C., & Moore, J. E. (2023). Antimicrobial Resistance (AMR). *British journal of biomedical science*, *80*, 11387.

31. Salam, M. A., Al-Amin, M. Y., Salam, M. T., Pawar, J. S., Akhter, N., Rabaan, A. A., & Alqumber, M. A. A. (2023). Antimicrobial Resistance: A Growing Serious Threat for Global Public Health. *Healthcare (Basel, Switzerland)*, *11*(13), 1946.

32. Prestinaci, F., Pezzotti, P., & Pantosti, A. (2015). Antimicrobial resistance: a global multifaceted phenomenon. *Pathogens and global health*, *109*(7), 309–318.

33. Komolafe O. O. (2003). Antibiotic resistance in bacteria - an emerging public health problem. *Malawi medical journal: the journal of Medical Association of Malawi*, *15*(2), 63–67.

34. Bashir, N., Dablool, A. S., Khan, M. I., Almalki, M. G., Ahmed, A., Mir, M. A., Hamdoon, A. A. E., Elawad, M. A., Mosa, O. F., Niyazov, L. N., Elkhalifa, M. E. M., Alghamdi, M. A., Anwar, A., & Ayaz, M. (2023). Antibiotics resistance as a major public health concern: A pharmaco-epidemiological study to evaluate prevalence and antibiotics susceptibility-resistance pattern of bacterial isolates from multiple teaching hospitals. *Journal of infection and public health*, *16 Suppl 1*, 61–68.

35. Ventola C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *P & T: a peer-reviewed journal for formulary management*, *40*(4), 277–283.

36. Chinemerem Nwobodo, D., Ugwu, M. C., Oliseloke Anie, C., Al-Ouqaili, M. T. S., Chinedu Ikem, J., Victor Chigozie, U., & Saki, M. (2022). Antibiotic resistance: The challenges and some emerging strategies for tackling a global menace. *Journal of clinical laboratory analysis*, *36*(9), e24655.

37. Oselusi, S. O., Dube, P., Odugbemi, A. I., Akinyede, K. A., Ilori, T. L., Egieyeh, E., Sibuyi, N. R., Meyer, M., Madiehe, A. M., Wyckoff, G. J., & Egieyeh, S. A. (2024). The role and potential of computer-aided drug discovery strategies in the discovery of novel antimicrobials. *Computers in biology and medicine*, *169*, 107927.

38. Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W., Jr (2013). Computational methods in drug discovery. *Pharmacological reviews*, *66*(1), 334–395.

39. Reddy, A. S., Pati, S. P., Kumar, P. P., Pradeep, H. N., & Sastry, G. N. (2007). Virtual screening in drug discovery -- a computational perspective. *Current protein & peptide science*, *8*(4), 329–351.

40. Lavecchia, A., & Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: a critical review. *Current medicinal chemistry*, *20*(23), 2839–2860.

41. Panwar, U., Murali, A., Khan, M. A., Selvaraj, C., & Singh, S. K. (2024). Virtual Screening Process: A Guide in Modern Drug Designing. *Methods in molecular biology (Clifton, N.J.)*, *2714*, 21–31.

42. Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. *Methods in molecular biology (Clifton, N.J.)*, *443*, 365–382.

43. Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design*, *7*(2), 146–157.

44. Agu, P. C., Afiukwa, C. A., Orji, O. U., Ezeh, E. M., Ofoke, I. H., Ogbu, C. O., Ugwuja, E. I., & Aja, P. M. (2023). Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. *Scientific reports*, *13*(1), 13398.

45. Sahu, D., Rathor, L. S., Dwivedi, S. D., Shah, K., Chauhan, N. S., Singh, M. R., & Singh, D. (2024). A Review on Molecular Docking As an Interpretative Tool for Molecular Targets in Disease Management. *Assay and drug development technologies*, *22*(1), 40–50.

46. Lindahl E. R. (2008). Molecular Dynamics simulations. *Methods in molecular biology (Clifton, N.J.)*, *443*, 3–23.

47. Hernández-Rodríguez, M., Rosales-Hernández, M. C., Mendieta-Wejebe, J. E., Martínez-Archundia, M., & Basurto, J. C. (2016). Current Tools and Methods in Molecular Dynamics Simulations for Drug Design. *Current medicinal chemistry*, *23*(34), 3909–3924.

48. Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. *Neuron*, *99*(6), 1129–1143.

49. Wu, X., Xu, L. Y., Li, E. M., & Dong, G. (2022). Application of Molecular Dynamics simulation in biomedicine. *Chemical biology & drug design*, *99*(5), 789–800.

50. Jayaram, B., Latha, N., Jain, T., Sharma, P., Gandhimathi, A., & Pandey, V. S. (2006). Sanjeevini: A comprehensive Active-Site directed lead design software. *Ind. J. Chem.*, *45A,* 1834–1837.

51. Jayaram, B., Singh, T., Mukherjee, G., Mathur, A., Shekhar, S., & Shekhar, V. (2012). Sanjeevini: a freely accessible web-server for target directed lead molecule discovery. *BMC bioinformatics*, *13 Suppl 17*(Suppl 17), S7.

52. Singh, A., Shekhar, S., & Jayaram, B. (2020). CADD: Some success stories from Sanjeevini and the way forward, in Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design. *Springer Nature*, Singapore Pte Ltd.

53. Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on Molecular Dynamics simulations. *Journal of chemical information and modeling*, *51*(1), 69–82.

54. Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *Journal of computational chemistry*, *32*(5), 866–877.

55. Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery*, *10*(5), 449–461.

56. Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., & Hou, T. (2019). End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chemical reviews*, *119*(16), 9478–9508.

57. Homeyer, N., & Gohlke, H. (2012). Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Molecular informatics*, *31*(2), 114–122.

58. Wang, E., Liu, H., Wang, J., Weng, G., Sun, H., Wang, Z., Kang, Y., & Hou, T. (2020). Development and Evaluation of MM/GBSA Based on a Variable Dielectric GB Model for Predicting Protein-Ligand Binding Affinities. *Journal of chemical information and modeling*, *60*(11), 5353–5365.

59. Doytchinova I. (2022). Drug Design-Past, Present, Future. *Molecules (Basel, Switzerland)*, *27*(5), 1496.

60. Torres, P. H. M., Sodero, A. C. R., Jofily, P., & Silva-Jr, F. P. (2019). Key Topics in Molecular Docking for Drug Design. *International journal of molecular sciences*, *20*(18), 4574.

61. Wang, Z., Sun, H., Shen, C., Hu, X., Gao, J., Li, D., et al. (2020). Combined strategies in structure-based virtual screening. *Phys. Chem. Chem. Phys.* 22, 3149–3159.

62. Liang, J., Tsui, V., Van Abbema, A., Bao, L., Barrett, K., Beresini, M., et al. (2013a). Lead identification of novel and selective TYK2 inhibitors. *Eur. J. Med. Chem.* 67, 175–187.

63. Shin, W. H., Zhu, X., Bures, M. G., & Kihara, D. (2015). Three-dimensional compound comparison methods and their application in drug discovery. *Molecules (Basel, Switzerland)*, *20*(7), 12841–12862.

64. von Behren, M. M., & Rarey, M. (2017). Ligand-based virtual screening under partial shape constraints. *Journal of computer-aided molecular design*, *31*(4), 335–347.

65. Li, Q., & Shah, S. (2017). Structure-Based Virtual Screening. *Methods in molecular biology (Clifton, N.J.)*, *1558*, 111–124.

66. Kireev D. (2016). Structure-Based Virtual Screening of Commercially Available Compound Libraries. *Methods in molecular biology (Clifton, N.J.)*, *1439*, 65–76.

67. Irwin, J. J., & Shoichet, B. K. (2005). ZINC--a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, *45*(1), 177–182.

68. Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., Moroz, Y. S., Mayfield, J., & Sayle, R. A. (2020). ZINC20-A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *Journal of chemical information and modeling*, *60*(12), 6065–6073.

69. Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., & Bryant, S. H. (2016). PubChem Substance and Compound databases. *Nucleic acids research*, *44*(D1), D1202–D1213.

70. Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., & Overington, J. P. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic acids research*, *40*(Database issue), D1100–D1107.

71. Cheng, T., Zhao, Y., Li, X., Lin, F., Xu, Y., Zhang, X., Li, Y., Wang, R., & Lai, L. (2007). Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *Journal of chemical information and modeling*, *47*(6), 2140–2148.

72. Scott A. Wildman and Gordon M. Crippen. (1999). *Journal of Chemical Information and Computer Sciences* **1999** *39* (5), 868-873.

73. Holderbach, S., Adam, L., Jayaram, B., Wade, R. C., & Mukherjee, G. (2020). RASPD+: Fast Protein-Ligand Binding Free Energy Prediction Using Simplified Physicochemical Features. *Frontiers in molecular biosciences*, *7*, 601065.

74. Mukherjee, G., & Jayaram, B. (2013). A rapid identification of hit molecules for target proteins via physico-chemical descriptors. *Physical chemistry chemical physics : PCCP*, *15*(23), 9107–9116.

75. Jain, T., & Jayaram, B. (2005). An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes. *FEBS letters*, *579*(29), 6659–6666.

76. Soni, A., Bhat, R., & Jayaram, B. (2020). Improving the binding affinity estimations of protein-ligand complexes using machine-learning facilitated force field method. *Journal of computer-aided molecular design*, *34*(8), 817–830.

77. Gupta, A., Gandhimathi, A., Sharma, P., & Jayaram, B. (2007). ParDOCK: an all atom energy based Monte Carlo docking protocol for protein-ligand complexes. *Protein and peptide letters*, *14*(7), 632–646.

78. Cerqueira, N. M., Gesto, D., Oliveira, E. F., Santos-Martins, D., Brás, N. F., Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2015). Receptor-based virtual screening protocol for drug discovery. *Archives of biochemistry and biophysics*, *582*, 56–67.

79. Panwar, U., Murali, A., Khan, M. A., Selvaraj, C., & Singh, S. K. (2024). Virtual Screening Process: A Guide in Modern Drug Designing. *Methods in molecular biology (Clifton, N.J.)*, *2714*, 21–31.

80. Pennello, G., Pantoja-Galicia, N., & Evans, S. (2016). Comparing diagnostic tests on benefit-risk. *Journal of biopharmaceutical statistics*, *26*(6), 1083–1097.

81. von Euler-Chelpin, M., Kuchiki, M., & Vejborg, I. (2014). Increased risk of breast cancer in women with false-positive test: the role of misclassification. *Cancer epidemiology*, *38*(5), 619–622.

82. Spyrakis, F., Benedetti, P., Decherchi, S., Rocchia, W., Cavalli, A., Alcaro, S., Ortuso, F., Baroni, M., & Cruciani, G. (2015). A Pipeline To Enhance Ligand Virtual Screening: Integrating Molecular Dynamics and Fingerprints for Ligand and Proteins. *Journal of chemical information and modeling*, *55*(10), 2256–2274.

83. Singh, K., Bhushan, B., & Singh, B. (2024). Advances in Drug Discovery and Design using Computer-aided Molecular Modeling. *Current computer-aided drug design*, *20*(5), 697–710.

84. Brooijmans, N., Sharp, K. A., & Kuntz, I. D. (2002). Stability of macromolecular complexes. *Proteins*, *48*(4), 645–653.

85. Wei, B. Q., Weaver, L. H., Ferrari, A. M., Matthews, B. W., & Shoichet, B. K. (2004). Testing a flexible-receptor docking algorithm in a model binding site. *Journal of molecular biology*, *337*(5), 1161–1182.

86. McConkey, B. J., Sobolev, V., & Edelman, M. (2002). The performance of current methods in ligand-protein docking. *Current Science*, *83*(7), 845-856.

87. Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews. Drug discovery*, *3*(11), 935–949.

88. Halperin, I., Ma, B., Wolfson, H., & Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, *47*(4), 409–443.

89. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of molecular biology*, *161*(2), 269–288.

90. Ewing, T. J., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *Journal of computer-aided molecular design*, *15*(5), 411–428.

91. Choi V. (2005). Yucca: an efficient algorithm for small-molecule docking. *Chemistry & biodiversity*, *2*(11), 1517–1524.

92. McGann, M. R., Almond, H. R., Nicholls, A., Grant, J. A., & Brown, F. K. (2003). Gaussian docking functions. *Biopolymers*, *68*(1), 76–90.

93. Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*, *261*(3), 470–489.

94. Jain A. N. (2003). Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of medicinal chemistry*, *46*(4), 499–511.

95. Miller, M. D., Kearsley, S. K., Underwood, D. J., & Sheridan, R. P. (1994). FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *Journal of computer-aided molecular design*, *8*(2), 153–174.

96. Welch, W., Ruppert, J., & Jain, A. N. (1996). Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chemistry & biology*, *3*(6), 449–462.

97. Goodsell, D. S., & Olson, A. J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins*, *8*(3), 195–202.

98. Totrov, M., & Abagyan, R. (1997). Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins*, *Suppl 1*, 215–220.

99. Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*, *267*(3), 727–748.

100. Liu, M., & Wang, S. (1999). MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. *Journal of computer-aided molecular design*, *13*(5), 435–451.

101. Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2002). A review of protein-small molecule docking methods. *Journal of computer-aided molecular design*, *16*(3), 151–166.

102. Reddy, S.V.G., Reddy, K.T., Kumari, V.V. (2015). The Computational Analysis of Protein – Ligand Docking with Diverse Genetic Algorithm Parameters. In: Satapathy, S., Govardhan, A., Raju, K., Mandal, J. (eds) Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1. Advances in Intelligent Systems and Computing, vol 337. Springer, Cham.

103. METROPOLIS, N., & ULAM, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, *44*(247), 335–341.

104. Marinova, V., Dodd, L., Lee, S. J., Wood, G. P. F., Marziano, I., & Salvalaglio, M. (2021). Identifying Conformational Isomers of Organic Molecules in Solution via Unsupervised Clustering. *Journal of chemical information and modeling*, *61*(5), 2263–2273.

105. Karplus, M., & McCammon, J. A. (2002). Molecular Dynamics simulations of biomolecules. *Nature structural biology*, *9*(9), 646–652.

106. McCammon, J. A., Gelin, B. R., & Karplus, M. (1977). Dynamics of folded proteins. *Nature*, *267*(5612), 585–590.

107. Levitt, M., & Lifson, S. (1969). Refinement of protein conformations using a macromolecular energy minimization procedure. *Journal of molecular biology*, *46*(2), 269–279.

108. Lifson, S., & Warshel, A.; Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and *n*-Alkane Molecules. *J. Chem. Phys.* 1 December 1968; 49 (11): 5116–5129.

109. Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.*, *118*, 11225–11236.

110. Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., Wang, L., Lupyan, D., Dahlgren, M. K., Knight, J. L., Kaus, J. W., Cerutti, D. S., Krilov, G., Jorgensen, W. L., Abel, R., & Friesner, R. A. (2016). OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of chemical theory and computation*, *12*(1), 281–296.

111. Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., & Mackerell, A. D., Jr (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry*, *31*(4), 671–690.

112. Vanommeslaeghe, K., & MacKerell, A. D., Jr (2012). Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *Journal of chemical information and modeling*, *52*(12), 3144–3154.

113. Yu, W., He, X., Vanommeslaeghe, K., & MacKerell, A. D., Jr (2012). Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of computational chemistry*, *33*(31), 2451–2468.

114. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of computational chemistry*, *25*(9), 1157–1174.

115. Wang, J., Wang, W., Kollman, P. A., & Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics & modelling*, *25*(2), 247–260.

116. Halgren, T. A. (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.*, *17*, 490–519.

117. Halgren, T. A. (1996) Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.*, *17*, 520–552.

118. Halgren, T. A. (1996) Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.*, *17*, 553–586.

119. Halgren, T. A., & Nachbar, R. B. (1996) Merck molecular force field. IV. conformational energies and geometries for MMFF94. *J. Comput. Chem.*, *17*, 587–615.

120. Halgren, T. A. (1996) Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.*, *17*, 616–641.

121. Daura, X., Mark, A. E., & Van Gunsteren, W. F. (1998) Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem.*, *19*, 535–547.

122. Schuler, L. D., Daura, X., & van Gunsteren, W. F. (2001) An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.*, *22*, 1205–1218.

123. Oostenbrink, C., Villa, A., Mark, A. E., & van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry*, *25*(13), 1656–1676.

124. Horta, B. A., Fuchs, P. F., van Gunsteren, W. F., & Hünenberger, P. H. (2011). New Interaction Parameters for Oxygen Compounds in the GROMOS Force Field: Improved Pure-Liquid and Solvation Properties for Alcohols, Ethers, Aldehydes, Ketones, Carboxylic Acids, and Esters. *Journal of chemical theory and computation*, *7*(4), 1016–1031.

125. Horta, B. A., Merz, P. T., Fuchs, P. F., Dolenc, J., Riniker, S., & Hünenberger, P. H. (2016). A GROMOS-Compatible Force Field for Small Organic Molecules in the Condensed Phase: The 2016H66 Parameter Set. *Journal of chemical theory and computation*, *12*(8), 3825–3850.

126. Vanommeslaeghe, K., Raman, E. P., & MacKerell, A. D., Jr (2012). Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *Journal of chemical information and modeling*, *52*(12), 3155–3168.

127. Malde, A. K., Zuo, L., Breeze, M., Stroet, M., Poger, D., Nair, P. C., Oostenbrink, C., & Mark, A. E. (2011). An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *Journal of chemical theory and computation*, *7*(12), 4026–4037.

128. van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W., & Vriend, G. (1996). PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *Journal of computer-aided molecular design*, *10*(3), 255–262.

129. Schüttelkopf, A. W., & van Aalten, D. M. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta crystallographica. Section D, Biological crystallography*, *60*(Pt 8), 1355–1363.

130. Yesselman, J. D., Price, D. J., Knight, J. L., & Brooks, C. L., 3rd (2012). MATCH: an atom-typing toolset for molecular mechanics force fields. *Journal of computational chemistry*, *33*(2), 189–202.

131. Zoete, V., Cuendet, M. A., Grosdidier, A., & Michielin, O. (2011). SwissParam: a fast force field generation tool for small organic molecules. *Journal of computational chemistry*, *32*(11), 2359–2368.

132. Izadi, S., & Onufriev, A. V. (2016). Accuracy limit of rigid 3-point water models. *The Journal of chemical physics*, *145*(7), 074501.

133. Toton, D., Lorenz, C. D., Rompotis, N., Martsinovich, N., & Kantorovich, L. (2010). Temperature control in molecular dynamic simulations of non-equilibrium processes. *Journal of physics. Condensed matter : an Institute of Physics journal*, *22*(7), 074205.

134. Shinoda, W., & Mikami, M. (2003). Rigid-body dynamics in the isothermal-isobaric ensemble: a test on the accuracy and computational efficiency. *Journal of computational chemistry*, *24*(8), 920–930.

135. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr, Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of computational chemistry*, *26*(16), 1668–1688.

136. Case, D. A., Aktulga, H. M., Belfon, K., Cerutti, D. S., Cisneros, G. A., Cruzeiro, V. W. D., Forouzesh, N., Giese, T. J., Götz, A. W., Gohlke, H., Izadi, S., Kasavajhala, K., Kaymak, M.

C., King, E., Kurtzman, T., Lee, T. S., Li, P., Liu, J., Luchko, T., Luo, R., … Merz, K. M., Jr (2023). AmberTools. *Journal of chemical information and modeling*, *63*(20), 6183–6191.

137. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of computational chemistry*, *26*(16), 1701–1718.

138. Singh, T., Adekoya, O. A., & Jayaram, B. (2015). Understanding the binding of inhibitors of matrix metalloproteinases by molecular docking, quantum mechanical calculations, molecular dynamics simulations, and a MMGBSA/MMBappl study. *Molecular bioSystems*, *11*(4), 1041–1051.

139. Rathore, R. S., Sumakanth, M., Reddy, M. S., Reddanna, P., Rao, A. A., Erion, M. D., & Reddy, M. R. (2013). Advances in binding free energies calculations: QM/MM-based free energy perturbation method for drug design. *Current pharmaceutical design*, *19*(26), 4674–4686.

140. Lawrenz, M., Baron, R., Wang, Y., & McCammon, J. A. (2012). Independent-Trajectory Thermodynamic Integration: a practical guide to protein-drug binding free energy calculations using distributed computing. *Methods in molecular biology (Clifton, N.J.)*, *819*, 469–486.

141. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic acids research*, *28*(1), 235–242.

142. Singh, T., Biswas, D., & Jayaram, B. (2011). AADS--an automated active site identification, docking, and scoring protocol for protein targets based on physicochemical descriptors. *Journal of chemical information and modeling*, *51*(10), 2515–2527.

143. Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. (2018). CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic acids research*, *46*(W1), W363–W367.

144. Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009). Fpocket: an open source platform for ligand pocket detection. *BMC bioinformatics*, *10*, 168.

145. Hassan Bulbul, M. R., Uddin Chowdhury, M. N., Naima, T. A., Sami, S. A., Imtiaj, M. S., Huda, N., & Uddin, M. G. (2022). A comprehensive review on the diverse pharmacological perspectives of *Terminalia chebula* Retz. *Heliyon*, *8*(8), e10220.

146. Jha, A. K., & Sit, N. (2023). Methods of extraction of bioactive compounds from *Terminalia chebula* (Haritaki) and their application in food and pharmaceutical industry: A review. *Food Bioengineering*, 2, 139–150.

147. World Health Organization (WHO). (2002). *Traditional medicine: Growing needs and potential*. World Health Organization.

148. Beuscher, N., Bodinet, C., Neumann-Haefelin, D., Marston, A., & Hostettmann, K. (1994). Antiviral activity of African medicinal plants. *Journal of ethnopharmacology*, *42*(2), 101–109.

149. Yin, W., & Wang, S. (2008). The character of *Pueraria lobata* starch, chemical composition and pharmacological function of *Pueraria lobata*. *Grain Processing*, 1, i84–i86.

150. Chowdhury, N. S., Alam, M. B., Haque, A. S. M. T., Zahan, R., Mazumder, M. E. H., & Haque, M. E. (2011). In vitro free radical scavenging and thrombolytic activities of Bangladeshi aquatic plant *Aponogeton undulatus Roxb*. *Global Journal of Pharmacology*, *5*(1), 27–32.

151. Chattopadhyay, R. R., & Bhattacharyya, S. (2007). Plant review *Terminalia chebula*: An update. *Pharmacognosy Reviews*, *1*(1), 151–156.

152. Rathinamoorthy, R., & Thilagavathi, G. (2013). Optimisation of process conditions of cotton fabric treatment with *Terminalia chebula* extract for antibacterial application. *Indian Journal of Fiber and Textile Research*, *38*, 293–303.

153. Saleem, A., Husheem, M., Härkönen, P., & Pihlaja, K. (2002). Inhibition of cancer cell growth by crude extract and the phenolics of Terminalia chebula retz. fruit. *Journal of ethnopharmacology*, *81*(3), 327–336.

154. Kumar, B. S., Lakshman, K., Jayaveera, K. N., Krishna, N. V., Manjunath, M., Suresh, M. V., & Naik, S. (2009). Estimation of rutin and quercetin in *Terminalia chebula* by HPLC. *Asian Journal of Research in Chemistry*, *2*(4), 388–389.

155. Kumar, N., Gangappa, D., Gupta, G., & Karnati, R. (2014). Chebulagic acid from Terminalia chebula causes G1 arrest, inhibits NFκB and induces apoptosis in retinoblastoma cells. *BMC complementary and alternative medicine*, *14*, 319.

156. Dash, V. B., & Tulku, V. D. (1991). Positive health in Tibetan medicine (p. 115). Sri Satguru Publications.

157. Aher, V., & Wahi, A. (2011). Immunomodulatory activity of alcohol extract of *Terminalia chebula Retz*. combretaceae. *Tropical Journal of Pharmaceutical Research*, 10(5), 567–575.

158. Sato, Y., Oketani, H., Singyouchi, K., Ohtsubo, T., Kihara, M., Shibata, H., & Higuti, T. (1997). Extraction and purification of effective antimicrobial constituents of Terminalia chebula RETS. against methicillin-resistant Staphylococcus aureus. *Biological & pharmaceutical bulletin*, *20*(4), 401–404.

159. Ahmad, I., Mehmood, Z., & Mohammad, F. (1998). Screening of some Indian medicinal plants for their antimicrobial properties. *Journal of ethnopharmacology*, *62*(2), 183–193.

160. Dutta, B. K., Rahman, I., & Das, T. K. (1998). Antifungal activity of Indian plant extracts. *Mycoses*, *41*(11-12), 535–536.

161. Mehmood, Z., Ahmad, I., Mohammad, F., & Ahmad, S. (1999) Indian medicinal plants: A potential source for anticandidal drugs. *Pharmaceutical Biology*, *37*(3), 237–42.

162. Vonshak, A., Barazani, O., Sathiyamoorthy, P., Shalev, R., Vardy, D., & Golan-Goldhirsh, A. (2003). Screening South Indian medicinal plants for antifungal activity against cutaneous pathogens. *Phytotherapy research: PTR*, *17*(9), 1123–1125.

163. Sohni, Y. R., Kaimal, P., & Bhatt, R. M. (1995). The antiamoebic effect of a crude drug formulation of herbal extracts against Entamoeba histolytica in vitro and in vivo. *Journal of ethnopharmacology*, *45*(1), 43–52.

164. Sohni, Y. R., & Bhatt, R. M. (1996). Activity of a crude extract formulation in experimental hepatic amoebiasis and in immunomodulation studies. *Journal of ethnopharmacology*, *54*(2-3), 119–124.

165. Pinmai, K., Hiriote, W., Soonthornchareonnon, N., Jongsakul, K., Sireeratawong, S., & Tor-Udom, S. (2010). In vitro and in vivo antiplasmodial activity and cytotoxicity of water extracts of Phyllanthus emblica, Terminalia chebula, and Terminalia bellerica. *Journal of the Medical Association of Thailand = Chotmaihet thangphaet*, *93 Suppl 7*, S120–S126.

166. Upadhyay, A., & Singh, D. K. (2011). Molluscicidal activity of Sapindus mukorossi and Terminalia chebula against the freshwater snail Lymnaea acuminata. *Chemosphere*, *83*(4), 468–474.

167. Kamaraj, C., & Rahuman, A. A. (2011). Efficacy of anthelmintic properties of medicinal plant extracts against Haemonchus contortus. *Research in veterinary science*, *91*(3), 400–404.

168. el-Mekkawy, S., Meselhy, M. R., Kusumoto, I. T., Kadota, S., Hattori, M., & Namba, T. (1995). Inhibitory effects of Egyptian folk medicines on human immunodeficiency virus (HIV) reverse transcriptase. *Chemical & pharmaceutical bulletin*, *43*(4), 641–648.

169. Yukawa, T. A., Kurokawa, M., Sato, H., Yoshida, Y., Kageyama, S., Hasegawa, T., Namba, T., Imakita, M., Hozumi, T., & Shiraki, K. (1996). Prophylactic treatment of cytomegalovirus infection with traditional herbs. *Antiviral research*, *32*(2), 63–70.

170. Lin, L. T., Chen, T. Y., Chung, C. Y., Noyce, R. S., Grindley, T. B., McCormick, C., Lin, T. C., Wang, G. H., Lin, C. C., & Richardson, C. D. (2011). Hydrolyzable tannins (chebulagic acid and punicalagin) target viral glycoprotein-glycosaminoglycan interactions to inhibit herpes simplex virus 1 entry and cell-to-cell spread. *Journal of virology*, *85*(9), 4386–4398.

171. Arora, S., Kaur, K., & Kaur, S. (2003). Indian medicinal plants as a reservoir of protective phytochemicals. *Teratogenesis, carcinogenesis, and mutagenesis*, *Suppl 1*, 295–300.

172. Reddy, D. B., Reddy, T. C., Jyotsna, G., Sharan, S., Priya, N., Lakshmipathi, V., & Reddanna, P. (2009). Chebulagic acid, a COX-LOX dual inhibitor isolated from the fruits of Terminalia chebula Retz., induces apoptosis in COLO-205 cell line. *Journal of ethnopharmacology*, *124*(3), 506–512.

173. Sabu, M. C., & Kuttan, R. (2002). Anti-diabetic activity of medicinal plants and its relationship with their antioxidant property. *Journal of ethnopharmacology*, *81*(2), 155–160.

174. Sharma, P., Prakash, T., Kotresha, D., Ansari, M. A., Sahrm, U. R., Kumar, B., Debnath, J., & Goli, D. (2011). Antiulcerogenic activity of Terminalia chebula fruit in experimentally induced ulcer in rats. *Pharmaceutical biology*, *49*(3), 262–268.

175. Gandhi, N. M., & Nair, C. K. (2005). Radiation protection by Terminalia chebula: some mechanistic aspects. *Molecular and cellular biochemistry*, *277*(1-2), 43–48.

176. Gupta, A. K., Mishra, A., Bansal, P., Singh, R., Kumar, S., & Gupta, V. (2010). Phytochemistry and pharmacological activities of Haritaki - A review. Journal of Pharmacy Research, 3, 417-424.

177. Howard, F. M., & Olopade, O. I. (2021). Epidemiology of Triple-Negative Breast Cancer: A Review. *Cancer journal (Sudbury, Mass.)*, *27*(1), 8–16.

178. de Leon-Boenig, G., Bowman, K. K., Feng, J. A., Crawford, T., Everett, C., Franke, Y., Oh, A., Stanley, M., Staben, S. T., Starovasnik, M. A., Wallweber, H. J., Wu, J., Wu, L. C., Johnson, A. R., & Hymowitz, S. G. (2012). The crystal structure of the catalytic domain of the NFκB inducing kinase reveals a narrow but flexible active site. *Structure (London, England: 1993)*, *20*(10), 1704–1714.

179. Jayaram, B., Bhushan, K., Shenoy, S. R., Narang, P., Bose, S., Agrawal, P., Sahu, D., & Pandey, V. (2006). Bhageerath: an energy based web enabled computer software suite for limiting the search space of tertiary structures of small globular proteins. *Nucleic acids research*, *34*(21), 6195–6204.

180. Jayaram, B., Dhingra, P., Lakhani, B., & Shekhar, S. (2012) Bhageerath - Targeting the Near Impossible: Pushing the Frontiers of Atomic Models for Protein Tertiary Structure Prediction. *Journal of Chemical Sciences, 124*, 83-91.

181. Jayaram, B., Dhingra, P., Mishra, A., Kaushik, R., Mukherjee, G., Singh, A., & Shekhar, S. (2014). Bhageerath-H: a homology/ab initio hybrid server for predicting tertiary structures of monomeric soluble proteins. *BMC bioinformatics*, *15 Suppl 16*(Suppl 16), S7.

182. UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic acids research*, *51*(D1), D523–D531.

183. Upadhyay, S., Tripathi, P. K., Singh, M., Raghavendhar, S., Bhardwaj, M., & Patel, A. K. (2020). Evaluation of medicinal herbs as a potential therapeutic option against SARS-CoV-2 targeting its main protease. *Phytotherapy research: PTR*, *34*(12), 3411–3419.

184. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269.

185. Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science (New York, N.Y.)*, *368*(6489), 409–412.

186. Yang, P., & Wang, X. (2020). COVID-19: a new challenge for human beings. *Cellular & molecular immunology*, *17*(5), 555–557.

187. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CL^{pro}) structure: basis for design of anti-SARS drugs. *Science (New York, N.Y.)*, *300*(5626), 1763–1767.

188. Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L. W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., & Yang, H. (2020). Structure of M^{pro} from SARS-CoV-2 and discovery of its inhibitors. *Nature, 582*(7811), 289–293.

189. Martyn, E., Eisen, S., Longley, N., Harris, P., Surey, J., Norman, J., Brown, M., Sultan, B., Maponga, T. G., Iwuji, C., Flanagan, S., Ghosh, I., Story, A., & Matthews, P. C. (2023). The forgotten people: Hepatitis B virus (HBV) infection as a priority for the inclusion health agenda. *eLife*, *12*, e81070.

190. Alberts, C. J., Clifford, G. M., Georges, D., Negro, F., Lesi, O. A., Hutin, Y. J., & de Martel, C. (2022). Worldwide prevalence of hepatitis B virus and hepatitis C virus among patients with cirrhosis at country, region, and global levels: a systematic review. *The lancet. Gastroenterology & hepatology*, *7*(8), 724–735.

191. Kim, S. K., Fujii, T., Kim, S. R., Nakai, A., Lim, Y. S., Hagiwara, S., & Kudo, M. (2022). Hepatitis B Virus Treatment and Hepatocellular Carcinoma: Controversies and Approaches to Consensus. *Liver cancer*, *11*(6), 497–510.

192. Chen, M. B., Wang, H., Zheng, Q. H., Zheng, X. W., Fan, J. N., Ding, Y. L., & Niu, J. L. (2019). Comparative efficacy of tenofovir and entecavir in nucleos(t)ide analogue-naive chronic hepatitis B: A systematic review and meta-analysis. *PloS one*, *14*(11), e0224773.

193. Kiruthika, S., Bhat, R., Dash, R., Rathore, A. S., Vivekanandan, P., & Jayaram, B. (2021). A novel piperazine derivative that targets hepatitis B surface antigen effectively inhibits tenofovir resistant hepatitis B virus. *Scientific reports*, *11*(1), 11723.

194. Prasad, S., Srivastava, A., Singh, N., Singh, H., Saluja, R., Kumar, A., Yadav, N., & Qidwai, T. (2022). Present and future challenges in therapeutic designing using computational approaches. *Computational Approaches for Novel Therapeutic and Diagnostic Designing to Mitigate SARS-CoV-2 Infection*, 489–505.

195. Rawat, S., Subramaniam, K., Subramanian, S. K., Subbarayan, S., Dhanabalan, S., Chidambaram, S. K. M., Stalin, B., Roy, A., Nagaprasad, N., Aruna, M., Tesfaye, J. L., Badassa, B., & Krishnaraj, R. (2024). Drug Repositioning Using Computer-aided Drug Design (CADD). *Current pharmaceutical biotechnology*, *25*(3), 301–312.

196. Niazi, S. K., & Mariam, Z. (2023). Computer-Aided Drug Design and Drug Discovery: A Prospective Analysis. *Pharmaceuticals (Basel, Switzerland)*, *17*(1), 22.

197. Zhou, Y., Peng, S., Wang, H., Cai, X., & Wang, Q. (2024). Review of Personalized Medicine and Pharmacogenomics of Anti-Cancer Compounds and Natural Products. *Genes*, *15*(4), 468.

198. Prachayasittikul, V., Worachartcheewan, A., Shoombuatong, W., Songtawee, N., Simeon, S., Prachayasittikul, V., & Nantasenamat, C. (2015). Computer-Aided Drug Design of Bioactive Natural Products. *Current topics in medicinal chemistry*, *15*(18), 1780–1800.

199. Wu, Z., Chen, S., Wang, Y., Li, F., Xu, H., Li, M., Zeng, Y., Wu, Z., & Gao, Y. (2024). Current perspectives and trend of computer-aided drug design: a review and bibliometric analysis. *International journal of surgery (London, England)*, 10.1097/JS9.0000000000001289. Advance online publication.

200. Bassani, D., & Moro, S. (2023). Past, Present, and Future Perspectives on Computer-Aided Drug Design Methodologies. *Molecules (Basel, Switzerland)*, *28*(9), 3906.