

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Role of Force Fields in Protein Function Prediction

Zaved Hazarika, Sanchaita Rajkhowa and Anupam Nath Jha

Abstract

The world today, although, has developed an elaborate health system to fortify against known and unknown diseases, it continues to be challenged by new as well as emerging, and re-emerging infectious disease threats with severity and probable fluctuations. These threats also have varying costs for morbidity and mortality, as well as for a complex set of socio-economic outcomes. Some of these diseases are often caused by pathogens which use humans as host. In such cases, it becomes paramount responsibility to dig out the source of pathogen survival to stop their population growth. Sequencing genomes has been finessed so much in the 21st century that complete genomes of any pathogen can be sequenced in a matter of days following which; different potential drug targets are needed to be identified. Structure modeling of the selected sequences is an initial step in structure-based drug design (SBDD). Dynamical study of predicted models provides a stable target structure. Results of these *in-silico* techniques greatly depend on force field (FF) parameters used. Thus, in this chapter, we intend to discuss the role of FF parameters used in protein structure prediction and molecular dynamics simulation to provide a brief overview on this area.

Keywords: homology modeling, force field (FF), molecular dynamics (MD) simulations, molecular docking

1. Introduction

What is a “disease”? A disease is any condition that harms the normal function of a body organ and/or system, of the psyche, or of the organism as a whole, which is associated with specific signs and symptoms. Factors that often lead to the damage of the function of organs and/or systems may be of two types, i.e., intrinsic and extrinsic. Those factors, that arise from within the host body interfering with the normal functioning processes of a body organ and/or system, as a result of genetic features of an organism or any disorder within the host are known as intrinsic factors [1]. Huntington’s disease is an example of genetic disease which causes uncontrolled movements, emotional problems and loss of thinking ability (cognition) owing to a progressive brain disorder, due to mutations in the *HTT* gene, involving a DNA segment known as CAG trinucleotide repeats [2]. When a host comes in contact with a pathogen from outside, the host’s system is accessed by extrinsic factors [3]. Microorganisms are the main causative agents which are responsible for causing infectious diseases. Their importance is determined from the type and extent of damage their causative agents inflict on organs and/or systems when they enter into a host. Entry into the host is mostly by routes such as the mouth, eyes, genital openings, nose and the skin. Damage to tissues mainly results from the growth and

metabolic processes of infectious agents intracellular or within body fluids, with the production and release of toxins or enzymes that interfere with the normal functioning of organs and/or systems [4]. An example of extrinsic factor is the infection caused by novel pathogen, such as SARS-CoV-2, which represents an extremely challenging and complex endeavor. Currently, several promising therapeutics are underway and also many vaccine candidates with promises to mitigate the catastrophic effects of COVID-19 pandemic are under clinical trials. Still, an effective and successful countermeasure to control this catastrophe is not available [5].

In December 2019, a kind of pneumonia having an unknown etiology was reported from the Wuhan city of China in the Hubei province [6]. Isolation and genomic characterization of the complete sequence of the virus using next-generation sequencing (NGS), identified it as a novel coronavirus (CoV) and named it as 2019-nCoV, now as SARS-CoV-2 [7]. Although the characterization of the complete sequence was completed in January 2020, yet till date, there is no definitive cure or vaccine available for this virus. With the availability of the sequence, the three-dimensional (3D) structures of many proteins belonging to SARS-CoV-2 are now available. These 3D-structures can be obtained using various experimental and computational techniques. X-ray crystallography and NMR spectroscopy are currently the two major experimental techniques for protein structure determination [8] which are deposited in both UniProt and Protein Data Bank (PDB) [9]. For computational modeling of the 3D structure of proteins, homology modeling technique is used. Homology modeling is a computational technique which uses the amino acid sequence to predict the 3D structure. It is one of the widely used computational structure prediction method.

Proteins are one of the most extensively studied and complex macromolecules within living organisms with a unique 3D structure. Usually this leads to a diversity in their spatial shape, structure and thus, leading to different biological functionalities in a living system [8]. Yet, very little is known about the process of protein folding leading to its specific tertiary structure from its primary structure. Till date, approximately 175,000 experimentally determined 3D structures of biological macromolecules are available in the PDB [9]. However, reference sequence (refseq) release of National Center for Biotechnology Information (NCBI) contains as many as 178,304,046 protein sequences. This signifies a huge difference between the number of sequences in the NCBI and the number of protein 3D structures in the PDB. The difference in the number is even higher due to the fact that the reference sequences in the NCBI are non-redundant, whereas, structures available in PDB contain redundancy. This has resulted in an alarming situation owing to the increasing gap between the available 3D structures and the protein sequences. Therefore, computational structural prediction methods such as homology modeling are much needed in covering this widening gap. Thus, this chapter discusses homology modeling in a holistic manner covering the principles and different types of structure prediction methods along with giving a flavor of the different force field (FF) parameters that are used in protein structure prediction. The chapter also includes a brief overview of the molecular dynamics (MD) simulations that are used in computational modeling of proteins along with discussion of some application examples in this field.

2. Protein structure prediction

Protein sequences are much easier to obtain as compare to their structures. This is due to advancements in the field of protein sequencing technology. As a result, an exponential growth in the accumulation of protein sequences can be observed. An amino acid sequence is a very important source of insight into proteins, its function, structure and history. This is mostly because, first, comparison of an unknown

sequence with a known sequence helps in deciding whether significant similarities exist between them, which in turn helps in establishing the class of protein and can give valuable information regarding its structure and function. Secondly, genealogical relationships can be studied by comparing the sequences of the same protein from different species. Thirdly, the presence of internal repeats in protein sequences reveals the history of the proteins. Also, sequencing of amino acids is very important for making DNA probes which can be used for encoding of its protein, as knowledge of the primary structure also allows the use of reverse genetics [10].

2.1 Amino acid sequence determination techniques

Determination of the amino acid sequence of all or part of a protein or peptide is known as prediction of protein sequence. It is used to categorize the protein and may help in characterizing its post-translational modifications. In a protein, determination of the amino acid sequence involves the following steps [10]:

- i. *Hydrolysis*: This procedure is required in order to hydrolyze the protein into its amino acid and includes the protein being heated in 6 M hydrochloric acid (HCl) at 100–110° C for 24 hours or longer.
- ii. *Separation*: Separation of amino acid from a peptide can be achieved by ion-exchange chromatography. The amino acids are eluted by mixing them with an acidic solution and passing a buffer steadily while increasing the pH through the chromatography column on sulfonated polystyrene. Accordingly, when an amino acid reaches its isoelectric point, it is separated. The buffer used is correlated to a specific amino acid type. Thus, the amino acid having the most acidic side chain will emerge first, while the amino acid having the most basic side chain will emerge last. The absorbance is used to determine the amount of similar type amino acid residues.
- iii. *Quantitation*: Once the separation of the amino acids is achieved, their respective quantities are determined by adding a reagent called ninhydrin which gives an intense blue color to the amino acids, except proline which, due to the presence of secondary amino group in its structure, gives it a yellow color. For very small quantities (nanogram), reagents like fluorescamine or ortho-phthaldehyde (OPA) are used to obtain fluorescent products. Therefore, the concentration of amino acids is directly proportional to either the absorbance of the resulting solution or the fluorescence emitted by the sample.

For determining the composition and the sequence of the protein, two direct methods can be used:

- a. *Edward Degradation Method*: This method uses phenyl iso-thio-cyanate to cleave the amino acids one by one starting from the amino terminal. The amino acids when treated with phenyl iso-thio-cyanate forms a phenyl-thio-hydantoin (PTH)-amino acid (e.g. PTH-lysine, etc.) terminal residue, which gets released under mild acidic conditions. The released terminal compound is then identified using chromatographic procedures.
- b. *Mass Spectrometry*: Another technique to determine protein sequence is the mass spectrometry which uses the time of flight of ionized proteins to calculate the mass of the ionized proteins. In this process, the protein is cleaved using specific enzymes. The ionized amino acids are triggered by a laser beam which

travels to the detector through a flight tube. The ions with lighter mass will reach the detector faster due to Newton's second law ($F = ma$) and hence, will be detected first. After the spectrum is recorded, it is further analyzed and compared against a database of sequenced proteins. A detailed sequence of protein fragments can be determined by repeating the process with different enzymes for cleavage. As a result, the fragments become much smaller with the fragments overlapping each other establishing the order of the protein.

2.2 Experimental determination of protein structure

The basic prerequisite for understanding the function of a protein is the knowledge of the protein 3D structure. The experimental methods used in the study of tertiary structure include:

- i. *Protein X-ray crystallography*: X-ray crystallography is presently the most sought-after technique for determination of biological macromolecule structures. In this method, the determination of protein structure is achieved by crystallization of the purified protein at high concentration and exposing the crystals to an X-ray beam. The resultant diffraction patterns, obtained from the diffraction spots, are then processed to get knowledge about the symmetry of the packaging of the crystal and the size of the repeating units forming the crystal. A map of the electron density is then calculated using the "structure features", which are determined from the intensities of the diffraction spots. The quality of the electron density map can be improved using various methods. This is done to get a definitive idea to build the molecular structure using the amino acid sequence. Finally, the structure that is obtained is further refined to fit the map more accurately and to assume a conformation which is thermodynamically more favorable. Protein crystallography is known to provide highly accurate protein structures by giving atomic resolution. However, this method is not always straightforward and may take a lot of time to complete, which is around 3–5 years [11].
- ii. *Nuclear magnetic resonance (NMR spectroscopy)*: Another useful technique to determine the protein structure is the NMR spectroscopy. It is a primary quantitative method which allows concentration determination of proteins in an aqueous environment that may resemble its actual physiological state more closely. In principle, the NMR spectroscopy is dependent on the electromagnetic radiation and the sample protein interaction. It is used to observe the local magnetic fields prevailing around the protein atomic nuclei. The NMR signal is obtained when sensitive radio receivers detect the excitation of the material nuclei with radio waves into the nuclear magnetic resonance. Thus, it provides access to the electronic structure of the sample protein. The major advantage of NMR over X-ray crystallography is that the protein in NMR spectroscopy can be examined in their native-like physiological state. However, NMR is not suitable for proteins with more than 150 amino acid and needs the protein under study to be stable in room temperature for a long time of data acquisition, which is a drawback of this technique [12].
- iii. *Electron microscopy (especially Cryo-electron microscopy)*: Electron microscopy (EM) and cryo-electron microscopy (cryo-EM) are used to study objects that are comparatively larger in size such as cellular organelles or large macromolecular complexes with higher resolution. EM and cryo-EM use a method known as single-particle reconstruction. In principle, the data set in EM and

cryo-EM is split randomly into half and the two averages (or 3D reconstructions) over rings (or shells, respectively) are compared, with increasing radius in Fourier space using an appropriate amount of reproducibility [13]. The protein sample in EM and cryo-EM does not require crystallization, saving a lot of time and effort, which is a major advantage over protein x-ray crystallography. Nevertheless, for membrane proteins, electron crystallography is used which require two dimensional (2D) crystals of the sample protein. Another advantage of cryo-EM is that it requires very less amount of sample materials. However, one of the limitations of cryo-EM is that it has to compromise with the resolution comparative to resolution obtained from x-ray crystallography and NMR spectroscopy [14].

2.3 Protein structure prediction

The field of structural biology is mostly dominated by experimental methods which are expensive and laborious in nature. However, since the last few decades, the application of computational techniques in structural biology has been widely used, with significant improvements in these techniques since last 10–20 years. This has helped to achieve substantial developments in protein structure prediction methods. *In-silico* protein structure prediction enables the prediction of 3D structures for proteins with known sequences and unknown structures. Prediction of the tertiary structure also helps in understanding the folding and unfolding of proteins. Also, protein engineering may help in incorporation of new functions in proteins thus facilitating drug design and discovery [15]. Protein structure prediction can be achieved by three different ways:

- i. Computer simulation-based on empirical energy minimization
- ii. Knowledge based-approaches using information derived from known sequences of experimentally determined protein 3-D structures
- iii. Hierarchical methods.

2.3.1 Approaches based on energy minimization

The energy minimization method is also known as the *ab-initio* (*de novo*) method for protein structure prediction and is based on the theory that the native structure of protein is always at thermodynamic equilibrium with minimum energy, which is calculated using basic laws of physics and chemistry (**Figure 1**). Energy minimization-based methods always attempt to detect the global minima in free energy surface of the protein molecule as it is thought that global minima correspond to the native conformation. This method is not very helpful to design protein sequence length of more than 150 amino acid residues. However, it can be used to design small stable peptides that can bind to any specific therapeutic targets [16]. Two types of energy minimization methods are broadly used in *de novo* structure prediction approach, namely static and dynamical minimization methods. Some of the major FF used for energy minimizations are GROMOS, AMBER, CHARMM and ECEPP [17, 18]. One of the *ab-initio* protein structure prediction software packages is ROSETTA. This software package is based on the postulation that local interactions lead the conformation of short segments while global interactions establish the 3D protein structure [19]. The advantage of *ab-initio* approach is that it is based on physicochemical principles, however, these principles are hampered by the vast number of degrees of freedom which are needed to be looked after and also the performance of energy functions are

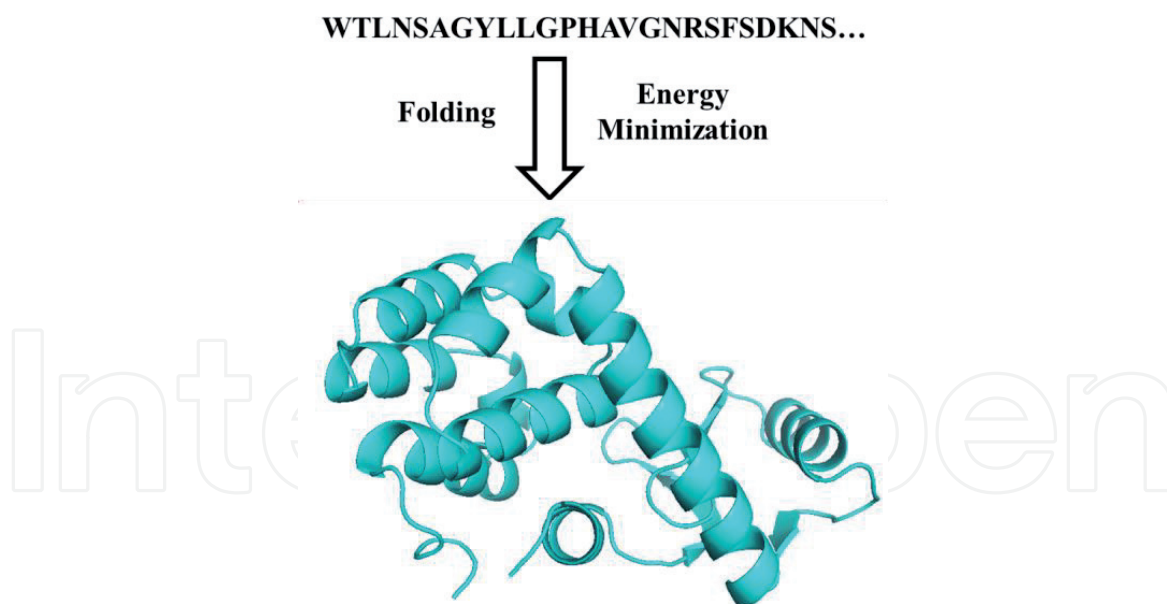


Figure 1.
An example of ab-initio structure prediction.

limited. The disadvantage of this method is that it requires high computations and for such studies there are no “good enough” interaction potentials which can model the native structure of a protein with atomic detail [20].

2.3.2 Approaches based on knowledge

The available protein structures are used to derive the knowledge based potentials [21, 22]. Further, these potentials are used to obtain the secondary structural information from amino acid sequence. The methods, based on the knowledge procured from known protein structures are of two types.

2.3.2.1 Homology modeling

One of the most powerful methods used to predict the 3D structure of proteins is the homology modeling. This method, also known as comparative modeling, uses a query protein having sequence similar with the target protein, having known tertiary structure [23–25]. The basis of this method lies on the observation that structures are more conserved than their sequences. Thus, if a target sequence has some degree of similarity with a protein sequence having known 3D structure, then that structure can be used to precisely model the target protein. A plethora of review articles are available on the strategies and challenges of computational protein structure prediction [8, 26].

For an accurate model building of a protein using homology modeling approach, the first step is template selection. The most crucial step involves the generation of a structure-based alignment between the query and the template protein sequence [27]. Models cannot be constructed for alignments having less than 20% identity. Additionally, the environment of the template such as the type of solvent, pH, presence of ligands, etc. and the quality of the experimentally-derived template structure must be taken into account. Once a desired template structure has been selected, a target-template alignment must be performed using standard sequence alignment techniques. After the creation of the template-target alignment, the 3D model of the target protein is created using several algorithms. Distance geometry is one of the commonly used methods to satisfy the spatial restraints obtained from the target-template alignment. MODELLER is one of the reliable homology

modeling program and it imposes spatial restraints that are derived from the bond distances and angles in the target structure that are based on its alignment with the template structure, and stereo-chemical restraints on bond distance and dihedral angle preferences that are obtained from a representative set of all known protein structures. Then the constructed model is getting minimized using molecular dynamics to follow the spatial restraints [28].

After the creation of 3D model, the next step is to perform the quality assessment of the predicted model. From last few decades, many methods have been developed to assess the quality and correctness of modeled protein structures which analyze their stereochemistry. Some of the programs for such analysis are PROCHECK [29] and WhatCheck [30]. Another method to analyze the modeled protein is to calculate a residue-by-residue energy profile, where a peak in the profile corresponds to an error in the model. But this method has a drawback considering that a section of residues may appear to be inaccurate, while in reality they will be interacting with an incorrectly modeled region. Thus, for the assessment of modeled proteins, energy profile should not be the only means of identifying a good model.

Homology modeling for the prediction of protein 3D structures consists of multiple steps (**Figure 2**). Although a number of tools and web-servers are available, but no single server or tool can be considered as best in every aspect as compared to others. The function of a protein is dependent on the 3D structure; therefore, it is very important to enhance the quality of the predicted model. Homology modeling has a wide variety of applications in structural biology and plays a vital role in drug discovery process, as because for the study of drug-receptor (protein) interaction, the structure of the receptor (protein) is of utmost importance. However, this approach does not work if homologous structures are not available.

2.3.2.2 Threading

Threading, also known as fold recognition is a method that searches the protein structure template in a library of folds with the lowest possible energy for a given query sequence [15]. Fold recognition of a sequence requires a precise alignment of the query sequence corresponding to the positions of the amino acid residues of a folding motif. A set of possible positions of the amino acids in 3D space is

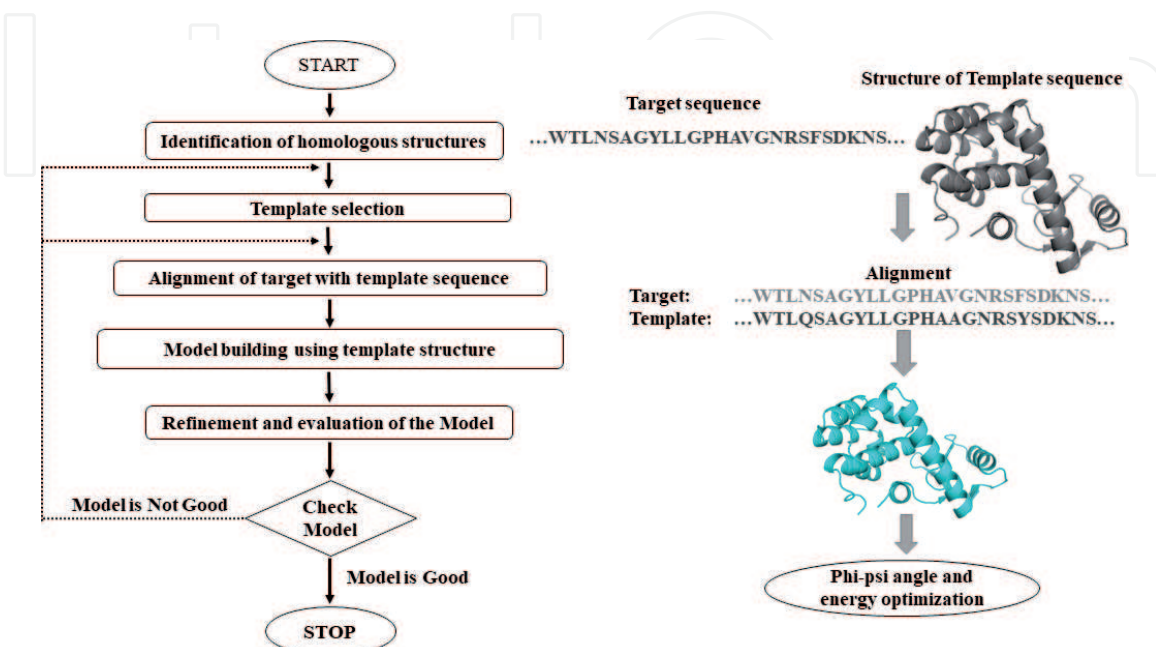


Figure 2.
 A scheme of homology modeling.

established by the known structure. This step is followed by making a similar structure by placing the amino acids of the query sequence into their aligned positions. The main goal of this method is either to choose the most probable fold for any given sequence or to find out the appropriate sequences that have the possibility to fold into a given structure. This method is heavily dependent on the knowledge of experimental atomic details of the recognized protein folds and is generally applicable for only those proteins whose amino acid sequences adopt one of the protein folds that have already been experimentally established.

2.3.3 Approaches based on hierarchy

The Hierarchical approach is another strategy for protein structure prediction from their sequences. In principle, this method uses the hierarchy of protein structure, i.e., from the primary to secondary structure and secondary to tertiary structure. Thus, in order to understand the relationship of the primary amino acid sequence and the tertiary 3D structure, the intermediate secondary structure is predicted. This intermediate structure is used to build the tertiary 3D structure. A number of algorithms are developed for the modeling of secondary structure, but, unfortunately, the precision for prediction of secondary structures from their sequences is only about 80%. Currently the methods that are available for the secondary structure modeling can be divided into methods based on statistics, physicochemical properties, evolutionary information, combinatorial analysis and artificial intelligence [31–33].

2.4 Structure prediction methods and benchmarking

The performance assessment of existing methods is one of the major setbacks in the field of protein structure prediction as methods have been and are still in the process of development using different proteins with various evaluation criterions. Thus, in 1994, an open experiment was conducted all over the world with the intention of helping the developers and users of these methods. The experiment was called the Critical Assessment of Protein Structure Prediction (CASP) (<https://predictioncenter.org/>) [34]. The CASP is a community-wide, worldwide experiment which is conducted every two years since 1994. CASP allows research groups to test their structure prediction algorithms and establish the current state of the art in protein structure prediction. They help to identify the current progress as well as highlight the efforts that are needed to be addressed in the future.

3. Proteins: structure and function

Proteins are simple polymers of amino acids. The short stretches of polymers join together and get folded to form secondary structures which in turn give rise to the 3D structure of proteins. The secondary structures can be recognized either by hydrogen-bonding (H-bond) patterns among the carbonyl and amide groups in a peptide backbone or from the dihedral angles viz. phi and psi. Mainly two known secondary structures in a protein are α -helices and β -sheets which tend to build up into small repeating arrangements in protein structures; termed as ‘supersecondary structures’ or ‘motifs’. These secondary structures assemble into larger subunits of structures termed as ‘domains’. Domains can be further understood as the smallest structural unit of proteins which can be folded autonomously such as serine protease which is made up of two β barrel domains. Proteins comprises either of a single domain or multiple domains. Protein structures were for the first time categorized into folds in 1976 [35]. Murzein et al. later incorporated the idea and developed the

publicly accessible database named SCOP (Structural Classification of Proteins) [36]. Folds in the SCOP were categorized by the class of secondary structure: all α , all β , α/β (wherein helices and sheets are mixed) and $\alpha + \beta$ (separate helices and sheets). Proteins are the most ubiquitous biomolecules and they accomplish the vast majority of functions in all the biological domains. The sequence-structure-function paradigm attracted the interests of scientists all over the world. As the proper functioning of all the biological processes depends on proteins and their non-functioning leads to grave diseases and disorder, biologists started working on them ever since. Way back in 1970s, Anfinsen have proposed that the 3D structure of native proteins comes from its sequence in a specified environment [37].

As proteins are dynamic in nature, experimental techniques fail to capture their different dynamical conformations and specially the transition between these conformations. One of the most widely utilized computational techniques, Molecular Dynamics (MD) Simulation tackles this challenge efficiently.

3.1 Molecular dynamics: the computational microscope

MD simulations assist us to comprehend and witness the time dependent behavior of proteins. As MD simulations have the ability to show the dynamic behavior of proteins at the level of atoms, it is also considered as computational microscope [38]. In this technique one requires an initial protein model which is obtained by either experimental methods or predictive modeling. As life sustains itself in water therefore one mimics simulation in explicit solvent. When the forces acting on all the atoms were acquired, Newton's laws of motion were utilized to compute the velocities and accelerations; besides updating the atom's positions. A time step of 2 fs (femtosecond) is usually applied for atomistic simulations while integrating the movement numerically. Finally, a trajectory of the system is generated by MD engine which can be further analyzed based on set objectives. The technique was first utilized in early 70's to study the most relevant biological challenge of the time; protein folding [39, 40]. The subsequent decades saw the application of MD simulations for investigating folding and unfolding mechanism of proteins [41]. Duan and Kollman were successful in 1998 to perform 1 μ s MD simulation for the first time on parallel supercomputer. They investigated the protein folding mechanism of villin with explicit solvation [42]. Apart from proteins, the technique has been extended to study other relevant biomolecules [43, 44] and protein-nanoparticle interactions [45–49].

Simulation of any system revolves around lot of factors. Earlier the system size comprises of few thousand of atoms. With the advancement of both experimental and computational techniques, availability of 3D data in regard to proteins, proteins complexes, membrane proteins etc. has been possible which made the system size amplified to several lakhs of atoms with explicit solvent in consideration [50]. Meanwhile the advent of high-performance computing (HPC) and algorithm parallelization made it possible to run long timescale simulations for the above-mentioned systems. Further advancements in the algorithms of MD engines and/or the implementation of GPUs (graphical processing units) along with CPUs have significantly improved the performance of MD simulations. Some of the most popular simulation engines are: AMBER, CHARMM, DESMOND, GROMACS and NAMD. They have been integrated with messaging passing interface (MPI), which made it possible to utilize all the available cores of the computer simultaneously during a MD run to reduce the computation time.

3.2 Workhorse of simulation: the force fields

Force fields (FF) lie at heart of the MD simulation. In order to perform simulation, one needs the parameters to deduce the potential energy function [51].

The FF is a group of equations and associated parameters designed to imitate molecular geometry and selected properties of some tested molecules. FF comprises primarily of two components; bonded and non-bonded terms. Any molecular feature can be basically represented with them. The bonded terms can be represented by springs for bond length and angles along with torsional angles; the non-bonded terms comprise of Lennard-Jones potentials for van der Waals (vdW) interactions and Coulomb's law for electrostatic interactions. They were primarily developed to reproduce structural properties and applied to predict other properties such as thermodynamic parameters. Further the energy functions utilized in molecular mechanics commonly comprise topological parameters which are obtained from experiments or quantum mechanical calculations. An important feature of FF is transferability of the parameters and the functional form. It means to model a series of related molecules; the same set of parameters can be utilized rather than defining a new set of parameters for each individual molecule. Even though most of the FF are additive, a number of them having higher order terms are called class II FF. Some of widely utilized FF for bio-molecular simulations are AMBER, CHARMM, GROMOS and OPLS [52]. Additionally it is noteworthy to mention the application of FF in predicting structures of proteins/RNA. FFs were developed and benchmarked against experimentally solved structures and these FF were later incorporated to predict the structure for the ones lacking experimental information. Another important aspect of the FF is to discriminate the near-native protein conformation among the generated 3D models [53]. FFs are subject to rigorous scrutinizing and they were refined to improve their accuracy over time. One such example is the improvement of the residue side-chain torsion potentials of the Amber ff99SB FF which is also validated with available NMR experimental datasets [54]. A number of benchmark studies were conducted time to time, to compare different FFs. One difference arises among the available variety of FF is the bias/overestimate towards particular secondary structure of proteins. Man et al. recently concluded from their comparative simulation study that FFs (AMBER94, AMBER99 & AMBER12SB) were not able to predict β -sheet formation whereas FFs (AMBER96, GROMOS45a3, GROMOS53a5, GROMOS53a6, GROMOS43a1, GROMOS43a2, and GROMOS54a7) were able to form β -sheets swiftly. Further they have showed that the best FFs for investigating amyloid peptide assembly based on their structure and kinetics were AMBER99-ILDN, AMBER14SB, CHARMM22*, CHARMM36, and CHARMM36m [55].

3.3 Application examples of MD simulations

MD simulations have immensely contributed to solve and hypothesize many biological research problems. The significance of the computational microscope can be well understood by observing the increase in the vast repertoire of literature in the recent decade. The technique of simulation along with other computational tools plays a significant role in the field of protein structure prediction. Using a set of seven small proteins Kato et al. have validated the application of MD simulations to predict the 3D structure of proteins. The set of small proteins were in the range of 10 to 46 residues. They have considered two properties; root mean squared deviation (RMSD) and occurrence of secondary structure to validate the predicted structures from simulation with that of the available experimental ones. AMBER12 simulation package with AMBER ff12SB have been utilized to carry out their simulations. With the help of MD simulations, they have shown the possibility of reproducing the secondary structures of small proteins [56]. Our group has also utilized the indispensable technique of simulation recently to investigate the dynamics and stability of *ab-initio* predicted structure of bacterial effector protein, HopS2. The importance of the

effector proteins lies with them conferring pathogenicity to bacteria. As the sequence similarity of the effector proteins lies in the twilight zone along with the few partially solved structures of effector proteins at disposal, it is a perplexing task to study the sequence-structure-function relationship of these proteins. With the assistance of MD simulations, our group was able to show the stability of local secondary structural elements of HopS2 which are vital to its overall structure and interaction. These investigation have been performed using Gromacs along with OPLS FF [57].

Another interesting aspect of human proteome is the intrinsically disordered proteins (IDP). There are many examples of proteins with folded domains but they feature disordered regions while some are entirely unstructured. Some IDPs fold upon interacting with their binding partners while other persists in unfolded state even in a bound complex. The IDPs plays a critical role in cell signaling and regulation. Pietrek et al. have carried out a recent work in this direction. They have considered a hierarchical algorithm to generate large ensembles of full length IDP structures and these structures can be further used as starting points for atomistic simulations. The IDP structures generated by their hierarchical approach implemented with all atom MD simulations were able to capture both local conformations compared with NMR experiments and also the gross dimension described by small angle X-ray experiments. Gromacs simulation package along with Amber03ws and Amber99SB*-ILDN-q FF were utilized by them to carry out the investigation [58]. The powerful computational microscope was also applied to investigate structure and dynamics of plasma membrane proteins. Mattedi et al. recently utilized MD simulations to study glucagon receptor, a class B GPCR. The glucagon-induced release of glucose from the liver into the bloodstream is facilitated by the glucagon receptor. There is scarce information about the mechanism of this receptor. They utilized extensive MD simulations and free energy landscape computation to elucidate the activation mechanism of the receptor. Through their simulation work, they identified an intermediate state of the glucagon receptor and decipher the mechanism of allosteric antagonists of the glucagon which locks transmembrane helix 6. They have employed AMBER14SB FF and LipidBook parameters for lipids with Gromacs package in their work [59].

4. Molecular docking

The plethora of diseases discovered ever since and being investigated tirelessly by scientists all over the world ultimately culminates to the sole objective of finding effective solutions. The therapeutic targets in most of the cases are proteins. After knowing their mechanism of actions, how the proteins works and what goes wrong during the diseased state, the next notion is to challenge their functionality with designing some inhibitors. It comes under the domain of drug discovery. And one of the most challenging fields of study is the drug design and development. The complete clinical trials take about 10–15 years of time with billions of dollars expenses for a single drug to reach market. With the completion of human genome project which leads to identification of ever-increasing number of new drug targets (mainly proteins); the efforts were strengthened to find solution to the diseases. Additionally, the availability of 3D structures of protein and protein-ligand complexes made it feasible to carry out research in this area. However, to experimentally screen millions of compounds and their conformers for a single therapeutic target requires enormous amount of time and resources which makes it quite challenging. With the application of computational techniques, the pre-clinical period can be reduced to save valuable assets. The *in-silico* approaches will significantly curtail the time needed for hit identification and also improve the chances of finding the anticipated drug molecules. To facilitate drug design and discovery, several modeling

techniques were available and mostly they are categorized into two main approaches viz. structure-based and ligand-based drug design approaches. The structure-based approach mainly relies on the 3D data of target and the ligand. The ligand-based approach is chiefly adopted in the absence of known experimental structure of the target. In ligand-based approach, the known ligands which were bound to the targets were investigated to decipher the physiochemical and structural properties of the ligands and these were correlated with the anticipated pharmacological activity of the ligands in hand [60].

One of the most extensively utilized computational techniques in the structure-based drug design is molecular docking. Molecular docking is usually achieved by first predicting the molecular orientation or pose of a ligand within the active site of a target and followed by assessing their binding affinity with the usage of a scoring function. The technique is exploited to decipher the interactions between a target and ligand at the atomic level allowing us to describe the behavior of ligands within the active sites of targets as well as to reveal fundamental biochemical processes. Since the first developments of docking algorithms in the 1980s, molecular docking became an indispensable tool in the field of drug discovery [61].

4.1 Types of molecular docking

Molecular docking can be basically categorized into three types: rigid docking, semi-flexible docking and flexible docking. In the rigid docking approach, both the structure of target and ligand does not change. The computation method is relatively modest and chiefly spans the degree of conformational matching, thus it is more apt for investigating macromolecular systems such as protein-nucleic acid and protein-protein systems. The semi/quasi flexible docking approach take flexibility into consideration while docking of the ligand and thus it is more appropriate to deal with the intermolecular interactions of small molecules and proteins. Usually the structure of the ligands can move freely while the target remains rigid or retain few rotatable residues ensuring computational efficiency during the docking process. In the flexible docking method, it is based on the idea that a protein is not always a rigid entity during the course of ligand binding and thus it considers both the protein and ligand as flexible entities. Over the years various methods have been introduced, based on induced fit model and/or conformational sampling.

4.1.1 Scoring function

One crucial element of any docking algorithm is the scoring function. The scoring function aids in the pose selection and it is involved in distinguishing putative precise binding modes and to filter out the non-binders from the N number of generated poses during a docking run. The speed and accuracy of docking programs is also dependent on scoring functions. Further computational efficiency and reliability are points kept in mind while developing any scoring function. There are three categories of scoring functions:

i. Force-field based scoring function

This scoring function is based on the concept of molecular mechanics which estimates the potential energy of a system with a mixture of intramolecular and intermolecular elements. In molecular docking, the intermolecular elements are usually considered, with the probable ligand-bonded terms, especially the torsional constituents. The non-bonded constituents include the van der Waals term which

is defined by Lennard-Jones potential, and the electrostatic term, specified by the Coulomb function. GoldScore [62], AutoDock [63] and GBVI/WSA [60] are few examples of the mentioned scoring function.

ii. Empirical scoring function

Empirical function is the sum of different empirical terms such as van der Waals, H-bond, electrostatic, entropy, desolvation, hydrophobicity, etc. Utilizing least square fitting method, they are optimized on a training set of target-ligand complexes to reproduce the binding affinity data. Empirical scoring functions compared to force-field ones are computationally much more efficient owing to their simple energy terms. The first example of empirical scoring function is the LUDI scoring function [64]. GlideScore [65] and ChemScore [66] are other examples of empirical scoring functions.

iii. Knowledge-based scoring function

Knowledge-based functions are directly obtained from the structural information of experimentally solved protein-ligand complexes. The frequencies of interatomic contact and/or distances between the target and the ligand are obtained. The premise for this criterion relies on the assumption that frequency of occurrences will be greater for the ones with more favorable interactions. Pairwise atom-type potentials were generated with the obtained frequency distributions. Further the score is computed by preferred interactions and imposing penalty for repulsive contacts between each pair of atoms in the target and ligand within a set cutoff. Examples of this scoring function are DrugScore [67] and GOLD/ASP functions [68].

With the advancement in the field of high-performance computing, scientists have also applied artificial intelligence based and machine learning based scoring functions in virtual screening which holds promising outcomes [69].

4.1.2 Sampling algorithms

Sampling plays the next crucial role in any molecular docking program. With a set therapeutic target, the sampling algorithm will generate a number of conformations (poses) of the small molecule within the docked site of the target. The knowledge of the docked site is considered either from experimental data or predicted with the aid of active site prediction software. As the speed and accuracy of molecular docking plays a role in large virtual screening research works, the area of developing and/or improving existing sampling algorithms have provided ample opportunities for computational scientists. The sampling algorithms can be categorized as: shape matching, systematic search algorithm and stochastic algorithm.

i. Shape matching

One of the earliest methods designed was the shape matching algorithm for sampling. The criterion implemented in this algorithm is that the molecular surface of the small molecule needs to complement the molecular surface of the binding region of the target. The three translational and three rotational (six degree of freedom) of the small molecule led to spans many probable orientations. Thus, the goal of this algorithm is to place as smoothly and quickly the small molecule into the binding site based on shape complementarity. In this method, the conformation of the small molecule is usually fixed and therefore, this method along with

flexible-docking is usually preferred rather than only shape matching. DOCK [61], LigandFit [70] and Surflex [70] are few examples of docking programs where shape matching algorithm is used.

ii. Systematic Search

With the help of systematic search algorithm, the ligand can explore all the degrees of freedom and it can generate all probable conformations. Unlike in shape matching algorithm, the conformations of ligands are not fixed here. Systematic search technique can be categorized into three types: exhaustive search, fragmentation and conformational ensemble.

In exhaustive search method, all the rotatable bonds of the small molecules are scanned in a systematic manner. However, to avoid a huge combinatorial explosion & to make the docking procedure practical, the search space is limited by geometric constraints criterion. Glide docking program implements this method.

The fragmentation method as the name suggests implements the idea of fragmenting the ligands into smaller rigid fragments. The incremental construction is one such mode wherein one fragment is placed first in the binding site and other fragments were attached incrementally. FlexX [70] utilizes this algorithm.

In the conformational ensemble algorithm, small molecule flexibility is signified by rigidly docking an ensemble of pre-generated conformers of the small molecule. Next the binding modes were collected from different docking runs then binding energy values are used to rank them. FLOG [71] and MS-DOCK [70] implements this algorithm.

iii. Stochastic Search

In the stochastic search, the sampling of the small molecule conformations is carried out by making random changes at every step in both the rotational/translation space and conformational space of the small molecule respectively. A probabilistic criterion is placed to either accept or reject the random change. Within stochastic search, there are four subtypes viz., Monte Carlo method, evolutionary algorithms (EA), Tabu search methods and swarm optimization (SO) methods. Genetic algorithm, one type of EA is implemented in AutoDock [63] and GOLD docking programs.

It is imperative to mention here that different docking programs/servers apply variety of algorithms in multi-phase wise in their docking pipeline.

4.2 Application examples of molecular docking

The molecular docking can be seen applied regularly in academic labs and pharmaceutical companies to find effective solutions and thwart deadly diseases [72]. The identification of hit molecules in the preliminary stage of drug discovery is today heavily relied upon high throughput screening. Moreover, the availability of small molecule databases such as PubChem, ZINC, MayBridge etc. along with the growth of experimental structures of targets (proteins, membrane proteins, protein-ligand complexes) have made the use of molecular docking to screen millions of compounds and made it possible to test only lead molecules.

G protein-coupled receptors (GPCR) are the attractive targets of drug design regimes because of their importance in cell signaling and functions. Kolb et al. have considered β 2-adrenergic receptor, a GPCR found in the smooth muscle tissue to investigate the structure-based approach for ligand discovery. In their study, they have utilized DOCK molecular docking program to screen approx. 1 million

compounds from ZINC database. They were able to test experimentally the resultant 25 high ranked molecules from docking; of which 6 molecules showed binding affinity $<4\ \mu\text{M}$. And the best compound showed 9 nM of inhibition constant against the receptor [73].

Rajkhowa et al. have utilized the structure-based drug design (SBDD) method along with MD simulations to design inhibitors against malaria, one of the most devastating infectious diseases. They have considered 178 compounds similar to known anti-malarial imidazopyrazine from the PubChem database to carry out the work. The target of the inhibitor is the phosphatidylinositol-4-OH kinase which is a lipid kinase involved in the membrane ingestion process of the erythrocytic stage of the life cycle of the plasmodium and recognized as a drug target. AutoDock 4.2 has been utilized in their work. They have reported three potential inhibitors based on molecular docking, MD simulations and ADMET studies [74].

Our group had worked in the direction of SBDD to tackle insulin resistance and type-2 diabetes (T2D). We have considered 142 anti-diabetic compounds spanning various categories of phytochemicals such as flavonoids, alkanoids, sulfonylurea and terpenes. The target of the study is A_{2A} adenosine receptor which had been shown in reports that it can be utilized to counteract insulin resistance and adipocyte inflammation. Numerous computational tools were utilized to carry out the work such as druglikeness filtering, QSAR modeling, ADMET profiling to molecular docking. The different level of screenings led to 6 molecules which were docked with the help of two different molecular docking approaches viz. AutoDock and AutoDockFR to get optimal receptor-ligand conformations. From the 142 compounds finally we got one molecule “indirubin-3'-monoxime” which is then followed by experimental validations [75].

5. Conclusion

In this era of high-performance computing technology, there is hardly any field of science which is not touched upon by some amount of significant computational works. The potential of computing power is much reliant on advancement in hardware and algorithms. Substantial number of computational tools and techniques were developed and applied in the fascinating area of proteomics also. Mathematical models were devised in the form of FF parameters and implemented in various algorithms. Here, we have discussed the inevitable role of FF in protein structure prediction/modeling, conformational dynamics and their functional aspects along with the applications in virtual screening programs. As discussed in the chapter, a lot of programs with variety of FFs are available for structure prediction, MD simulations etc., but there is still a scope of further developments. For example, till now it is a challenge for accurately predicting protein structures of larger sizes or the protein sequences having low amount of similarity with sequences of known structures. Also, the existing software are in use for trans-membrane protein structure prediction but it is an hour need to develop different program to model the trans-membrane segments. Although MD simulations were utilized for validating predicted structures of membrane proteins and/or for getting insights of their mechanism, challenge remains in the forms of FF as at times it is difficult to get the parameters for membrane proteins, lipids in which they were embedded, any bound coordinated metal ions in a single FF. The accuracy of models depends upon pH and dynamic charge environment instead of static electrostatic charges, and polarizable water models, requires further development and testing of polarizable force fields. The existing FF were designed with aid of experimental data for globular proteins and applied for studying IDPs whereas disordered

proteins are having non-structural segments. Thus, it necessitates designing and developing different set of FF parameters for simulating exclusively IDPs. In summary, there is always space for improvement in existing ones and developing new models with higher accuracy in any field of science.

Acknowledgements

Authors would like to acknowledge Department of Biotechnology (DBT) (project number BT/COE/34/SP28408/2018), Govt. of India for providing computational facilities.

Author details

Zaved Hazarika¹, Sanchaita Rajkhowa² and Anupam Nath Jha^{1*}

1 Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur, Assam, India

2 Centre for Biotechnology and Bioinformatics, Dibrugarh University, Dibrugarh, Assam, India

*Address all correspondence to: anjha@tezu.ernet.in

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Nii-Trebi NI. Emerging and Neglected Infectious Diseases: Insights, Advances, and Challenges. *Biomed Res Int.* 2017;2017:5245021.
- [2] Genetics Home Reference Available from: <https://ghr.nlm.nih.gov/condition/huntington-disease#genes>. Accessed on 2020-07-16.
- [3] World Health Organization, "Infections and infectious diseases: a manual for nurses and midwives in the WHO European Region," Available from: http://www.euro.who.int/__data/assets/pdf_file/0013/102316/e79822.pdf. Accessed on 2020-07-16.
- [4] Understanding Emerging and Re-emerging Infectious Diseases. National Institutes of Health (US); National Institutes of Health (US); Biological Sciences Curriculum Study. NIH Curriculum Supplement Series 2007.
- [5] Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology.* 2020;5(4):536-44.
- [6] Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *J Med Virol.* 2020;92(4):401-2.
- [7] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.* 2020;382(8):727-33.
- [8] Deng H, Jia Y, Zhang Y. Protein structure prediction. *Int J Mod Phys B.* 2018;32(18).
- [9] Berman HM, Vallat B, Lawson CL. The data universe of structural biology. *IUCrJ.* 2020;7(4):630-8.
- [10] Berg JM, Tymoczko JL, Stryer L. *Biochemistry*, 7th ed: W.H. Freeman and Company; 2012.
- [11] Smyth MS, Martin JH. x ray crystallography. *Mol Pathol.* 2000;53(1):8-14.
- [12] Marion D. An introduction to biological NMR spectroscopy. *Molecular & cellular proteomics : MCP.* 2013;12(11):3006-25.
- [13] Frank J. Single-particle reconstruction of biological macromolecules in electron microscopy--30 years. *Q Rev Biophys.* 2009;42(3):139-58.
- [14] Milne JL, Borgnia MJ, Bartesaghi A, Tran EE, Earl LA, Schauder DM, et al. Cryo-electron microscopy--a primer for the non-microscopist. *FEBS J.* 2013;280(1):28-45.
- [15] Kumar P, Halder S, Bansal M. Biomolecular Structures: Prediction, Identification and Analyses. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C, editors. *Encyclopedia of Bioinformatics and Computational Biology*. Oxford: Academic Press; 2019. p. 504-34.
- [16] Chevalier A, Silva DA, Rocklin GJ, Hicks DR, Vergara R, Murapa P, et al. Massively parallel de novo protein design for targeted therapeutics. *Nature.* 2017;550(7674):74-9.
- [17] Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications.* 1995;91(1):1-41.

- [18] van Gunsteren WF, Berendsen HJC. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. Angewandte Chemie International Edition in English. 1990;29(9):992-1023.
- [19] Simons KT, Bonneau R, Ruczinski I, Baker D. Ab initio protein structure prediction of CASP III targets using ROSETTA. Proteins. 1999;Suppl 3:171-6.
- [20] Bonneau R, Baker D. Ab initio protein structure prediction: progress and prospects. Annu Rev Biophys Biomol Struct. 2001;30:173-89.
- [21] Jha AN, Vishveshwara S, Banavar JR. Amino acid interaction preferences in proteins. Protein Sci. 2010;19(3):603-16.
- [22] Jha AN, Vishveshwara S, Banavar JR. Amino acid interaction preferences in helical membrane proteins. Protein Eng Des Sel. 2011;24(8):579-88.
- [23] Blundell TL, Sibanda BL, Sternberg MJ, Thornton JM. Knowledge-based prediction of protein structures and the design of novel molecules. Nature. 1987;326(6111):347-52.
- [24] Šali A, Overington JP, Johnson MS, Blundell TL. From comparisons of protein sequences and structures to protein modelling and design. Trends in Biochemical Sciences. 1990;15(6):235-40.
- [25] Sutcliffe MJ, Haneef I, Carney D, Blundell TL. Knowledge based modelling of homologous proteins, part I: three-dimensional frameworks derived from the simultaneous superposition of multiple structures. Protein Engineering, Design and Selection. 1987;1(5):377-84.
- [26] Kc DB. Recent advances in sequence-based protein structure prediction. Briefings in Bioinformatics. 2016;18(6):1021-32.
- [27] Pascarella S, Argos P. A data bank merging related protein structures and sequences. Protein Eng. 1992;5(2):121-37.
- [28] Farokhirad S, Bradley RP, Sarkar A, Shih A, Telesco S, Liu Y, et al. 3.13 Computational Methods Related to Molecular Structure and Reaction Chemistry of Biomaterials☆. In: Ducheyne P, editor. Comprehensive Biomaterials II. Oxford: Elsevier; 2017. p. 245-67.
- [29] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography. 1993;26(2):283-91.
- [30] Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature. 1996;381(6580):272.
- [31] Rost B. PHD: predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol. 1996;266:525-39.
- [32] Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. JPred: a consensus secondary structure prediction server. Bioinformatics. 1998;14(10):892-3.
- [33] Kaur H, Raghava GP. Prediction of alpha-turns in proteins using PSI-BLAST profiles and secondary structure information. Proteins. 2004;55(1):83-90.
- [34] Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. Proteins: Structure, Function, and Bioinformatics. 2019;87(12):1011-20.
- [35] Levitt M, Chothia C. Structural patterns in globular proteins. Nature. 1976;261(5561):552-8.

- [36] Andreeva A, Howorth D, Chothia C, Kulesha E, Murzin AG. SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Res.* 2014;42(Database issue):D310-4.
- [37] Anfinsen CB. Principles that govern the folding of protein chains. *Science.* 1973;181(4096):223-30.
- [38] Lee EH, Hsin J, Sotomayor M, Comellas G, Schulten K. Discovery through the computational microscope. *Structure.* 2009;17(10):1295-306.
- [39] Levitt M, Warshel A. Computer simulation of protein folding. *Nature.* 1975;253(5494):694-8.
- [40] McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature.* 1977;267(5612):585-90.
- [41] Karplus M, Sali A. Theoretical studies of protein folding and unfolding. *Curr Opin Struct Biol.* 1995;5(1):58-73.
- [42] Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science.* 1998;282(5389):740-4.
- [43] Das S, Hazarika Z, Sarmah S, Baruah K, Rohman MA, Paul D, et al. Exploring the interaction of bioactive kaempferol with serum albumin, lysozyme and hemoglobin: A biophysical investigation using multi-spectroscopic, docking and molecular dynamics simulation studies. *Journal of Photochemistry and Photobiology B: Biology.* 2020;205:111825.
- [44] Das S, Sarmah S, Hazarika Z, Rohman MA, Sarkhel P, Jha AN, et al. Targeting the heme protein hemoglobin by (–)-epigallocatechin gallate and the study of polyphenol–protein association using multi-spectroscopic and computational methods. *Physical Chemistry Chemical Physics.* 2020;22(4):2212-28.
- [45] Hazarika Z, Jha AN. Computational Analysis of the Silver Nanoparticle–Human Serum Albumin Complex. *ACS Omega.* 2020;5(1):170-8.
- [46] Nayak PS, Borah SM, Gogoi H, Asthana S, Bhatnagar R, Jha AN, et al. Lactoferrin adsorption onto silver nanoparticle interface: Implications of corona on protein conformation, nanoparticle cytotoxicity and the formulation adjuvanticity. *Chemical Engineering Journal.* 2019;361:470-84.
- [47] Arakha M, Borah SM, Saleem M, Jha AN, Jha S. Interfacial assembly at silver nanoparticle enhances the antibacterial efficacy of nisin. *Free Radical Biology and Medicine.* 2016;101:434-45.
- [48] Saikia N, Jha AN, Deka RC. Interaction of pyrazinamide drug functionalized carbon and boron nitride nanotubes with pncA protein: a molecular dynamics and density functional approach. *Rsc Advances.* 2013;3(35):15102-7.
- [49] Asthana S, Hazarika Z, Nayak PS, Roy J, Jha AN, Mallick B, et al. Insulin adsorption onto zinc oxide nanoparticle mediates conformational rearrangement into amyloid-prone structure with enhanced cytotoxic propensity. *Biochimica et Biophysica Acta (BBA) - General Subjects.* 2019;1863(1):153-66.
- [50] Bora N, Jha AN. An integrative approach using systems biology, mutational analysis with molecular dynamics simulation to challenge the functionality of a target protein. *Chemical Biology & Drug Design.* 2019;93(6):1050-60.
- [51] Rajkhowa S, Jha AN. Molecular Dynamics Simulations: A Tool to Investigate the Interactions between Biomolecules and Nanoparticles. Nova Science Publishers. 2019:65-108.

- [52] Fluitt Aaron M, de Pablo Juan J. An Analysis of Biomolecular Force Fields for Simulations of Polyglutamine in Solution. *Biophysical Journal*. 2015;109(5):1009-18.
- [53] Rubenstein AB, Blacklock K, Nguyen H, Case DA, Khare SD. Systematic Comparison of Amber and Rosetta Energy Functions for Protein Structure Evaluation. *Journal of chemical theory and computation*. 2018;14(11):6015-25.
- [54] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 2010;78(8):1950-8.
- [55] Man VH, He X, Derreumaux P, Ji B, Xie XQ, Nguyen PH, et al. Effects of All-Atom Molecular Mechanics Force Fields on Amyloid Peptide Assembly: The Case of Abeta16-22 Dimer. *Journal of chemical theory and computation*. 2019;15(2):1440-52.
- [56] Kato K, Nakayoshi T, Fukuyoshi S, Kurimoto E, Oda A. Validation of Molecular Dynamics Simulations for Prediction of Three-Dimensional Structures of Small Proteins. *Molecules*. 2017;22(10).
- [57] Borah SM, Jha AN. Identification and analysis of structurally critical fragments in HopS2. *BMC Bioinformatics*. 2019;19(Suppl 13):552.
- [58] Pietrek LM, Stelzl LS, Hummer G. Hierarchical Ensembles of Intrinsically Disordered Proteins at Atomic Resolution in Molecular Dynamics Simulations. *Journal of chemical theory and computation*. 2020;16(1):725-37.
- [59] Mattedi G, Acosta-Gutierrez S, Clark T, Gervasio FL. A combined activation mechanism for the glucagon receptor. *Proc Natl Acad Sci U S A*. 2020;117(27):15414-22.
- [60] Corbeil CR, Williams CI, Labute P. Variability in docking success rates due to dataset preparation. *J Comput Aided Mol Des*. 2012;26(6):775-86.
- [61] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. *J Mol Biol*. 1982;161(2):269-88.
- [62] Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*. 2003;52(4):609-23.
- [63] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*. 1998;19(14):1639-62.
- [64] Böhm H-J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*. 1994;8(3):243-56.
- [65] Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem*. 2006;49(21):6177-96.
- [66] Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*. 1997;11(5):425-45.
- [67] Velec HF, Gohlke H, Klebe G. DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand

poses and better affinity prediction. J Med Chem. 2005;48(20):6296-303.

[68] Mooij WT, Verdonk ML. General and targeted statistical potentials for protein-ligand interactions. Proteins. 2005;61(2):272-87.

[69] Sieg J, Flachsenberg F, Rarey M. In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening. Journal of Chemical Information and Modeling. 2019;59(3):947-61.

[70] Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. Biophysical reviews. 2017;9(2):91-102.

[71] Miller MD, Kearsley SK, Underwood DJ, Sheridan RP. FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. J Comput Aided Mol Des. 1994;8(2):153-74.

[72] Rajkhowa S, Jha AN, Deka RC. Anti-tubercular drug development: computational strategies to identify potential compounds. J Mol Graph Model. 2015;62:56-68.

[73] Kolb P, Rosenbaum DM, Irwin JJ, Fung JJ, Kobilka BK, Shoichet BK. Structure-based discovery of beta2-adrenergic receptor ligands. Proc Natl Acad Sci U S A. 2009;106(16):6843-8.

[74] Rajkhowa S, Borah SM, Jha AN, Deka RC. Design of *Plasmodium falciparum* PI(4)KIII β Inhibitor using Molecular Dynamics and Molecular Docking Methods. ChemistrySelect. 2017;2(5):1783-92.

[75] Choudhary SA, Bora N, Banerjee D, Arora L, Das AS, Yadav R, et al. A novel small molecule A2A adenosine receptor agonist, indirubin-3'-monoxime, alleviates lipid-induced inflammation and insulin resistance in 3T3-L1 adipocytes. Biochem J. 2019;476(16):2371-91.