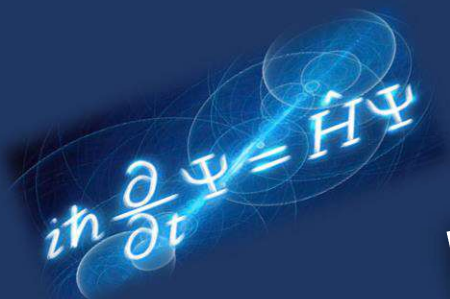
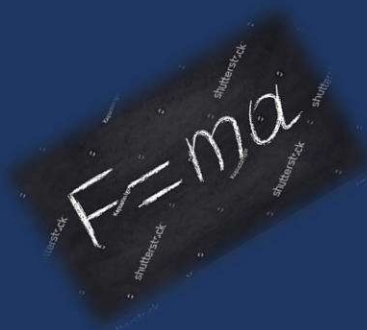


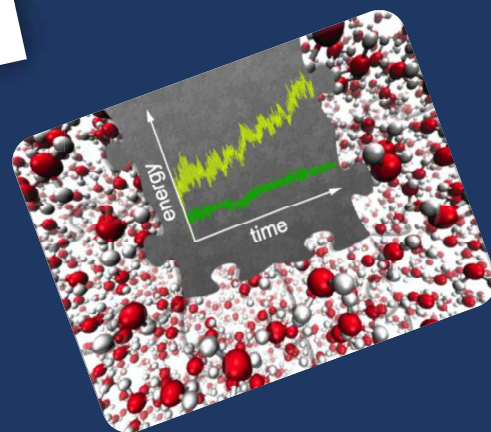
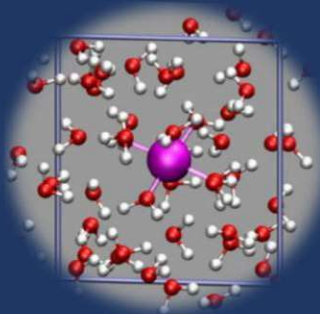
# Chapter 2

## Basic Framework of Density Functional Theory and Classical Molecular Dynamics Simulation

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi$$
A glowing blue particle with a trail of light is shown next to the Schrödinger equation.



$$F = ma$$
The equation F = ma is written on a dark, chalkboard-like surface with some faint, illegible text in the background.



Numerous scientific and technological fields, including biosensing, drug delivery, and nanomedicine, rely on the complex interaction between biomolecules with carbon nanoparticles. Understanding the underlying molecular interactions that govern the binding process is crucial in the quest to build efficient biomolecular systems. While experimental procedures can provide valuable insights, they frequently necessitate significant time and financial investment. Computational approaches provide a powerful and cost-effective way to investigate these interactions at the atomic level.

In this chapter, we present a full explanation of the computational methodologies used in this thesis. The planned computational approach is combining Gaussian 09[1] based Density Functional Theory (DFT) calculations and GROMACS[2] based Classical Molecular Dynamics (MD) simulations. This intersection provides a way to access the system's dynamic properties. It enables us to investigate time-dependent perturbation reactions and dig into the world of rheological characteristics. These include a variety of features such as electronic complexities, structural dynamics, vibrational characteristics, and key transport coefficients. This chapter thus serves as a bridge to a complete understanding, containing the approaches used to unravel the molecular complexities controlling complicated interactions inside the atomic domain.

## 2.1. Theory of Electronic Structure

### 2.1.1. Quantum Mechanics, Schrödinger Equation

The many-body wave function  $\Psi(\{r_i\}, \{R_A\})$  is used to model a composite system with  $N$  interacting electrons and  $M$  interacting ions. There is a wave function  $\Psi$  that depends on both the electronic coordinates  $\{r_i\}$  and the nuclear coordinates  $\{R_A\}$ , and is the solution to the Schrödinger equation[3] that is time-independent. Its non-relativistic form can be expressed as

$$\hat{H}\psi = E\psi \text{ ---- (2.1)}$$

$\hat{H}$ : Hamiltonian operator

$E$ : total energy of the system

The Hamiltonian is divided into five parts:

$$\hat{H} = \hat{T}_e + \hat{T}_A + \hat{V}_{ae} + \hat{V}_{ee} + \hat{V}_{AA} \text{ ---- (2.2)}$$

Where,

$$\widehat{T}_e = -\sum_{i=1}^N \frac{\hbar^2 \nabla_i^2}{2m_e}: \text{kinetic energy of electrons}$$

$$\widehat{T}_A = -\sum_{A=1}^M \frac{\hbar^2 \nabla_A^2}{2M_A}: \text{kinetic energy of ions}$$

$$\widehat{V}_{Ae} = -\sum_{i=1}^N \sum_{A=1}^M \frac{e^2 Z_A}{|r_i - R_A|}: \text{attractive energy between electrons and ions}$$

$$\widehat{V}_{ee} = \sum_{i=1}^N \sum_{j>1}^N \frac{e^2}{|r_i - r_j|}: \text{repulsive energy between electrons}$$

$$\widehat{V}_{AA} = \sum_{A=1}^M \sum_{B>A}^M \frac{e^2 Z_A Z_B}{|R_A - R_B|}: \text{repulsive energy between ions}$$

Now, the Schrodinger equation becomes,

$$\left[ -\sum_{i=1}^N \frac{\hbar^2 \nabla_i^2}{2m_e} - \sum_{A=1}^M \frac{\hbar^2 \nabla_A^2}{2M_A} - \sum_{i=1}^N \sum_{A=1}^M \frac{e^2 Z_A}{|r_i - R_A|} + \sum_{i=1}^N \sum_{j>1}^N \frac{e^2}{|r_i - r_j|} + \sum_{A=1}^M \sum_{B>A}^M \frac{e^2 Z_A Z_B}{|R_A - R_B|} \right] \psi = E\psi \text{ ---- (2.3)}$$

### 2.1.2. Born-Oppenheimer Approximation

Since equation (2.3) is difficult to solve, it becomes necessary to introduce some kind of approximation to solve it. The first one is called ‘‘Born-Oppenheimer approximation[4]’’ (BO approximation). It relies upon the fact that the nucleons are about 1800 times heavier than the electrons. However, the kinetic energy of nucleons would be much lesser as compared to that of the electrons. Therefore, the BO approximation neglects the kinetic energy term of nucleons, and allows us to dissociate the kinetic energies of electrons and ions. This implies that the nuclei are frozen in such configuration, which reduces the difficulty of solving the electron dynamics. The wave function then becomes:

$$\psi(\{r_i\}, \{R_A\}) = \psi(\{r_i\}) \times \psi(\{R_A\}) \text{ ---- (2.4a)}$$

$$H_e \psi(\{r_i\}) = E_e \psi(\{r_i\}) \text{ ---- (2.4b)}$$

$$\text{Where, } H = -\sum_{i=1}^N \frac{\hbar^2 \nabla_i^2}{2m_e} - \sum_{i=1}^N \sum_{A=1}^M \frac{e^2 Z_A}{|r_i - R_A|} + \sum_{i=1}^N \sum_{j>1}^N \frac{e^2}{|r_i - r_j|} + \sum_{A=1}^M \sum_{B>A}^M \frac{e^2 Z_A Z_B}{|R_A - R_B|}.$$

However, there are more simplifications needed as equation set (2.4) is difficult to solve.

The **Slater Determinant** [5] is introduced to for the free electrons of the system which satisfies the Pauli’s exclusion principle. The wave function  $\psi$  is not an observable anymore and

$|\psi|^2 dx$  represents the probability of finding an electron in a given point in space  $d\mathbf{x}_i$  ( $i=1, \dots, N$ ). The electron exchange will only invert the sign as electrons are identical:

$$\psi(\mathbf{x}_i) = -\psi(\mathbf{x}_i) \text{ ---- (2.5)}$$

However, the probability remains the same:

$$|\psi(\mathbf{x}_i)|^2 = |\psi(\mathbf{x}_i)|^2 \text{ ---- (2.6)}$$

A set of wave functions for a collection of electrons can be written as  $X_i(\mathbf{x}_i)$ . The Slater determinant for N-electron system is given by:

$$\phi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} X_1(\mathbf{x}_1) & \cdots & X_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ X_i(\mathbf{x}_N) & \cdots & X_N(\mathbf{x}_N) \end{vmatrix} \text{ ---- (2.7)}$$

### 2.1.3. Hartree-Fock Method

Whenever the N-body wave function and determination of energy of a system becomes complicated to solve, the ‘‘Hartree-Fock (HF) method’’ is introduced[6]. The Fock Hamiltonian is given by:

$$f_i = -\frac{1}{2}\nabla_i^2 - \sum_k^{core} \frac{Z_k}{r_{ik}} + V_i^{HF} \text{ ---- (2.8)}$$

$V_i^{HF}$  is the HF potential and it represents the average repulsive potential of one electron to the other  $N-1$  electrons. It is given by:

$$V_i^{HF}(\mathbf{x}_1) = \sum_i^N (\hat{J}_i(\mathbf{x}_1) - \hat{K}_i(\mathbf{x}_1)) \text{ ---- (2.9)}$$

Where,  $\mathbf{x}_1$  is the position of single electron,  $\hat{J}_i$  is the Coulomb operator, and  $\hat{K}_i$  is the exchange contribution to the HF potential, it is defined when it is operating on spin orbital.

$\hat{J}_i$  and  $\hat{K}_i$  are defined as:

$$\hat{J}_i(\mathbf{x}_1) = \int \left| X_j(\mathbf{x}_2) \right|^2 \left( \frac{1}{r_{12}} \right) d\mathbf{x}_2 \text{ ---- (2.10a)}$$

$$\hat{K}_j(\mathbf{x}_1) X_i(\mathbf{x}_1) = \int \left( X_j^*(\mathbf{x}_2) \right)^2 \left( \frac{1}{r_{12}} \right) (X_i(\mathbf{x}_2)) d\mathbf{x}_2 X_i(\mathbf{x}_1) \text{ ---- (2.10b)}$$

This makes the HF potential  $V_i^{HF}$  non-local and dependent on initial unknown spin orbitals. A Hartree-Fock Self-Consistent Field (HF-SCF) technique is utilised to determine the  $V_i^{HF}$ . A new  $V_i^{HF}$  is produced using an initial set of spin orbitals ( $X_i$ ), which are



mathematically conjectured. Then, new spin orbitals are calculated using this HF potential. The cycle is continued until the outcome satisfies the convergence criteria. The HF energy is more than the total energy because the HF approach does not take into account the dynamic correlation between electronic movements.

### 2.1.4. Density Functional Theory

The total density of electrons is used as a fundamental parameter in Density Functional Theory (DFT), rather than the electrons' N-body wave functions. A physical property is the electron density. It is simpler to calculate as the number of electrons rises. More significantly, this approach gets beyond the Hartree-Fock method's main drawback of ignoring electron correlation. As a result, DFT considerably increases computation accuracy, whereas the exchange part in HF method is well defined.

The first fundamental theorem of DFT was proposed by Hohenberg and Kohn[7], and their key premise was that "the external potential  $V_{ext}$  applied on the system is an external potential to the system which is due to the presence of the nuclei is a unique functional of the electronic density  $\rho(\mathbf{r})$ ."

"The functional  $F_H[\rho]$  that provides the ground state energy of the system delivers the lowest energy if and only if the input density corresponds to the actual ground state density", as per Hohenberg and Kohn's second fundamental theorem of DFT.

### 2.1.5. Kohn-Sham Equations

In 1965, Walter Kohn and Lu Jeu Sham came up with a method [8] to solve the Schrodinger equation. Considering they hypothesised that the electronic ground states remained the same, they replaced the interactive system with a non-interactive one. The density  $\rho(\mathbf{r})$  can be used to explain all of the contributions to the total energy. The energy is at its lowest and corresponds to the ground state only when the density is at its fundamental level,  $\rho(\mathbf{r}) = \rho_0(\mathbf{r})$ . The electron clouds' overall energy can be expressed as:

$$E[\rho] = T[\rho] + V_{ee}[\rho] + E_{XC}[\rho] + \int V_{ext}(\mathbf{r})\rho(\mathbf{r})d^3\mathbf{r} \text{ ---- (2.11)}$$

Where,

$T[\rho]$ : kinetic energy of electrons without interactions

$V_{ee}[\rho]$ : classical Coulombic interaction term

$E_{XC}[\rho]$ : exchange correlation term

$\int V_{ext}(\mathbf{r})\rho(\mathbf{r})d^3r$ : interaction between the electrons and the nuclei's external potential

As currently understood, the Density Functional Theory is exact since it is devoid of approximations.  $E[\rho]$ , the exchange-correlation term, is the only undetermined term. It must be approximated, and the next section describes various approximations.

## 2.2. Exchange-Correlation Functionals

The exchange and correlation term,  $E[\rho]$  is determined using a number of methods, including the Generalised Gradient Approximation (GGA) and the Local Density Approximation (LDA).

### 2.2.1. Local Density Approximation (LDA)

The LDA approach [9] suggests that the charge density ranges gradually on an atomic scale and that inhomogeneous electronic systems can be viewed as locally homogeneous by treating the electrons as a uniform electron gas. The exchange-correlation energy for a spin-unpolarized system can be expressed as:

$$E_{XC}^{LDA}[\rho] = \int \rho(\mathbf{r})\epsilon_{XC}(\rho)d\mathbf{r} \text{ ---- (2.12)}$$

As one of the instances of the LDA functional, Vosko, Wilk, and Nusair (VWN) produced a Monte Carlo computation for a uniform electron gas. The local spin density approximation (LSDA) for the spin-polarized system can be used to express the exchange-correlation energy as follows:

$$E_{XC}^{LSDA}[\rho, \rho_B] = \int \rho(\mathbf{r})\epsilon_{XC}(\rho, \rho_B)d\mathbf{r} \text{ ---- (2.13)}$$

Where,  $\rho, \rho_B$  are electronic densities. For computing molecular attributes like structures, vibrational frequencies, charge and dipole moments, etc., equation (2.13) offers an acceptable degree of precision. However, it turns out to be a poor approximation for energetic properties, such as binding energies, energy barriers, etc., when the density varies quickly.

### 2.2.2. Generalized Gradient Approximation (GGA)

The Generalized Gradient Approximation (GGA)[10] has been suggested as a solution to the prior issue. By treating the system as a non-uniform electron gas and accounting for non-local electron effects in the functional, it enhances the LDA technique. As a function of  $\rho$  and  $\nabla\rho$ , the exchange-correlation term can be expressed as follows:

$$E_{XC}^{GGA}[\rho] = \int \varepsilon_{XC}(\mathbf{r})[\rho(\mathbf{r}) + \nabla\rho(\mathbf{r})]d^3\mathbf{r} \text{ ---- (2.13)}$$

Different approximations for the aforementioned equation can be used to construct the GGA family members, which includes BLYP [11], PW91 [12], PBE[10], and other functionals. GGA typically produces better results than LDA in terms of total energies, atomization energies, energy barriers, and structural energy differences since it allows for a broader range of variation. A better gradient-corrected GGA functional is the PBE functional (the exchange-correlation functional developed by Perdew, Burke, and Ernzerhof). All of the parameters in its straightforward derivation are fundamental constants. This functional, which is used in the current work, is most frequently used to both liquid molecules and solids.

### 2.2.3. Meta-GGA

With additional dependency on the Kohn-Sham kinetic energy density, the meta-GGA functionals[13] slightly improve the GGA functionals:

$$\tau(\mathbf{r}) = \sum_i \frac{1}{2} |\nabla\varphi_i(\mathbf{r})|^2 \text{ ---- (2.14a)}$$

$$E_{XC}^{meta-GGA}[\rho] = \int \varepsilon_{XC}(\mathbf{r})[\rho(\mathbf{r}) + \nabla\rho(\mathbf{r})]d^3\mathbf{r} \text{ ---- (2.24b)}$$

One of the most popular meta-GGA functionals is the TPSS functional, which was created by Tao, Perdew, Staroverov, and Scuseria.

### 2.2.4. Hybrid Functionals

The exchange terms of the previously mentioned functionals are poorly defined because of the issue of electronic self-interaction, but the exchange part in HF is precisely defined[14]. However, the results are unsatisfactory when the exchange part from HF and the correlation part from DFT are combined directly. This issue would be avoided by hybrid functionals, which incorporate DFT and HF and group the exchange and correlation

components of each approaches. For instance, the B3LYP hybrid functional[15] is the most popular hybrid functional. It is spelt out as:

$$E_{XC}^{B3LYP} = E_X^{LDA} + a_0(E_X^{HF} - e_X^{LDA}) + a_X(E_X^{GGA} - E_X^{LDA}) + E_C^{LDA} + a_C(E_C^{GGA} - E_C^{LDA}) \quad (2.25)$$

Where  $E_X^{GGA}$  and  $E_C^{GGA}$  are the GGA exchange and correlation functionals,  $E_C^{LDA}$  is the LDA to the correlation functional,  $a_0$  is set to 0.20,  $a_X$  is set to 0.72,  $a_C$  is set to 0.81.

In this thesis, we have used the B3LYP hybrid to replicate the structural geometries of C<sub>24</sub>, C<sub>36</sub>, C<sub>50</sub> and C<sub>70</sub> fullerenes.

### 2.3. Basis Sets

Atom-localized basis functions can be used to generate atomic or molecule orbitals. As a linear combination of these functions with certain coefficients, the orbitals are enlarged. They can be divided into two categories: Slater- and Gaussian-type orbitals (STOs and GTOs)[16].

$$\eta^{STO} = N r^{n-1} e^{\zeta r} Y_l(\theta, \varphi) \quad (2.26)$$

Where N is a normalization coefficient,  $\zeta$  is the exponent,  $r$ ,  $\theta$ ,  $\varphi$  are the spherical coordinates,  $Y_{lm}$  represents the angular momentum component, n symbolises the principal quantum number,  $l$  defines the angular quantum number, and  $m$  represents the magnetic quantum number.

They can be expressed in the following way for the GTOs:

$$\eta^{GTO} = N x^l y^m z^n e^{-r^2} \quad (2.26)$$

$N$  denotes the normalisation factor, while  $x$ ,  $y$ , and  $z$  are Cartesian coordinates.

In general, the STOs have a spike at  $r=0$  and a nice exponential decay for greater values of  $r$ . These characteristics aid in accurately describing the hydrogen atomic orbitals. In contrast, GTOs do not have a peak at  $r=0$  and decay fast as  $r$  increases. The product of two GTOs, on the other hand, is a third one located between them, which is not the case for STOs. This trait eliminates the computationally expensive challenge of dealing with four-center-two-electron integrals. Combining numerous GTOs to simulate an STO is frequently more efficient than employing STOs directly.

The degree of complexity, or precision, of a basis set is defined as the number of contracted functions (CGF) necessary to represent each atomic orbital[17]. The STO-3G basis

set, for example, is constructed by the linear combination of three CGF to represent a STO. Typically, double-zeta and triple-zeta basis sets provide adequate precision and system insight. The valence electrons are of great importance since they are the most involved in chemical reactions. Hence, a flexible description of valence electrons is required. To achieve this goal, split valence basis sets are developed to treat the core and valence orbitals differently. The most widely used split valence basis set is the 6-31G basis set[18], which follows the nomenclature X-YZG. X is the number of primitive GTOs employed to describe one single contracted Gaussian function of the core. Y and Z are the number of primitive GTOs employed to describe the valence orbitals. By increasing the angular momentum, polarisation and diffuse functions can be introduced to the basis sets. For example, the 6-31G(d,p) basis set combines p and d polarisation functions. The diffuse functions are added to the 6-31+G basis set.

The GTO valence basis set is coupled with pseudopotentials in the SVP (split valence polarisation) type basis set[19]. The pseudopotentials will be discussed in the following section. This basis set includes a single basis function to describe the inner shell atomic orbitals, two basis functions to describe each valence shell atomic orbital, and a collection of polarisation functions. The basis set namely 6-31G (d,p) is mainly used in the present thesis.

#### 2.4. Time Dependent Density Functional Theory (TD-DFT)

The DFT has been simplified to include time-dependent potentials and is now known as time-dependent density functional theory (TD-DFT). In the context of the time-dependent electron density  $\eta(\mathbf{r},t)$ , the TD-DFT is considered time-dependent quantum mechanics[20]. In 1984, Runge and Gross proposed the Runge-Gross theorem[21], which is related to the Hohenberg-Kohn theorem for time-dependent electron density. TD-DFT, unlike DFT, does not have a minimum energy concept. Rather, it is based on the notion of stationary action. It can offer an imaginary system of non-interacting electrons with the same density  $\eta(\mathbf{r},t)$  as a real system that hops in time-dependent effective potential  $v_{eff}(\mathbf{r},t)$  as DFT. This yields the time-dependent Kohn-Sham equation, which is identical to the time-dependent Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} \psi_i(\mathbf{r}, t) = \left[ \frac{-\hbar^2}{2m_e} + v_{eff}(\mathbf{r}, t) \right] \psi_i(\mathbf{r}, t) \text{ ---- (2.27)}$$

Here,  $v_{eff}(\mathbf{r}, t) = v_{ext}(\mathbf{r}, t) + v_{XC}(\mathbf{r}, t) + e^2 \int k \frac{\eta(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}'$ , where  $v_{XC}(\mathbf{r}, t)$  indicates the time-dependent exchange-correlation potential and the unknown term has to be determined. The density therefore is given by:

$$\eta(\mathbf{r}, t) = \sum_i |\psi_i(\mathbf{r}, t)|^2 \text{---- (2.28)}$$

The adiabatic local density approximation [22] (ALDA), which uses uniform gas with the instantaneous density, is the simplest approximation for  $v_{XC}(\mathbf{r}, t)$ .

$$v_{XC}^{ALDA}(\mathbf{r}, t) = v_{XC}^{unif}[\eta(\mathbf{r}, t)] \text{---- (2.29)}$$

Equation (2.29) calculates the excited state energies as well as the photo-absorption cross-section of molecules and clusters.

## 2.5. Properties from DFT for the Present Thesis

### 2.5.1. Vibrational Spectra Calculation

The optimization of molecular structural geometry aids in the prediction of IR and Raman spectra [23]. The frequency computation, which is independent of experimental investigation, provides the position and intensity of vibration of bands. The harmonic model is used to calculate the frequencies [24]. In practice, however, they are anharmonic. As a result, this describes the differences between experimental and calculated frequencies. The total energy of a molecule with N atoms adjacent to its equilibrium structure can be calculated as follows:

$$E = T_k + V_p = \frac{1}{2} \sum_{i=1}^{3N} q_i^2 + V_{eq} + \sum_{i=1}^{3N} \sum_{j=1}^{3N} \left( \frac{\partial^2 V}{\partial q_i \partial q_j} \right)_{eq} q_i q_j \text{---- (2.30)}$$

Where,  $q_i$  represents the mass-weighted Cartesian displacements and written as:

$$q_i = M_i^{\frac{1}{2}} (X_i - X_{ieq}) \text{---- (2.31)}$$



Where  $X_i$  is the position of the nuclei in relation to their equilibrium locations,  $M_i$  and  $X_{ieq}$  are the masses of the nuclei. The potential energy at the equilibrium nuclear conformation is denoted by  $V_{eq}$  in equation (2.30), and the extension of a power series is succinct at the second order. Therefore, the equation for classical mechanics can be expressed as:

$$Q_i = \sum_{i=1}^{3N} f_{ij} q_i \text{ ---- (2.32)}$$

Here,  $f_{ij}$  stands for the quadratic force constants calculated through:  $f_{ij} = \left( \frac{\partial^2 V}{\partial q_i \partial q_j} \right)_{eq}$

It can be calculated through numerical second differentiation:  $\frac{\partial^2 V}{\partial q_i \partial q_j} = \frac{\Delta(\Delta V)}{\Delta q_i \Delta q_j}$

The current thesis investigates Raman intensities as well as their vibrational modes in relation to potential energy distribution. The Raman intensities were computed in Gaussian 09 using the obtained Raman activities ( $S_i$ ) and the intensity theory of Raman scattering [25] relationship:

$$I_i = \left[ \frac{f(v_0 - v_i)^4 S_i}{v_i \{1 - \exp(-\frac{hcv_i}{kT})\}} \right] \text{ ---- (2.33)}$$

In equation (2.33),  $v_0$  represents the exciting wavenumbers in  $cm^{-1}$ , and  $v_i$  represents the vibrational wavenumber of the  $i^{th}$  normal mode. The universal constants  $h$ ,  $c$ , and  $k$  are used, while  $f$  is a suitable common normalisation factor for all peak intensities.

### 2.5.2. Frontier Molecular Orbitals (FMO)

The development of a new set of orbitals comes from the overlap of two orbitals [26]. If the superimposition is overlooked, the difference in energy before and after the overlap is given by the second-order perturbation expression.

$$E = 2 \sum_{l=1}^{n/2} \lambda_l \text{ ---- (2.34a)}$$

$$E = 2 \sum_{i=1}^{(n-1)/2} \lambda_i + \lambda_{(n+1)/2} \text{ ---- (2.34b)}$$

If  $n$  is an even number, the  $(n/2)$  eigenvector indicates the highest (doubly) occupied molecular orbital (the HOMO), with an energy of  $\lambda_{n/2}$ . The following eigenvector is  $(\lambda_{n/2}+1)$  for the lowest unoccupied molecular orbital (the LUMO), whose energy is. The HOMO-LUMO separation is then stated as:

$$\Delta HL = \frac{\lambda_n}{2} - \frac{\lambda_n}{2} + 1 \text{ ---- (2.35)}$$

### 2.5.3. Mulliken Charge Transfer Analysis

Mulliken charge transfer analysis is critical for understanding the mechanisms of binding or interaction in a variety of chemical and physical processes [27]. It sheds light on the nature of bonding, electron transfer, and the overall stability of the system by providing vital insights about the redistribution of electron density during these interactions. This is especially relevant when studying noncovalent interactions like hydrogen bonding or van der Waals interactions. Researchers can predict the reactivity of molecules and get insights into reaction mechanisms by analysing the degree of charge transfer. Large charge transfer values frequently suggest important interactions and may assist to project reaction pathways. The change in charge ( $\Delta Q$ ) for an atom  $i$  due to charge transfer can be expressed as:

$$\Delta Q_i = q_i(\text{final}) - q_i(\text{initial}) \text{ ---- (2.36)}$$

Where,  $q_i(\text{final})$  is the final Mulliken charge on atom  $i$  after the interaction;  $q_i(\text{initial})$  is the initial Mulliken charge on atom  $i$  before the interaction. Positive  $\Delta Q$  values imply electron flow from atom  $i$  to other atoms, whereas negative values suggest electron transport towards atom  $i$ . Large positive or negative values imply considerable charge transfer, which can indicate chemical bond formation or destruction.

### 2.5.4. Natural Bond Orbital (NBO) Analysis

Natural Bond Orbital (NBO) analysis is a dependable and strong approach for investigating conjugative or charge transfer interactions. It provides a useful foundation for understanding intra and intermolecular bonding in molecules. The intensity of the interaction between electron donors and acceptors is proportional to the value of the stabilization energy. The effect of second order micro disable theory of the Fock matrix[28] on the nanomaterials

(as mentioned in the title) has been studied, taking into account all conceivable interactions. Delocalization of electron density between occupied Lewis-type NBOs and formally vacant non-Lewis type orbitals corresponds to a stabilizing Donor accepted interaction and demonstrates  $\pi$ -resonance. It has been established that non-covalent interactions play an important role in the structural stability of biomolecules. NBO further validates the interaction of hyperconjugative charge transfer from filled lone-pairs into the empty anti-bond. In the case of a hydrogen-bonded system. The stabilization energy ( $E^2$ ) associated with electron delocalization between donor and acceptor is given for each filled donor NBO ( $i$ ) and empty acceptor NBO ( $j$ ):

$$E^2 = -n_{\sigma} \left[ \frac{\langle \sigma | F | \sigma \rangle^2}{\varepsilon_{\sigma}^* - \varepsilon_{\sigma}} \right] = -n_{\sigma} \left[ \frac{F_{ij}^2}{\Delta E} \right] \text{---- (2.37)}$$

Where,  $n_{\sigma}$  is the population of donor NBO,  $\langle \sigma | F | \sigma \rangle^2$  is the Fock matrix element between  $i$  and  $j$  orbitals,  $\varepsilon_{\sigma}^*$  and  $\varepsilon_{\sigma}$  are the energies of  $\sigma$  and  $\sigma^*$  NBOs.

### 2.5.5. Molecular Electrostatic Potential (MESP)

The molecular electrostatic potential is a very important parameter for explaining charge distribution and thus charge-related phenomena. It is useful in predicting the reactivity of hydrogen bonds and the relative polarity of molecules [29]. The charge distribution around the molecule in space helps visualise the reactive zone for electrophilic and nucleophilic attack. The MESP at a specific point in space around the molecule is expressed by:

$$V(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}') d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \text{---- (2.38)}$$

Where,  $Z_A$  is the charge on the nucleus A, located at  $\mathbf{R}_A$ , and  $\rho$  is the molecular electronic density function. The first term on the right hand side of the equation (2.35) indicates nuclei effects, whereas the second term represents electron effects. The mapping of an electron density iso-surface with a molecule's electrostatic potential surface illustrates the size, shape, charge density, and location of chemical reactivity (active site) of the molecules and offers the molecular electrostatic potential surface. The red colour symbolises low potential areas (excess of electrons), the blue colour represents high potential areas (void of electrons), and the green colour represents the neutral region (zero potential).

### 2.5.6. Global Reactivity Parameters

In chemistry, reactivity is a significant concept that functions as a stimulus for chemical reactions and increases the physiochemical properties of molecules. Electrophilicity and nucleophilicity, which affect the reactivity and stability of molecules, are predicted using conceptual quantum mechanical descriptors or global reactivity parameters [30]. While Fukui functions and other local reactivity factors are employed to grasp active sites suitable for electrophilic and nucleophilic attacks in the molecular system [31].

The **chemical potential ( $\mu$ )** of a molecule is defined as the amount of energy (U) received or released when the number of particles in a specific molecule varies at a constant external potential (V). It can also be expressed as the inverse of electronegativity  $\chi$ . It is defined as follows:

$$\mu = \left( \frac{\partial U}{\partial N} \right)_V = -\chi \text{ ---- (2.39)}$$

Parr and Pearson postulated that chemical potential offers resistance to deform or modify the quantity of electrons during the charge transfer process. The **global hardness** can be defined as the second derivative of the energy (E) versus the number of electrons at a fixed external potential  $V(\mathbf{r})$ :

$$\eta = \frac{1}{2} \left( \frac{\partial^2 E}{\partial N^2} \right)_{V(\mathbf{r})} \text{ ---- (2.40)}$$

The global softness (S) is inversely proportional the global hardness.

$$S = \frac{1}{2\eta} \text{ ---- (2.41)}$$

The **electrophilicity index ( $\omega$ )** is a chemical reactivity metric used to understand global reactivity trends. A molecule with a lower  $\omega$  value is a good nucleophile that is more reactive, whereas a molecule with a higher  $\omega$  value is a good electrophile. When the system receives electrons from the environment,  $\omega$  evaluates the donating tendency of the electrons.

$$\omega = \frac{\mu^2}{2\eta} \text{ ---- (2.42)}$$

### 2.5.7. Local Reactivity Parameters (Fukui Functions)

The Fukui function[31] is defined for a molecular system as the derivative of electron density with respect to the number of electrons. Fukui functions are local reactivity parameters that are used to predict the nucleophilic and electrophilic reactivity sides of compounds, and thus their reactivity. A Fukui function is mathematically defined as:

$$f(\mathbf{r}) = \frac{\partial \rho(\mathbf{r})}{\partial N_{electrons}} \text{ ---- (2.43)}$$

The following equations can be used to create three types of condensed Fukui functions for nucleophilic, electrophilic, and radical attacks on an atom  $k$  in a molecule, depending on the type of charge transfer.

$$f_k^+(\mathbf{r}) = [q_k(N + 1) - q_k(N)] \text{ for nucleophilic attack}$$

$$f_k^-(\mathbf{r}) = [q_k(N) - q_k(N - 1)] \text{ for electrophilic attack}$$

$$f_k^0(\mathbf{r}) = \frac{1}{2} [q_k(N + 1) - q_k(N - 1)] \text{ for radical attack}$$

Where  $q_k$  is the total electronic population of the molecular atom  $k$ .

### 2.5.8. Solvent Effects

Solvent effects[32] can play a crucial role in accurately modelling and predicting the properties of molecules and chemical processes that occur in solution in Density Functional Theory (DFT). When researching processes and properties involving molecules interacting with a surrounding solvent environment, it is critical to account for solvent effects. The Polarizable Continuum Model (PCM), developed by Mennucci and Tomasi, is the most extensively used continuum solvation model. It calculates the molecule's solvation energy using the solvent's dielectric constant and refractive index. The PCM represents the solvent environment as a simple continuum with a dielectric constant. Here's a quick rundown of the equations:

**Solvation Energy** ( $\Delta G_{solv}$ ): The interaction between the solute (molecule of interest) and the solvent is accounted for by the solvation energy. The following equation can be used to compute it:

$$\Delta G_{solv} = \Delta G_{sol} - \Delta G_{vac} \text{ ---- (2.44)}$$

Where,  $\Delta G_{solv}$  denotes the solvation energy;  $\Delta G_{sol}$  is the solute's Gibbs free energy in the solvent;  $\Delta G_{vac}$  is the isolated solute's Gibbs free energy in vacuum.

The **Gibbs free energy** change can be related to the enthalpy change ( $\Delta H$ ) and the entropy change ( $\Delta S$ ):

$$\Delta G = \Delta H - T\Delta S \text{ ---- (2.45)}$$

The **enthalpy change** can be approximated as the difference in the electronic energies between the solvated and gas-phase states:

$$\Delta H = E_{solv} - E_{vac} \text{ ---- (2.46)}$$

The Poisson equation can be used to compute the **electrostatic interaction energy** between the solute's electron density and the electric potential owing to the solvent:

$$\Delta E_{el} = \int \rho(\mathbf{r})\varphi(\mathbf{r}) dV \text{ ---- (2.47)}$$

Where,  $\Delta E_{el}$  is the electrostatic interaction energy;  $\rho(\mathbf{r})$  is the solute's electron density;  $\varphi(\mathbf{r})$  is the electric potential due to the solvent;  $dV$  represents the volume integration over the entire system.

**Integral Equation Formalism Polarizable Continuum Model (IFPCM)**, is a PCM extension that adds a more stringent treatment of the solvent response. It describes the influence of the solvent on the solute using integral equations. The solvent reaction field equation is the main equation in IFPCM:

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{dis} \text{ ---- (2.48)}$$

Where,  $\Delta G_{cav}$  shows the contribution from the solute cavity formation;  $\Delta G_{dis}$  represents the contribution from the solute-solvent interaction.

In order to investigate the solubility of the biomolecules listed in the present thesis in a water-based medium, we calculated the solvation energies ( $\Delta E_{solv}$ ) using the following equation:

$$\Delta E_{solv} = E_{sol} - E_{gas} \text{ ---- (2.49)}$$

Where,  $E_{sol}$  and  $E_{gas}$  are the system energies in the water solvent and gas, respectively.



### 2.5.9. Dipole Moments

The dipole moment ( $D$ ) and polarizability ( $\alpha$ ) were calculated using the Gaussian 09 software and the finite field (FF) technique[33]. Buckingham's Cartesian coordinate definitions can offer them:

$$D = (D_x^2 + D_y^2 + D_z^2)^{\frac{1}{2}} \text{ ---- (2.50)}$$

$$\langle \alpha \rangle = \frac{1}{3} (\alpha_{xx} + \alpha_{yy} + \alpha_{zz}) \text{ ---- (2.51)}$$

## 2.6 Molecular Dynamics Simulation (MDS)

An additional sector in the realm of theoretical and computational physics delves into the intricacies of molecular dynamics (MD) simulations. Rather than examining the intricate specifics of the electronic makeup of atoms and molecules along with their usually unchanging characteristics, MD simulations adopt a classical perspective toward the molecular arrangement. These simulations explore the temporal progression of the system by employing classical equations of motion to elucidate its dynamic evolution, or how it changes over time. The classical approximation enables the contemplation of significantly larger systems in contrast to those manageable through conventional quantum chemical techniques, although it typically entails a greater infusion of empiricism into the theoretical framework [34]. Molecular dynamics techniques emulate the motion of individual particles in a deterministic manner, as opposed to Monte Carlo (MC) techniques. MC techniques explore the configuration space through successive probabilistically accepted snapshots generated stochastically. Though both MD and MC strategies offer insights into the current system structure, only MD is suited for delving into the genuine system dynamics [35-36].

In a strictly classical technique, where the behaviour of atomic nuclei and the potential energy of a system are estimated using a parametrized force field, a significant effort is devoted to choosing this force field and figuring out its precise parameters. The necessity to develop complex potential energy expressions to faithfully imitate the underlying physics occurs when situations occur where the electrical constitution of species within the system leads to elaborate forms of interaction[37-39].

Following, we will give a broad introduction of molecular dynamics, including its underlying theories and applications, while discussing the specific potential energy function that was developed and used to support this thesis in the Results section.

### 2.6.1 The Model System:

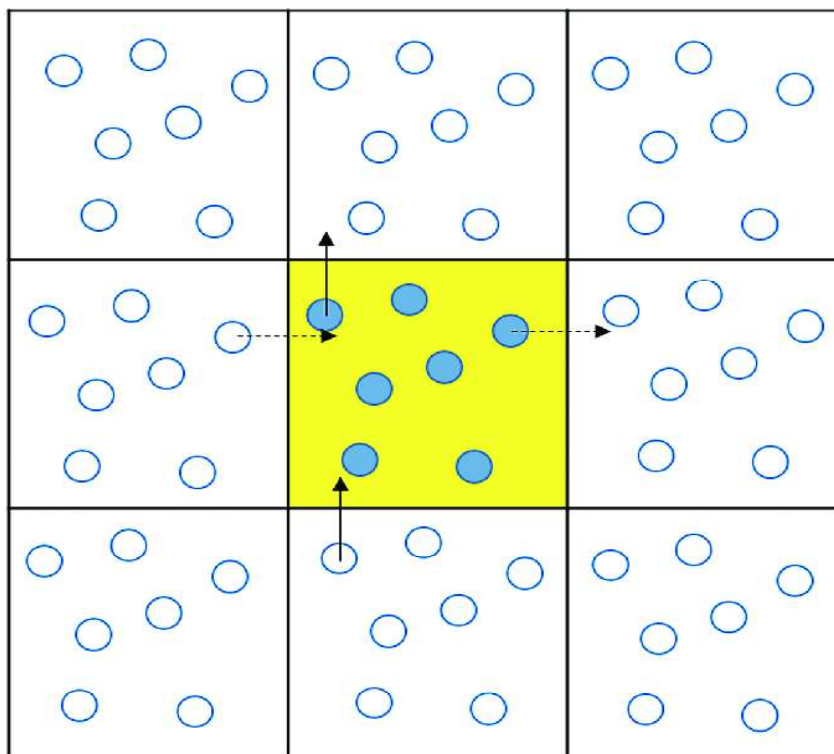
The initial component of an MD simulation involves establishing the model system. Broadly, within the realm of MD, there exist various potential levels of detail that can be designated for the system, contingent upon the intended application. The utmost level of detail is achieved by explicitly designating every atom within the system, resulting in the highest count of degrees of freedom. In contrast, simpler models are also possible, for example, when dealing with particularly large molecules like proteins or biomolecules[40]. In these cases, parts of the system can be viewed as single entities rather than individual atoms, reducing the overall number of degrees of freedom; for example, functional groups or solvent molecules can be combined into a single particle. For the sake of the present, we shall proceed on the assumption that the entire system is made up solely of atoms.

In the realm of conventional atomistic simulations, the clearly defined atoms are contained within molecules, which are treated as indivisible units, prohibiting the occurrence of any chemical reactions throughout the simulation. This makes sure that throughout the computation, the system's variety of molecular species stays the same [41].

The molecules comprising the system are positioned within a simulation box, also known as a cell, which can adopt various geometries, often taking the form of cubes or parallelepipeds. Accompanying the simulation cell are the boundary conditions that define its behaviour. Two fundamental choices exist: PBC and OBC. While OBC may find utility in specific scenarios, the primary objective of conducting MD simulations lies in acquiring the collective characteristics of a molecular system [42]. This aspiration mandates the management of a vast multitude of molecules, a challenge addressed by the implementation of PBC. The simulation box is populated with a finite quantity of molecules and subsequently duplicated throughout the entirety of space, spanning all three Cartesian dimensions. This facilitates the handling of an effectively boundless system, devoid of any boundary-related influences. A visual depiction of a model system within a two-dimensional context, as expounded thus far, is presented in **Figure 2.1** [36], mirroring the principles applicable to the three-dimensional scenario.

Evidently, all attributes of atoms can be arbitrarily assigned, encompassing aspects like mass, charge, polarizabilities, and more. Furthermore, it's also possible to introduce imaginary particles, which might possess attributes like zero mass but non-zero charge [43].

In brief, the model configuration encompasses several molecules, perceived as inseparable entities, each constructed from constituent atoms, each harbouring distinct atomic



**Figure 2.1:** *An illustrative model system set within a two-dimensional square container utilizing periodic boundary conditions. The blue discs denote the particles' positions, while the black arrows depict their velocity vectors. Evidently, the central highlighted box is duplicated in both dimensions.*

characteristic. These molecules find their position within a simulation box, defined with precise dimensions, form, and boundary constraints. Now, we delve into the mechanics of how atoms maintain their cohesion within molecules throughout calculations, alongside an exploration of how these molecular entities interrelate.

### 2.6.2 Classical Force Fields:

The behaviour of the particles within the modelled system is regulated by a traditional force field, dictating their overall configuration energy, denoted as  $V_{\text{tot}}$ . The all-encompassing

expression for this force field takes the shape of the combination of two components, specifically,

$$V_{\text{tot}} = V_b + V_{\text{nb}} \text{----} (2.52)$$

where  $V_b$  and  $V_{\text{nb}}$  represent the bonded and non-bonded interactions, respectively. As every molecule is meticulously specified within an MD simulation, it is crucial to reiterate that any chemical transformations involving bond rupture or formation are precluded [44]. As a result, the designation  $V_b$  represents an intramolecular potential, efficiently encompassing the stored potential energy within each molecular unit and contingent upon the arrangement of atoms. The functional expression of  $V_b$  generally undergoes additional dissection into various components, typically accounting for interactions between pairs, triplets, and quadruplets of atoms [45]. These components characterize the potential energy associated with bonds, bond angles, and dihedral angles, respectively.

The mathematical formulation of the generic expression for  $V_b$  is as follows[46]:

$$V_b(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \sum_{i=1}^{N_{\text{bond}}} V_{\text{bond}}(\mathbf{r}_{\text{ai}}, \mathbf{r}_{\text{bi}}) + \sum_{i=1}^{N_{\text{angle}}} V_{\text{angle}}(\mathbf{r}_{\text{ai}}, \mathbf{r}_{\text{bi}}, \mathbf{r}_{\text{ci}}) + \sum_{i=1}^{N_{\text{dihe}}} V_{\text{dihe}}(\mathbf{r}_{\text{ai}}, \mathbf{r}_{\text{bi}}, \mathbf{r}_{\text{ci}}, \mathbf{r}_{\text{di}}) + \sum_{i=1}^{N_{\text{other}}} V_{\text{other}}(\mathbf{r}_{\text{ai}}, \mathbf{r}_{\text{bi}}, \mathbf{r}_{\text{ci}}, \mathbf{r}_{\text{di}}) \text{----} (2.53)$$

Here, the aggregations of distinct potential types traverse across all existing bonds ( $V_{\text{bond}}$ ), bond angles ( $V_{\text{angle}}$ ), dihedral angles ( $V_{\text{dihe}}$ ), and any other potential avenues for intramolecular structural adaptations ( $V_{\text{other}}$ ). It's important to recognize that the designation  $V_{\text{other}}$  possesses a broad scope, exemplified by its ability to encompass constraints on molecular arrangement that might not neatly fit into a clearly defined chemical conformation category. An illustrative case is the inversion through a point, a phenomenon exhibited by molecular isomers, yet not rigidly categorized as a standard chemical conformation. The arrangement of these expressions is specifically chosen in accordance with the unique chemical entities being analyzed, with the intention of achieving the most faithful representation of the genuine physical interactions. By way of illustration, let's consider the  $V_{\text{bond}}$  term. A frequently employed strategy for characterizing bond energy encompasses the utilization of a harmonic oscillator potential energy function, as illustrated by the equation -

$$V_{\text{bond}}(\mathbf{r}_{\text{ai}}, \mathbf{r}_{\text{bi}}) = \frac{1}{2} k (r_{\text{ab}} - r_e)^2 \text{----} (2.54)$$

where  $r_e$  denotes the equilibrium distance of the bond and  $r_{ab}$  denotes the separation between the two atoms that make up the bond.  $K$ , a harmonic (force) constant, controls the bond's rigidity. The equation labelled as (3) represents just one potential configuration for the expression of the energy in a chemical bond; numerous alternative formulations are available[47]. However, we won't enumerate all the other possibilities here. Instead, we will elucidate this concept through the provided example. In the end, the specific selection of these expressions involves balancing between precision and computational expense. However, it should be acknowledged that on numerous occasions, it could be computationally advantageous, and potentially even physically justified, to constrain certain or all intrinsic molecular degrees of freedom. This step helps in circumventing the necessity for an intramolecular potential. While this simplification generally yields satisfactory outcomes for highly rigid systems, it's advisable to exercise caution when handling exceptionally flexible molecules.

The  $V_{nb}$  term represents the potential energy arising from non-bonded interactions among molecules within a system. Defining this term proves to be particularly challenging, given its variability stemming from the diverse nature of intermolecular interactions. Just as with intramolecular potentials, over time, various functional forms have been suggested, differing in accuracy and the computational resources they demand[48].

In essence, the overarching representation of  $V_{nb}$  can be concisely outlined as follows:

$$V_{nb}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \sum_{i=1}^N \sum_{j>i}^N V_{2-body}(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i=1}^N \sum_{j>i}^N \sum_{k>j}^N V_{3-body}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \sum_{i=1}^N \sum_{j>i}^N \sum_{k>j}^N \sum_{n>k}^N V_{4-body}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k, \mathbf{r}_n) + V_{nonadd}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) + \sum_{i=1}^N V_{ext}(\mathbf{r}_i) \quad (2.55)$$

Here, we observe a transition in atom indices from a, b, c... to i, j, k... This shift is intentional, highlighting that the involved centers exist within distinct molecules. Of significant prominence within Equation (4) is the summation encompassing the two-body factors, responsible for capturing both Van der Waals (VdW) forces and electrostatic Coulomb interactions. While instances might arise where 3- and 4-body potentials come into play, they are frequently omitted due to their computational intensity stemming from additional nested summations. Crafting the potential energy function  $V_{nonadd}$  presents a distinct challenge, for it encompasses interactions of a non-additive nature, including the intricate induction interaction. This term defies a universal structure, its applicability remaining notably restricted. On

occasion, especially when dealing with symmetrically spherical interacting entities, the effects of non-additivity might find integration into the two-body facet of the potential. The final component, denoted as  $V_{\text{ext}}$ , delineates the influence of an extrinsic potential exerting its effect upon the system, akin to the impact of an electric field.

The structure of the Van der Waals potential energy function holds paramount significance, encompassing both a far-reaching attractive segment and a close-range repulsive barrier. Undoubtedly, the most renowned among these potentials is the Lennard-Jones (LJ) potential, characterized by the subsequent mathematical representation[49].

$$V_{LJ}(r_{ij}) = 4\epsilon \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right] \text{ ---- (2.56)}$$

Here,  $\epsilon$  represents the depth of the potential well, while  $\sigma$  signifies the position where the repulsive element intersects the zero-energy threshold. To offer a visual portrayal of the conventional LJ potential, a representative instance featuring  $\epsilon = 4$  and  $\sigma = 3$  is depicted in **Figure 2.2**. Assigning values to the parameters  $\epsilon$  and  $\sigma$  is contingent upon the particular interacting entities, often determined through fitting techniques against experimental observations or from initial ab initio simulations.

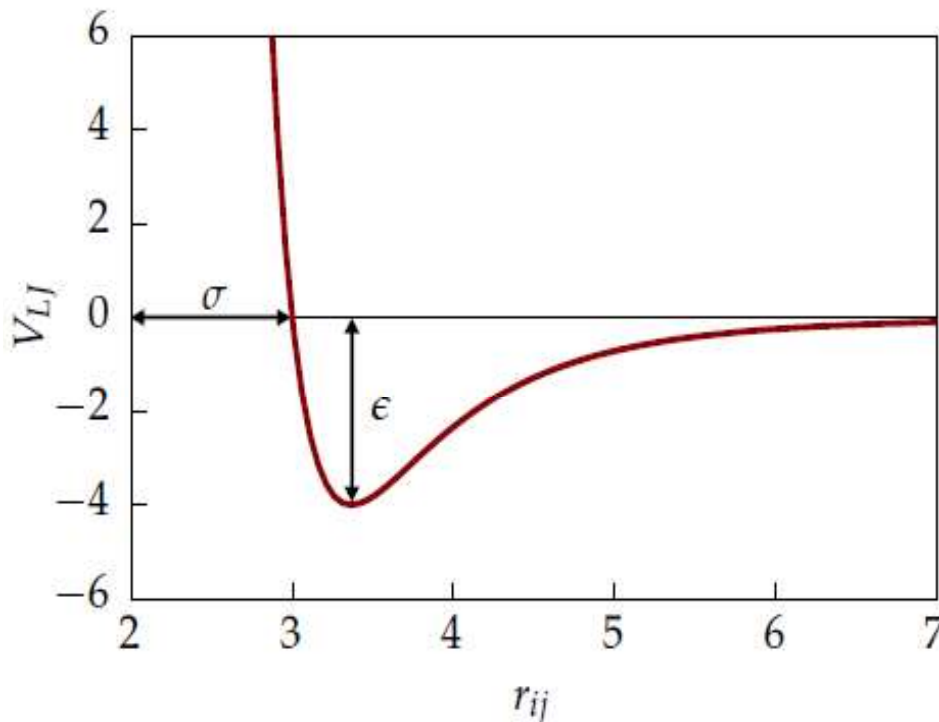
Numerous alternate functional formulations for the Van der Waals potential are in existence. Particularly noteworthy, we will delve into a modified iteration of the LJ potential, referred to as the Improved Lennard-Jones (ILJ). The ILJ potential represents an endeavour to enhance the LJ potential in regions where it falls short, while preserving its efficacy where it shines. In practical terms, this entails refining the potential curve's behaviour over extended distances and its characteristics at exceedingly close ranges. This adjustment aims to minimize the impact around the equilibrium position, a point where the accuracy of the LJ potential is established. These enhancements are realized through the utilization of the subsequent expression [50]:

$$V_{ILJ}(r_{ij}) = \epsilon \left[ \frac{m}{n(r_{ij})-m} \left( \frac{r_m}{r_{ij}} \right)^{n_{ij}} - \frac{n(r_{ij})}{n(r_{ij})-m} \left( \frac{r_m}{r_{ij}} \right)^m \right] \text{ ---- (2.57)}$$

$$\text{Where, } n(r_{ij}) = \beta + 4.0 \left( \frac{r_{ij}}{r_m} \right) \text{ ---- (2.58)}$$



The parameter  $r_m$  signifies the position where the well depth  $\epsilon$  is situated. The coefficient  $\beta$  stands as an unconstrained factor that correlates with the system's rigidity, typically varying within the range of 7 to 9.



**Figure 2.2:** An illustration of the Lennard-Jones potential, with all measurements expressed in arbitrary units.

The parameter  $m$ , explicitly influencing the exponent in the attractive segment of the potential, finds its value determined by the charge characteristics of the interacting entities. Specifically, for neutral-neutral interactions,  $m$  assumes the value of 6. In contrast, for ion-neutral interactions, it becomes 4, and for ion-ion interactions,  $m$  takes on the value of 1. Embedding factors and exponents reliant on distance within the function  $n(r)$  enhances both the attractive and repulsive facets of the potential, all the while retaining the foundational LJ form in the vicinity of the equilibrium separation. Noteworthy is the augmentation of a solitary parameter,  $\beta$ , in the ILJ potential when contrasted with the time-honoured LJ potential[51].

An essential prerequisite for the viability of a potential function lies in possessing a mathematically expressible initial derivative concerning the interatomic distance. This criterion proves indispensable since, as elaborated in the ensuing subsection, these derivatives are essential for calculating the forces exerted on the particles. Notably, the numerical computation

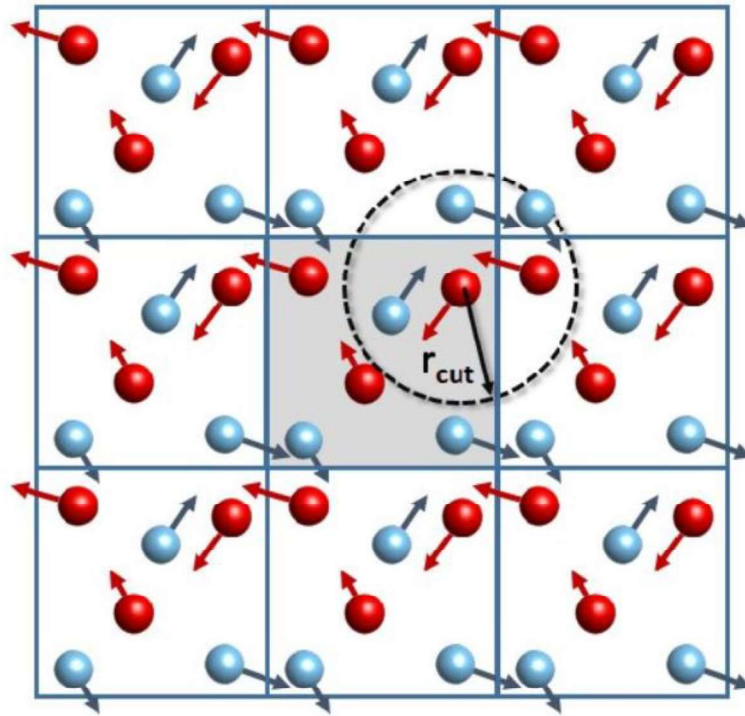
of forces necessitates considerably more computational resources compared to the utilization of analytical expressions, which can be seamlessly incorporated directly into molecular dynamics (MD) software. The overall computational efficiency bears profound significance within the MD framework, given the execution of millions of steps in simulations, likely contributing to the LJ potential's remarkable efficacy. Despite its more intricate mathematical structure, the ILJ potential, too, yields analytical first derivatives, rendering it highly suitable for these simulations [52].

Another significant form of two-body potential arises from the Coulombic interaction between charged particles. Often, atoms within a molecule can be attributed fractional charges, characterizing an apparent buildup or reduction in electron density. In such instances, the interaction between two centres, each carrying a charge of  $q_i$  and  $q_j$ , is governed by the following expression:

$$V_{els}(r_{ij}) = \left(\frac{q_i q_j}{4\pi\epsilon_0 r}\right) \text{---- (2.59)}$$

Because of the  $1/r$  dependency, the contribution from this electrostatic component is often fairly high and of a long-range character.

Usually, when particles are distant from each other, their non-bonded interactions tend to approach zero. Explicitly computing all of these interactions becomes computationally wasteful due to their collectively minor impact. To address this, it's customary to introduce a predefined cutoff radius, which dictates the maximum distance over which the intermolecular potential needs computation. For the absent long-range portion, an estimation is often computed and then incorporated into the overall total [53]. The cutoff radius's extent is inherently constrained; it can't surpass half the dimensions of the simulation cell. Its role is pivotal in delineating the interaction scope for all particles. Notably, in the realm of periodic boundary conditions (PBC), this parameter serves to elucidate the rationale behind the minimum image convention utilized for determining distances among interacting elements. To provide a visual depiction of both the cutoff radius and the minimum image convention, refer to **Figure 2.3**. Once all interactions within the system are precisely defined, the subsequent subsection will elucidate the mechanisms by which the system's temporal evolution is steered.



**Figure 2.3:** A visual representation that elucidates the notions of the cutoff radius and the minimum image convention.

### 2.6.3 Newton's Equation of Motion:

As introduced at the outset of this section, the primary objective of molecular dynamics simulations is to scrutinize how a specific system evolves over time. This objective is attained through the integration of Newton's equations of motion. Given an assemblage of positions denoted as  $\{\mathbf{r}_i\}$  for all particles, the progression of dynamics materializes by implementing Newton's equations of motion, expressed as follows:

$$\frac{d\mathbf{v}_i}{dt} = \frac{1}{m_i} \mathbf{f}_i \text{ ---- (2.60)}$$

$$\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i \text{ ---- (2.61)}$$

For indices  $i = 1, 2, \dots, N_{\text{tot}}$ , where  $m_i$  signifies the mass of the  $i$ -th particle and  $\mathbf{f}_i$  represents the comprehensive force (in vector form) exerted on it. The force vector  $\mathbf{f}_i$  is derived by differentiating the potential energy function  $U$ . To elaborate, the specific vector's components are derived as follows:

$$\mathbf{f}_i = \begin{pmatrix} f_{i,x} \\ f_{i,y} \\ f_{i,z} \end{pmatrix} \text{---- (2.62)}$$

is calculated for each element of the Cartesian equation  $w = x, y, \text{ and } z$  to produce.

$$f_{i,w} = - \sum_{j \neq i} \frac{\partial U(r_{ij})}{\partial w_i} \text{---- (2.63)}$$

$$\text{Where, } \frac{\partial U(r_{ij})}{\partial w_i} = \frac{\partial U(r_{ij})}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial w_i} \text{---- (2.64)}$$

Here,  $W_{ij} = w_i - w_j$ . Consequently, the force is represented in vector notation.

$$\mathbf{f}_i = \sum_{j \neq i} \mathbf{f}_{ij} = - \sum_{j \neq i} \frac{1}{r_{ij}} \frac{\partial U(r_{ij})}{\partial r_{ij}} \mathbf{r}_{ij} \text{---- (2.65)}$$

As highlighted in the earlier section, the potential energy function's derivative is expressible in a closed form, facilitating its straightforward computation. The equations of motion presented in Equation (9) require integration for the purpose of conducting dynamic simulations and deducing particle positions over time. In practical application, this process entails discretizing the temporal dimension into discrete steps of size  $\Delta t$ , followed by the numerical integration of the equations. Various integration techniques are available, with the velocity Verlet and leapfrog algorithms being the most renowned. Both of these algorithms are grounded in the Taylor expansion of the equations of motion. For example, within the leapfrog approach, forces and positions are known at integer timesteps, while velocities are known at half-integer timesteps, as demonstrated by<sup>46</sup>:

$$\mathbf{v} \left( t + \frac{1}{2} \Delta t \right) = \mathbf{v} \left( t - \frac{1}{2} \Delta t \right) + \frac{\mathbf{f}(t)}{m} \Delta t \text{---- (2.66)}$$

$$\text{and } \mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v} \left( t + \frac{1}{2} \Delta t \right) \Delta t \text{---- (2.67)}$$

This technique operates with speed and efficiency, as it avoids the necessity of preserving the entire sequence of positions and velocities; solely one set needs storage<sup>49,53</sup>. While velocities are established at half-integer intervals, having knowledge of them at integer positions is often advantageous, attainable by maintaining an additional collection of velocities from a preceding timestep and subsequently averaging between the two sets. Nonetheless, a drawback of the leapfrog approach is that the Taylor expansion employed to deduce Equations (15) and (16) is truncated prematurely. This truncation results in positions accurate to terms of

up to  $\Delta t^4$  and velocities accurate to  $\Delta t^2$ , marking a limitation. This signifies the need for careful consideration when selecting the timestep. The typical magnitude of  $\Delta t$  becomes a pivotal factor in ensuring the stable integration of Equation (9), usually opting for values in the sub-femtosecond range. The constraint of a brief timestep necessitates that an MD simulation extend across millions of steps to yield meaningful outcomes, essentially spanning a timeframe substantial enough to glean pertinent insights about the system. Lastly, a noteworthy aspect concerning the leapfrog algorithm is its minimal propensity for long-term energy deviations. This attribute is attributed to its time-reversible nature, as expounded upon by Frenkel and Smit.

### 2.6.4 Statistical Thermodynamics:

Lastly, we will succinctly present a pivotal concept indispensable in the realm of molecular dynamics simulations: the statistical ensemble. The bridge between the microscopic perspective of a system, operating at the molecular scale, and the macroscopic attributes we commonly gauge in laboratory settings – like temperature, pressure, and such – is established through the framework of statistical physics [54]. Fundamentally, the shift from analysing properties of individual particles to comprehending the thermodynamic traits of substantial molecular assemblies is achieved by acknowledging that the latter emerges from collective interactions. This emergence is encapsulated by deriving averages from the microscopic arrangement, effectively encapsulating the collective behaviour. To be more specific, a measurable macroscopic attribute  $A$  can be determined by computing the weighted mean of the corresponding microscopic factors  $a_i$ , expressed as follows[55]:

$$\langle A \rangle = 1/N \sum_{i=1}^N (p_i, a_i) \text{ ---- (2.68)}$$

where  $p_i$  is the probability, in terms of the number  $N$ , that the system is in microstate  $i$ .

In broad terms, a significant assumption is formulated concerning the generation of the microscopic state  $i$  and, correspondingly, the microscopic variable  $a_i$ . In both molecular dynamics (MD) simulations and real-world experiments, mean properties are derived by averaging over states at different temporal instances, as opposed to averaging over an independent collection of systems. This assumption, which equates time-averaging to ensemble-averaging, is commonly referred to as the ergodic hypothesis. In a general sense, an ensemble represents the repetitive occurrence of a closed system possessing specific

thermodynamic characteristics, all interacting with one another. This concept closely aligns with the scenario we initially established in the first subsection when we introduced the model system with periodic boundary conditions (PBC)[56]. Within such an ensemble, various microscopic states can manifest themselves with varying probabilities. Nevertheless, a prevailing configuration will exist, ensuring that the ensemble predominantly adopts that configuration for the majority of its existence. Diverse ensembles can be created based on the specific factor or variable we aim to hold constant [57-58]. These variables encompass the system's macroscopic attributes such as temperature (T), volume (V), pressure (P), chemical potential of particles ( $\mu$ ), particle count (N), and energy (E)[59]. These ensembles come in various forms, including the canonical ensemble (NVT), isothermal-isobaric ensemble (NPT), microcanonical ensemble (NVE), and grand canonical ensemble ( $\mu$ VT). Among these four ensembles, the most commonly employed ones are the NVT [60] and NPT [61] ensembles. A constant volume and temperature are maintained by the NVT ensemble, whereas a constant pressure and temperature are maintained by the NPT ensemble. Unrealistic theories include the NVE ensemble [62], which maintains constant energy and volume. Because the constant energy ensemble cannot reach the appropriate temperature during the equilibration phase, its use is not advised. As a result, in order to guarantee algorithmic stability for the selected system, in this thesis we have chosen the NVT ensemble, followed by the NPT ensemble. We used the Martyna-Tuckerman-Klein barostat to control the system's temperature, while the Nose-Hoover thermostat was used to control the pressure[63]. However, the intent of this section was to furnish the fundamental prerequisites for conducting an MD simulation, thereby barely delving into the fundamental underpinnings of atomistic simulations and the realm of statistical physics.

### 2.6.5 Molecular Dynamics Algorithm:

We will conclude the theoretical aspect of MD simulations by providing an outline of the algorithm that drives them. Broadly speaking, an algorithm for conducting a molecular dynamics simulation can be summarized as follows:

1. Specify the ensemble, relevant thermodynamic properties, time step, number of steps, sampling frequency, and any other necessary parameters.
2. Choose initial positions and velocities for all particles in the system to initiate the simulation.



3. Calculate the forces acting on all particles.
4. Integrate the equations of motion to obtain new positions and velocities.
5. Compute the required and interesting system properties.
6. Return to step 3 and repeat the process until the desired number of steps for the dynamics has been completed.

**References:**

1. M. J. Frisch, Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., *Gaussian 09, Revis. A.02, Gaussian, Inc.*, 2016, 1–2.
2. H. Bekker, H. Berendsen, E. Dijkstra, S. Achterop, R. Van Drunen, D. Van der Spoel, A. Sijbers, H. Keegstra, B. Reitsma and M. Renardus, *Phys. Comput.*, 1993, **92**, 252–256.
3. E. Schrödinger, *Phys. Rev.*, 1926, **28**, 1049–1070.
4. M. Born and R. Oppenheimer, *Ann. Phys.*, 1927, **389**, 457–484.
5. J. C. Slater, *Phys. Rev.*, 1929, **34**, 1293–1322.
6. D. R. Hartree and W. Hartree, *Proc. R. Soc. London. Ser. A - Math. Phys. Sci.*, 1935, **150**, 9–33.
7. P. Hohenberg, W. Kohn, *Phys. Rev.*, 1964, **136**, B864.
8. W. Kohn, L. J. Sham, *Phys. Rev.*, 1964, **140**, A1133.
9. J. W. Negele, *Phys. Rev. C*, 1970, **1**, 1260–1321.
10. J. P. Perdew and K. Burke, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 1996, **54**, 16533–16539.
11. P. M. W. Gill, B. G. Johnson, J. A. Pople and M. J. Frisch, *Chem. Phys. Lett.*, 1992, **197**, 499–505.
12. K. Burke, J. P. Perdew and Y. Wang, *Electron. Density Funct. Theory*, 1998, 81–111.
13. C. Adamo, M. Ernzerhof and G. E. Scuseria, *J. Chem. Phys.*, 2000, **112**, 2643–2649.
14. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372–1377.
15. K. Raghavachari, *Theor. Chem. Acc.*, 2000, **103**, 361–363.
16. H. Stoll, G. Wagenblast and H. Preuß, *Theor. Chim. Acta*, 1980, **57**, 169–178.
17. A. D. McLean and G. S. Chandler, *J. Chem. Phys.*, 1980, **72**, 5639–5648.

- 18 V. A. Rassolov, J. A. Pople, M. A. Ratner and T. L. Windus, *J. Chem. Phys.*, 1998, **109**, 1223–1229.
- 19 C. Chemistry, C. Chemistry and M. Modeling, *Comput. Chem. Mol. Model.*, 2008, 115–138.
- 20 K. Burke, J. Werschnik and E. K. U. Gross, *J. Chem. Phys.*, 2005, **123**, 1–34.
- 21 E. Runge and E. K. U. Gross, *Phys. Rev. Lett.*, 1984, **52**, 997–1000.
- 22 J. Harris, .
- 23 J. Neugebauer, M. Reiher, C. Kind and B. A. Hess, *J. Comput. Chem.*, 2002, **23**, 895–910.
- 24 R. M. LEVY, *Ann. N. Y. Acad. Sci.*, 1986, **482**, 24–43.
- 25 R. A. Cowley, *Proc. Phys. Soc.*, 1964, **84**, 281–296.
- 26 Z. Cui, X. Wang, Y. Ding, E. Li, K. Bai, J. Zheng and T. Liu, *Appl. Surf. Sci.*, 2020, **530**, 147275.
- 27 E. Transfer and O. Reactions, *Heterocycl. Commun.*, 2007, **13**, 2–8.
- 28 E. D. Glendening, C. R. Landis and F. Weinhold, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 1–42.
- 29 J. S. Murray and P. Politzer, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 153–163.
- 30 H. Chermette, *J. Comput. Chem.*, 1999, **20**, 129–154.
- 31 P. Bultinck, R. Carbó-Dorca and W. Langenaeker, *J. Chem. Phys.*, 2003, **118**, 4349–4356.
- 32 S. Canuto, 2008, 1–9.
- 33 D. Tzeli and A. Mavridis, *J. Chem. Phys.*, 2005, **122**, 7–8.
- 34 T. Morishita and A. M. Ito, *Phys. Rev. Res.*, 2019, **1**, 33032.
- 35 E. Paquet and H. L. Viktor, *Biomed Res. Int.*, 10.1155/2015/183918.
- 36 R. S. Katiyar and P. K. Jha, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2018, **8**, 1–18.
- 37 C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2021, **11**, 1–49.
- 38 O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K. R. Müller, *Chem. Rev.*, 2021, **121**, 10142–10186.
- 39 L. Monticelli and D. P. Tieleman, 2013, **924**, 197–213.

- 40 A. Hospital, J. R. Goñi, M. Orozco and J. L. Gelpí, *Adv. Appl. Bioinforma. Chem.*, 2015, **8**, 37–47.
- 41 M. O. Steinhauser, *Computational Multiscale Modeling of Fluids and Solids Theory and Applications*, Springer, 2017.
- 42 S. Z. Myron W Evans, G Wilse Robinson, Surjit Singh, *Water In Biology, Chemistry And Physics: Experimental Overviews And Computational Methodologies*, World Scientific Publishing Company, 1996.
- 43 M. O. Steinhauser, *Molecular Dynamics Studies of Synthetic and Biological Macromolecules*, IntechOpen, 2012.
- 44 Rappé, Anthony K., *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.
- 45 J. Å. and A. Warshel, *Chem. Rev.*, 1993, **93**, 2523–2544.
- 46 M. A. González, *École thématique la Société Française la Neutron.*, 2011, **12**, 169–200.
- 47 A. Gavezzotti, *Molecular Aggregation Structure Analysis and Molecular Simulation of Crystals and Liquids*, OUP/International Union of Crystallography, 2006.
- 48 I. Chandrasekhar, M. Kastenzholz, R. D. Lins, C. Oostenbrink, L. D. Schuler, D. P. Tieleman and W. F. Van Gunsteren, *Eur. Biophys. J.*, 2003, **32**, 67–77.
- 49 D. J. T. Michael P. Allen, *Computer Simulation of Liquids*, Oxford university press, 2017.
- 50 F. Pirani, S. Brizi, L. F. Roncaratti, P. Casavecchia, D. Cappelletti and F. Vecchiocattivi, *Phys. Chem. Chem. Phys.*, 2008, **10**, 5489–5503.
- 51 A. Trokhymchuk and J. Alejandre, *J. Chem. Phys.*, 1999, **111**, 8510–8523.
- 52 D. J. Price and C. L. Brooks, *J. Chem. Phys.*, 2004, **121**, 10096–10103.
- 53 G. J. H. Adams, David J., Eveline M. Adams, *Mol. Phys.*, 1979, **38**, 387–400.
- 54 G. J. M. Tuckerman, Mark E., *J. Phys. Chem. B*, 2000, **104**, 159–178.
- 55 R. Lustig, *J. Chem. Phys.*, 1994, **100**, 3060–3067.
- 56 Binder, Kurt, *J. Phys. Condens. Matter*, 2004, **16**, S429.
- 57 L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner and R. Abel, *J. Am. Chem. Soc.*, 2015, **137**, 2695–2703.
- 58 D. R. Hill, *Theory of Modelling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems: by BP Zeigler, H. Praehofer, TG Kim, Academic*

*Press, San Diego, CA, 2000, 2002.*

- 59 M. K. McCammon, J. Andrew, Bruce R. Gelin, *Nature*, 1977, **267**, 585–267.
- 60 J. M. Richardson and S. R. Brinkley, *J. Chem. Phys.*, 1960, **33**, 1467–1478.
- 61 D. S. Corti and G. Soto-Campos, *J. Chem. Phys.*, 1998, **108**, 7959–7966.
- 62 H. Q. Beard, Daniel A., *Chemical biophysics: quantitative analysis of cellular systems*, Cambridge: Cambridge University Press, 2008.
- 63 S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.