Chapter- 2

Methodology

Our investigation has embraced a diverse set of computational strategies effectively. In the following sections, we delve into each methodology with meticulous detail, supplemented by an extensive analysis of outcomes, that are elaborated in the corresponding chapters. Through this comprehensive approach, we aim to offer a nuanced understanding of our research findings and their implications.

2.1. Selection of protein targets

The selection of protein targets is guided by multiple criteria, including the protein's indispensability for the pathogen, its mechanism of action, its non-homology with host proteins, and its distinctiveness from host proteins in terms of identity and similarity. In Chapter 3, we have curated a dataset comprising 3D protein structures of *Leishmania donovani* obtained from the Protein Data Bank (PDB) [1]. From this dataset, we have identified 10 proteins that constitute a non-redundant subset characterized by favorable attributes such as high resolution and X-ray diffraction structure [2]. The two proteins are associated with distinct metabolic pathways in the organism *L. donovani*, namely the purine salvage pathway and the pyrimidine pathway. The availability of a three-dimensional structure of a protein provides a foundation for subsequent structural analyses, facilitating the exploration of their structural and functional interplay with various candidate inhibitors.

In Chapter 4, following an extensive examination of the literature pertaining to drug targets and essential genes in *L. donovani*, a comprehensive list of therapeutic targets and essential proteins linked to the essential genes of *L. donovani* was collated. Common proteins from both lists were selected, that includes the vitamin B6 pathway and sterol biosynthesis of *L. donovani*. While one target protein from the vitamin B6 pathway has a known three-dimensional structure, the other protein from the sterol biosynthesis pathway lacks a three-dimensional structure.

In Chapter 5, a subtractive genomics approach was applied, utilizing the proteomes of five different *Leishmania* species obtained from the NCBI database. Key selection criteria included the identification of orthologous groups, non-homologous sequences, druggability analysis, and the prediction of protein active sites, which together facilitated the identification of common proteins across the five *Leishmania* species. Among these, Glutamate Dehydrogenase (GDH) emerged as a promising drug target for combating *Leishmania* infections.

2.2. Protein structure preparation

2.2.1. Details of Homology Modeling

Homology modeling is an indispensable in structural biology, facilitating the prediction and analysis of protein structures, and advancing research in drug discovery, protein engineering, and other comprehension of fundamental biological processes [3]. Many proteins remain structurally unknown due to challenges in crystallization or the expense of experimental methods like crystallography and Nuclear Magnetic Resonance (NMR) techniques. The number of experimentally determined structures of protein is notably smaller compared to the number of known protein sequences. Homology modeling enables the prediction of three-dimensional structure of proteins by utilizing its amino acid sequence and the known structures of similar proteins. It aids in extrapolating structural insights from characterized proteins to related, uncharacterized ones, substantially enriching our structural knowledge base. By employing homology modelling, we can forecast the consequences of amino acid mutations or structural alterations on protein structure and function. Along with that scientists can engineer proteins for a myriad of purposes, including enzyme catalysis, protein-protein interactions, and therapeutic protein design [4].

Prior to initiating the homology modeling process, preliminary steps are necessary. These steps entail conducting a search for a suitable template protein sequence by utilizing Basic Local Alignment Search Tool (BLAST) [5] against the PDB, which houses a repository of protein structures. This approach yields dependable outcomes when the query sequence closely matches the template sequence, exhibiting both high sequence identity and similarity, coupled with substantial query coverage. The homology modeling process comprises several key stages: sequence alignment, template selection, model building, model refinement, and validation. In the initial step of sequence alignment, the amino acid sequence of the target protein is aligned with that of one or more template proteins, leading to the identification of similarities and differences between them. Following this, template selection involves choosing appropriate template structures based on their sequence similarity and overall structural compatibility with the target protein. Models are then built in the Model Building stage, wherein the aligned sequences serve as a foundation for constructing the structure of the target protein. During this process, atoms from the template structure are mapped onto corresponding positions in the target sequence, while regions lacking template structures are constructed anew. Subsequently, model refinement involves

refining the initial model through energy minimization and molecular dynamics simulations to optimize its geometry and eliminate any structural irregularities. Finally, validation assesses the quality of the homology model using various criteria, including stereochemical quality and structural integrity, often utilizing graphs such as Ramachandran plots [6].

AlphaFold, developed by DeepMind, stands at the forefront of structural biology advancements. This cutting-edge deep learning system significantly enhances protein folding prediction [7]. Employing an intricate neural network architecture, AlphaFold achieves remarkable precision in predicting protein 3D structures, often surpassing conventional methods. It adopts a pioneering strategy, amalgamating multiple sequence alignment with deep learning techniques, to produce highly precise structural predictions. Leveraging extensive genomic datasets and insights from experimental protein structures, AlphaFold offers invaluable insights into protein folding dynamics, function, and interactions [8].

UniProt, known as the Universal Protein Resource, serves as an extensive database offering comprehensive insights into proteins. It acts as a centralized hub for curated protein sequence and functional data, drawn from diverse origins including literature, experimental findings, and computational analyses [9]. Covering a wide spectrum of organisms, UniProt hosts a vast array of protein sequences, spanning both familiar model organisms and less explored species. Each entry in UniProt undergoes detailed annotation, providing key details of protein names, gene identifications, functions, subcellular locations, biological processes, molecular functions, and associated ailments. Utilizing sequence similarities and functional attributes, UniProt categorizes proteins into families, aiding researchers in tracing evolutionary connections and inferring protein functions. Moreover, UniProt facilitates seamless integration with external databases and resources through cross-references, empowering researchers to conduct comprehensive analyses and investigations [10].

In this research, we have delineated a holistic strategy that amalgamates a range of computational methodologies to pinpoint potential protein drug targets and regions of structural importance. This entails utilizing an array of computational techniques, such as *ab initio* structure prediction and molecular dynamics simulation, to identify and analyze crucial proteins. In chapter 4, the chosen pdb structure of pyridoxal kinase (PK) forms a homodimer having molecular weight of 36.2 kDa. However, the protein structure exhibits missing residues, which were addressed through modeling techniques with MODELLER

10.1 [11]. A monomeric form of the protein was generated through modeling and subsequently subjected to minimization using GROMACS [12]. Likewise, another protein target, Sterol 14-alpha demethylase (SDM), lacks a three-dimensional structure. To acquire its structure, homology modeling techniques were applied. In this process, the sterol alpha-14 demethylase of *L. infantum* with pdb structure served as the template. Subsequently, the monomeric structure was refined and utilized as the definitive structure for conducting MD simulations. In Chapter 5, the glutamate dehydrogenase (GDH) protein target was identified for five *Leishmania* species through a subtractive genomics approach. Due to the lack of experimental three-dimensional structures for these protein targets, AlphaFold structures of GDH proteins from the five *Leishmania* species were acquired from the UniProt database. Subsequently, all five structures underwent minimization using the GROMACS software.

2.2.2. Binding site prediction of protein structure

The prediction of binding sites in protein structures is a crucial aspect of structural bioinformatics, aiding in the understanding of protein-ligand interactions and facilitating drug discovery efforts [13]. These methods leverage the three-dimensional structure of proteins to predict binding sites. They often rely on the geometric and physicochemical properties of protein structures. Web servers like CASTp [14] and DeepPocket [15] utilize algorithms to identify likely binding regions on protein surfaces, including those proteins that do not have an experimental three-dimensional structure. These are basically protein structures which were developed with the help of homology modelling approach. It is necessary to obtain the binding site of these proteins as it will help in binding of the ligands into it which will further affect the mechanism of the protein in the organism. In Chapter 5, the principal ligand-binding pocket within the GDH protein of all five *Leishmania* species was delineated through CASTp predictions. This crucial identification allows for the precise binding of suitable ligands to the GDH protein, facilitating deeper insights into its functional mechanisms across different species.

2.3. Screening of Ligands

In the realm of bioinformatics, ligand screening stands as a foundational step within the drug discovery pipeline, harnessing computational techniques to navigate extensive libraries of chemical compounds in pursuit of potential drug candidates. This pivotal process underpins drug discovery and design endeavors, aiming to identify compounds capable of modulating

target protein activity to achieve therapeutic objectives. By screening extensive compound databases against target proteins, researchers can accelerate the discovery of lead molecules for subsequent experimental validation, thereby streamlining the drug discovery trajectory [16]. Moreover, these methodologies leverage *in-silico* models to shift through and prioritize compounds within large chemical repositories, based on their projected propensity to bind with the target protein.

2.3.1. Collection of ligands

The PubChem database, maintained by the National Center for Biotechnology Information (NCBI), is a comprehensive resource that catalogs information on substances, chemical compounds, including their structures, properties, bioactivities, and biological targets. Additionally, PubChem provides access to genes, pathways, cell lines, bioassay data, allowing researchers to evaluate the biological effects of compounds and prioritize candidates for further investigation [17,18]. The PubChem database encompasses an extensive array of chemical information, including data on the periodic table elements numbering 118, compounds totalling approximately 118,022,280, and substances numbering around 317,731,311, among other valuable datasets. It acts as a primary repository for chemical details, gathering data from diverse origins such as literature, patents, and chemical vendors. With millions of compounds indexed, PubChem offers extensive search and analysis tools, enabling to explore chemical space, identify potential drug candidates, and study structure-activity relationships [19]. Moreover, it supports data deposition and sharing, fostering collaboration and transparency in chemical research.

DrugBank is a comprehensive online database that provides detailed information about drugs, their mechanisms of action, pharmacology, interactions, and therapeutic indications [20]. It provides valuable information for scientists, healthcare practitioners, and pharmaceutical developers interested in both approved and investigational drugs. DrugBank contains extensive data on drug structures, chemical properties, pharmacokinetics, and pharmacodynamics, curated from a wide range of literature sources and databases [21]. One of the key features of DrugBank is its comprehensive coverage of drug targets and their associated pathways. Containing over 500,000 drugs and drug products, this database also encompasses a wide array of approved drugs, numbering 13,113, in addition to providing data on 19,535 drug targets sourced from diverse organisms. This database helps to explore detailed information about the molecular targets of drugs, including proteins, enzymes,

receptors, and transporters. Additionally, DrugBank provides insights into the mechanisms of drug action, including how drugs interact with their targets to produce therapeutic effects. It also includes information on drug metabolism, including details about drug metabolizing enzymes and metabolic pathways involved in the biotransformation of drugs in the body [22]. This information is crucial for understanding drug metabolism, drug-drug interactions, and the potential for drug toxicity.

The DrugCentral database serves as a comprehensive repository of drug information, offering a wealth of data on various aspects related to drug discovery, development, and usage. It contains a vast collection of information on approved drugs, investigational drugs, and experimental compounds, providing detailed profiles of each drug's pharmacological properties, mechanisms of action, indications, adverse effects, and regulatory status [23]. Additionally, DrugCentral includes data on drug targets, drug-drug interactions, drug metabolism pathways, and drug formulations. This database is a valuable asset for scientists and clinicians engaged in drug discovery, development, and clinical practice. [24].

In Chapter 3, we accessed the PubChem database to procure the Simplified Molecular Input Line Entry System (SMILES) notation of the compounds necessary for our study. Subsequently, these compounds were subjected to additional filtration based on various parameters. For our second objective, we utilized both DrugBank and DrugCentral databases to retrieve approved drugs utilized in diverse treatments. These approved drugs underwent additional filtration to isolate the desired ligands. In our final objective, outlined in Chapter 5, the PubChem database served as the source for obtaining the ligands.

2.3.2. Ligands filtration

In chapter 3, the initial molecule selection process, a primary screening was conducted to assess the drug-likeness characteristics. Lipinski's rule of five and Veber's rule, each with distinct parameters, were taken into consideration for selection of ligands. Lipinski's rule encompasses criteria such as molecular weight (≤500 Da), hydrogen bond donors (≤5), hydrogen bond acceptors (≤10), and partition coefficient (logP) (≤5) [25]. Conversely, Veber's rule incorporates parameters such as the count of rotatable bonds (≤10) and the polar surface area (≤140 $Å^2$) [26]. Both these rules serve as valuable filters in the early stages of drug discovery, helping to prioritize compounds with optimal drug-like properties for further development. These rules are often applied in combination with other screening criteria to

identify potential drug candidates with improved pharmacokinetic profiles and reduced risk of toxicity. They contribute to the rational design and optimization of new therapeutic agents, ultimately enhancing the efficiency and success rate of drug discovery programs. The computation of Lipinski's rule of five and Veber's rule properties was conducted utilizing SwissADME, an accessible online resource [27]. This screening helps in the selection of compounds which can show poor absorption in cells. SwissADME, an open web source was used to calculate the properties of Lipinski's rule of five and Veber's rule. In addition to the aforementioned filtering parameters employed for ligand screening, our study also utilized ADMET and PASS analyses for further evaluation.

ADMET, encompassing Absorption, Distribution, Metabolism, Excretion, and Toxicity, stands as a pivotal aspect in the assessment of potential drug candidates throughout the drug discovery and development journey [28]. It plays a crucial role in predicting the pharmacokinetic and toxicological profiles of these candidates, thereby optimizing their efficacy and safety while mitigating the risk of adverse effects in human subjects. Utilizing a blend of *in vitro* assays, animal models, and computational methodologies, ADMET studies offer comprehensive insights into the drug's behavior within the body. Specifically, drug absorption denotes the process whereby a substance enters the bloodstream from its point of administration, which could be oral, intravenous, or topical. Various factors affect this process, including the physical and chemical properties of the drug, its formulation, and the characteristics of the administration route. Distribution phenomena encompass the movement of a drug throughout the body, from the bloodstream to different tissues and organs. This transport is influenced by factors such as blood flow dynamics, tissue permeability, and the drug's ability to bind to plasma proteins or tissue receptors. Metabolism, also known as biotransformation, refers to the chemical alteration of a drug by enzymes in the body, primarily occurring in the liver. This process can lead to the formation of active or inactive metabolites with different pharmacological or toxicological properties compared to the parent drug. Excretion involves the removal of drugs and their metabolites from the body, primarily through the kidneys (via urine) and the liver (via bile), with additional routes including sweat, saliva, faeces, and exhaled air. The rate of drug excretion affects its plasma concentration and duration of action. Toxicity assessment evaluates potential adverse effects of a drug on biological systems, including acute and chronic toxicity, as well as organ-specific toxicity. This testing aims to identify safety concerns and guide dose selection in clinical trials [29].

PASS (Prediction of Activity Spectra for Substances) analysis is a computational method used in drug discovery to predict the biological activity profile of chemical compounds based on their chemical structure. This approach aids researchers in prioritizing compounds for experimental assessment by offering insights into their potential pharmacological impacts [30]. During the analysis, the software assesses the structural attributes of a compound and juxtaposes them with a database containing verified compounds with known biological activities. Through the utilization of statistical algorithms and machine learning methodologies, PASS determines the probability of the compound manifesting specific biological activities across diverse targets or pathways. The outcome of the analysis typically furnishes a catalog of anticipated biological activities alongside corresponding probability scores, indicating the likelihood of the compound interacting with various biological targets or eliciting particular pharmacological responses [31].

2.4. Molecular Docking

Molecular docking, a prevalent computational method in drug discovery and structural biology, forecasts the favoured orientation and binding affinity of a small molecule (ligand) within the binding site of a large molecule target (receptor), often a protein. It simulates ligand-receptor complex formation, assessing their stability and affinity [32]. In drug discovery, molecular docking is vital for identifying potential drug candidates by predicting their interactions with target proteins, thereby elucidating molecular mechanisms of ligand-protein recognition and facilitating the design of new therapeutics with enhanced potency and specificity [33]. Molecular docking includes diverse methods designed for specific research requirements and computational capacities.

Molecular docking encompasses several distinct methodologies tailored to specific research needs and computational capacities. Rigid docking assumes that both the ligand and receptor maintain fixed conformations, which enhances computational efficiency but may overlook conformational changes that occur upon binding. Conversely, flexible docking permits flexibility in either or both ligand and receptor, accommodating conformational changes and thereby enhancing accuracy, though this comes at the expense of increased computational demands. Induced fit docking combines rigid docking with limited receptor flexibility post-initial docking, capturing induced fit effects for more precise binding predictions. Ligand-based docking involves docking ligands into predetermined protein binding sites without considering receptor flexibility, serving well in virtual screening and predicting ligand

binding modes. Protein-protein docking predicts interactions between protein molecules, crucial for comprehending protein complex structures and functions in drug discovery and biological research. Blind docking involves docking ligands across the entire protein surface without specifying a binding site, useful for uncovering potential binding sites and understanding protein-ligand interactions in new systems. Cross-docking docks ligands into multiple receptor structures to probe ligand selectivity, assess binding affinities across diverse protein conformations or elucidate structure-activity relationships [34,35].

The docking process comprises several essential stages. Initially, the ligand and receptor structures undergo preparation, involving optimization of their geometry, protonation states, and addition of necessary hydrogen atoms [36]. The protein molecule undergoes preparation by adding hydrogen atoms and Kollman charges, while the ligand is prepared by adding Gasteiger charges. Subsequently, a three-dimensional grid is generated to delineate the binding site of the receptor, specifying the region where ligand binding occurs. The docking algorithm is then executed, systematically orienting and translating the ligand within the search space to explore diverse binding modes. During this phase, docking algorithms assess the complementarity between the ligand and receptor, considering factors such as steric clashes, hydrogen bonding, and hydrophobic interactions [37]. Following docking, protein-ligand complex poses are scored based on their predicted binding affinity, with lower energy scores indicating more favourable binding modes. Finally, docking results undergo analysis and validation to elucidate the molecular interactions between ligand and receptor. Visual inspection and comparison with experimental data contribute to validating the reliability of the predicted binding modes [38].

There are several molecular docking software programs available, each with its own unique features and capabilities. Some commonly used ones: AutoDock, AutoDock Vina, GOLD (Genetic Optimization for Ligand Docking) and GLIDE [39]. AutoDock 4.2 used Lamarckian genetic algorithmic (LGA) to obtain the proper protein-ligand complex along with the proper orientation of ligands. AutoDock Vina uses Broyden-Fletcher-Goldfarb-Shanno algorithm for its operation. CB-Dock is a web server which works on density-peak-based clustering algorithm and uses blind docking approach to form protein-ligand complex.

2.5. Method of Subtractive genomics

Subtractive genomics, a bioinformatics methodology, aims to discern distinctive genetic components or proteins within a target organism through comparative analysis of its genome or proteome against that of a reference organism [40]. This approach enables the identification of novel elements crucial for the target organism's biology, potentially offering valuable insights for various biomedical applications.

2.5.1. Collection of proteomes

The foundational data required for conducting subtractive genomics analysis includes genomes or proteomes of a specific organism, sourced from databases like the National Center for Biotechnology Information (NCBI) database [41], Ensembl [42] and others. These databases serve as rich repositories of genomic and proteomic information, providing researchers with access to comprehensive data sets essential for comparative genomic studies and the identification of unique genetic elements or proteins within target organisms. Such analyses offer valuable insights into the genetic basis of various biological processes and can inform further research in fields such as pathogenesis, drug discovery and evolutionary biology. Genome files offer extensive details about an organism, encompassing the sequence of nucleotides forming its DNA, crucial for its growth, development, and functionality. These files typically incorporate annotations specifying gene locations, regulatory elements, and additional genomic attributes. In chapter 5, five proteomes were taken for the work. In case of proteome, the files provide comprehensive data about the entire set of proteins expressed by an organism. This includes details such as the amino acid sequences of proteins, their functions, subcellular localization, etc. Additionally, proteome files often contain annotations that describe the biological roles of proteins, their involvement in pathways and processes and any relevant experimental evidence. Moreover, proteome files include data on hypothetical proteins and paralogous proteins, aiding in the identification of functionally relevant proteins within the proteome of a specific organism [43]. This information assists in distinguishing between proteins with known functions and those that require further characterization, contributing to a more comprehensive understanding of the organism's biology.

2.5.2. Data Filtration

Proteome of five *Leishmania* species retrieved contains information such as functional proteins, hypothetical proteins and paralogous proteins. Initially, the hypothetical proteins are excluded from the proteome. Hypothetical proteins are products of genes anticipated from genomic sequences but lack experimental confirmation of their expression or roles. These proteins are typically detected through computational means like gene prediction algorithms for analyzing DNA sequences. They're labelled as "hypothetical" because their presence and functions haven't been experimentally verified using different methods. Their existence may result from diverse factors, such as errors in genome annotation, anomalies in sequencing, or the presence of genuinely new genes whose functions remain unidentified. Later on, removal of paralogous proteins in the proteome was done by CD-HIT software [44]. CD-HIT is a widely used bioinformatics program designed for clustering and comparing protein or nucleotide sequences. It employs an efficient algorithm to reduce redundancy in sequence databases by clustering highly similar sequences into representative subsets, thereby accelerating subsequent sequence analysis. CD-HIT operates by calculating pairwise sequence similarities and grouping sequences into clusters based on a user-defined similarity threshold. The program offers various parameters to control clustering stringency, such as sequence identity and alignment coverage. Additionally, CD-HIT provides options for generating cluster representatives, which are non-redundant sequences representing each cluster, and for removing redundant sequences from input datasets [45]. It is widely utilized in genomic and proteomic research for tasks such as sequence database compression, sequence annotation, and identifying sequence homologs. After filtering, the proteome sequences were subjected to orthology analysis using OrthoFinder software, a bioinformatics tool utilized for comparative genomics and orthology inference across diverse species [46]. This software identifies clusters of orthologous genes by comparing protein sequences from various organisms, leveraging sequence similarity to infer evolutionary relationships. OrthoFinder utilizes a sophisticated algorithm incorporating sequence similarity, evolutionary divergence, and gene tree reconciliation to precisely detect orthologs and deduce gene orthology connections [47]. Furthermore, to identify regions of non-homologous proteins in *Leishmania* species compared to the human proteome, BLAST (Basic Local Alignment Search Tool) searches and the Markov Cluster Algorithm were employed [48]. This program works by conducting pairwise alignments between the query protein sequence and sequences in the target database, searching for regions of local

similarity. BLAST utilizes a scoring system to assess the significance of sequence matches, taking into account factors such as sequence identity, similarity and alignment length.

2.5.3. Druggability analysis and identification of novel drug targets

In this segment of the methodology, established drug targets sourced from the DrugBank database [49] were considered. DrugBank comprises an extensive repository of FDA-approved drug targets, encompassing proteins or macromolecules targeted by therapeutics endorsed for clinical use by the U.S. Food and Drug Administration (FDA). These targets are pivotal in numerous physiological processes and disease pathways, rendering them attractive candidates for drug discovery and development endeavors. The utilization of the DrugBank database facilitates exploration into the pharmacological characteristics of FDA-approved drug targets, identification of novel drug-target interactions, and prioritization of drug discovery initiatives aimed at specific diseases or therapeutic targets [50]. To ascertain the druggability of shared non-homologous proteins from *Leishmania* species, they underwent assessment for their potential to strongly bind with therapeutic ligands [51]. This evaluation entailed BLASTp searches against FDA-approved drug targets retrieved from the DrugBank database. The objective of this procedure was to determine the suitability of the selected proteins as potential drug targets, thereby identifying novel candidates for further investigation in drug discovery endeavors. This approach aimed to uncover proteins exhibiting characteristics akin to established drug targets, laying the groundwork for subsequent in-depth analyses and potential therapeutic interventions.

2.6. Insights from Molecular Dynamics Simulations

Biomolecules exhibit inherent dynamism, a characteristic crucial for their proper functioning. Their ability to undergo dynamic changes is fundamental to various biological processes. The demand for computational methods such as molecular dynamics simulations, capable of offering dynamic insights into biomolecular behavior over time, is becoming increasingly imperative [52]. This necessity arises from the inadequacy of conventional experimental techniques like X-ray crystallography, which are unable to capture dynamic information pertaining to molecular motions resulting from changes in atomic or bonded interactions [53].

The MD Simulation process involves the time-dependent behavior of three-dimensional macromolecular structures, which can be accessed through various databases. Following Newtonian laws of motion, MD enables the observation of atomic behavior within biological entities, offering detailed insights into macromolecular motion at the atomic level [54]. Examples of biological phenomena studied using molecular dynamics simulations include protein stability [55], protein-ligand binding [56], protein-nanoparticle interactions [57], protein folding [58] and protein-protein interactions [59]. Experimental methods like X-ray crystallography and NMR spectroscopy, which elucidate the structures of proteins and other macromolecules, offer initial coordinates for utilization in the simulation procedure. The primary objective of MD simulation is to mimic life processes. To achieve this, an explicit system is constructed, incorporating water, which is essential for life [60]. MD simulations allow for the observation of molecular motions at the femtosecond timescale, facilitating the examination of subtle changes in the behavior of macromolecules [61]. This capability enables researchers to gain more in-depth details into the changing behavior of biomolecules in aqueous environments.

MD simulation process operates at the atomistic level, it initially records the position and movement of each atom throughout the entire simulation, providing a comprehensive depiction of the molecule's behavior. This level of detail is challenging to achieve with experimental techniques, as they cannot observe such changes continuously [62]. A significant advantage of MD simulations lies in their ability to operate under diverse conditions, facilitated by meticulous control and understanding of simulation parameters. These include regulation of initial coordinates of proteins and ligands, characterization of molecular composition within the system, imposition of constraints and charges, as well as management of temperature and pressure variables [63]. This benefit enables the tailored design of simulations to address various inquiries and interpret the resulting simulation outcomes effectively. This comprehensive control empowers researchers to explore molecular dynamics across a spectrum of scenarios, facilitating nuanced investigations into biological systems.

2.6.1. Understanding force fields in molecular dynamics simulations

A force field is a mathematical model used in molecular dynamics simulations to calculate the forces and energies acting on atoms and molecules within a system. It comprises mathematical functions that describe the interactions between atoms, including bonded and

non-bonded interactions such as covalent bonds, angles, dihedrals, and van der Waals and electrostatic interactions. In molecular dynamics simulations, the force field is applied to the initial coordinates of atoms to simulate their movements over time [64]. The equations of motion, typically derived from Newton's laws of motion, are integrated numerically to predict the trajectory of each atom. By iteratively calculating the forces acting on each atom and updating their positions and velocities, the simulation can simulate the behavior of the system over time [65]. Various force fields are tailored for specific types of molecules and may incorporate parameters calibrated against experimental data. The accuracy of simulation results depends on the optimization of force field parameters and simulation conditions, such as temperature, pressure and solvent environment. Protein force fields utilize the following energy terms:

$$V(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} k_\phi[\cos(n\phi + \delta) + \sum_{\substack{nonbond \\ pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}\right]$$

---------- eq. 2.1

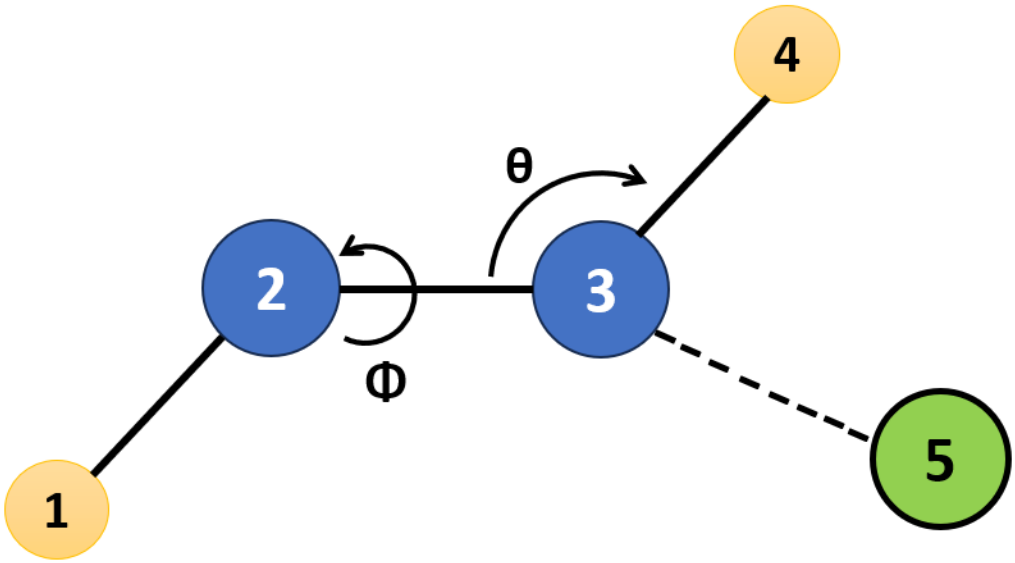

Figure 2.1: Diagrammatic representation of force field interactions: covalent bonds are shown with bold solid lines, while non-bonded interactions are depicted with light dashed lines.

The development of force fields is a complex process that involves optimization of parameters based on different systems and problems. T. Halgren has remarked that this process remains as much an art as it is a science, highlighting the intricate nature of force

field development [66]. Comparing one force field with another for the same system proves challenging due to the significant influence of various parameters inherent to each force field. Several force fields commonly employed in molecular dynamics simulations of biomolecules include CHARMM [67], AMBER [68] and GROMOS [69], while for condensed phase simulations, OPLS and COMPASS are frequently utilized. The utilization of these diverse force fields enables researchers to explore a wide range of biological problems and phenomena. Each force field possesses unique characteristics that make it a valuable option for solving specific problems. However, they also have shortcomings related to the data and procedures employed in their parameterization.

2.6.1.1. Additive Force Fields used in MD Simulations

- **CHARMM Force Field:** The CHARMM (Chemistry at HARvard Molecular Mechanics) additive all-atom force field is one of the most widely used force fields in molecular dynamics simulations. Its development and parameterization have been ongoing since the early 1980s, leading to its broad applicability in studying biomolecular systems. It encompasses a broad spectrum of systems, including proteins with extensive amino acid support, nucleic acids covering both DNA and RNA, lipids for membrane simulations and carbohydrates [70]. CHARMM is also widely applied to inorganic materials, making it a valuable tool in materials design. Its adaptability allows for the simulation of complex systems, enabling advancements in areas such as nanotechnology and material science. For instance, the CHARMM force field boasts various versions such as CHARMM19, CHARMM22, and CHARMM27. CHARMM36 force field is the latest and widely used all-atom force field that has been adapted for use in GROMACS, a popular molecular dynamics simulation package used for different systems [71].
- **AMBER Force Field:** AMBER (Assisted Model Building with Energy Refinement) force field is a family of molecular mechanics force fields widely used for biomolecular simulations. The initial development of AMBER force field began in the 1980s as part of the AMBER software package. The AMBER force field was designed to work seamlessly for proteins and nucleic acids, enabling the simulation of a diverse array of molecules [72]. It includes various versions (e.g., ff94, ff99, ff03, ff14SB, ff19SB), each improving parameterization and accuracy. AMBER

force fields are not limited to the AMBER software package and are also compatible to GROMACS and other platforms [73].

- **GROMOS Force Field:** The GROMOS (GROningen MOlecular Simulation) force field is a family of molecular mechanics force fields primarily designed for the study of biomolecules, such as proteins, lipids, and carbohydrates, in molecular dynamics simulations. In the realm of biomolecular molecular dynamics simulations, force fields are in a constant state of evolution, while the GROMOS force field is represented by versions like GROMOS96, GROMOS45a3, GROMOS53a5, GROMOS53a6 and GROMOS54a7 [69]. GROMOS54a7 is specifically parameterized for simulating proteins, nucleic acids, and small organic molecules, ensuring accurate modeling of biomolecular systems. It provides stability to protein backbones and is also compatible with SPC and SPC/E water models. GROMOS force field parameters include bond lengths, angles, backbone flexibility and stability, water models, and other parameters [73].

2.6.2. Ensemble

Ensembles play a crucial role in representing how atoms or molecules interact within a system and respond to external factors like temperature, pressure, and chemical potential. Simulating molecular systems across diverse ensemble conditions enables the exploration of numerous phenomena in thermodynamics and kinetics, encompassing phase changes, interactions between proteins and ligands, and the folding of biomolecules. Commonly used ensembles in MD simulations are Microcanonical Ensemble (NVE), Canonical Ensemble (NVT), Isothermal-Isobaric Ensemble (NPT) and Grand Canonical Ensemble (μVT) [74]. In the microcanonical ensemble, the system is isolated and conserves its total energy (E), volume (V), and number of particles (N). NVE simulations are useful for studying the dynamics of isolated systems and energy conservation. In the canonical ensemble, the system exchanges energy with a heat bath at a constant temperature (T) while maintaining a fixed volume (V) and number of particles (N). NVT simulations are suitable for studying the equilibrium properties of systems at a given temperature, such as phase transitions and thermodynamic properties. The isothermal-isobaric ensemble maintains a constant temperature (T), pressure (P) and number of particles (N) while allowing for volume fluctuations. NPT simulations are essential for studying systems under constant external pressure, such as liquids and gases. The grand canonical ensemble enables the exchange of

particles between the system and a particle reservoir at constant chemical potential (μ), temperature (T) and volume (V). μVT simulations are valuable for studying systems with variable particle numbers, such as adsorption processes and chemical reactions.

2.6.3. Understanding Periodic Boundary Conditions

Periodic boundary conditions (PBC) are a computational technique used in molecular dynamics simulations to simulate an infinite system within a finite simulation box. In this method, molecules interact not only with neighboring molecules within the simulation box but also with periodic images of these molecules, which are positioned at regular intervals in all three dimensions. This approach effectively eliminates edge effects and simulates bulk behavior, making it suitable for studying systems with periodicity, such as crystals, liquids, and other extended structures. Under periodic boundary conditions, when a molecule traverses the border of the simulation box, it appears again on the opposite side as a periodic image [75]. The interactions between molecules and their periodic images are computed using minimum image convention, ensuring that only the nearest periodic image is considered for each molecule. This allows for accurate calculation of intermolecular forces while avoiding duplication of interactions. Periodic boundary conditions are essential for studying large systems and systems with long-range interactions, as they enable efficient computation of interactions while preserving the periodicity of the system [76]. They are widely used in molecular dynamics simulations of biological molecules, materials, and complex fluids to study equilibrium properties, transport phenomena, and phase transitions.

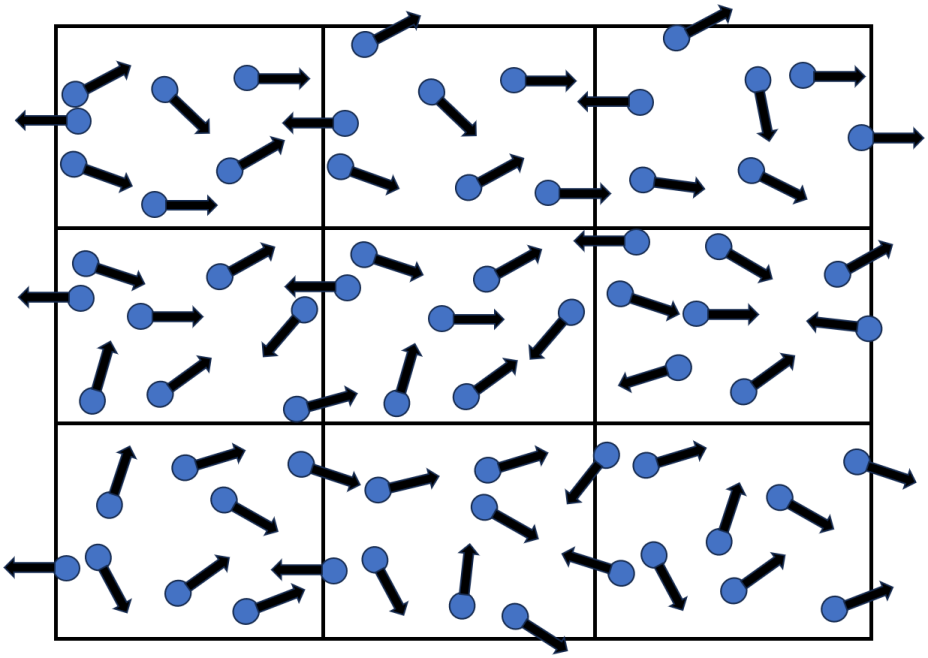

Figure 2.2: Visualization of periodic boundary conditions in two dimensions.

2.6.4. Particle mesh Ewald Summation

Particle mesh Ewald (PME) summation is a computational technique used to efficiently compute long-range electrostatic interactions in molecular dynamics simulations. PME works by representing the charge distribution on a three-dimensional grid, known as the charge grid. This grid is typically much coarser than the molecular system itself, allowing for significant computational savings. The charge grid is used to calculate the electrostatic potential at each grid point using Fast Fourier transform (FFT) techniques [77]. The resulting potential is then interpolated back onto the molecular system to compute the forces acting on each particle. One of the key advantages of PME is its ability to efficiently handle periodic boundary conditions, allowing accurate treatment of long-range electrostatic interactions in periodic systems. Additionally, PME is highly parallelizable, making it suitable for large-scale simulations on modern computing architectures [78]. Overall, PME summation is a potent method for precisely simulating electrostatic interactions in molecular dynamics simulations, allowing researchers to investigate various biological, chemical, and materials systems with exceptional accuracy.

$$U = \frac{1}{2}\sum_{i,j,n}\frac{q_i q_j}{|r_i - r_j + nL|} \quad \text{---------- eq. 2.2}$$

Where, i and j = atoms, q_i and q_j = charges on the atom, r_i and r_j = coordinates and nL = displacement vector

Further the calculation of energy (U_{Ewald})equation can be termed:

$$U_{Ewald} = U^r + U^m + U^{self} \quad \text{---------- eq. 2.3}$$

Where, U^r = summation in real space, U^m= Fourier term and U^{self}= constant term

2.6.5. Structure preparation

Proteins with available three-dimensional structures from the Protein Data Bank (PDB) were selected for the molecular dynamics (MD) simulation process. For proteins with missing residues or atoms in their structures, or those lacking a structure entirely, modeling was performed using MODELLER software. The resulting modeled structure underwent

minimization using GROMACS 2020.1 [79]. Following minimization, the modeled structure was subjected to docking with the ligand to form the protein-ligand complex. The GROMOS54a7 force field was utilized for the simulation setup of the system. To prepare the ligand, the Automated Topology Builder (ATB) [80] was employed to generate the topological parameters compatible with the GROMOS54a7 force field, facilitating the simulation process. All of the atom types, their partial charges, bond, angle, dihedral angle, and torsional angle information are all included in the ligand's parameter file.

2.6.6. System preparation

All simulations were conducted under periodic boundary conditions in all dimensions. The LINCS algorithm was applied to maintain bond constraints throughout the simulation process [81]. The systems were solvated using the SPC (simple point charge) water model, specifically employing the three-site SPC water model for MD simulations [82]. The protein or protein-ligand complex was positioned at a distance of 1.0 nm from the box edge. Prior to simulation initiation, counterions were introduced into the system to neutralize any charges that might be present. We enforce PBC across all spatial dimensions and employ the LINCS algorithm to uphold bond constraints. Additionally, we utilized the Steepest Descent method to minimize the energy of the systems. Equilibration involves two stages: initially, the NVT canonical ensemble is applied for 2 ns, maintaining constant volume, temperature, and number of atoms. Subsequently, the system is equilibrated for 5 ns under the NPT isobaric–isothermal ensemble to balance the pressure and achieve equilibrium [83].

2.6.7. Binding energy calculations

After performing simulations, binding energy calculations are necessary to measure the contribution of the residues in forming interaction between protein and ligand. This approach helps to decode the biomolecular interactions and components. MMPBSA (Molecular Mechanics Poisson–Boltzmann Surface Area) technique was used with the help of g_mmpbsa tool [84]. This tool helps to compute the Gibbs free energy of binding. Gibbs free energy equation is as follows:

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand}) \quad \text{---------- eq. 2.4}$$

Where total binding free energy of the protein-ligand, protein and ligand are represented by $G_{complex}$, $G_{protein}$ and $G_{molecule}$ in the solvent, respectively.

$$G_x = E_{MM} + (TS) + G_{solvation}$$ ---------- eq. 2.5

Where x represents complex, protein and ligand, E_{MM} represents both bonded and non-bonded interactions energy in vacuum, $G_{solvation}$ represents electrostatic & non-electrostatic contributions to solvation energy and TS represents temperature and entropy. All the parameters of energy were calculated by MMPBSA technique and finally gives total free binding energy.

Bibliography

[1] Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H. and Velankar, S. Protein Data Bank (PDB): the single global macromolecular structure archive. *Protein crystallography: methods and protocols*, 627-641, 2017.

[2] Fewster, P. F. X-ray diffraction from low-dimensional structures. *Semiconductor science and technology*, 8(11): 1915, 1993.

[3] Muhammed, M. T. and Aki-Yalcin, E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical biology & drug design*, 93(1): 12-20, 2019.

[4] Hameduh, T., Haddad, Y., Adam, V. and Heger, Z. Homology modeling in the time of collective and artificial intelligence. *Computational and Structural Biotechnology Journal*, 18: 3494-3506, 2020.

[5] Dey, F., Cliff Zhang, Q., Petrey, D. and Honig, B. Toward a "structural BLAST": Using structural relationships to infer function. *Protein Science*, 22(4): 359-366, 2013.

[6] Carugo, O. and Djinović-Carugo, K. Half a century of Ramachandran plots. *Acta Crystallographica Section D: Biological Crystallography*, 69(8): 1333-1341, 2013.

[7] Sala, D., Engelberger, F., Mchaourab, H. S. and Meiler, J. Modeling conformational states of proteins with AlphaFold. *Current Opinion in Structural Biology*, 81: 102645, 2023.

[8] Fersht, A. R. AlphaFold–A personal perspective on the impact of machine learning. *Journal of molecular biology*, 433(20): 167088, 2021.

[9] Hinz, U. and UniProt Consortium. From protein sequences to 3D-structures and beyond: the example of the UniProt knowledgebase. *Cellular and molecular life sciences*, 67: 1049-1064, 2010.

[10] Wang, Y., Wang, Q., Huang, H., Huang, W., Chen, Y., McGarvey, P. B., ... and UniProt Consortium. A crowdsourcing open platform for literature curation in UniProt. *PLoS biology*, 19(12): e3001464, 2021.

[11] Webb, B. and Sali, A. *Protein structure modeling with MODELLER* (pp. 239-255). Springer US, 2021.

[12] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. and Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1: 19-25, 2015.

[13] Zhao, J., Cao, Y. and Zhang, L. Exploring the computational methods for protein-ligand binding site prediction. *Computational and structural biotechnology journal*, 18: 417-426, 2020.

[14] Tian, W., Chen, C., Lei, X., Zhao, J. and Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic acids research*, 46(W1): W363-W367, 2018.

[15] Aggarwal, R., Gupta, A., Chelur, V., Jawahar, C. V. and Priyakumar, U. D. DeepPocket: ligand binding site detection and segmentation using 3D convolutional neural networks. *Journal of Chemical Information and Modeling*, 62(21): 5069-5079, 2021.

[16] Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug discovery today*, 11(13-14): 580-594, 2006.

[17] Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J. and Bryant, S. H. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic acids research*, 37(suppl_2): W623-W633, 2009.

[18] Wang, Y., Bolton, E., Dracheva, S., Karapetyan, K., Shoemaker, B. A., Suzek, T. O. ... and Bryant, S. H. An overview of the PubChem BioAssay resource. *Nucleic acids research*, 38(suppl_1): D255-D266, 2010.

[19] Kim, S. Getting the most out of PubChem for virtual screening. *Expert opinion on drug discovery*, 11(9): 843-855, 2016.

[20] Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., ... and Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research*, 36(suppl_1): D901-D906, 2008.

[21] Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Wilson, A. ... and Wishart, D. S. Drugbank 6.0: the drugbank knowledgebase for 2024. *Nucleic Acids Research*, 52(D1): D1265-D1275, 2024.

[22] Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P. ... and Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research*, 34(suppl_1): D668-D672, 2006.

[23] Avram, S., Bologa, C. G., Holmes, J., Bocci, G., Wilson, T. B., Nguyen, D. T., ... and Oprea, T. I. DrugCentral 2021 supports drug discovery and repositioning. *Nucleic acids research*, 49(D1): D1160-D1169, 2021.

[24] Ursu, O., Holmes, J., Bologa, C. G., Yang, J. J., Mathias, S. L., Stathias, V. ... and Oprea, T. DrugCentral 2018: an update. *Nucleic acids research*, 47(D1): D963-D970, 2019.

[25] Chagas, C. M., Moss, S. and Alisaraie, L. Drug metabolites and their effects on the development of adverse reactions: Revisiting Lipinski's Rule of Five. *International journal of pharmaceutics*, 549(1-2): 133-149, 2018.

[26] Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W. and Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*, 45(12): 2615-2623, 2002.

[27] Daina, A., Michielin, O. and Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports*, 7(1): 42717, 2017.

[28] Kharkar, P. S. Two-dimensional (2D) *in silico* models for absorption, distribution, metabolism, excretion and toxicity (ADME/T) in drug discovery. *Current topics in medicinal chemistry*, 10(1): 116-126, 2010.

[29] Sanders, J. M., Beshore, D. C., Culberson, J. C., Fells, J. I., Imbriglio, J. E., Gunaydin, H. ... and Bueters, T. Informing the selection of screening hit series with in silico absorption, distribution, metabolism, excretion, and toxicity profiles: miniperspective. *Journal of medicinal chemistry*, 60(16): 6771-6780, 2017.

[30] Parasuraman, S. Prediction of activity spectra for substances. *Journal of pharmacology & pharmacotherapeutics*, 2(1): 52, 2011.

[31] Lagunin, A., Stepanchikova, A., Filimonov, D. and Poroikov, V. PASS: prediction of activity spectra for biologically active substances. *Bioinformatics*, 16(8): 747-748, 2000.

[32] Morris, G. M. and Lim-Wilby, M. Molecular docking. *Molecular modeling of proteins*, 365-382, 2008.

[33] Pinzi, L. and Rastelli, G. Molecular docking: shifting paradigms in drug discovery. *International journal of molecular sciences*, 20(18): 4331, 2019.

[34] Fan, J., Fu, A. and Zhang, L. Progress in molecular docking. *Quantitative Biology*, 7: 83-89, 2019.

[35] Agarwal, S. and Mehrotra, R. J. J. C. An overview of molecular docking. *JSM chem*, 4(2): 1024-1028, 2016.

[36] Guedes, I. A., de Magalhães, C. S. and Dardenne, L. E. Receptor–ligand molecular docking. *Biophysical reviews*, 6, 75-87, 2014.

[37] Dias, R., de Azevedo, J. and Walter, F. Molecular docking algorithms. *Current drug targets*, 9(12): 1040-1047, 2008.

[38] Jakhar, R., Dangi, M., Khichi, A. and Chhillar, A. K. Relevance of molecular docking studies in drug designing. *Current Bioinformatics*, 15(4): 270-278, 2020.

[39] Pagadala, N. S., Syed, K. and Tuszynski, J. Software for molecular docking: a review. *Biophysical reviews*, 9: 91-102, 2017.

[40] Asalone, K. C., Nelson, M. M. and Bracht, J. R. Novel sequence discovery by subtractive genomics. *JoVE (Journal of Visualized Experiments)*, (143): e58877, 2019.

[41] Federhen, S. The NCBI taxonomy database. *Nucleic acids research*, 40(D1), D136-D143, 2012.

[42] Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J. ... and Flicek, P. Ensembl 2018. *Nucleic acids research*, 46(D1): D754-D761, 2018.

[43] Shahid, F., Shehroz, M., Zaheer, T. and Ali, A. Subtractive Genomics Approaches: Towards Anti-Bacterial Drug Discovery. *Front. Anti-Infect. Drug Discov*, 8: 144-158, 2020.

[44] Fu, L., Niu, B., Zhu, Z., Wu, S. and Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23): 3150-3152, 2012.

[45] Saravanan, S. and Shylaja, G. Genome subtraction to identify the novel therapeutic targets in Mycobacterium tuberculosis. *Drug Invention Today*, 12(8): 2019.

[46] Emms, D. M. and Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology*, 20: 1-14, 2019.

[47] Rodrigues, T. C. V., Jaiswal, A. K., de Sarom, A., de Castro Oliveira, L., Oliveira, C. J. F., Ghosh, P. ... and de Castro Soares, S. Reverse vaccinology and subtractive genomics reveal new therapeutic targets against Mycoplasma pneumoniae: a causative agent of pneumonia. *Royal Society Open Science*, 6(7): 2019.

[48] Kaur, H., Kalia, M. and Taneja, N. Identification of novel non-homologous drug targets against Acinetobacter baumannii using subtractive genomics and comparative metabolic pathway analysis. *Microbial Pathogenesis*, 152: 104608, 2021.

[49] Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Wilson, A. ... and Wishart, D. S. Drugbank 6.0: the drugbank knowledgebase for 2024. *Nucleic Acids Research*, 52(D1): D1265-D1275, 2024.

[50] Fatoba, A. J., Okpeku, M. and Adeleke, M. A. Subtractive genomics approach for identification of novel therapeutic drug targets in Mycoplasma genitalium. *Pathogens*, 10(8): 921, 2021.

[51] Kaur, H., Kalia, M. and Taneja, N. Identification of novel non-homologous drug targets against Acinetobacter baumannii using subtractive genomics and comparative metabolic pathway analysis. *Microbial Pathogenesis*, 152: 104608, 2021.

[52] Karplus, M. and McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature structural biology*, 9(9): 646-652, 2002.

[53] Acharya, K. R. and Lloyd, M. D. The advantages and limitations of protein crystal structures. *Trends in pharmacological sciences*, 26(1): 10-14, 2005.

[54] Durrant, J. D. and McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC biology*, 9: 1-9, 2011.

[55] Pikkemaat, M. G., Linssen, A. B., Berendsen, H. J. and Janssen, D. B. Molecular dynamics simulations as a tool for improving protein stability. *Protein engineering*, 15(3): 185-192, 2002.

[56] Shukla, R. and Tripathi, T. Molecular dynamics simulation of protein and protein–ligand complexes. *Computer-aided drug design*, 133-161, 2020.

[57] Khan, S. and Nandi, C. K. Optimizing the underlying parameters for protein-nanoparticle interaction: advancement in theoretical simulation. *Nanotechnology Reviews*, 3(4): 347-359, 2014.

[58] Scheraga, H. A., Khalili, M. and Liwo, A. Protein-folding dynamics: overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.*, 58: 57-83, 2007.

[59] Rakers, C., Bermudez, M., Keller, B. G., Mortier, J. and Wolber, G. Computational close up on protein–protein interactions: how to unravel the invisible using molecular dynamics simulations?. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 5(5): 345-359, 2015.

[60] Kadaoluwa Pathirannahalage, S. P., Meftahi, N., Elbourne, A., Weiss, A. C., McConville, C. F., Padua, A. ... and Christofferson, A. J. Systematic comparison of the structural and dynamic properties of commonly used water models for molecular dynamics simulations. *Journal of Chemical Information and Modeling*, 61(9): 4521-4536, 2021.

[61] Zwier, M. C. and Chong, L. T. Reaching biological timescales with all-atom molecular dynamics simulations. *Current opinion in pharmacology*, 10(6): 745-752, 2010.

[62] Hospital, A., Goñi, J. R., Orozco, M. and Gelpí, J. L. Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry*, 37-47, 2015.

[63] Hollingsworth, S. A. and Dror, R. O. Molecular dynamics simulation for all. *Neuron*, 99(6): 1129-1143, 2018.

[64] Guvench, O. and MacKerell, A. D. Comparison of protein force fields for molecular dynamics simulations. *Molecular modeling of proteins*, 63-88, 2008.

[65] González, M. A. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique*, 12: 169-200, 2011.

[66] Halgren, T. A. Potential energy functions. *Current opinion in structural biology*, 5(2): 205-210, 1995.

[67] MacKerell Jr, A. D., Bashford, D., Bellott, M. L. D. R., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J. ... and Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, 102(18): 3586-3616, 1998.

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

[69] Oostenbrink, C., Villa, A., Mark, A. E. and Van Gunsteren, W. F. 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, 2004.

[70] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., ... and Mackerell Jr, A. D. 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, 2010.

[71] Lopes, P. E., Guvench, O. and MacKerell, A. D. Current status of protein force fields for molecular dynamics simulations. *Molecular modeling of proteins*, 47-71, 2015.

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

[73] Ponder, J. W. and Case, D. A. Force fields for protein simulations. *Advances in protein chemistry*, 66: 27-85, 2003.

[74] Itoh, S. G., Okumura, H. and Okamoto, Y. Generalized-ensemble algorithms for molecular dynamics simulations. *Molecular Simulation*, 33(1-2): 47-56, 2007.

[75] Makov, G. and Payne, M. Periodic boundary conditions in ab initio calculations. Physical Review B, 51(7):4014, 1995.

[76] Barclay, P. L. and Zhang, D. Z. Periodic boundary conditions for arbitrary deformations in molecular dynamics simulations. *Journal of Computational Physics*, 435: 110238, 2021.

[77] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. and Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of chemical physics*, 103(19): 8577-8593, 1995.

[78] Rackers, J. A., Liu, C., Ren, P. and Ponder, J. W. A physically grounded damped dispersion model with particle mesh Ewald summation. *The Journal of chemical physics*, 149(8): 2018.

[79] Kutzner, C., Kniep, C., Cherian, A., Nordstrom, L., Grubmüller, H., de Groot, B. L. and Gapsys, V. GROMACS in the cloud: A global supercomputer to speed up alchemical drug design. *Journal of Chemical Information and Modeling*, 62(7): 1691-1711, 2022.

[80] Malde, A. K., Zuo, L., Breeze, M., Stroet, M., Poger, D., Nair, P. C. ... and Mark, A. E. An automated force field topology builder (ATB) and repository: version 1.0. *Journal of chemical theory and computation*, 7(12): 4026-4037, 2011.

[81] Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of chemical theory and computation*, 4(1): 116-122, 2008.

[82] Saha, D., Borah, N. J. and Jha, A. N. Molecular scaffold recognition of drug molecules against essential genes of Leishmania donovani using biocomputing approach. *South African Journal of Botany*, 162: 52-63, 2023.

[83] Quraishi, S., Saha, D., Kumari, K., Jha, A. N. and Roy, A. S. Non-covalent binding interaction of bioactive coumarin esculetin with calf thymus DNA and yeast transfer RNA: A detailed investigation to decipher the binding affinities, binding

location, interacting forces and structural alterations at a molecular level. *International Journal of Biological Macromolecules*, 257: 128568, 2024.

[84] Kumari, R., Kumar, R., Open Source Drug Discovery Consortium. and Lynn, A. g_mmpbsa A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of chemical information and modeling*, 54(7):1951-1962, 2014.