Chapter 2
Models and Methodology

Computational microscopy[82] has become an essential technique to investigate molecular and atomic-scale phenomena. Using atomic-molecule interaction models, the authors enable detailed analysis without relying on advanced experimental instruments. The technique spans various spatial and temporal scales, from nanometers to micrometres and picoseconds to microseconds. A central challenge in computational microscopy is balancing the physics of the problem with the computational resources available. For instance, researchers may study time-dependent processes such as self-assembly, examining changes in structure and dynamics alongside thermodynamic considerations. Alternatively, some studies emphasize equilibrium configurations where time dependency is less critical. Typically, computational models treat atoms and molecules as classical particles obeying Newton's laws of motion. This approach, often called all-atom modelling[83], accounts for individual atomic-level interactions. However, such models can become computationally expensive, especially for complex systems like biomolecular interactions with membranes. To address these challenges, coarse-graining techniques [84] simplify models by grouping atoms into larger units, reducing computational demands while preserving essential system behaviours. The choice of model and simulation method depends on the complexity of the system and the relevant length and timescales.

Due to enormous advancements in the development of supercomputing machines, theoretical modelling and simulation have proved extremely useful and essential to interpreting the experimental results and getting information at the atomistic level. Theoretical modelling and simulation have emerged as a powerful tool for investigating the structure and dynamics of bulk and interfacial water or aqueous solutions. Theoretical modelling and simulation can yield information at the atomistic level and provide detailed insight, sometimes beyond the scope of even modern, state-of-the-art experimental techniques. Among the different available simulation techniques, molecular dynamics (MD) simulation has

the advantage of not only predicting the structure of the fluid at the atomistic scale resolution but also providing information about the dynamic time history and, hence, dynamics of the system.[85] A strong foundational understanding of MD simulation and modelling techniques is critical for conducting advanced research. This ensures appropriate methodologies are applied when exploring specific scientific questions or phenomena.

2.1 Classical Molecular Dynamics

Molecular Dynamics (MD) simulation, first introduced by B. J. Alder and T. E. Wainwright in 1957, was initially employed to study phase transitions in systems modelled as hard spheres.[86, 87] This pioneering work laid the foundation for the development of MD as a computational tool for understanding the behaviour of complex systems. In 1964, A. Rahman significantly advanced the field by using MD to investigate the local structure and dynamics of condensed-phase argon systems, where interactions were modelled using effective potentials. His work introduced equilibrium autocorrelation functions to estimate transport properties, such as diffusion coefficients and viscosity, marking a milestone in applying MD simulations.[88] The advent of high-performance supercomputing machines has significantly enhanced the capabilities of MD simulations, enabling their use across diverse disciplines, including Physics, Chemistry, Biology, Material Science, Chemical Engineering, and Mechanical Engineering. This versatility has made MD an indispensable tool for exploring a wide range of phenomena, from protein folding and drug binding in biophysics to designing advanced materials and nanostructures.

Classical MD simulation is a deterministic method that follows the laws of classical mechanics to depict the time evolution of the phase space of a set of interacting atoms or molecules. For example, by integrating Newton's equations of motion, this method generates a set of coordinates and momenta (velocities) of constituent particles of the system as a function of time. For a simple atomic system, the force (F) experienced by any par-

ticle within the system, as per Newton's second law of motion, can be written as

$$\overrightarrow{F} = m\frac{d^2\overrightarrow{r}}{dt^2} \tag{2.1}$$

where, m is the mass of the atom and r is the positional coordinate. By integrating the above equation of motion from a set of positions and velocities at a particular time step, the consecutive positions and velocities of the next time frames separated by a short time interval dt can be calculated. The abovementioned equation can be solved using different numerical integration algorithms, giving the new positions and velocities of the atoms or particles at each time step. Among these methods, Verlet algorithm and velocity Verlet algorithm[89], leap-frog algorithm [90], to name a few.

According to the Velocity-Verlet algorithm[89]; the position $r(t+dt)$ and velocity $v(t+dt)$ of each particle constituting the system at time $(t+dt)$ can be obtained from position $r(t)$, velocity $v(t)$ and force $(F(t))$ of the same at time t such that

$$r(t+dt) = r(t) + v(t)dt + \frac{F(t)}{2m}dt^2 \tag{2.2}$$

$$v(t+dt) = v(t) + \frac{F(t)+F(t+dt)}{2m}dt \tag{2.3}$$

The force acting on each constituent particle in the system can be calculated from the negative gradient of the total potential energy of the system, i.e.

$$\overrightarrow{F}_i = -\overrightarrow{\nabla}_i U(r_{ij}) \tag{2.4}$$

Hereby, $U(r_{ij})$ is the potential energy between any two particles i and j, often calculated from an empirical force field corresponding to the simulation system. Thus, once the empirical force field is defined, the above set of equations can be solved successively to obtain positions and velocities as a function of time, commonly known as the system's trajectory.

2.2 Molecular Modeling

In a classical MD simulation, the system characteristics are hidden in its potential energy functions, commonly known as the force field. Depending on the bonding nature in a molecule, such a site-site interaction potential function consists of two distinct types of interactions, namely (a) nonbonded interactions and (b) bonded interactions:

$$U = U_{bonded} + U_{nonbonded} \tag{2.5}$$

By suitably choosing the potential functions for these two types of interactions, the total potential energy of the system can be calculated by considering various intra- and inter-molecular interactions among different particles within the system. We have used a standard force field from the literature to model a molecular system.[91]

U_{bonded} represents the bonded interactions that arise from bond stretching (U_{bond}), bond bending (U_{angle}), and torsion $(U_{dihedral})$ and improper $(U_{improper})$, whichever may be applicable to describe the system suitably.

$$U_{bond} = \sum_{bonds} K_b(r - r_0)^2 \tag{2.6}$$

$$U_{angle} = \sum_{angles} K_\theta(\theta - \theta_0)^2 \tag{2.7}$$

$$U_{dihedral} = \sum_{dihedrals} f(\phi) \tag{2.8}$$

This $f(\phi)$ has different mathematical forms.

$$f(\phi) = k_1[1 + cos(\phi)] + k_2[1 - cos(2\phi)] + k_3[1 + cos(3\phi)] + k_4[1 - cos(4\phi)] \tag{2.9}$$

$$f(\phi) = k_n cos^{n-1}(\phi) \tag{2.10}$$

$$U_{improper} = \sum_{impropers} k_\omega(\omega - \omega_0)^2 \tag{2.11}$$

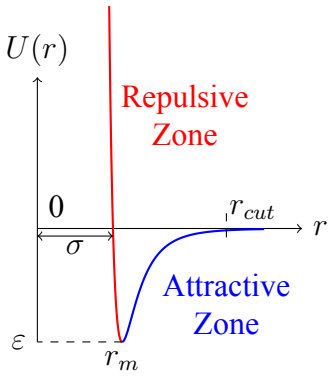

Figure 2.1: Schematic illustration of Lennard-Jones potential

Here, r is the bond distance between two atoms, and r_0 is the equilibrium bond length. θ defines the angle between three atoms and θ_0 is the equilibrium angle. ϕ defines a dihedral of four atoms. ω defines the improper torsion/dihedral angle and its equilibrium value ω_0. k_b, k_θ, k_1, k_2, k_3, k_4, k_n, k_ω area the constants.

The nonbonded potential, $U_{nonbonded}$, arises from the van der Waals and Coulombic interactions. We used the Lennard-Jones 12-6 (LJ-12-6) potential to model the van der Waals interactions between the particles. The LJ pair potential is defined as –

$$U_{ij} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.12}$$

where ϵ_{ij}, σ_{ij}, and r_{ij} are the potential well depth, atomic diameter and the distance between two atoms i and j, respectively.

As shown in Figure 2.1, when the particles are far from each other, they always experience an attractive force. Of course, it is a function of distance; therefore, at a larger r, the attractive force is weak. As the particles approach each other, the force becomes more potent as the slope of the curve becomes steeper, according to Equation (2.4). As they come into contact, a repulsive force emerges due to the repulsion of their electron clouds. By differentiating (2.12) with respect to r_{ij} and setting the result to zero, we find that the optimal distance, where attractive force equals repulsive force, i.e., $U = 0$, is $\sigma = 2^{-1/6} r_m$. If $r_{ij} > \sigma$, the attraction will be dominating; else it is repulsion. In most cases, the Lennard-Jones (LJ) interaction is truncated at a specific cut-off distance, r_{cut}, because its influence

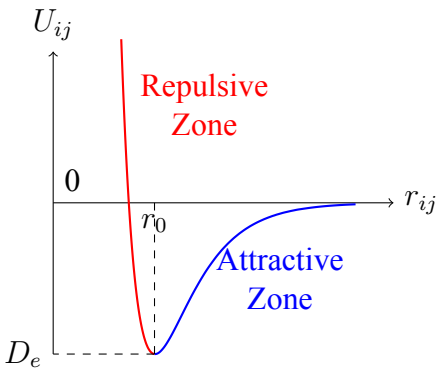

Figure 2.2: Schematic illustration of Morse Potential

is much less significant than bonded interactions. Nevertheless, it remains essential because short-range van der Waals interactions govern many physical properties of materials, regardless of their phase. The cut-off distance is chosen such that the interaction energy at this point is comparable to the thermal energy of the system – $\Delta U_{ij}(r_{cut}) = \epsilon_t \approx k_B T$, where k_B is Boltzmann's constant and $T$is the absolute temperature. To avoid discontinuities in force calculations at $r = r_{cut}$, the LJ potential is often shifted/truncated at r_{cut} as – $\Delta U_{ij,truncated} = \Delta u_{ij,actual} - \Delta u_{ij}(r = r_{cut})$. The Lennard-Jones potential is only applicable to represent the non-bonded van der Waals type of interactions for the atoms/particles that are spherically symmetric. For nonsymmetric particles, some other potentials, such as $GayBarne$ and $Gaussian - overlap$ potentials,[92], can be helpful when the particle's aspect ratio becomes essential. Apart from them, the Morse potential is widely used for the atoms that tend to create diatomic molecules (primarily applicable to metals). It is beneficial for modelling bond stretching in molecular systems, providing a more realistic representation of interatomic interactions than the harmonic potential. The Morse potential is defined by,

$$U_{ij}^{Morse} = D_e \left[e^{-2\alpha(r_{ij}-r_0)} - e^{-\alpha(r_{ij}-r_0)} \right] \tag{2.13}$$

where r_{ij} is the distance between particles i and j, and D_e, r_0, and α are empirical parameters, describing the depth of the potential well at the lowest point, position of the minimum energy and width of the potential well, respectively. Figure 2.2 depicts the Morse potential as a function of the interparticle distance, r_{ij}, which describes the potential energy of

interaction between the two particles.

To calculate the electrostatic interaction, the Coulombic potential is used to address the interactions due to partial charges present in molecules/atoms.

$$U_{ij,coulomb} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.14}$$

where, q_i and q_j are the magnitude of charges of atoms i and j respectively, ϵ_0 is the permittivity of free space, r_{ij} is the distance between atoms i and j. Hence, the total potential energy due to the molecular interactions can be expressed as follows:

$$U_{Total} = U_{bond} + U_{angle} + U_{dihedral} + U_{improper} + \sum_{i\neq j}^{natoms} U_{ij,LJ} + \sum_{i\neq j}^{natoms} U_{ij,Coulomb} \tag{2.15}$$

Force field parameters are primarily derived from experimental data and quantum mechanical calculations. Over the past three decades, various force fields have been developed and are widely used for simulations across diverse systems of interest. In our work, we have predominantly employed the GROMOS96 53a6 [93] and CHARMM36 [94] force fields for non-solvent molecules, along with the SPC/E and TIP3P [95] water models for simulating aqueous environments. A key prerequisite for any MD simulation is the proper modelling of the system, which involves selecting an appropriate length scale and force field or model potential, followed by the creation of an initial configuration. The initial velocities of atoms are typically assigned using the Maxwell-Boltzmann distribution, ensuring consistency with the desired average temperature of the system. Before starting the simulation, suitable boundary conditions should be employed to mimic the system of interest. A periodic boundary condition[96] is such a condition that allows particles that exit one side of the simulation box to reappear on the opposite side, creating an infinite system by removing the edge effects, as shown in Figure 2.3a. While calculating the force between the particles, a minimum distance between a particle and its periodic image is considered, which is known as the Minimum Image Convention.[97]. As in Figure 2.3b, ${r^2}_{ij}$ is considered for the distance calculation between the deep blue and the grey particle *i* and *j*, respectively. In our simulations, periodic boundary conditions and minimum im-

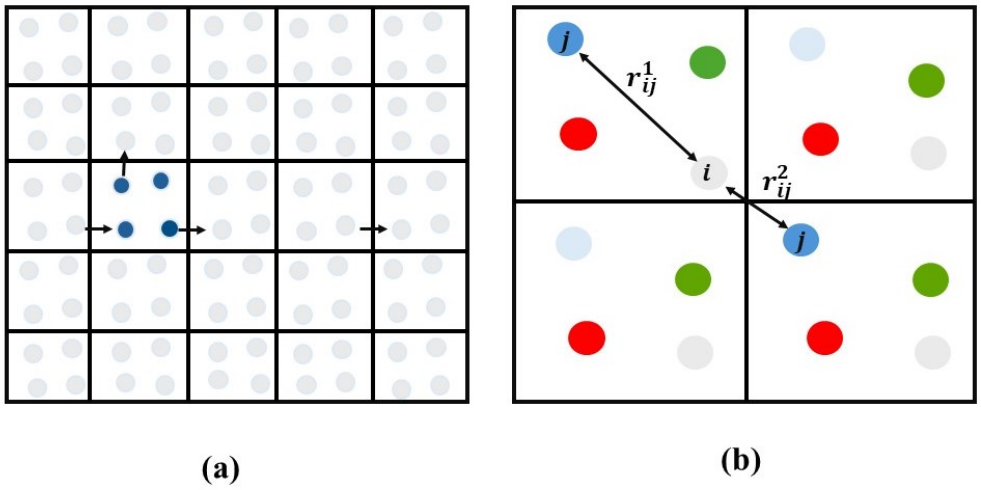

Figure 2.3: Schematic representation of (a) Periodic Boundary Condition, (b) Minimum Image convention

age conventions are applied in all three directions. Because of the long-range nature of the Coulomb potential, the particle-particle particle-mesh (PPPM) method[98] is used to evaluate the charge-charge interaction correctly. The equation of motion is generally integrated by applying finite difference methods. The basic criteria of a reliable integrator are that it should have high accuracy (i.e. follow the true trajectory), stability (conservation of energy) and robustness (allow larger time in the propagation of the system in phase space). Several algorithms have been developed for integrators in MD simulations. Throughout these works, we have used the Velocity-Verlet algorithm for our simulations.[99] One usual way of avoiding larger computational requirements is to avoid simulating very fast motions like the one due to bond or angle vibration by constraining these bonds or angles to their equilibrium value during the simulation. Here, we have used SHAKE to constrain the OH bonds of the water molecules.[100]

2.3 Thermodynamic Ensembles

A thermodynamic ensemble is a fundamental concept in statistical mechanics and molecular simulations. It refers to a collection of hypothetical systems, each representing a possible state of the system being studied, subject to specific macroscopic conditions (e.g., temperature, volume, or pressure). These ensembles compute averages of physical prop-

erties, linking microscopic interactions to macroscopic thermodynamic behaviour. Thermodynamic ensembles allow us to study a system at equilibrium by sampling all possible configurations it can adopt under given constraints. Each configuration (or microstate) has a specific probability determined by the ensemble's governing thermodynamic conditions. We obtain macroscopic quantities like pressure, energy, and entropy by averaging properties over the ensemble. Three types of thermodynamic ensembles are the miniaturization of the classical thermodynamic systems.

2.3.1 Microcanonical Ensemble

The microcanonical ensemble is a fundamental thermodynamic ensemble in which the total number of particles (N), volume (V), and energy (E) of the system are kept constant. The system is isolated in this ensemble, meaning it does not exchange energy or particles with its surroundings, and its total energy remains conserved.

2.3.2 Canonical Ensemble

In the canonical ensemble, the total number of particles (N), volume (V), and temperature (T) of the system are held constant. This ensemble is used to study systems under controlled temperatures and fixed spatial boundaries. The temperature is typically maintained using a thermostat, such as the velocity rescaling,[101] Nosé-Hoover,[102] Berendsen,[103] or Langevin thermostat,[104] which ensures that the kinetic energy distribution of the particles follows the Maxwell-Boltzmann distribution. The NVT ensemble is especially useful for investigating equilibrium properties at a constant temperature, such as structural configurations, density, and internal energy. In the NPT ensemble, the total number of particles (N), pressure (P), and temperature (T) are kept constant. The pressure is regulated using a barostat, such as the Parrinello-Rahman,[105] Berendsen,[106] barostat, which adjusts the simulation box dimensions to maintain the desired pressure. These ensembles represent a closed system in thermal contact with an external heat bath, allowing heat exchange while maintaining a constant volume or pressure.

2.3.3 Grand-Canonical Ensemble

The grand canonical ensemble ($\mu VT/\mu PT$) is a thermodynamic ensemble in which the chemical potential (μ), volume (V), and temperature (T) of a system are kept constant. In this ensemble, the system can exchange energy and particles with a reservoir, meaning that the number of particles (N) in the system is not fixed but fluctuates. This is related to the open system. Most physical experiments are generally done at a constant temperature or under constant pressure conditions. Hence, in most cases, NVT/NPT ensembles are chosen to simulate any system. Average thermodynamic properties, like temperature, pressure, potential energy, kinetic energy (total energy), etc., are updated in every timestep. As the system is *ergodic*, calculating any time-average quantity will give the average quantity of the particular system of interest.

2.4 Property Calculation

Molecular dynamics simulations offer a powerful tool to calculate various properties of a system, providing insights into its behaviour. Here, we discuss the specific properties relevant to this thesis.

2.4.1 Density Profile

The density profile describes how the density of atoms or molecular groups changes with distance from the centre of the micelle. For a system with a solid-liquid interface, calculating the density distribution of the molecules plays a vital role in understanding where the molecules are getting self-assembled on the surface. Suppose a person is interested in investigating how hairy nanoparticle droplets spread on a solid surface in the dimension normal to the surface. This is done by discretizing the simulation box into several bins in the dimension normal to the surface, and calculating the number of molecules or atoms of interest within every bin. Finally, to estimate the density profile, the number of molecules in each bin is divided by the bin's volume to find the density. The number density is multiplied by the mass of the tin molecule or group of interest for mass density. It indicates

whether the molecules get strongly or weakly adsorbed onto the surface. The sharpness of the peaks means the strong layering of molecules.

2.4.2 Radial Distribution Function

The radial distribution function, $g(r)$, quantifies the likelihood of finding a particle at a distance r from a reference particle. It provides a spatial density profile relative to the system's bulk density, expressed as $g(r) = \rho(r)/\rho_{bulk}$. The characteristics of $g(r)$ vary significantly across different states of matter. In solids, where particles are arranged in a highly ordered structure and oscillate around their equilibrium positions, $g(r)$ exhibits sharp peaks at regular intervals. These peaks correspond to the first, second, and subsequent coordination shells, reflecting the spatial arrangement of neighbouring particles within the solid lattice. In contrast, liquids adhere to the hard-sphere model, where molecules repel each other to prevent overlap. At larger distances r, molecules behave independently, and the local density approaches the bulk density of the liquid. The RDF of a real gas typically displays a single coordination sphere that decays quickly beyond the first shell, aligning with the bulk density of the gas. This is classically calculated by considering the domain to be spherical in three dimensions. In some cases, the in-plane radial distribution function (RDF) also becomes important to understand the extent of correlation/interaction of molecules/particles within a layer. It also tells about the *phase the system belongs to* based on the correlation between particles. For a *solid*-like behaviour, the RDF should have consecutive peaks without any decay along r (blue line). In contrast, for *liquids*, initially, RDF may have higher peaks (black line), but eventually, it reaches unity.For *gases*, as the correlation is the least, the RDF decays the fastest (ash line) as shown in Figure 2.4 [107].

2.4.3 Radius of Gyration

The radius of gyration, R_g, measures the overall size of the micelle and how mass is distributed from the centre. It gives an estimate of the shape of the micelle and compactness, with smaller values indicating a more compact structure. We monitored the size and shape

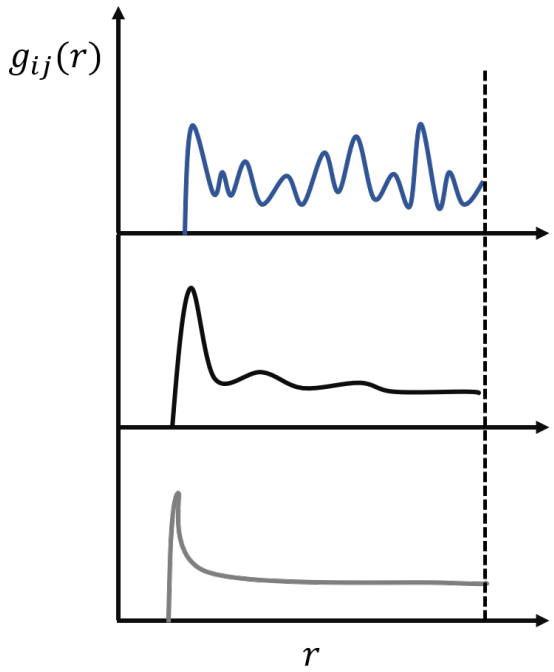

Figure 2.4: Radial distribution functions at different phases. Colour code: blue - solid, black - liquid, and grey - gas.

of the micelle by calculating the components of the radius of gyration (R_g) tensor, $R_{g_{xx}}$, $R_{g_{yy}}$, $R_{g_{zz}}$, respectively. The radius of gyration R_g of the micelle was calculated as – R_g=$\sqrt{\Sigma_{(i=1)}^{N}(R_i - R_{COM})^2/N}$, where, N is the total number of atoms in the micelle, R_{COM} is the centre of mass (COM) of that micelle, and R_i is the position of the i^{th} atom in the micelle. We calculated the order parameter to understand the orientation of the molecules within the micelle on the gold surface.

2.4.4 Potential of Mean Force

The Potential of Mean Force (U_{PMF}) provides a free energy landscape for a process along its reaction coordinate, offering insights into the directionality and energetics of the process. PMF captures the energy landscape of molecular binding, surface interactions, and structural transitions. A negative U_{PMF} indicates that the process can occur spontaneously. In contrast, a positive U_{PMF} implies that the process is energetically unfavourable and requires an additional input of energy or work to proceed. The U_{PMF} profile can quantify this energy requirement. Several methods can be employed to derive U_{PMF} from molecular dynamics simulations, including umbrella sampling [108] and

steered molecular dynamics [109], offering valuable insights for material design, drug delivery, and nanotechnology. Moreover, the Potential of Mean Force can be directly related to properties such as the radial distribution function (RDF) [110], density profiles [111] depending on the degree of correlation between isotropic particles as ΔU_{PMF} = - k_B T ln[g_{ij}(r)], where, k_B is Boltzmann's constant, and T is the absolute temperature of the system.

2.4.5 Steered Molecular Dynamics

Steered Molecular Dynamics (SMD) is a computational technique used to simulate and explore the energetics and dynamics of molecular systems when subjected to an external force. It is particularly useful for studying processes such as ligand unbinding, protein folding and unfolding, and conformational transitions that may not occur within the typical timescale of unbiased molecular dynamics (MD) simulations. SMD is widely applied to find out U_{PMF} by inducing an external force between two different groups of atoms/molecules to find out the free energy profile of their interactions by dragging one group towards another. One can express the force acting in terms of spring constant, k and relative displacement as – $F = \frac{-k(z-z_0)m}{M}$, where, z_0 represents the equilibrium distance between two groups, m denotes the mass of a single atom or particle, and M is the total mass of the group being dragged. As the position of the group is changed while being dragged toward another, the corresponding U_{PMF} profile is generated. If at any position z, the U_{PMF} value is negative, it indicates that the process or the dragged group is energetically stable at that position. Conversely, suppose U_{PMF} is positive at any position z. In that case, one needs to supply some external work/energy, indicated from U_{PMF} profile, to drag that particular group to that position and beyond.

2.4.6 Umbrella Sampling

Umbrella Sampling (US) is an enhanced sampling technique used in molecular dynamics simulations to calculate free energy profiles along a reaction coordinate. It is particularly effective for overcoming energy barriers and sampling rare events inaccessible in stan-

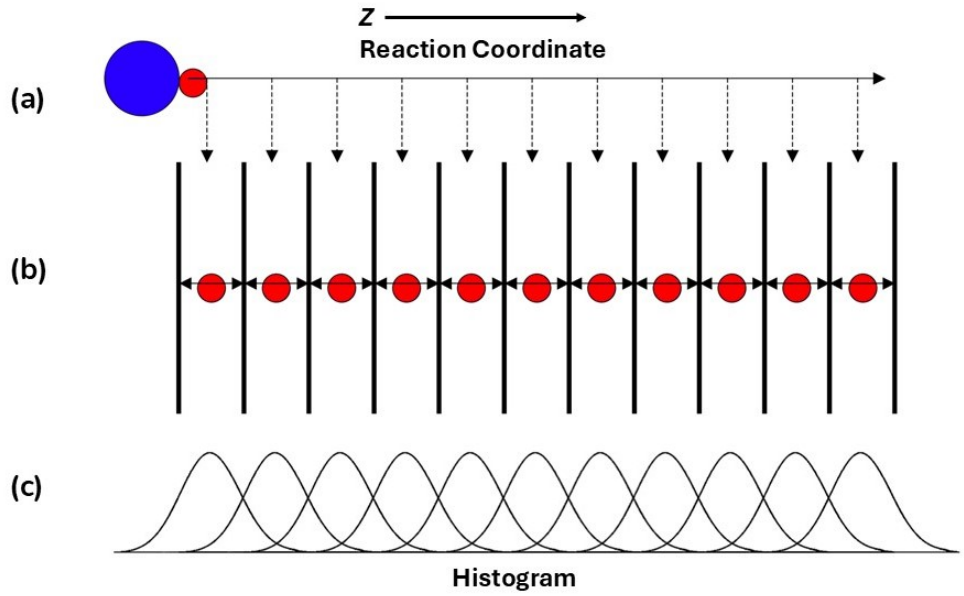

Figure 2.5: Umbrella Sampling (a) Pulling Simulation, (b) Simulation Window, (c) Histogram. (Image is taken from GROMACS Tutorial[112])

dard simulations. To perform umbrella sampling, a series of configurations must first be generated along a reaction coordinate, z. These configurations serve as the initial configurations for the Umbrella Sampling windows, each of which is run as an independent simulation.

This process is illustrated as follows: Figure 2.5(a) depicts the pulling simulation, which is conducted to generate a series of configurations along the reaction coordinate. After the pulling simulation is completed, specific configurations are extracted (represented by dashed arrows connecting between (a) and (b)). To eliminate the biasing potential and reconstruct the potential of mean force (PMF), we applied the Weighted Histogram Analysis Method (WHAM).[113]. Figure 2.5(b) shows the independent simulations performed within each sampling window. In WHAM, a set of linearly spaced windows along the reaction coordinate, the distance between the centre of mass of the nucleate and the apex of the gold surface, was subjected to an external biasing potential to accomplish extensive sampling in these windows. A harmonic potential, $u_i = k_i(r - r_i)^2$ was used as the external biasing potential at window i, where r_i is the reference location and k_i is the harmonic force constant. In WHAM, the unbiased probability distribution is obtained by iteratively

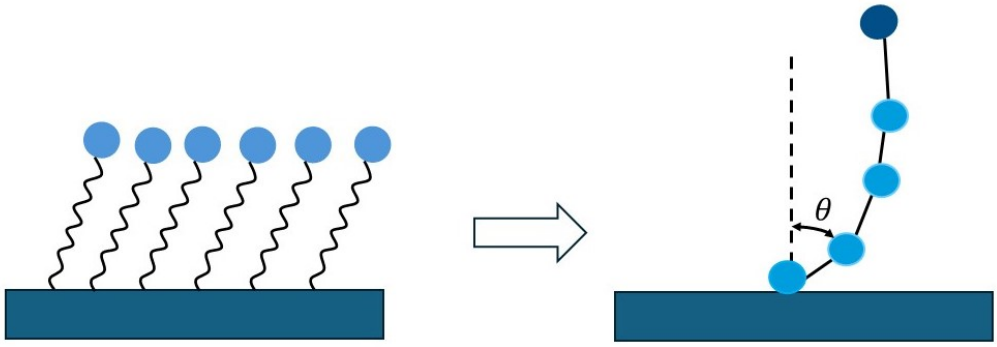

Figure 2.6: The tilt angle of self-assembled molecules on a solid surface.

solving a set of equations that balance the contributions from all sampled windows. Each histogram is weighted based on its sampling frequency and the biasing potential used in that particular window. The algorithm ensures that the reconstructed probability distribution is continuous and consistent across the reaction coordinate, provided there is sufficient overlap between adjacent histograms, as shown in Figure 2.5(c).

2.4.7 Order Parameter

The order parameter of the surfactant molecules is a measure of the degree of ordering and alignment of CTAB molecules within a micelle structure. The order parameter is expressed as $S_z = (3cos^2\theta_i - 1)/2$, where, θ_i is the angle between the end-to-end vector of the ith molecule and the surface normal vector, Z. The tilt angle is calculated by considering the end-to-end distance [36] between the sky blue atom near the surface and the blue atom shown in Figure 2.6 and the director with respect to which the molecule is tilted. Its value ranges from 0 to 1, where S_z = 0 corresponds to a completely disordered system with random molecular orientations, and S_z = 1 represents perfect alignment of molecules along the reference axis.

In the subsequent chapters, detailed methodology of creating the initial configurations and simulation techniques are discussed, along with results and discussions on the micelle-mediated nucleation growth of gold nanocrystals on the surface and the effects of co-surfactants such as oleylamine and hexadecanethiols.