Chapter 3
Materials and Methods

Materials and Methods:

3.1. Tools and techniques:

Numerous computational methods and approaches have been employed to examine key structural and dynamic aspects of the α-Syn protein, including its dimerization and oligomerization, fibril characteristics, polymorphism of its amyloid fibrils, and suppression of α-Syn aggregation. MD simulation has been our primary tool of choice for handling these analyses. The following diagram illustrates the fundamental idea of the MD simulation technique:

3.1.1. Molecular Dynamics (MD) Simulation:

MD has been widely used to study the structure and dynamics of macromolecules, such as proteins or nucleic acids. Due to revolutionary advances in computer technology and algorithmic improvements, MD has subsequently risen to prominence in many fields of physics and chemistry. Two primary families of MD approaches may be identified based on the mathematical formulation employed to represent a particular physical system. Molecules are represented as classical objects in the "classical" mechanics method used in MD simulations, corresponding to the "ball and stick" paradigm [184-186]. Information on several biological issues is provided by means of quantum MD simulations, representing a significant advancement over the classical method [187]. Experiments like NMR and X-ray diffraction can be used to determine the structure and understand the function of large biological molecules [188]. However, experiments can only be carried out in combination with theories and models. The relationship between theory and experiment has changed as a result of computer simulations. Using a computer to simulate a physical system is the core of simulation. The machine performs calculations suggested by a mathematical model and interprets the findings in terms of physical attributes. A theoretical technique might describe computer simulation since it works with models. One method of statistical mechanics is MD. Physical quantities are represented by averages across microscopic states (configurations) of the system, dispersed in accordance with a certain statistical ensemble, according to statistical mechanics [189]. Energy conservation is implied by Newtonian dynamics, and MD trajectories offer a range of configurations dispersed by the micro-canonical ensemble [190]. Consequently, an arithmetic average for the present value of a physical parameter derived from the trajectories may be used to measure it using MD simulation.

3.1.1.1. History of simulation:

In the past decade that followed, computers have increased in speed and power and become more accessible to researchers worldwide. The first simulation findings on a realistic model system, a liquid argon MD simulation, were first reported in a published work by Rahman in 1964 [191]. Also, in another study reported by Alder and Wainright, earlier MD simulations were used to simulate systems made out of hard spheres, a discontinuous potential that was more generic than realistic [192]. The revolutionary applications of MD to liquid dynamics by Rahman, Wainwright, and Alder in the late 1950s and early 1960s led to the development of MD as one of the first simulation methods. In the 1950s, computer simulations of condensed matter systems were first used. Then, the MD approach and the Monte Carlo (MC) sampling methodology [193] were introduced, considered to be the two main domains of MD simulations. Another noteworthy work by Rahman and Stillinger [194], who first published an MD research on a model of liquid water in 1971, a system made up of molecules, not simply atoms, was a significant advancement in this regard. This work was a major accomplishment because it showed that the structure of liquid water was made up of a random network of hydrogen bonds and did not resemble the solid phases in any way. It also showed that the diffusion of water molecules occurred continuously rather than through a hopping mechanism, as was previously widely believed. Although the first MD simulation of a simple protein was done by Karplus and collaborators in 1977 [195], by integrating multiple time scales, it was possible to enhance the maximum time step in biomolecular simulations to around 5-10 fs while maintaining the complete physics inherent in contemporary atomistic force fields [196]. The study of Ceperley and Alder to handle systems like the electron gas [197] and hydrogen at high pressures [198] became computationally possible in the late 1970s and early 1980s. In a 1985 publication in Physical Review Letters, Car and Parrinello [199] proposed a method to integrate MD with density functional theory (DFT)-based direct electronic structure computation. In the 1980s and 1990s, condensed matter MD simulation research gained popularity, partly because of previous achievements but also because of the increasing power and accessibility of computers [200].

3.1.1.2. Theory of MD simulation:

The MD simulation method is based on the equation of motion. The equation of motion, or Newton's second law of motion, is expressed as F=ma, where 'm' is the particle's mass, 'a' is the acceleration, and 'F' is the force acting on it. Nonetheless, one should be able to determine the acceleration for each atom in the system if the force acting on each atom is known. The

spatial location of an atom as a function of time could be predicted using Newton's equations of motion. The generated trajectory essentially represents the system's atomic-level configuration at each instant of the simulated time period as a three-dimensional motion. Specifically, a method of updating the location and velocity of each atom is to iteratively calculate the forces acting on each atom as it advances through time. In the case of MD modeling approaches, time and resource contributions may be higher. However, with advancements, the cost and speed of computers are decreasing. Millisecond-level temporal computations are made in the simulations for solvated proteins. However, millisecond simulations have also been documented in other research works.

The equation of motion based on Newton's law can be stated as:-

$$F_i = m_i a_i \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.1)$$

$$\vec{F} = ma \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.2)$$

$$F = -\frac{d}{dr}\mu \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.3)$$

In this equation, F_i defines the force exerted on particle **i**, m_i as the mass of particle m_i and a_i as the acceleration of particle **i**, derived from the potential energy **μ**.

The Newton's force, F_i can also be expressed as the potential energy gradient V, ∇_i is del or nabla symbol is a vector differential operator:

$$F_i = -\nabla_i V \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.4)$$

By combining the two equations given,

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.5)$$

The system's potential energy, or **V**, is defined here. There is a connection between Newton's equation of motion and the potential energy derivative that is utilized to explain position changes in time.

The main objective of the numerical integration for Newton's equation of motion is to find a formula that may define position r_i at the time **t+Δt** in relation to the locations that are known at time t. However, the Velocity Verlet technique employs the positions and accelerations at time t as well as the positions from time t-Δt to get the new locations at time **t+Δt**. The velocities in this procedure are not produced directly. Although their knowledge is not required for evolution, it is required to calculate the kinetic energy K. Throughout the route, the entire

energy E=K+U should be preserve. $\boldsymbol{a_i}$ as the acceleration of particle **i**. The position of atom **r$_i$** in every Δt time step: -

$$r_i(t + \Delta t) = \frac{r_i(t) + v_i(t)\Delta t \ + \ 1 \ a_i(t)\Delta t)^2}{t} \quad \dots \dots \dots \dots \dots \dots . (3.6)$$

The velocity $\boldsymbol{v_i}$ is used as a half time step: -

$$r_i(t + \Delta t) = r_i\left(t - \frac{\Delta t}{2}\right) + r_i(t)\frac{\Delta t}{2} \dots \dots \dots \dots \dots \dots \dots \dots . (3.7)$$

The velocities can be computed from the Δt time step: -

$$r_i(t) = \frac{r_i(t + \frac{\Delta t}{2})^2 + r_i(t - \frac{\Delta t}{2})^2}{2} \quad \dots \dots \dots \dots \dots \dots \dots . (3.8)$$

It is significant that when kinetic energy is required at time t, then the velocity rescaling is required. Also, the required atomic positions are obtained from:

$$r_i(t + \Delta t) = r_i(t) + r_i(t + \Delta t)\frac{\Delta t}{2} \dots \dots \dots \dots \dots \dots \dots . (3.9)$$

Force fields are used to illustrate how bond lengths, bond angles, torsion, non-bonding van der Waals, and electrostatic interactions between atoms change over time. This force field, which aims to replicate the molecular geometry and some of the unique properties of the tested structures, is composed of a set of linked constants and equations.

3.1.1.3. Force field:

A mathematical equation that are used to describe the dependency of the energy system on the locations of the constituent particles is called a force field. The analytical interatomic potential energy form U (r1, r2,... rN) and other contributing factors. The parameters are usually acquired by fitting to experimental data from many sources, including neutron, X-ray, and electron diffraction, NMR, infrared, Raman, and neutron spectroscopy, etc., or from ab initio or semi-empirical quantum mechanical calculations. A collection of atoms bound together by basic elastic (harmonic) forces is the simplest definition of a molecule, and the FF substitutes a simplified model that is valid in the simulation zone for the genuine potential.
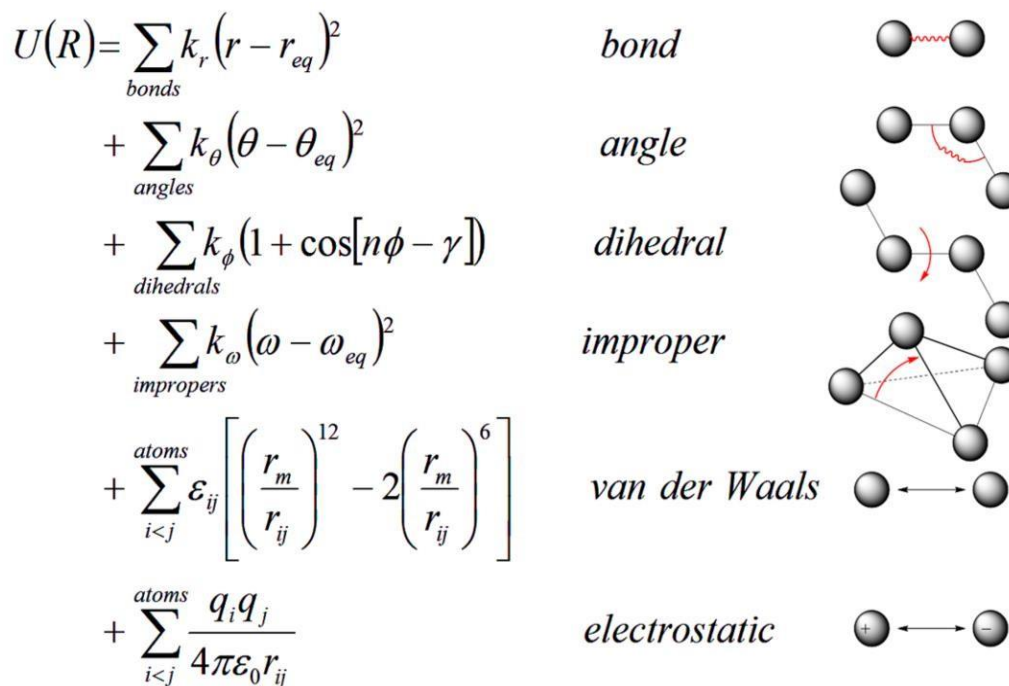
$$V(rN) = \sum bonds \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum angles \frac{k_i}{2} (\Theta_i - \Theta_{i,0})^2$$

$$+ \sum torsions \frac{V_n}{2} (1 + \cos(n\emptyset - \emptyset_o))$$

$$+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} 4\varepsilon_{ij} [\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6]$$

$$+ \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots . \ldots \quad (3.10)$$

Here,

V (rN): potential energy as a function of the positions (r) of N atoms;

k_i: force constant;

l_i ,l_0: current and reference bond lengths;

θ_i, θ_0: current and reference valence angle;

Vn: barrier height of rotation;

$\emptyset$: torsion angle;

n : the number of energy minima throughout a complete rotation is determined by the multiplicity.;

σ_{ij}: collision diameter for i and j, the two atoms interacting;

q_i, q_j: partial atomic charges on the atoms i and j;

ε_{ij}: well depth of the Lennard-Jones potential for the i-j interaction;

r_{ij}: existing distance of atoms i and j;

ε_0, ε_r: relative permittivity of the surroundings and the vacuum's permittivity, respectively;

$\emptyset_0$: phase factor that establishes the energy minimum of the torsion angle at each point.

The bulk of the potential energy function is composed of bond lengths, angles, and rotations as well as non-bonded interactions including van der Waals and electrostatic interactions. A graphical illustration of the many forms of interactions may be seen in **Figure 3.1**.

$$U(R) = \sum_{bonds} k_r (r - r_{eq})^2 \qquad \text{bond}$$

$$+ \sum_{angles} k_\theta (\theta - \theta_{eq})^2 \qquad \text{angle}$$

$$+ \sum_{dihedrals} k_\phi (1 + \cos[n\phi - \gamma]) \qquad \text{dihedral}$$

$$+ \sum_{impropers} k_\omega (\omega - \omega_{eq})^2 \qquad \text{improper}$$

$$+ \sum_{i<j}^{atoms} \varepsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^6 \right] \qquad \text{van der Waals}$$

$$+ \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \text{electrostatic}$$

Figure 3.1. *Diagrammatic illustration of the main contribution to the potential energy function (Taken from [201])*

The bond stretching, angle bending, dihedral and incorrect torsions, and other local or intramolecular contributions to the total energy are represented by the first four components in the equation. The last two elements of the equation, in this example using a 12-6 Lennard-Jones potential, represent the repellent van der Waals interactions and the Coulombic contacts.

3.1.1.4. Long range interaction: Ewald sum:

In computer simulations of condensed-matter systems, the Ewald Summation is widely used for predicting electrostatic interactions [202]. In constructing the Ewald Sum, the inaccuracies arising from truncating the infinite real and Fourier-space lattice sums are examined. An ideal option with a screening parameter of 7 is retrieved for the Fourier-space cut-off. Generally speaking, one observes that the quantity of Fourier space vectors needed to get a certain degree of accuracy scales with 7/3. However, by controlling the effective computing parameters for Ewald sums, this suggested approach may be used to assess the caliber of Ewald-sum implementations and contrast different implementations. This approach may be used the most frequently in MD simulations to assess long-range interactions. Analyzing a charge distribution for each site's opposite sign is the core idea behind the Ewald sum. The extra charge distribution shows the interactions between the surrounding atoms. The cut-off scheme can manage the interactions, despite their limited range. To make up for the excess charge distribution, the identical charge distribution with the opposite sign and short-range interaction is made up in

the reciprocal space. It is easy to get the input for the electrostatic potential at a given location r_i generated by a collection of screened charges because the electrostatic potential induced by the screened charge is a rapidly diminishing function of r. The following formula provides the total potential energy for the long-range Coulomb ($\mathbf{\mu}_c$) interaction:

$$\mu_c = \mu_q(\alpha) - \mu_{self}(\alpha) + \Delta\mu(\alpha) \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots. (3.11)$$

Sharper distributions are produced by larger values of '**α**' in the equation; to increase accuracy for such large numbers, K summations are added. A larger value for '**α**', on the other hand, reduces the filtered potential range ($\mathbf{\mu}_q$) and enables us to use a lower cut-off radius. $\mathbf{\mu}_{self}$ denotes self-correction subtracts for each of the charge independent of configuration. Each terms ($\mathbf{\mu}_q$, $\mathbf{\mu}_{self}$, and $\Delta\mathbf{\mu}$) depends on **α** .Consequently, there is room for optimization of the value of α between the two factors in order to maximize both efficiency and accuracy. The Ewald summation is represented by N2 solely in the aforementioned scales. However, by choosing the right values for α and K for the k-space summation cut-off, Finchman was able to optimize the summing that scales as N (3/2). Additionally, by evaluating the reciprocal summation with the Fast Fourier Transform (FFT), this Ewald summation method can be enhanced. On the other hand, the particle mesh-based solution relies on the usage of an estimated reciprocal space sum based on FFT that grows as N log (N) and a set cut-off for the direct space sum. The infinite-range Coulomb interaction is efficiently calculated by this method under periodic boundary conditions (PBC). Furthermore, there is a version called Particle Mesh Ewald (PME) that accelerates the Ewald reciprocal sum to almost linear scaling by using the three-dimensional fast Fourier transform (3DFFT). Under PBC, particle 'i' inside the unit cell interacts electrostatically with every other particle j inside the cell, with every periodic image of '**j**', and with each of its own periodic pictures because of the infinite range of the Coulombic interaction. The total Coulomb energy of a system '**U**' consisting of N particles in a size L cubic box and all of their infinite duplicates in PBC is given by:

$$U = \frac{1}{2} \sum_{n}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{r_{ij,n}} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.. (3.12)$$

Where $r_{ij,n}$ is partial atomic charges on the atoms **i** ,**j** and **n** and $q_i q_j$ is partial atomic charges on the atoms **i** and **j**.

Ewald reformulated a single slowly and conditionally converging series, as the product of two rapidly convergent series plus a constant term,

$$U_{Ewald} = U^r + U^m + U^0 \quad .... (3.13)$$

Therefore, the total of these three elements—the real (direct) space sum (U^r), the reciprocal (imaginary, or Fourier) sum (U^m), and the constant term (U°), also known as the self-term— denotes the Ewald sum.

3.1.1.5. SHAKE algorithm: Dealing with molecules:

The choice of time step is limited by the many time scales related to vibrational degrees of freedom, such as bond vibration, angle stretching, or torsional modes inside a molecular system. The integration time step is limited to 1 fs due to the bonds involving hydrogen atoms having a faster vibrational state. To necessitate a longer time period, however, one can restrain these rapid degrees of freedom while addressing the unconstrained degrees of freedom. Since hydrogen bonds have the greatest frequency, the SHAKE technique, created by Ryckaert et al., can restrict dynamics for these types of bonds [203]. Relaying the unconstrained equations of motion of the atomic system is the first step in the SHAKE algorithm. The fundamental idea behind the SHAKE method is to use the Lagrange multiplier formalism to enforce bonding distances to remain constant. Given **Nc,** the constraint $\propto_k$ is provided by:

$$\propto_k = r^2_{k1\,k2} - R^2_{k1\,k2} = 0, where\ k = 1,2,3 ...... N_C ...... (3.14)$$

The distance between the atoms of k_1 and k_2 is thought to be limited by the parameters r_{k1k2} and R_{k1k2}. The following defines the modified constrained equation of motion:

$$m_i \frac{d^2 r_i(t)}{dt^2} = -\frac{\partial}{\partial r_i}[V\,(r_1....r_N) + \sum_{k=1}^{Nc} \tau k\,(t)\,\alpha_k(r_1 ...... r_N)] ....(3.15)$$

In this case, τk represents the unknown Lagrange multiplier for the kth constraint, and m_i is the mass of the ith particle, r_i is the distance between the atom i and r_N is the distance between the N atoms. **Nc** quadratic coupled equations can be used to solve this modified restricted equation of motion for an unknown multiplier. The distance of the atom k_1 by the parameter **r**. Ultimately, the motion that follows has been determined.

$$r_{k1}(t + \Delta t) = r_{k1}{}^{uc}(t + \Delta t) - 2(\Delta t)^2 \, m_{k1}{}^{-1} \, \tau k\,(t)r_{k1k2}(t) .... (3.16)$$

The position updates with unconstrained force solely are represented by r^{uc} and Δt is the step time of the time interval **t**. But this method is repeated if the specified tolerance is not provided.

Instead of explicitly inverting the matrix, the SHAKE algorithm approach modifies the particle coordinates periodically until the system satisfies all requirements within a certain tolerance. Constraint decay, or the increase in departure from the ideal lengths caused by the accumulation of numerical errors, is a factor that constraint algorithms must take into consideration in addition to maintaining the rigid bonds. But since the convergence of each time step must happen Iterative algorithms enable accurate constraint decay automatically, within a certain tolerance. The constrained distance deviations from the initial values undergo frequent inspections and fixes. Non-iterative algorithms required a deliberate strategy to counter constraint deterioration since they lacked a natural feedback mechanism for detecting changes in distance.

3.1.1.6. Periodic Boundary conditions:

To explain the periodic boundary conditions, we will construct a system consisting of N interacting particles in a volume V and at a temperature T. The system must be tied by copies of itself in order for us to guarantee that periodic limits on boundaries that are similar to the 2D Using system. Consequently, it can be observed that, given a system of particles, a particle must rejoin the central box on the opposite side that it departs. The atoms of the molecules are organized in a hypothetical box bounded by translated copies of their coordinates. Particle 1 inside the middle box may potentially interact with many copies of particle 3 that are present there, as shown in **Figure 3.2**. Furthermore, considering a particular interaction between particles 1 and 3 is suitable, and selecting the interaction that results in the smallest interatomic distance makes sense. This method is referred to as the closest image convention. There is evidence that a periodic 3-dimensional array encircles the inner cell. An atom gets replaced when it passes through a barrier and enters the opposite side at the same speed. The particles in the core box have a fixed volume after this. However, in order to handle non-bonded interactions, a non-bonded cut-off is typically used, allowing each atom in the system to interact with just one image of each and every other atom in the system.
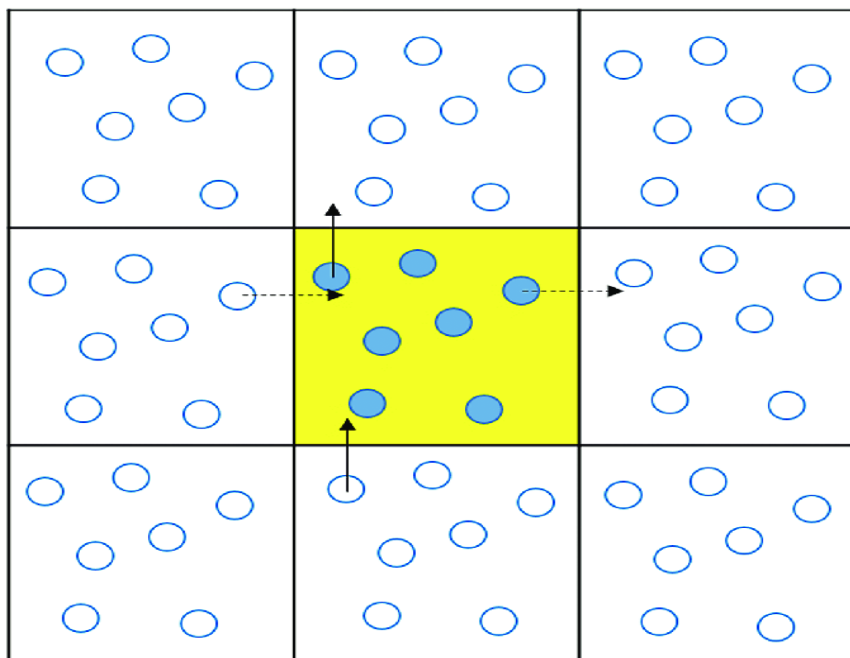
Figure 3.2. *The two dimensional projection of Periodic boundary conditions. The simulation cell (dark color) is surrounded by translated copies of itself (light color) (Taken from [204])*

3.1.1.7. Temperature and Pressure computation control:

Multiple strategies are being investigated at the moment to achieve the isothermal MD simulations. Adding a fictitious heat bath to the system in order to maintain the average temperature **T** at a certain target temperature **T₀** is theoretically equivalent to utilizing a thermostat. However, the heat bath still matters for the m_i denotes mass of the particular particle i, F_i defines the force exerted on particle i, r_i is the distance between the atom i since it might alter the particle's velocity or modify Newton's equation of motion:

$$m_i \, \frac{\mathrm{d}^2 r_i}{\mathrm{dt}^2} = F_i(r_i) \; \dots \dots \dots \dots \dots .. \, (3.17)$$

The system usually evolves according to the above-described, which has a micro-canonical (NVE) energy distribution, in the absence of temperature control. This micro-canonical ensemble provides the value of "real" dynamics, such as classical Newtonian dynamics, for a system described by an individual force field, at the precision level limited by the integration method and the force calculations. By attaching the system to a Berendsen thermal bath, the starting temperature of the system is determined [205]. The bath serves as a source of thermal energy by adding or removing heat from the system as needed. The following equation is used to adjust the system temperature **T(t)** that differs from the bath temperature **T₀**:

$$\frac{dT(t)}{dt} = \frac{1}{\tau} \, \{T_0 - T(t)\} \dots \dots \dots \dots \dots \dots \dots \dots \, (3.18)$$

where the intensity of the coupling between the bath and the system is determined by the time constant, or τ. The system's temperature is adjusted by dividing each step's atom velocities by a fraction χ, which is given by:

$$\chi = \left[1 + \frac{\Delta t}{\tau T} \left[\frac{T_0}{T(t)} - 1 \right] \right] \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.19)$$

A change in the time constant τ can alter the coupling's strength. The temperature control method and the pressure control approach are comparable. The system may be connected to a barostat, and by periodically scaling the atomic locations and simulation cell size by μ, the pressure can be kept constant:

$$\mu = 1 - \omega \frac{\Delta t}{\tau p} (P - P_0) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.20)$$

Here, ω is the isothermal compressibility, P_0 is the barostat pressure, P is the momentary pressure at time t, Δt is the step time, and τp is the relaxation constant. The AMBER 18 standard simulation software is used. The MD are carried out using the PMEMD one module of AMBER [205].

3.1.1.8. Water molecule models:

Three-site water models were followed by the development of four-site water models. The Bernal and Fowler model is the first of the four-site models. It was created in 1933 and is currently only sometimes used, but it is significant historically [206]. As on the negative charge is moved from the oxygen and towards the hydrogens in that particular model at the bisector of the HOH angle, which is 0.15 Å from the oxygen atom. In either case, the centre of the Lennard Jones interaction site contains the oxygen atom. Ten distances, as opposed to nine, are needed to evaluate the interaction function for a three-site model. Numerous levels of approximation (such as quantum vs classical, flexible versus rigid) and the intricacy of water characteristics have resulted in the suggestion of hundreds of theoretical and computational models for water [207]. Among the classical water models, the rigid non-polarizable models that represent water as a collection of point charges at fixed locations relative to the oxygen nucleus are the most straightforward and computationally efficient. These models are the class that is employed in the great majority of biomolecular investigations conducted today. The most often used models in this class, such as the triple point charge (TIP3P) [208] and SPCE [209] 3-point models, the four point charge (TIP4Pew) [210] 4-point model, and the TIP5P [211] 5-point model, have generally succeeded in striking a fair balance between speed and

accuracy, albeit they are far from ideal. The basic TIP4P water model is re-parameterized for use with Ewald methods, offering a worldwide increase in water characteristics overall when compared to a number of widely-used polarizable and non-polarizable water potentials. The new TIP4Pew potential has a density maximum at approximately 1°C and reproduces experimental bulk densities and the enthalpy of vaporization with an absolute average error of less than 1%, from -37.5 to 127 °C at 1 atm, using high precision simulations and careful application of standard analytical corrections [211] **(Figure 3.3)**.

(a)

(b)
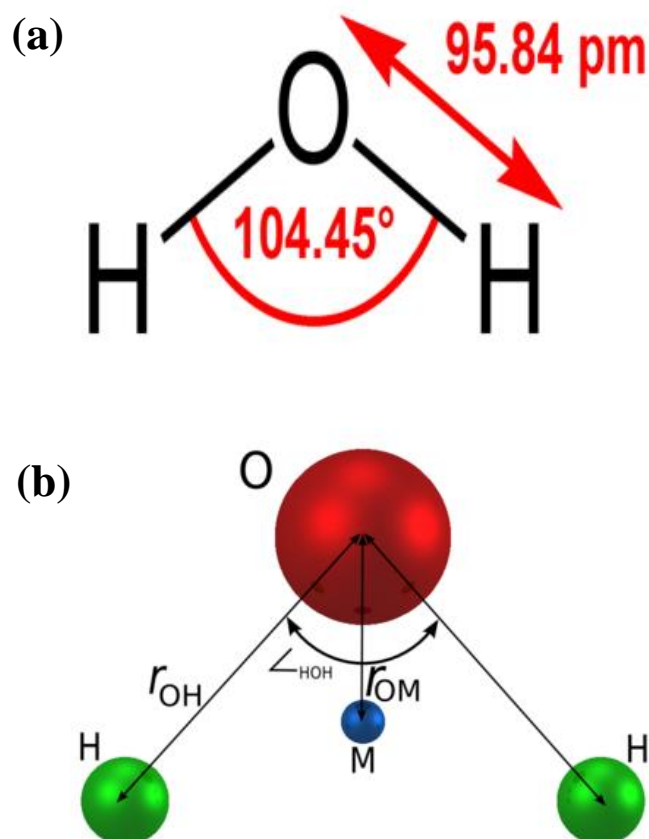
Figure 3.3. Schematic representation of TIP3P water model (Taken from [211])

It takes 10 distances instead of nine to evaluate the interaction function for a three-site model. As an example, the ST2 model requires 17 participants in a five-site strategy. The TIPS2 and TIP4P potentials were later developed by Jorgensen and associates using just the BF water model as a basis [212]. A negative charge is further provided to a site that is located on the bisector of the HOH angle, in a manner similar to the TIP4P water model. Both the hydrogen sites and the oxygen are impacted by the LJ encounter unlike the other two water models. In order to accurately depict the ice and liquid water surrounding the melting point, a six-site water model was created. Ice and water close to the melting point were properly modelled in

terms of their structural and thermodynamic properties.

3.1.1.9. Molecular Dynamics steps:

To establish the molecular system involved in the four phases:

1. Energy Minimization

2. Heating Dynamics

3. Equilibration Dynamics

4. Production Dynamics
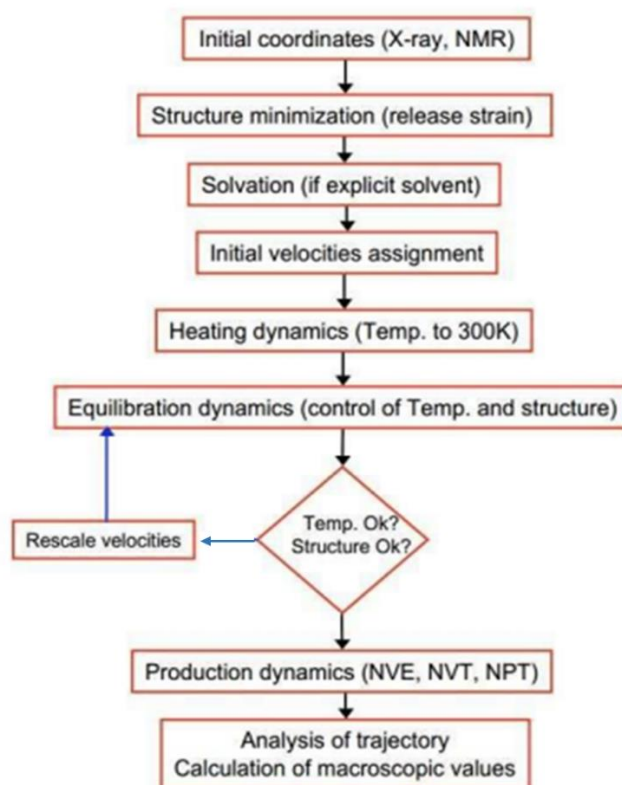
Figure 3.4. *Schematic representation of the steps involved in MD simulation (Taken from [201])*

1. Energy Minimization

Energy minimization is used to replicate the data from MC or MD simulations. Despite being crucial for determining entropy and thermodynamic averages, MC structures and dynamical ensembles are too numerous to thoroughly examine at the microscopic level. The reduced structures are an important and helpful place to start for structural study, even though they depict the underlying configurations related to the fluctuations that happen during dynamics [213, 214]. Locating a stable point or minimum on the potential energy surface using the force

field assigned to the system's atoms is required to start dynamics. The net force acting on each atom disappears on the surface of least potential energy. Limitations can be applied in the dynamics as well as minimization processes. These restrictions can be imposed by a template to force a ligand to find the structure that is structurally closest to a target molecule, or they can be obtained from data, such as NOEs from an NMR experiment. A function (provided by the force field) and an initial estimate or set of coordinates are needed for minimization. The direction and size of a step (i.e., change in coordinates) required to approach a minimal configuration may be determined using the magnitude of the first derivative. Convergence can be rigorously described by its magnitude in addition to the size of the first derivative. There are two steps involved in the energy reduction of a molecule structure. The first stage is to create and evaluate an equation for a given conformation that describes the energy of the system as a function of its coordinates.
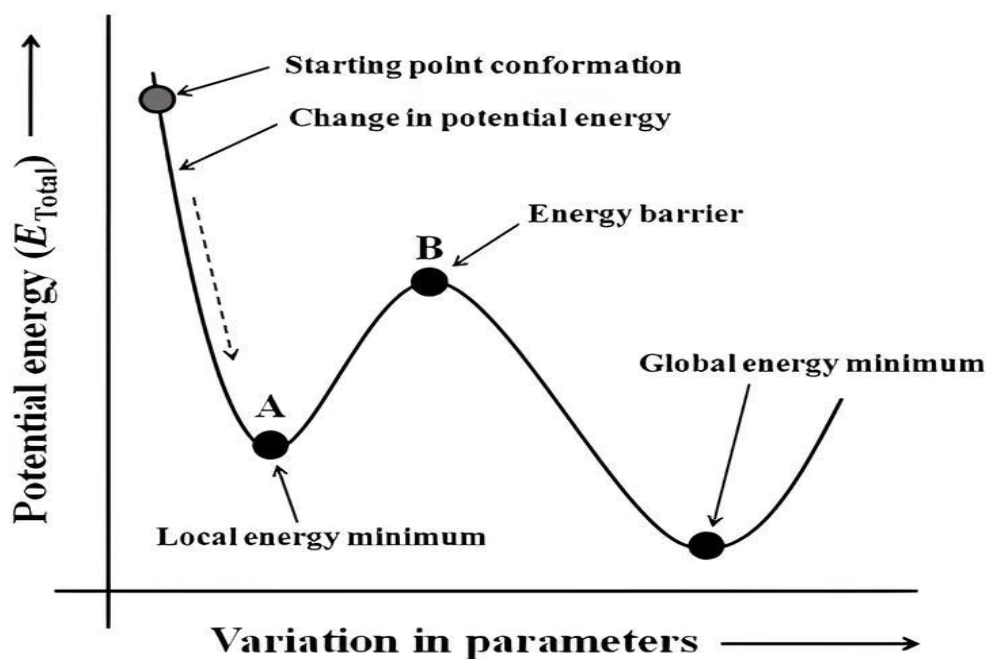
Figure 3.5. A schematic representation of the Different phases of a molecule during minimization of its energy (Taken from [215])

Figure 3.4 illustrates the several stages that an energy-saving operation might go through. Our goal is to identify the bioactive conformer since we are curious about the behavior of bioactive compounds. Research has indicated that the bioactive conformer may not be the same as the most active conformer, despite the latter appearing to be physiologically potent. The bioactive conformer, however, continues to be in a zone in proximity to the most active conformer. The conformation of a ligand attached to a receptor pocket, as established by experimental studies such as X-ray crystallographic analysis, is generally regarded the bioactive molecular

arrangement or conformation if the cocrystal geometry of the molecule is present. The bioactive conformer can be regarded as the most stable conformer when there is no cocrystal geometric structure.

But the following formulation of the energy minimization problem might be used. It must be demonstrated that, given a function '**f** ' and one or more independent variables (x_1, x_2,..... x_i), the values of each independent variable can be found by taking the minimum value of **f**. For any variable, the first derivative of the function at its lowest point is 0, while its second derivatives are positive:

$$\frac{\partial f}{\partial x_i} = 0;\ \frac{\partial^2 f}{\partial^2 x_i} > 0 \ \text{... ... ... (3.21)}$$

The direction for the energy of the initial derivative determines the position of the minimum, while the gradient's magnitude informs the local slope is steep. By allowing each atom to move in response to the force applied to it, it is feasible to lower the system's energy when the force equals minus the gradient. Together with information that may be used to predict when the function will change direction (by passing through a minimum or another stationary point), the second derivatives provide information about the curvature of the function. The two techniques most commonly used in molecular modeling for the first-order minimization processes are the steepest descents and conjugate gradient approaches.

When the derivatives are near zero, minimum energy converges. It is essential to carry out energy reduction on the structure before starting an MD simulation to get rid of bad connections that might otherwise cause structural deformation. The three main minimization techniques are Newton-Raphson, conjugate gradient, and steepest descent.

(i) **The Steepest Descents Method**:

This approach determines which path leads to the minimum by taking the first derivative. It travels in a direction parallel to the net force. This direction is represented as a 3N-dimensional unit vector with 3N Cartesian coordinates, namely after determining the direction of travel, the next step is to determine the distance to be travelled along the gradient. In **Figure 3.5**, the two-dimensional energy surface [215]. The gradient's direction from the beginning is along the line. In the study [216], it started from the point $x_0 = (-9,5)$, where all three methods converge towards the minimum at (0,0) as depicted in **Figure 3.6**. Three techniques are employed on the quadratic function $(x) = xT(2\ 11\ 6)x$. Plots depict the progression of these methods on the contour plot of f. It's noteworthy that the steepest descent method fails to attain the solution

within a finite number of steps. The function will pass through a minimum and then rise if we visualise slicing through the surface along the line [217] as shown in **Figure 3.7**.
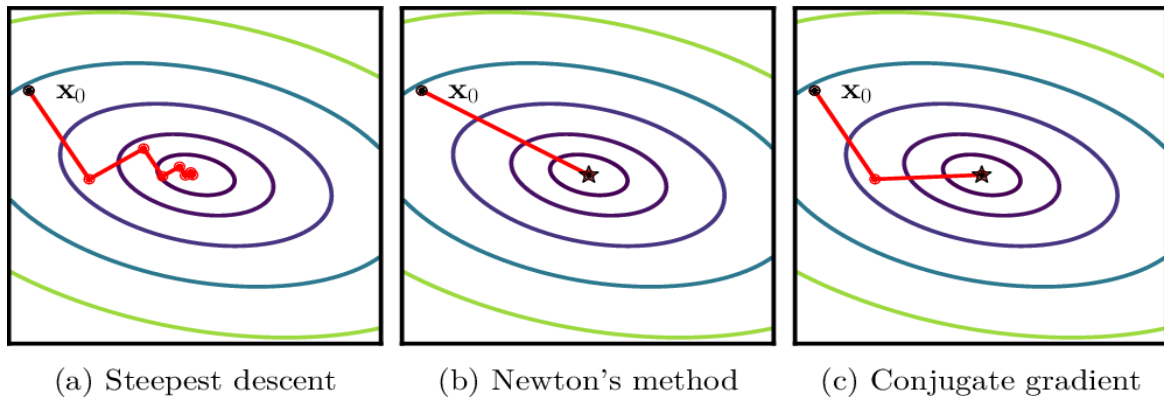

(a) Steepest descent   (b) Newton's method   (c) Conjugate gradient

Figure 3.6. The three different approaches for non-linear gradient method. Plots on the contour plot of demonstrate different iterations of the techniques (Taken from [216])
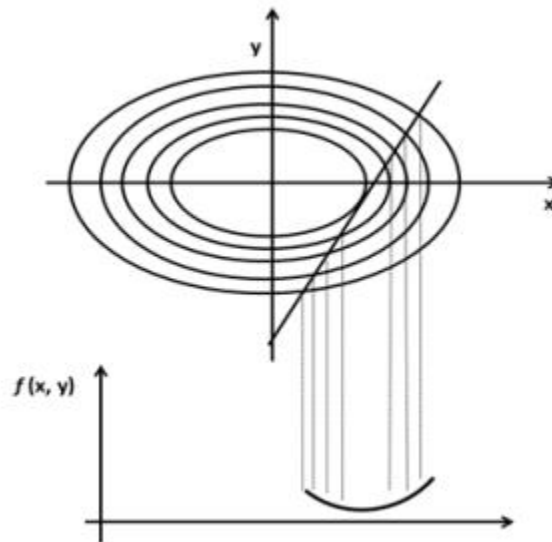

Figure 3.7. A line search is used to locate the minimum in the function in the direction of the gradient [217]

(ii) Minimization of Conjugate Gradients:

In the energy minimization methodology, the conjugate gradient approach yields a set of directions that do not display the steepest descents or oscillate behaviour in the narrow valleys. In the conjugate gradient approach, the gradients are orthogonal at each point even if the directions are conjugate. Given a quadratic function with M variables, the minimum will be found in M steps from a collection of conjugate directions. Beginning at point $\mathbf{g}k$, the conjugate gradient approach proceeds in the direction $\mathbf{v}k$. The gradient at the point and the preceding direction vector $\mathbf{v}k{-}\mathbf{1}$ are used to determine $\mathbf{v}k$ [218].

$$vk = \text{gk} + \gamma kvk - 1 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots..(3.22)$$

The scalar constant γk in the above equation is provided by:

$$\gamma k = \frac{\text{gk}.\text{gk}}{\text{gk}-1.\text{gk}-1} \dots\dots\dots\dots\dots..(3.23)$$

(iii) The Newton-Raphson method:

In the Newton-Raphson method, both the first and second derivatives are used. The curvature is utilized not just to use gradient information, but also to predict where the direction will change along the function's gradient. In order to achieve energy reduction, this technique requires the greatest amount of computing power. If more water molecules are required to increase the system's solubility prior to minimization, they are introduced. An expansive container of water that has been preheated to the identical temperature is utilized for solvation. All of the system is covered by the water box, which removes any water molecules that come into contact with proteins. The mathematical model of Newton-Raphson equation is as follows:

$$r_{min} = r_0 - A_0{}^{-1.\nabla V(r_0)} \dots\dots\dots\dots\dots.(3.24)$$

where $\mathbf{A_o}$ is the matrix of second partial derivatives of the energy with respect to the coordinates at $\mathbf{r_o}$ (also known as the Hessian matrix), r_{min} is the anticipated minimum, r_o is an arbitrary starting point, and $\nabla V(r_o)$ is the gradient of the potential energy at $\mathbf{r_o}$.

2. Heating Dynamics

During the heating phase, when each atom in the system is allocated a beginning velocity (at 0 K), the Newton's equations of motion, which show the temporal growth of the system, are numerically integrated. Following the assignment of new velocities at short, predefined intervals that correspond to marginally higher temperatures, the simulation is set to continue until the goal temperature of 300 K is attained. Heating causes structural stresses to relax, which in turn releases force limitations on different subdomains of the simulation system. Constant volume (NVT) is the typical working condition for thermal dynamics.

3. Equilibration Dynamics

During the equilibrium phase, a system reaches equilibrium as it changes from its starting state. Equilibration should continue continuously, or at least until the values of the set of monitored attributes are settled. Together with structural features, the main measured attributes are

thermodynamic variables like energy, temperature, and pressure. Nonetheless, the initial structure of the liquid state simulations is similar to a solid lattice. In actuality, it is imperative to arrange the components so that the lattice has "melted" prior to the start of the manufacturing process. The reaching point of the liquid state can be ascertained by using the order parameters. This order parameter refers to the assessment of a system's degree of order. The atoms might, however, remain mostly in the same location throughout, preserving a high degree of order while mimicking a crystal lattice. Translational dysfunction may result from the species' propensity for frequent movement inside a liquid. Solving the equations of motion for an atomic system is known as MD. The equation of molecular motion can be solved to determine its trajectory and the temporal development of its motions. MD allows for the bridging of barriers and the examination of several alternatives, depending on the temperature at which the simulation is run. First, velocities need to be assigned in order to start the MD. The Maxwell-Boltzmann distribution limitation for the random number generator is used to achieve this. The average kinetic energy of the system establishes the temperature, according to the kinetic theory of gases. The internal energy of the system is measured in $U = 3/2\ NkT$.

$U = 1/2\ Nmv2$ gives the kinetic energy of the system. However, the temperature may be determined by taking the average of all the atoms' velocities in the system. Throughout the simulation, the Maxwell-Boltzmann distribution may be maintained after the starting set of velocities is determined. The temperature can be considered to be strictly zero Kelvin after energy reduction. Before the system can be heated to the required temperature, the dynamics must be initialized. In order to assign velocities at low temperatures, dynamics is performed in compliance with the equations of motion. The temperature is raised after a number of dynamics rounds, though. A 20 ps timeline of progressive heating dynamics is conducted from 0 to 300 K under atomic constraints. Velocity scaling is the most often used approach to temperature scaling. A run of at least 5 ps (5000 time steps) and frequently 10 or 20 ps is necessary for equilibration time steps of 1 fs. After heating, dynamic equilibration lasts for 100 ps.

4. Production Dynamics:

For the dynamics of the time period of interest is mostly employed to compute thermodynamic averages or sample new configurations. During this step, calculations are being performed about the thermodynamic properties and further data. The simulated system determines most of the parameters, including the kinetic, potential, and total energy, velocities, temperature, and pressure, which are used to determine whether or not equilibrium has been attained. However, in a simulation of the micro canonical ensemble, the total energy remains constant, even though

the kinetic and potential energies could change. Each of the three directions x, y, and z should possess an equivalent quantity of kinetic energy, and the velocity components need to fall inside the Maxwell-Boltzmann distribution. The system's variable during the production phase is the temperature. The system is left to evolve when all counts are reset to zero at the beginning of the manufacturing phase. The temperature of the system is now estimated because no velocity scaling is carried out while creating a micro canonical ensemble. The characteristics are precisely calculated and retained for additional processing and analysis throughout the manufacturing phase. If problems occur, it could be required to restart the simulation if it's being closely observed based on its behavior. Additionally, it is standard procedure to save configurations' energies, positions, and velocities throughout time in order to retrieve the other properties when the simulation is finished. Nevertheless, the MD simulation may be performed while the thermodynamic parameters are being calculated. The production run is created using a few hundred ps to ns or even more.

3.1.2. CHARMM-GUI - Construction of membrane-protein model:

The CHARMM-GUI website (http://www.charmm-gui.org) was developed to enable the biomolecular simulation program CHARMM7 (Chemistry at Harvard Macromolecular Mechanics; http://www.charmm.org) to use common and sophisticated simulation techniques more easily and uniformly [219-220]. Utilized extensively in macromolecular mechanics and dynamics, CHARMM is an academic research program that offers possible energy functions for proteins, nucleic acids, lipids, carbohydrates, and more [221-225]. Furthermore, CHARMM may be used for a wide range of conformational and chemical free energy computations under varied constraints. Additionally, it offers flexible tools for manipulating and analyzing dynamical trajectories and atomic coordinates. CHARMM-GUI has been created to offer a graphical user interface (GUI) for generating different input files and molecular systems for CHARMM in an interactive manner by utilizing online environments. CHARMM-GUI, a Java-integrated molecular visualization software, is the basis for a novel online visualization tool that allows one to interactively inspect the generated molecular system at every step. If an issue is discovered, users can return to the previous setup and regenerate the entire system before closing the web browser. CHARMM inputs and molecular systems for biomolecular simulations and modeling may be produced in this manner using approved standards and competent methodologies. The primary goal of the CHARMM-GUI development project is to offer customers with optimum, acceptable CHARMM input files in a GUI so they may

download and utilize the inputs locally. By avoiding numerous small errors in the input, automated input generation can regulate the accuracy of calculations using recognized benchmarks and approved techniques. Users should be able to read and edit the input for their own research needs with the assistance of such input creation. Hence, the key component of CHARMM-GUI is the input generator. The process entails creating inputs and molecular systems using guidelines and modifiable parameters, after which the resulting files are made available for download. Using the integrated online visualization tool based on Marvin Space, users may interactively see the created structures on the web browser at every stage of the input production process [224]. One of the biggest benefits of utilizing the online environment is that users may easily return to the previous webpage and restart the entire system by changing the settings till it works. For instance, one can go back to earlier phases to change the system size if visual examination reveals that the created system is larger or smaller than anticipated. Short computations are run on the CHARMM-GUI main server at each stage of input production to verify that systems and produced inputs are operating as intended. Users can either (1) solvate their molecule or (2) create a water box exclusively for other uses with the aid of the Solvator module. With water box forms that are orthorhombic, hexagonal, truncated octahedral (octagonal), or spherical, users are directed in estimating the dimensions of the solvated system. Users can modify the default value of 10 Å, which represents the separation between a biomolecule's edges and a solvation box's edges. If users don't provide every dimension, Solvator creates a cubic water box with the longest axis. The default ion concentration, 0.15 M KCl, is in proximity to the ion concentration found in the human body. One of the most crucial elements of contemporary biomolecular modeling and simulation is a molecular mechanics force field (FF) [225]. Physicochemical characteristics of (small) target molecules, which can be determined experimentally or computationally, are usually used to develop and validate these FFs. Molecular mechanics FFs can be difficult to parameterize since many parameters need to be optimized while also carefully selecting experimental and computational data. The CHARMM general force field (CGenFF) [226] and the general AMBER force field (GAFF) [227] are two examples of generalized FFs for small molecules that have been constructed. Based on such a generalized FF, algorithms that can recommend topologies and FF parameters have also been created.

Step 1: Upload the pdb file

○ **Membrane Only System**

References for Lipid Force Fields:
J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell, Jr., and R. W. Pastor (2010)
Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types *J Phys Chem B. 114(23): 7830–7843*
R.M. Venable, A.J. Sodt, B. Rogaski, H. Rui, E. Hatcher, A.D. MacKerell, Jr., R.W. Pastor, and J.B. Klauda. (2014)
CHARMM All-Atom Additive Force Field for Sphingomyelin: Elucidation of Hydrogen Bonding and of Positive Curvature. *Biophys J. 107(1):134-45*
R.M. Venable, S. Swaroop, A.D. MacKerell, Jr., and R.W. Pastor.
Evaluation of a CHARMM Force Field for Phosphoinositol Lipids Based on Comparisons to Experiments. *(In preparation)*
K.C. Song, A.J. Sodt, R.M. Venable, A. Maliniak, W. Im, K. Gawrisch, and R.W. Pastor
What is the Shape of Cardiolipin? *(tentative title, in preparation)*

Next Step:
Select Model/Chain ➤

Step 2: Orientation of the protein along Z axis
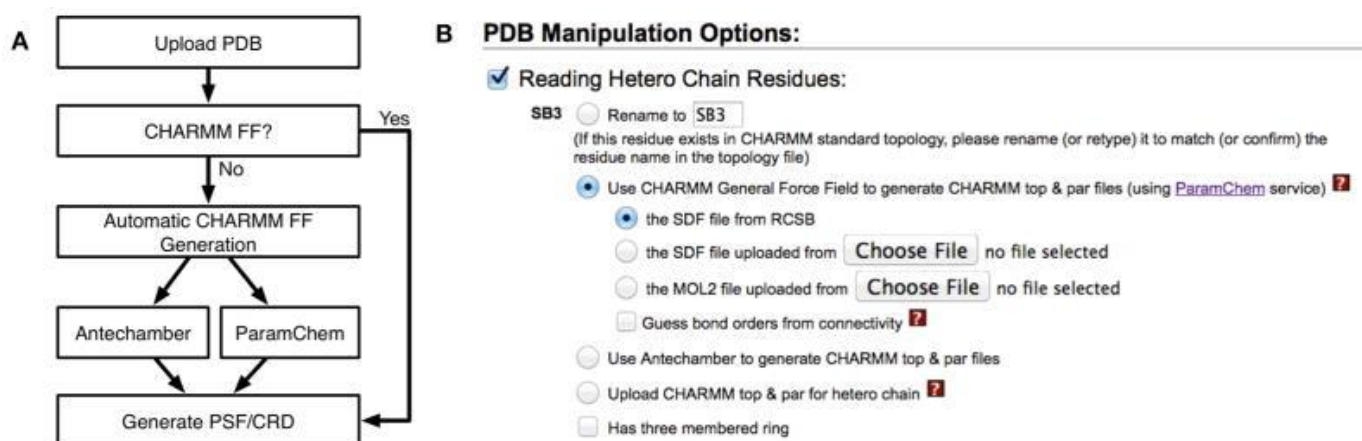
Step 3: Determining and calculating the size of the system in the Upper and Lower leaflets

Length of X and Y: 190 (initial guess)
(The system size along the X and Y must be the same)

Show the system info | click this once you fill the following table:

Lipid Type	Charge [e]	Tail Info. [sn1/sn2]	Images	Upperleaflet Ratio (Integer)	Lowerleaflet Ratio (Integer)	Surface Area

Calculated Number of Lipids:

Lipid Type	Upperleaflet Number	Lowerleaflet Number
DOPC	108	104
DOPE	270	260
DOPS	162	156

Calculated XY System Size:

	Upperleaflet	Lowerleaflet
Protein Area	0	1445.13
Lipid Area	36244.8	34902.4
# of Lipids	540	520
Total Area	36244.8	36347.53
Protein X Extent	32.01	
Protein Y Extent	91.98	
Average Area	36296.16	
A	190.52	
B	190.52	

The upperleaflet can have more lipids

If there are errors on the next page, the XY system size may need to be increased.

Step 4: Neutralization of the ions using MC method

System Building Options:

○ Insertion method   Build system using insertion method
● Replacement method   Build system using replacement method
   ☑ Check lipid ring (and protein surface) penetration

For this system, insertion method can not be used. Replacement method will be used instead.

Component Building Options:

☑ Include Ions
   ● 0.15 M ✔ KCl ion concentration) Calculate number of ions
   ○ Add neutrali NaCl
86 positive ions a CaCl2 ions will be generated. Note that this is the estimated ion numbers, so the actual ion numbers may differ.
        MgCl2
☑ Ion Placing Method: Distance

Step 5: Assembling of the complex system

Force Field Options:

AMBER

AMBER Force Fields

Protein	DNA	RNA	Glycan	Lipid	Water	Ligand
FF19SB	OL15	OL3	GLYCAM_06j	LIPID17	TIP4PEW	GAFF2

☐ Hydrogen mass repartitioning
☐ 12-6-4 ion
Glycolipids and lipoglycans are not supported in current CHARMM-GUI Amber FF implementation.

3.1.3. Molecular docking:

The computerized prediction of interactions between tiny molecules and proteins is one of the most challenging challenges in structural biology. Reliable and precise prediction of interactions might be beneficial to a lot of biological research, both in academia and industry.

The difficulty in protein-protein docking is precisely associating two interacting molecules. The exact prediction is based on the interactions between residues involved in the target interaction. There are several docking techniques available [228-232].

3.1.3.1. CB-Dock2 tool Prediction:

CB-Dock2 is a protein-surface-curvature-based cavity identification approach (CurPocket) [233, 234] that uses AutoDock Vina (version 1.1.2) to guide the molecular docking process [235, 236]. CB-Dock has been employed by several studies to examine the molecular mechanism and binding properties of medications. In four stages, CB-Dock2 accomplishes highly automated protein–ligand blind docking. The stages include data input, data processing, cavity identification and docking, and visualization and analysis (**Figure 3.8**). The MOL2/SDF/PDB file for the query ligand and the PDB file for the query protein make up the data input. The provided protein will be examined by CB-Dock2, which will also add the side-chain atoms [237, 238] and hydrogen atoms that are absent, notify about any missing residues in the protein [239], and remove co-crystallized waters and other het groups. Template matching is the first step in cavity identification and docking. It searches through the prepared complex database for existing complexes that have comparable proteins and ligands. If any homologous complexes are found, CB-Dock2 will execute docking simulation using two parallel pipelines, namely structure-based and template-based blind docking. Given the protein and ligand that the user has provided, CB-Dock2 will search our server's protein-ligand complexes database for template ligands that have a high topological similarity (FP2 ≥ 0.4) the first time around. If it exists, a comparison will be made between the query protein and the proteins that have been complexed with the chosen template ligands. For the next template-based cavity identification and molecular docking, the complexes with more than 40% sequence identity (and pocket RMSD < 4Å) in the template ligand binding site will be kept. This process uses FitDock, an in-house designed docking method that fits an initial conformation using a hierarchical multi-feature alignment approach to the supplied template. It then explores potential conformations and generates enhanced docking poses.
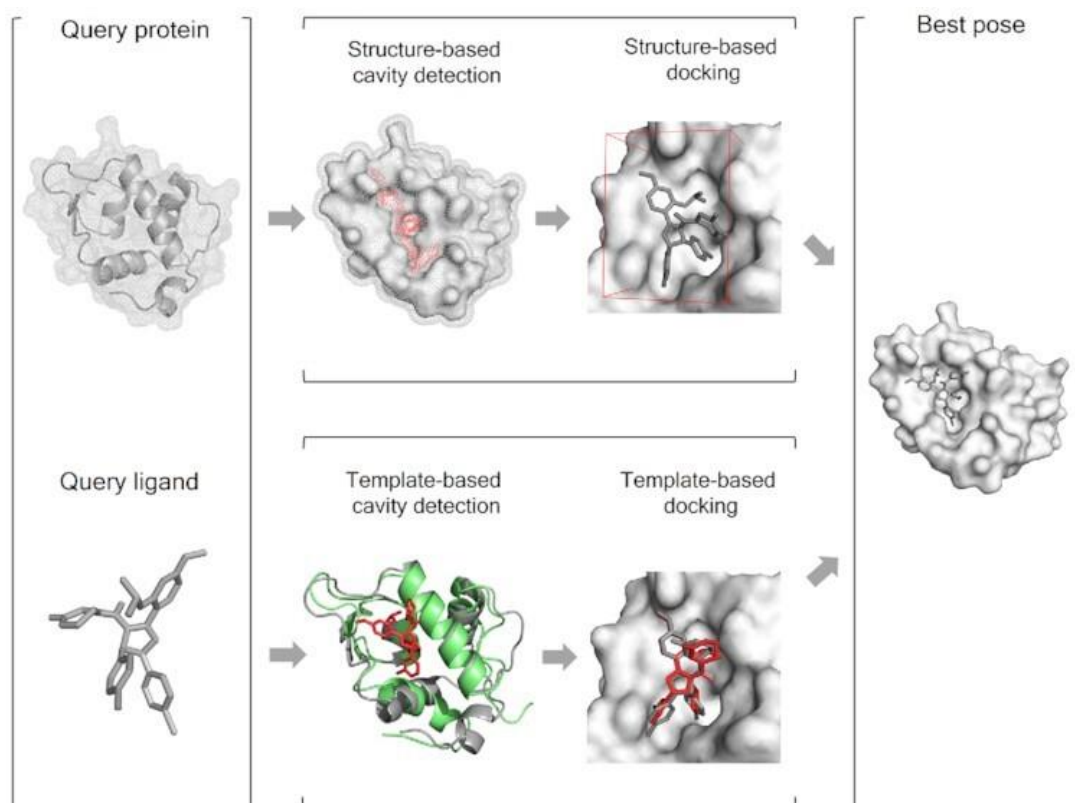
Figure 3.8. *The structure-based and template-based blind docking algorithms performed by CB-Dock2 server integration (Taken from [240])*

3.1.4. Prediction of protein-protein / protein-ligand interaction using *in-silico* tools:

Many cellular physiological processes as well as numerous illnesses depend on protein-protein interactions (PPIs) [241-243]. Protein interfaces need to be closely studied since protein-protein interactions vary. Protein contact size determines specificity and stability in protein-protein interactions. A typical protein interface comprises a surface area of 1500-3000 Å^2 that is immersed in each protein [244-246]. Van der Waals interactions between nonpolar protein residues produce protein-protein interaction sites, which are characterized by hydrophobic effects [247] and considerable shape complementarity [248-250] in proteins. The electrostatic complementarity of the interacting protein surfaces of the two proteins encourages the development and stability of the complex. The complex grows and becomes more stable as protein surfaces come into contact. Predicting protein-protein interactions is essential for creating novel treatments. Although extrinsic substances can obstruct it, protein interaction is necessary for many biological activities, both beneficial and detrimental. The two primary processes in the contemporary drug discovery process are choosing a potential pharmacological

target, gathering additional information about it, and creating a suitable ligand [251]. Thus, knowledge of protein-protein interactions can aid in the creation of modulators that specifically target protein complexes.

3.1.4.1. PDBsum server:

A web-based database called PDBsum provides a visual summary of all the important information pertaining to every macromolecular structure that has been submitted to the Protein Data Bank (PDB) (http://www.ebi.ac.uk/pdbsum) [252]. Included are extensive structural pictures, annotated plots of each protein chain's secondary structure, thorough structural analyses, a synopsis of the PROCHECK findings, and schematic diagrams of the interactions between proteins and small molecules, proteins and DNA, and other molecules. Important structural elements like the protein's domains, PROSITE patterns, and interactions between proteins and ligands are highlighted using RasMol scripts. Publicly accessible at http://www.biochem.ucl.ac.uk/bsm/pdbsum, PDBsum is updated whenever new structures are made available by the PDB. This server produced a PDB structural information library, making it the first webserver to utilize the new World Wide Web technology. Its main goal was to offer a comprehensive visual encyclopedia of the proteins and their complexes included in the PDB. It was first developed in 1995 at University College London (UCL). These images are made up of many structural studies that are either not available or not easily accessible elsewhere. Providing each 3-D model's structural information in the most visually appealing way possible is the main goal of this server. The molecules that make up each PDB entry, such as ligands, metals, protein/DNA/RNA chains, and their interactions, are thus shown graphically. Over time, a growing number of new features have been introduced. One may painstakingly compile this kind of material for oneself, in addition to the references to literature and other links to databases, however, it's preferable to deliver it right away. The following features of PDBsum server include:

a) **Wiring Diagram**

PDBsum offers a "protein page" with a schematic representation of the protein's secondary structure, or "wiring diagram," for each distinct protein chain of a structural model (**Figure 3.9**).
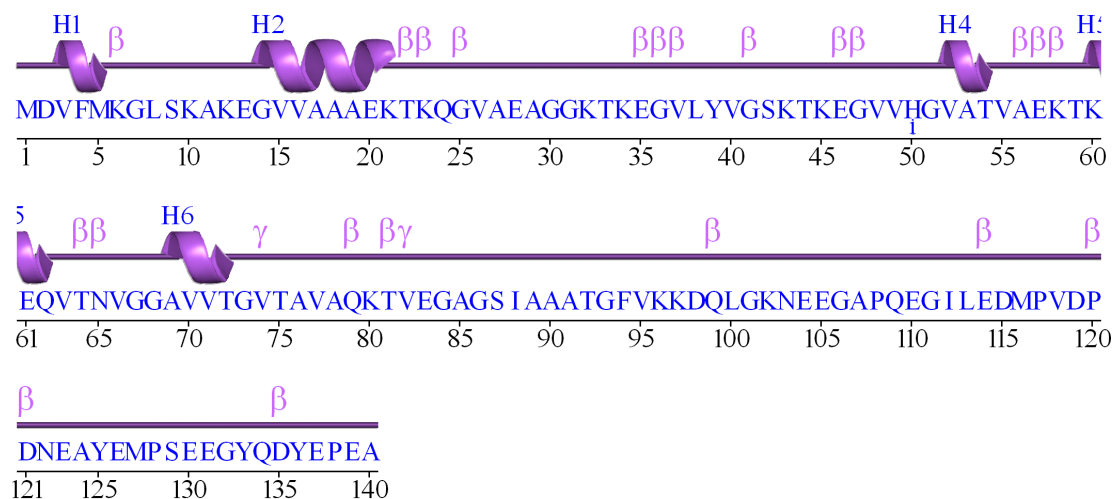
Figure 3.9. Wiring diagram showing secondary structures present in α-Syn protein (1XQ8 pdb). Helices are labelled as H1, H2, H3, H4, H5, and H6

b) Surface topology

The protein page also includes a topology diagram (**Figure 3.10**) that illustrates the connections and arrangements between the protein's strands and helices. When a protein chain includes many domains, the wiring diagram's domain shading is followed in the creation of each domain's diagram, which is done independently. Hydrogen bonding plots may be converted into topology diagrams using Gail Hutchinson's HERA programme [253].
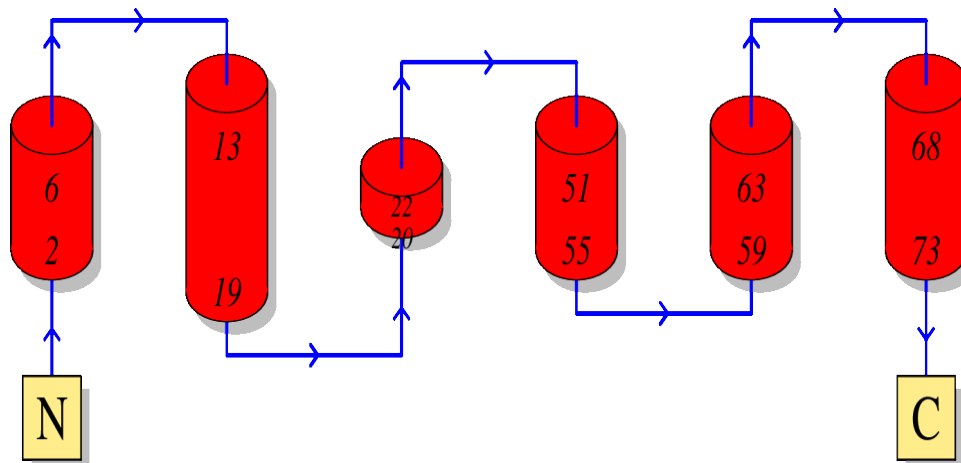
Figure 3.10. Surface topology diagram showing number of helices, beta turns, and gamma turns in α-Syn protein (1XQ8 pdb)

3.1.4.2. LigPlot server:

The LigPlot+ programme [254] creates a 2-D visual representation of the hydrogen bonds and non- bonded interactions between the protein residues that the ligand interacts with when a protein-ligand complex in the PDB format is uploaded (**Figure 3.11**). Additionally, a standalone version of the LigPlot+ programme named LigPlot+ is available for download and

installation in order to construct the protein-ligand interaction profile. The result is a PostScript (PS) file that shows intermolecular interactions such as hydrophobic interactions, hydrogen bonds, and atom accessibility as well as their intensities in either color or black and white. The programme is fully universal for all ligands. There are further specialized servers for the prediction of residue interactions in nucleic acids [255-256].
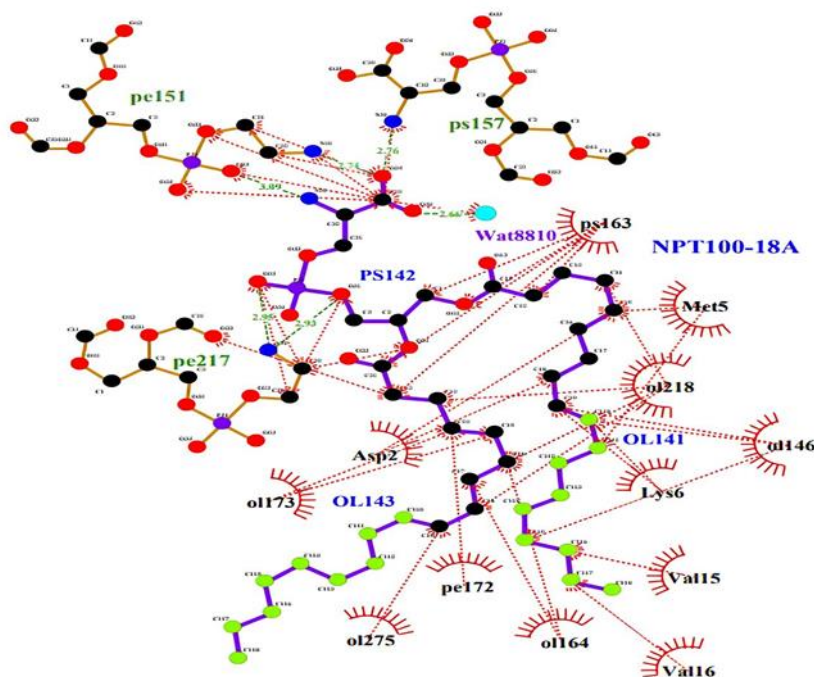

Figure 3.11. *Ligplot analysis showing hydrophobic interaction, hydrogen bonds between α-Syn protein (1XQ8 pdb) and ligand molecule (NPT100-18A)*

3.1.5. YASARA tool:

YASARA tool [257, 258] is an application designed for protein molecular modelling is needed for this kind of computational work. Molecular or homology-based modelling are two computer approaches that can produce structural information about the reaction of the system. Two particular applications are demonstrated, encompassing both homology modelling and molecular modelling techniques like energy minimization, molecular docking simulations, and MD simulations. The applications have been selected to provide concrete illustrations of how structural information obtained from homology and molecular modelling is applied to direct protein modelling research studies.

3.1.6. ESBRI server:

A web-based programme called ESBRI [259] (http://bioinformatica.isa.cnr.it/ESBRI/) is designed to examine the salt bridges in a protein structure by beginning with the atomic coordinates. Salt bridges can be crucial for both protein structure and function, and they can also have stabilising and destabilising effects on protein folding.

Input of the program: The user can input the protein name, upload a PDB format coordinates file, and select from five analysis options: (1) all salt bridges in the structure, (2) those in a specific chain, (3) between two specified chains, (4) involving positively charged residues, and (5) involving negatively charged residues. For the latter two options, the user can specify the type, chain, and residue number. A default 4.0 Angstrom threshold is set for salt bridge identification. Additional options allow for adjusting the distance between oppositely charged residues and highlighting differently charged amino acids in the results table using a color code (blue for positive, red for negative).

Output of the program: ESBRI offers various options to display salt bridge data for protein structures, such as showing all putative salt bridges in the analyzed structure, focusing on salt bridges in a specific chain chosen by the user, or displaying salt bridges between different protein chains. Users can quickly obtain information on protein binding sites, interfaces, and structural characteristics. Additionally, ESBRI identifies salt bridges between charged residues specified by the user and the surrounding chain or other chains in the structure file to assess stabilization effects and protein interactions. The tool is valuable for mutation studies and protein interaction research, providing results instantly in an easily accessible HTML format and enabling users to save results as text files.

3.1.7. Binding Free Energy (BFE) Analysis:

By combining the AutoDock Vina [236] and AutoDock-GPU [261] docking programmes in the AutoDock Suite [261], various structure-truncated MM/PB(GB)SA free energy calculation procedures, and multiple poses based per-residue energy decomposition analysis strategy, the fastDRH server was created to predict and analyse protein–ligand complex structure [262]. Compared to the two webservers we previously constructed, HawkDock and farPPI are primarily designed for the analysis and prediction of PPI target-small molecule complexes and protein–protein complexes, respectively. Theoretically, any protein-small molecule combination may be predicted and analysed using FastDRH. Users may detect the right binding mode from the docking positions more often by employing fastDRH. Furthermore, it facilitates the prediction of the protein's hotspot residue for ligand binding. The fastDRH server has fully automated prediction and analysis features, with interactive results shown through an intuitive user interface. The fastDRH server provides a workflow consists of five main phases. (1) Adding ligand and protein files to the input structures list. Second, using molecular docking to create ligand binding sites. (3) and (4) To find native/near-native binding pose and possible hotspot residues, rescoring binding poses and breaking down energies to per-residue utilising shortened protein structure using MM/PB(GB)SA techniques. (5) Examining the data and creating a zipped output file that can be downloaded that contains the raw data (**Figure 3.12**).
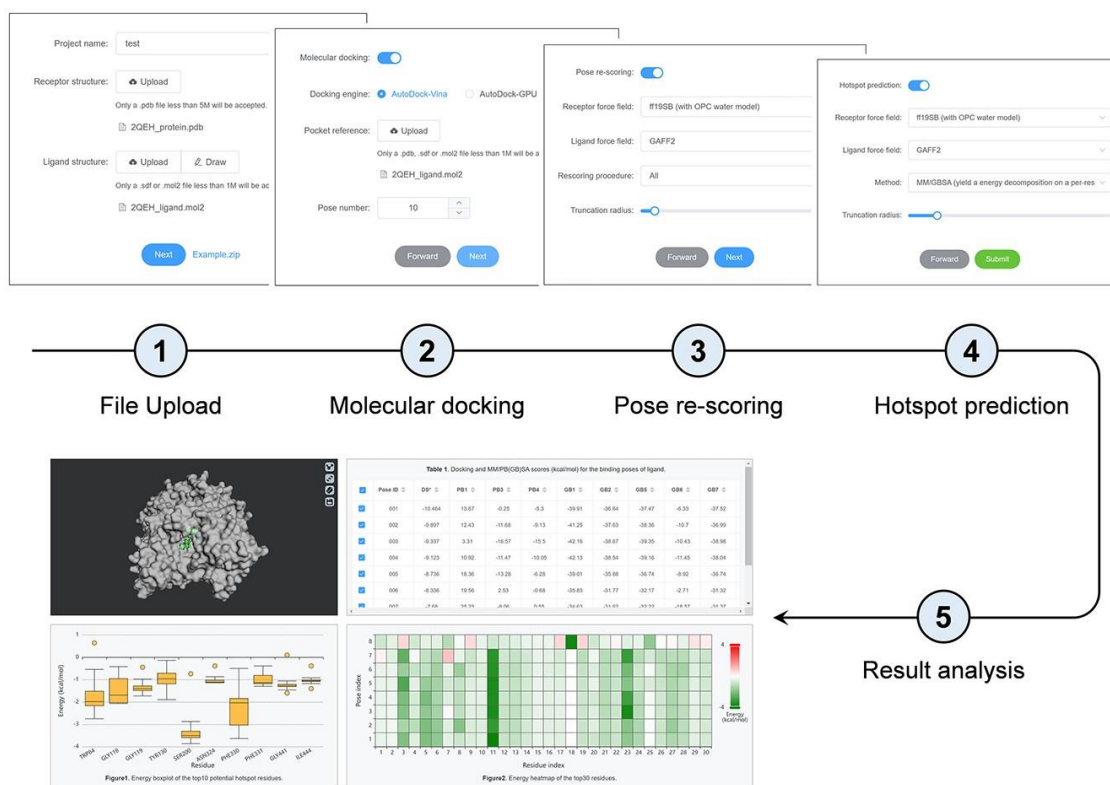

Figure 3.12. *Flowchart of the main steps of FastDRH server to calculate the hotspot prediction*

thereby calculate the binding free energy analysis [262]

3.1.8. Prediction of binding pocket analysis using CASTp Server:

Protein structures are intricate, with several interior cavities, cross pathways, and surface pockets. For proteins to perform their many tasks, including ligand binding, DNA association, and enzymatic activity, these topographic characteristics offer a structural foundation and microenvironments. Protein topographic characteristics need to be identified and quantified in order to comprehend the link between protein structures and function [263], modify proteins to have desired attributes [264], and create treatments that target protein targets [265]. The alpha shape approach [266], which was established in computational geometry, is used by the Computed Atlas of Surface Topography of proteins (CASTp) service to identify topographic features, measure area and volume, and compute imprint [267-273]. The server provides the key functions such as the ability to identify and describe protein structure channels, pockets, and cavities. Furthermore, by including pre-computed topography characteristics of biological assemblies in the PDB database together with imprints of negative volumes of these topographic features that significantly expanded its functionalities. In addition, a restoration of the user interface has improved its informatics and readability. The default probe radius for pre-computed findings is 1.4 A, which is the industry standard for calculating solvent accessible surface area. Any desired probe radius may be specified by users for customised calculation requests. The CASTp service gives a precise definition of all the atoms involved in the development of a protein structure, including all surface pockets, inner cavities, and cross channels. In addition, it measures the precise areas and volumes of their mouth holes, if any **(Figure 3.13)**. Both the solvent accessible surface model (Richards' surface) [272] and the molecular surface model (Connolly's surface) [273] are used to compute these metrics analytically. Topographic feature imprints are also provided by the CASTp server [274]. The CASTp server provides these findings for direct download, and they may be seen with the UCSF Chimera [274].
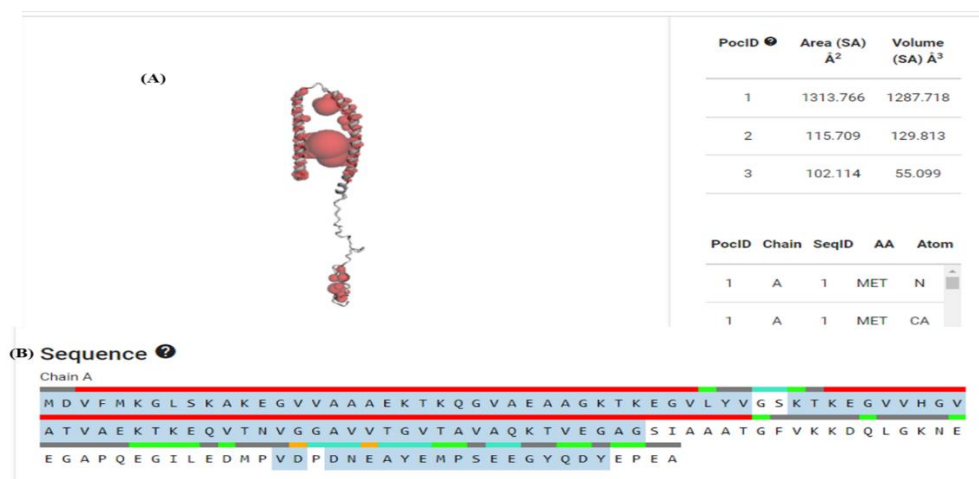
Figure 3.13. *The primary CASTp server user interface. (A)The panel in the pocket and (B) sequence panels*

3.1.9. Prediction of effect of mutants on the stability of the protein:

The effect of single point mutation on the stability and functioning of the WT α-Syn was predicted using online servers such as

I-Mutant 2.0: I-Mutant [275] is a support vector machine based tool for predicting protein stability changes resulting from single point mutations

CUPSAT: CUPSAT [276] predicts protein stability changes in terms of the unfolding free energy difference (ΔΔG) after point mutations using structural environment-specific atom potentials and torsion angle potentials between WT and mutant proteins.

MuPro: Mupro [277] predicts protein stability changes without tertiary structure knowledge.

Dynamut: DynaMut [278] is a web application that employs two normal mode approaches, sample conformations and evaluate protein dynamics and stability due to vibrational entropy changes.

Auto-mute: Auto-mute [279] web-based programmes estimate stability changes after single residue substitutions in proteins with known native structures using cutting-edge supervised classification and regression methods.

Site directed mutator: The Site Directed Mutator (SDM) [280] is a computational technique that analyses the variation of amino acid replacements occurring at specific structural environments that are tolerated within the family of homologous proteins of known 3-D structures and transforms them into substitution probability tables. These tables are used as a quantitative measurement to forecast protein stability following mutation.

Prem PS: PremPS [281] predicts the effects of single-point mutations on protein stability using evolutionary and structure-based features and a symmetrical dataset of stabilising and destabilising mutations.

SIFT: SIFT [282] predicts the effects on the protein functioning for a single point mutation that are analysed based on its tolerance/intolerance score.

PONDR VLXT: The percentage of disorderliness can be determined by PONDR VLXT [283]. The term PONDR VL-XT describes the combination of three predictors: two trained on X-ray characterised Terminal disordered areas and one trained on variously characterised long disordered regions.

ExPASy ProtParam online tool: The physiochemical properties of instability index, aliphatic index can be determined from ExPASy online tool [284].

3.1.10. Helanal Server:

HELANAL server [285] performs a least square 2D-line and circle fit to the local helix origin points (LHOP) in order to determine the geometry of the associated helix. For each pair of four successive Cα atoms within the helix, local helix axes are used to derive local helical parameters. For every set of four Cα atoms, HELANAL server computed the local helical parameters: rise, twist, virtual torsion angle, LHOP, and local helix axis. After LHOP were reoriented in the X-Y plane, the best-fit circle and straight line were determined. Recent research has shown that, depending on the protein's orientation in three dimensions, a helix's categorization as linear or curved may rely on this orientation. As a result, helix geometry may vary for the same helix when coordinates are retrieved from two separate databases. To address this issue, we have created an enhanced version called HELANAL-Plus, which assigns the geometry to the helix as linear or curved using least square 3D line and sphere fitting to LHOP.

Dataset Preparation: Protein structures of various kinds have been used to test the HELANAL geometry method. Structural Classification of Proteins (SCOP) database is the standard fold benchmark used by the server. In a system of four levels—class, fold, superfamily, and family—SCOP organises all protein domains with known structures. The proteins in our study are grouped at the fold level, which is a topologically connected grouping with the same key secondary structures.

3.1.11. SOPMA server:

The self-optimized prediction multiple alignment method (SOPMA) [286], a novel technique, has recently been reported to increase the success rate of protein secondary structure prediction. In this work, we present gains achieved by forecasting every sequence in a group of aligned proteins from the same family. In a complete database of 126 chains of non-homologous (less than 25% identical) proteins, this enhanced SOPM approach (SOPMA) accurately predicts 69.5% of amino acids given a three-state description of the secondary structure (alpha-helix, beta-sheet, and coil). For 74% of the co-predicted amino acids, joint prediction using SOPMA and a neural network approach accurately predicts 82.2% of residues. The predictions can be obtained via email.

3.1.12. Density Functional Theory (DFT):

Density functional theory (DFT) calculations are conducted to optimize the geometry of the small molecules examined for the study without imposing any constraints. The B3LYP hybrid functional [287], coupled with the def2-TZVP basis set [288, 289], was utilized for these calculations. Grimme's dispersion correction in the D3 version [290] and the conductor-like polarizable continuum model (CPCM) were employed to simulate the aqueous solvent environment [291], denoted as B3LYP+D3/def2-TZVP (CPCM), with a dielectric constant ε = 80.0. Frequency calculations were performed at the same level of theory to ascertain the correct stationary ground state, ensuring that all minima possessed real vibrational frequencies. The calculations are conducted using Gaussian 16 version C.01 [292]. Additionally, natural bond orbital (NBO) analysis was conducted using Gaussian 16, within the B3LYP+D3/def2-TZVP (CPCM) environment [293], to gain insight into the nature of the frontier molecular orbitals (MO) of the inhibitors. Subsequently, the xyz coordinates of the small molecule structure are converted to PDB format using the Open Babel server [294].

3.1.13. Analysis of MD trajectory:

(i) Root Mean Square Deviation (RMSD):

RMSD is a measurement used to determine a structure's deviation from a certain conformation. It is described as:

$$RMSD = \left[\frac{\sum_N \left(R_i - R_i{}^0 \right)^2}{N} \right]^{1/2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.25)$$

where, R_i is the vector location of particle i (the target atom) in the snapshot, $R_i{}^0$ is the coordinate vector for reference atom i, and **N** is the total number of atoms/residues taken into account in the computation. Using backbone atoms and the simulation's first frame as a

reference, the RMSD was calculated. The RMSD is the product of the number of locations (i), the number of strands (j), and the number of angular parameters (k). The value of N in the above equation denotes the total number of variables needed to compute the RMSD.A radial vector of length r in the structure space denoted by the RMSD absolute magnitude is the calculated RMSD. The radial issue is predicated on the idea that there is more configurational space volume between a given r and r+dr, the broader the radius. The same RMSD value might capture both equivalent and different structures at larger r values. Techniques that use significantly lower RMSD values offer a more precise way to quantify difference. The existence of two or more structural substrates creates a second important problem with the use of RMSD because of the molecule's intrinsic flexibility. However, a technique is required to accurately define the dynamical properties without compromising the information.

(ii) Root Mean Square Fluctuation (RMSF):

The measure of divergence between the particle location i and a reference position is defined by Root Mean Square Fluctuation, or RMSF:

$$\text{RMSF} = \left[\frac{1}{T}\sum_{t=1}^{T}(r_i(t) - r_i^{ref})^2\right]^{1/2} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots (3.26)$$

T is the desired average time, and r_i^{ref} is the particle i reference location. The time-averaged location will serve as the reference location of the same particle i, that is, r_i^{ref} = i. Root mean square deviation (RMSD) and root mean square fluctuation (RMSF) are metrics used to quantify the spatial fluctuations of biomolecules in MD simulations. For a given collection of atoms, RMSD is the difference between two structures; on the other hand, RMSF is the variation around an average, per atom or residue, over a series of structures (e.g. from a trajectory).

(iii) Radius of gyration (Rg):

The radius of gyration is computed to determine the structure's compactness:

$$R_g = \left[\frac{\sum_i |r_i| 2m_i}{\sum_i m_i}\right]^2 \quad \dots\dots\dots\dots\dots\dots\dots\dots (3.27)$$

The mass of atom i is denoted by m_i, and its location in relation to the molecule's centre of mass is indicated by r_i.

(iv) DSSP plot analysis : Secondary structural content analysis:

For most of the proteins in the PDB, Kabsch and Sander (1983) created a database of the ASA

by using a method called Dictionary of Secondary Structure for Protein (DSSP) to ascertain the solvent accessibility of the residues [295]. This programme is frequently used to create the ASA values for use in prediction algorithms [296-298]. It works by primarily classifying secondary structures of proteins based on their backbone H-bonds. It also gives information on bond and Cα-pseudo dihedral angles; only the latter is required by DSSP in order to determine the LSS of a residue. The electrostatic hydrogen bond detection criteria set it apart. Consequently, the unique hydrogen-bond patterns are being used to designate the elements of secondary structure. This method is often used to assign secondary structures as a gauge. DSSP is used by many software applications to assign secondary structures as needed. For example, a widely used visualization tool such as Rasmol assigns repeating structures in a fast way that is comparable to the DSSP. Classification of β-bridges is based on non-recurring H-bonds, whereas classification of helix or strands is based on repeating patterns of the same type of H-bonds [299]. Residues that share the same secondary structure pattern are clustered together quite tightly in a Ramachandran plot because the relative orientations of the nitrogen and oxygen atoms in the backbone are reflected in the corresponding (φ, ψ) backbone tilt angles. The DSSP offers information on the secondary structure of the protein on two levels [295]. The one-character secondary structure information (1CSSI) code, which at the higher level characterizes the LSS of a residue mostly based on the H-bond arrangement of the protein backbone into eight classes, summarises the secondary structure from DSSP analysis. The angle that exists between the vectors Cα (i) − Cα (i−2) and Cα (i) − Cα (i+2) for each residue is known as the C-pseudo bond angle, and it is instead calculated using DSSP. A gap for better discrimination is created by DSSP in the event that none of the previously specified conditions are satisfied; this gap is represented by the letter "C." These residues, however, correspond to a usually straight section of the protein backbone structure and do not contain the backbone H-bonds required for the formation of secondary structures.

(V). Membrane bilayer stabilization analysis:

To determine if the adopted simulation protocol is correct, lipid profile analyses were carried out using MD trajectory files. The main parameters includes- area per lipid (APL) analysis, membrane thickness and Electron density profiles (EDP).

The average APL was determined by multiplying the simulation box's x/y dimensions and dividing the result by the quantity of lipid and cholesterol molecules found in a single bilayer leaflet. The distance in the EDP between the phosphate peaks was used to compute the bilayer thickness. The 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) analysis

yielded an average APL value of 65.32 Å^2 to 66.32 Å^2 as per the experimental average [299-302]. Then, based on how far each atom is from a flat plane, often denoted as z = 0, each atom is binned along the z dimension of a periodic simulation box. EDPs calculated using this flat-patch assumption are referred to as coming from the z-bin technique. The broad need is clear: simulated systems must be small enough such that the z-direction undulations are minor in comparison to the average fluctuations inherent in the EDP. The exact size of a simulated patch that must exist for the flat-patch assumption to hold is not evident. To calculate the average membrane thickness, the distance that is the shortest between a phosphorus atom in one leaflet of a lipid layer and every phosphorus atom in the opposite leaflet of the lipid layer is determined. The development of the membrane thickness for mixed lipid bilayers, for a simulated duration can be calculated.

(VI). Principal Component Analysis:

Principal Component Analysis (PCA) in MD trajectories identifies coordinated motion by eigenvectors of the mass-weighted covariance matrix, based on Cα positional fluctuations [303, 304]. It's used to analyze a-Syn complexes with and without inhibitors, studying protein shape changes. All MD trajectory inputs were assessed, showing system stabilization after a period of time. PCA creates a comprehensive reduced space basis set using eigenvectors, aiding in matching structures in new coordinates. Large-weight components can highlight key structural features. Free energy profiles for α-Syn were calculated based on point likelihood in a bin.

$$\Delta G = -k_B T \ln(\rho) \dots\dots\dots\dots\dots\dots\dots\dots\dots (3.28)$$

Where, k_B is dissociation rate, ρ is reaction quotient. Free energy profiles for α-Syn were calculated using Amber's CPPTRAJ module [305].

## 3.1.14. 3-D structure visualization tool:

UCSF Chimera:

University of California, San Francisco (UCSF) Chimera: A highly adaptable instrument known as UCSF Molecular structures and related data, including as conformational ensembles, density maps, sequence alignments, supramolecular assemblies, and docking results, may be interactively seen and analysed using Chimera [306] as shown in **Figure 3.14**. Chimera is an application that was created by the Resource for Biocomputing, Visualisation, and Informatics (RBVI) with support from the National Institutes of Health (NIH).
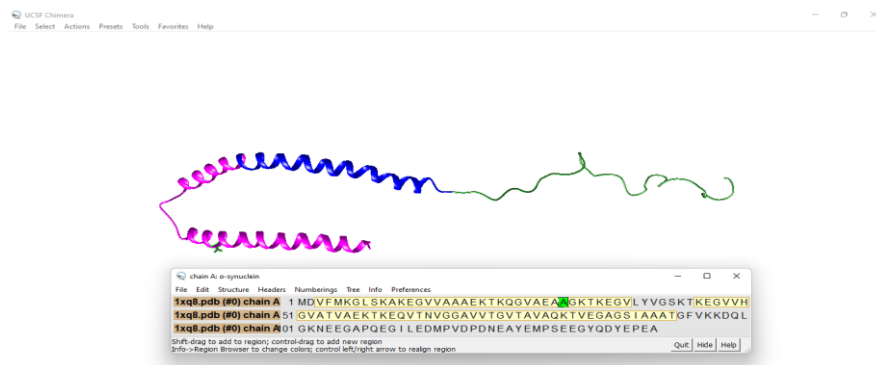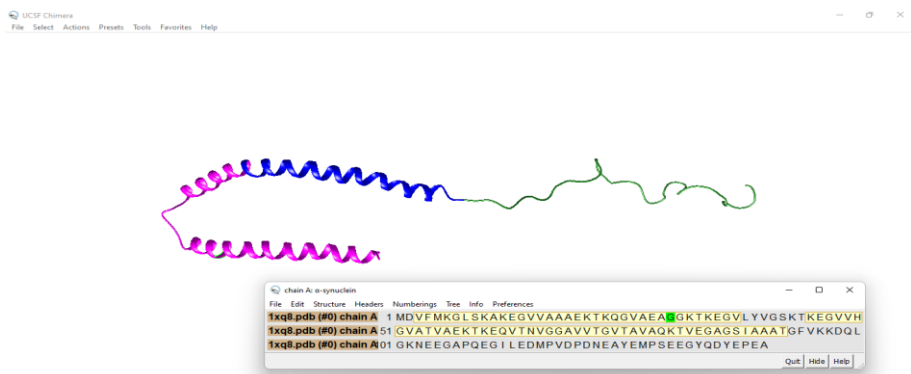
STEP 1: Tools → Structure editing → Rotamers →Select the residue (Cltr + shift + left click)
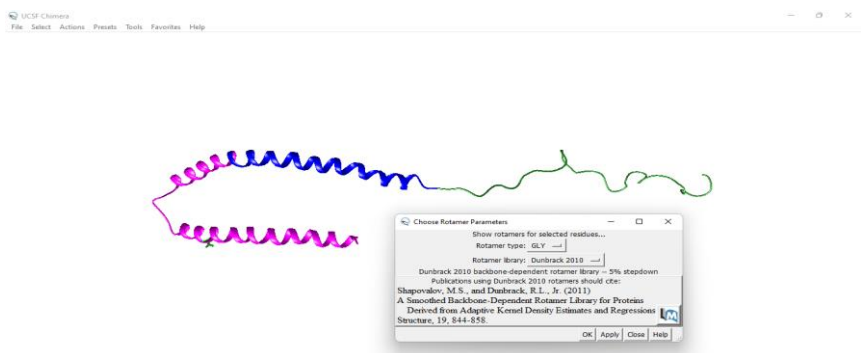
STEP 2: Tools → Binding Analysis → Find clashes/contacts → Designate → Apply

STEP 3: Favorites → Sequence→ mutated residue

Figure 3.14. *Construction of α-Syn mutants using UCSF Chimera software [306]*