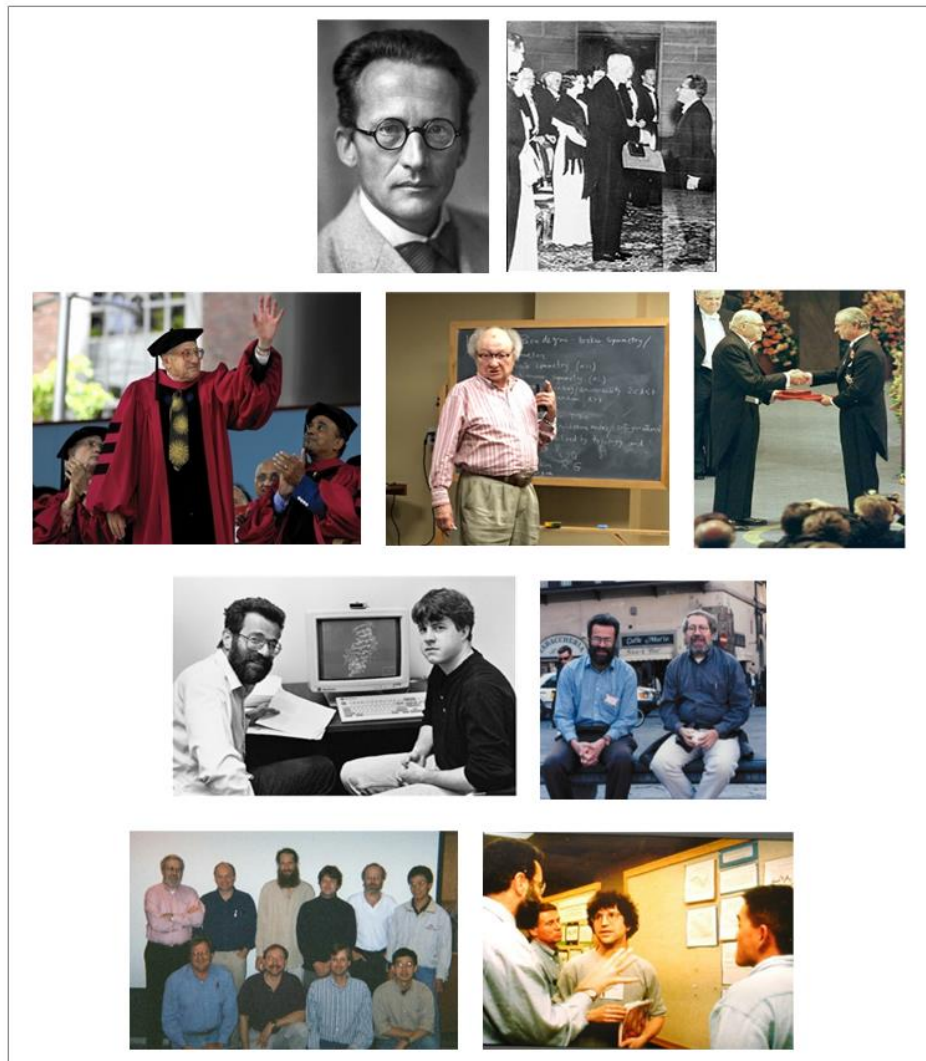




## Chapter 2

### Overview of Computational Tools & Methodologies



## 2.1. Computational Chemistry: A General Introduction

Computational chemistry is a branch of chemistry that uses methods of theoretical chemistry, incorporated into various computer programs and softwares, to calculate the structures and properties of small molecules, large biomolecules as well as group of molecules [1-2]. The phrase "computational chemistry" is mentioned long back in Sidney Fernbach & Abraham Haskell Taub's 1970 book "Computers and Their Role in the Physical Sciences," as in "It seems, therefore, that 'Computational Chemistry' can finally be more and more of a reality." [3]. Over the years, advancements in computing processing capability eventually shrank the gap between theoretical and experimental results while resolving issues in a variety of scientific fields involving both macroscopic and microscopic phenomena.

Results from computational simulations assist in solving the information obtained during chemical experiments, to complement the experimental results with the theoretical models, and in some cases predicts hitherto unobserved chemical phenomenas as well [2]. One can readily establish the mechanism of any chemical reaction with the aid of molecular modelling tools, which is one example of how computational chemistry can help one grasp the fundamentals of any type of chemical problem that is occasionally not directly supported by experimental evidence. Techniques in computational chemistry focuses mainly on investigating the electronic structures, vibrational frequencies, relative energies, chemical reactivities, electron and charge distributions, various physical and chemical properties, potential energy surfaces (PES), transition states and reaction pathways, host-substrate interactions, thermodynamic and kinetic studies and calculations of many other molecular and bulk physical and chemical properties. It is a tool widely used in the design of new drug molecules and materials. However, the idea of molecular modelling has little bearing on the objectives of chemistry. Therefore, it is important to realize that molecular modelling creates a representation of the real world, and that we study the model rather than the real world. A model is legitimate as long as it accurately depicts the real world.

Computational chemistry also paves the door for research into complex biological problems and designing of drugs [4-5]. The traditional process of drug discovery and development is highly complicated, time-consuming, resource-intensive and expensive with only a handful of drug moieties getting the approval from the FDA making the process of discovering and developing new pharmaceuticals exceedingly

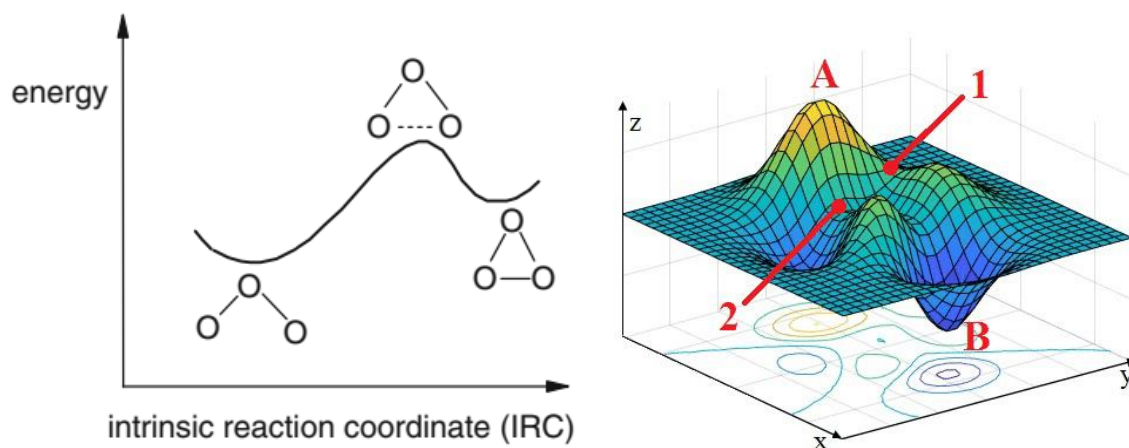
arduous. Although the structure-based drug-design method was developed in the 1980s, it wasn't until the last decade that a variety of computational tools and models emerged for a better understanding of the processes at a fundamental molecular level. In recent times, a more popular computer-aided drug design (CADD) has been the tool that has given the drug delivery processes a greater proficiency [6-7]. Such methods consist of numerical techniques of Ab-initio method, semi-empirical molecular mechanics (MM) methods as Monte Carlo (MC), Molecular Dynamics (MD) simulations as well as Quantum Mechanical (QM) methods as density functional theory (DFT) and hybrid QM/MM techniques to determine both qualitative and quantitative informations. These are the most important tools of computational chemistry, including the concept of a potential energy surface (PES) understanding of which is crucial to learn molecular modeling in chemistry [8]. Accordingly, in this chapter, we will provide a brief summary of the tools that have been used to carry out the computational calculations included in this thesis.

## 2.2. Computational Chemistry Tools

Using computers to uncover chemistry problems, computational chemistry opens the door to research into complex chemical and biological problems. In the peripheral area of computational chemistry, the primary emphasis is on methods for obtaining structural information, physical & chemical characteristics, energetics, thermochemistry, and kinetics through computer simulations. There are five main branches that make up the core of computational chemistry. These include Molecular Mechanics (MM), Molecular Dynamics (MD), Ab-initio methods, Semi-empirical methods and Density Functional Theory (DFT). The tools used to complete the work described in this thesis are briefly detailed in the following subsections, with a focus on MD and DFT.

### 2.2.1. Concept of Potential Energy Surface (PES)

A key idea in computational chemistry is the potential energy surface (PES), which is the mathematical or graphical link between energy and molecule geometry. The foundation of computational chemistry is the idea of a PES, which expresses the connection between a species' energy and shape (Figure 2.1) [8]. For instance, in a di-atomic molecule A-B, the geometry of the molecule can only be changed by one parameter, the bond length, which results in a one-dimensional PES known as 1D-PES. Similarly, 2D-PES is produced if the energy of the molecule is dependent on the bond length and bond angle.



**Figure 2.1:** Example of points on potential energy surface (PES) of a random molecule.

The stationary point is a location on the potential energy surface (PES) at which the surface is flat, i.e., parallel to the horizontal line (or plane) corresponding to the geometrical parameter(s), is another essential feature of the PES. In mathematical terms, the first derivative of energy with respect to each geometric parameter represents a stationary point, and it always equals zero.

$$\frac{\partial E}{\partial q_1} = \frac{\partial E}{\partial q_2} = \dots = \frac{\partial E}{\partial q_i} = 0 \quad (1)$$

The minima and the saddle point are the two different categories of stationary places on the potential energy surface. A saddle point is a maximum along the reaction coordinate and a minimum along all other directions, whereas a minimum is a minimum along all directions. On the PES, certain minima are referred to as local minima, while the lowest energy minimums are known as global minimums (Figure 2.1). The transition state, which connects two minimums in the reaction coordinate, is the first order saddle point. They can be differentiated mathematically by their second derivatives:

For a minimum,  $\frac{\partial^2 E}{\partial q^2} > 0$  for all q.

For a transition state,  $\frac{\partial^2 E}{\partial q^2} > 0$  for all q, except along the reaction coordinate,

$\frac{\partial^2 E}{\partial q^2} < 0$  for all q, along the reaction coordinate.

The fact that the second derivative of energy with respect to a single geometric coordinate is negative is a crucial feature of a first-order saddle point. Higher-order saddle points, often known as hilltops, are the sites on PES where the second derivative of energy is negative with respect to many coordinates.

Several computational tools and models for a deeper understanding of the processes at a fundamental molecular level have only recently evolved, despite the fact that the first structure-based drug-design strategy was created in the 1980s. CADD based on the various computational tools is the present day go-to technique for improving the drug discovery and delivery process's competency.

### 2.2.2. Molecular Mechanics (MM)

The late seventeenth-century discovery of the laws of motion by Sir Isaac Newton set the groundwork for classical mechanics and opened the door to understanding the characteristics and behaviour of macroscopic objects [9]. The same theory of classical mechanics serves as the foundation theory of molecular mechanics (MM) [10]. MM is expressed in terms of a mathematical model of a molecule called the "ball and spring" model, which is set for optimization and views atoms and bonds as a collection of balls held together by springs. A molecule's potential energy can be expressed as the product of terms pertaining to bond stretching, angle bending, dihedral angles and nonbonded interactions. The mathematical expression and its parameters together form a force field, utilized to minimize energy. Creating a forcefield involves giving these terms precise mathematical representations, and parameterizing the field involves assigning numerical values to the forcefield's constants. The type of force field that is used determines the accuracy of this procedure.

The potential energy of the molecule is calculated by the parametric function of coordinates and is expressed mathematically as

$$E = E_{stretch} + E_{bend} + E_{torsion} + E_{vdw} + E_{HB} + E_{electro} \quad (2)$$

Here,  $E_{stretch}$  = energy term for bond stretching;  $E_{bend}$  = energy term for bond angle bonding;  $E_{torsion}$  = energy term for dihedral angle rotation;  $E_{vdw}$  = energy term for van der Waals energy;  $E_{HB}$  = energy term for hydrogen bonding and  $E_{electro}$  = electrostatic interactions between atoms or groups which are nonbonded.

A common application of MM is to generate suitable geometries as input for other computations. As long as the forcefield has been meticulously specified for the different kinds of molecules being studied, these computations may be completed quickly and accurately. In addition to being quick and accurate, MM requires little processing resources and offers suitable initial geometries for quantum mechanical computations. Since it disregards electrons, it can only supply parameters via analogy, such as the dipole moment. Regarding the suitability of MM parameters care must be taken. Even in cases when they are relative minima, stationary points from MM might not be global minima. Therefore, calculating the energy and geometries of transition states is a limited advantage of using MM. For polar compounds, ignoring the effects of the solvent can lead to inaccurate conclusions. MM provides strain energies, which can be parameterized to provide enthalpies of formation. For structurally comparable isomers, the difference of these strain energies represents enthalpy differences. Relative isomer concentrations are, strictly speaking, dependent on free energy differences. Even when accurately recognized, the main conformation is not always the reactive one.

One of the most significant uses of MM is in the pharmaceutical industry's drug design process. One such use of MM is in the analysis of potential drug candidates' fit into biomolecules active sites (molecular docking) and the associated QSAR models. MM is currently widely used in organic synthesis, allowing chemists to predict which products would be preferred and to map out more practical paths to a target molecule than were previously feasible. MM is used in Monte Carlo simulations to determine the energies of the several randomly produced states, and it is also used in molecular dynamics (MD) to generate the forces acting on molecules and hence to calculate their motions.

### **2.2.3. Molecular Dynamics (MD)**

The flexible and dynamic state of the molecules in the biological environment is crucial in defining their functions [11]. Conventional experimental techniques, like X-ray crystallography, NMR can produce static images of proteins and nucleic acids at a particular moment in time, but they are unable to produce dynamic information about the motions that take place in real time [12]. Molecular Dynamics (MD) is a computational tool used to examine the physical movements of atoms and molecules through the elaboration of N-body simulation, and is averaged throughout time [13-14].

MD in theory is governed by the same classical Newton's second law of motion, which states  $F = ma$ , where 'F' stands for the force applied individually to each atom, 'm' for mass, and 'a' for acceleration [9]. For a predetermined amount of time, the atoms and molecules are permitted to interact, providing insight into the system's dynamic evolution. In essence, MD can ascertain the atoms' positions in relation to time and a trajectory is produced by the atomic coordinates and velocities moving through time. The equations of motion derived from Newton's laws of motion, are solved numerically to predict the trajectory of each particle in the system. The negative gradient of potential energy is used to calculate forces and is expressed as:

$$m_i \frac{d^2 r_i(t)}{dt^2} = F_i \quad (3)$$

Here,  $m_i$  is the mass,  $r_i(t)$  signifies the position vector for the  $i^{\text{th}}$  particle and  $F_i$  denotes force acting on it [15].

By computing forces on each atom repeatedly and using the results to determine velocities and position coordinates, it is possible to determine the average motions of the system. This allows the simulation to capture the dynamic behavior of molecules as they move and interact. The structure, dynamics, thermodynamics, oscillations and structural changes of biological molecules as proteins, nucleic acids and their complexes can be thoroughly studied using MD simulations. It's particularly powerful for studying the dynamic aspects of molecular systems at the atomic level. The key concepts behind Molecular Dynamics include:

### **2.2.3.1. Integration Algorithms to induce dynamic behaviour**

The availability of initial atomic coordinates and velocities is a prerequisite for starting MD simulations. After obtaining the atoms' position and velocity data, the trajectory's motion can be calculated by integrating the equations of motion. Among other well-known techniques, numerical integration algorithms like Verlet [16] and Leap-frog [17] are used to obtain the positions and velocities of particles at each time step. In Verlet Algorithm positions are obtained in reference to a previously defined position using the numerical integration formula. Because there is less requirement for storage, the Verlet theorem is nevertheless widely used, despite its reduced precision. Determining position and speed simultaneously is not possible. The key feature of the leapfrog algorithm is that it separates the updates of position and velocity, which can be advantageous in certain simulations. Consequently, because the computations of potential energy and



kinetic energy are done separately, it is impossible to determine the total energy. The choice of integration algorithm affects the accuracy and stability of the simulation.

### 2.2.3.2. Ensemble Averages

The calculations describing a macromolecular system's dynamics are performed at a fixed energy. The temperature and pressure are continuously kept at levels that closely mimic the biological environment in order to closely simulate it. A group of objects that are phase-space independent yet thermodynamically same is called an ensemble, or a macroscopic quantity. MD simulations are typically run for an ensemble of particles, representing a statistical sample of possible configurations [18]. Properties of interest, such as temperature, pressure, and energy, are often calculated as ensemble averages over time. There exist four categories of ensembles, contingent upon the variable that is fixed constant:

- ✚ **Micro canonical ensemble (NVE):** The isolated system with constant atom number (N), volume (v) and equal energy (E) represents the NVE ensemble.
- ✚ **Canonical ensemble (NVT):** This ensemble is characterised by a fixed number of atoms (N), same volume (V) and temperature (T).
- ✚ **Isobaric-Isothermal ensemble (NPT):** This ensemble constitutes system with equal number of atoms (N), constant pressure (P) and temperature (T).
- ✚ **Grand Canonical ensemble ( $\mu$ V T):** This ensemble is characterised by constant chemical potential ( $\mu$ ) along with constant volume (V) and temperature (T).

Usually, the number of particles in the system selected for simulation is much lower than in the comparable real macroscopic system. The particles in MD simulations encounter boundary surfaces that are inconsistent with events seen in bulk environments since the simulations are carried out inside a solvation box. Furthermore, particles will evaporate from the system as atoms that are getting closer to the box's edge leave the system.

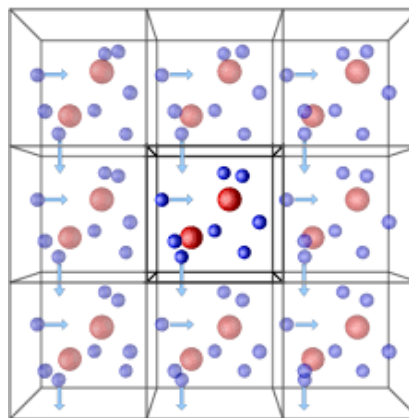
### 2.2.3.3. Boundary Conditions

MD simulations are conducted in a periodic box, meaning that when a particle exits one side of the simulation box, it re-enters from the opposite side. This avoids artifacts related to the simulation box's boundaries and allows for a more realistic representation of an infinite system. Periodic boundary conditions are the favoured boundary conditions among many researchers in the study of various simulation studies.



### 2.2.3.3.1 Periodic Boundary Conditions (PBC)

An alternate method for carrying out bulk conditions in simulations is PBC (Figure 2.2) [19]. As seen in Figure 2.2, the main simulation box is encircled by translated copies of the boxes on both sides in such a way that, each time an atom exits the box, it is replaced by an atom from the twin box traveling at the same speed, keeping the initial particle count constant. The impression of a bulk solvent devoid of surface effects is created by the particles close to the edge constantly interacting with the particles in the remaining compartments.



**Figure 2.2:** A representation of PBC. The arrows mark that as the blue-coloured particles in the primary box go out of the system, adjacent particles from image box replace them.

### 2.2.3.4. Thermostats and Barostats

To control the temperature and pressure of the simulated system, thermostats and barostats are often employed. These algorithms adjust the velocities and positions of particles to maintain desired thermodynamic conditions. MD simulations employ the Berendsen thermostat [20] to regulate temperature. V-rescale, also known as velocity rescale [21], is a more advanced version of the Berendsen thermostat. Every time the Berendsen thermostat steps, the velocity is updated continuously, and the rate of temperature change is proportionate to the difference between the system and bath temperatures. A barostat like the Parinello-Rahman barostat, which is also used to control temperature, is used to control pressure [22].

### 2.2.3.5. Time Step

The time step is critical in MD simulations, representing the interval between successive updates of particle positions and velocities. Most significant biological changes occur in the nano-, micro-, and millisecond time frames. MD modeling today shows femtosecond motions, which guarantee numerical stability in the field of structural biology [23]. Most of the MD integrations were carried out using a 2.0 fs time step, employing the SHAKE algorithm on all the bonds involving hydrogen atoms. For non-bonding interactions, a cut-off distance of 10 Å has been used and the pair-list was updated at every 1000 steps.

### 2.2.3.6. Force Fields and Development of Force Field Parameters

Force fields are mathematical models that describe the interactions between particles in a molecular system. These interactions include bond forces, angle forces, dihedral forces, and non-bonded forces (van der Waals and electrostatic interactions) [24-25].

As already discussed, the potential energy of the molecule is calculated by the parametric function of coordinates and is expressed mathematically as

$$E = E_{stretch} + E_{bend} + E_{torsion} + E_{vdw} + E_{HB} + E_{electro} \quad (1)$$

Where,  $E_{stretch}$  = energy term for bond stretching;  $E_{bend}$  = energy term for bond angle bonding;  $E_{torsion}$  = energy term for dihedral angle rotation;  $E_{vdw}$  = energy term for van der Waals energy;  $E_{HB}$  = energy term for hydrogen bonding and  $E_{el}$  = electrostatic interactions between atoms or groups which are nonbonded.

The MM force field, a model that combines quantum computations with experimental data, is used to determine the forces involved in a simulation and are crucial for determining the forces acting on each particle at any given moment. Bonded interactions can be used to summarize a system's bond stretching, bond bending and torsional terms. On the other hand, weak van der Waals interactions and electrostatic interactions are determined using the Coulombic and van der Waals interactions (as Lennard-Jones potentials) [77].

Most simulations use a pair potential term to mimic  $V$ , which can be further classified into coulombic and non-coulombic terms as

$$V(\mathbf{r}_i, \mathbf{r}_j) = \frac{q_i q_j}{r_{ij}} + \phi(\mathbf{r}_{ij}). \quad (4)$$

The first term represents long-range electrostatic interactions between a pair of atoms with effective charges  $q_i$  and  $q_j$  while the latter term is a two-body short-range interaction. Long range interactions are given by Coulomb's law as:  $V(\mathbf{r}) = \sum \frac{q_i q_j}{r_{ij}}$ .

Non-bonded interactions consist of either a Lennard-Jones 6-12 potential or a Buckingham potential, where the  $r^{-12}$  repulsion is replaced by an exponential term, and a Coulomb term. The Coulomb interaction has a fixed dielectric constant and can be modified by a reaction field, mimicking the effect of a homogeneous dielectric environment beyond the cut-off radius, including the effect of ionic strength. The non-bonded functions can be modified by a shift function or a switch function that causes forces and their derivatives to be continuous at the cut-off radius.

In order to more closely resemble biological processes and produce more accurate results, the force fields have undergone substantial evolution over the past few years and are regularly modified. Force-fields for water are made to mimic the experimental values of vaporization enthalpy and water density [26]. Bonded term parameters are generally found using QM computations, using either *ab initio* or semiempirical techniques [27]. Existing force fields are empirically matched to liquid or solid-state systems in order to derive the van der Waals parameters [28]. In most models, partial charges were used to create electrostatic interactions with water that were well-adjusted. In AMBER [29], *Cornell et al.* used *Ab-initio* Hartree Fock computations. In most models, the torsional potential term of force fields is computed using QM computations in addition to empirical data [30]. Approaches like the Ewald technique that provide more accurate long-range interactions should be preferred when it comes to nucleic acids [31]. A few previously developed force-fields include AMBER [32], CHARMM [33], ENCAD [34], GROMOS96 [35] and OPLS-AA [36].

Every effort is being made in the current situation to get beyond the limitations placed on force fields. Through the integration of QM/MM, groups are attempting to address the current limitations in electronic information potential [37]. Calculating the MM component of the QM/MM technique remains challenging. To address chemical reactivity, reactive force fields are currently being actively developed. Simulations have become more computationally efficient as a result of the growing usage of polarization, even though it is still not fully developed to characterize non-bonded interactions. The use of GPUs and the development of high-performance computers (HPCs) supercomputers are examples of recent technological developments that have made it feasible to provide efficiency while lowering processing costs.

Over the past few decades, MD techniques have improved to recognize the biological significance using pre-existing 3D structures [38-40]. This technique, in contrast to X-ray crystallography and NMR, enables us to simulate the motion of atoms inside a chemical system throughout time. Protein folding, ligand binding, conformational changes in proteins, and protein-protein binding are just a few of the biomolecular processes for which MD simulations have shown to be incredibly helpful in capturing key motions [41-43]. They provide valuable insights into the behavior of molecular systems, helping researchers understand complex phenomena such as protein folding, chemical reactions, and material properties at the atomic level.

### 2.2.4. Ab-initio Methods

Classical mechanics is however unable to explain the behavior of microscopic entities like protons, neutrons, electrons and the elementary particles as quarks, leptons (such as the electron) and bosons (such as the photon). Theory of Quantum Mechanics (QM) governed by certain laws thus evolved as a solution to overcome the drawbacks of classical mechanics. These laws could accurately describe the structural and chemical characteristics of atomic and subatomic particles in the microscopic world. The laws of QM upon which the ab-initio approaches are founded, have shown to improve computation accuracy for microscopic systems [44-45]. Therefore, in computational chemistry, the ab initio method is an essential tool for examining the stability, physicochemical characteristics, and structure of various molecular systems at the microscopic level. Leading figures in the development of QM includes prominent scientist as Born, Pauli, Bohr, Schrödinger, Dirac and Heisenberg. Since QM covers the structural and physico-chemical characteristics of atoms, molecules, nanoparticles, and nanoclusters, it has attracted a lot of attention in the field of computational chemistry.

#### 2.2.4.1 Hartree-Fock Method

Hartree-Fock theory is a fundamental approach in quantum mechanics, used to approximate the wave function of a many-body quantum system, typically atoms and molecules. Initially put forth by Hartree and Fock, popularly known as the HF theory, is the simplest method that interprets that every electronic motion may be described by a single electron wavefunction known as the orbital, and it is not directly dependent upon the other electrons instantaneous motions [46-47]. Hartree-Fock method approximates the total molecular wavefunction  $\psi$  as a Slater determinant made up of occupied spin orbitals, where each spin orbital is the result of multiplying a spin function by a conventional spatial orbital  $\psi$ . The goal of HF theory is to find a solution to the electronic Schrödinger equation such that the system's electronic energy can be excluded.

The time-independent Schrödinger equation is expressed as:

$$\hat{H}\psi_i(\vec{\chi}_1, \vec{\chi}_2, \dots, \vec{\chi}_N, \vec{R}_1, \vec{R}_2, \dots, \vec{R}_M) = E_i\psi_i(\vec{\chi}_1, \vec{\chi}_2, \dots, \vec{\chi}_N, \vec{R}_1, \vec{R}_2, \dots, \vec{R}_M) \quad (5)$$

where,  $\hat{H}$  stands for the Hamiltonian operator of a molecular system having of M nuclei and N electrons.  $E_i$  is the energy Eigen value.

One may compute the wave function for electrons moving in the potential field of positively charged fixed nuclei using the Born-Oppenheimer approximation. The Born-Oppenheimer approximation is based on the notion that nuclear and electronic mobility in molecules can be distinguished from one another which results in a molecular wave in terms of nuclear locations and electron positions.

Therefore, electronic Hamiltonian  $\hat{H}$  is expressed as

$$\hat{H}_{\text{elec}} = -\frac{1}{2}\sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>1}^M \frac{1}{r_{ij}} \quad (6)$$

The electron's kinetic energy is represented by the first term in the equation, the attractive potential between electron I and nucleus A, which is separated by a distance r, is represented by the second term, and inter-electronic repulsion is represented by the last term. The nuclear factor is added independently to obtain the total energy of the system, and hence it is not included in this equation. The equation is in atomic units.

By using the Schrodinger equation, the molecular energy can be written as the wavefunction's expectation value ( $E = \langle \psi | \hat{H} | \psi \rangle$ ). One may obtain the HF equations by differentiating E with regard to the spin orbitals that make up the wavefunction. The spatial orbitals are represented as a linear combination of basis functions in order to be used in real-world computations. Though they can actually be any mathematical functions that produce a sensible wavefunction that is, a wavefunction that produces reasonable results when we perform the calculations-these are typically associated with atomic orbitals.

The main idea of the HF theory is to replace the many electrons problem with a one-electron problem in which electron-electron repulsion is treated in an average way

$\left(\sum_{\alpha} \frac{Z_{\alpha}}{r_{1\alpha}}\right)$  i.e.,

$$\hat{H}^{core}(1) = -\frac{1}{2}\nabla_1^2 - \sum_{\alpha} \frac{Z_{\alpha}}{r_{1\alpha}} \quad (7)$$

The primary flaw in the HF method is the improper treatment of electron correlation. Each electron is thought to move in an electrostatic field, which is represented by the average positions of the other electrons. However, in reality, electrons avoid one another more effectively than this model suggests because each electron actually perceives the other as a moving particle, and they both adjust (correlate) their motions to reduce the energy of their interactions.

Post-HF techniques including the Møller-Plesset (MP), configuration interaction (CI), and coupled cluster (CC) approaches handle electron correlation more well. By permitting the electrons to live in both formally vacant MOs (virtual MOs) and conventionally occupied MOs (the  $n$  lowest MOs for a  $2n$ -electron species), these techniques reduce the energy of electron-electron interactions.

The ab initio method is mostly used to calculate molecular energies, vibrational frequencies, spectra (IR, UV, and NMR), ionization energies, electron affinities, and features such as dipole moments that are directly related to electron distribution or distribution. These calculations have both theoretical and practical applications; for instance, spectroscopy is crucial in identifying and comprehending novel molecules, enzyme-substrate interactions rely on shapes and charge distributions, and reaction equilibria and rates depend on energy differences. When evaluating calculation results, it can be crucial to see estimated phenomena such as molecular vibrations, charge distributions, and molecular orbitals shown. However, the computational cost of using ab-initio computations to attain accuracy is high for comparatively large molecular systems.

### **2.2.5. Semi-empirical Method**

Semi-empirical methods were created to offset the high computing expense of utilizing Ab-initio simulations to attain accuracy for somewhat large molecular systems. To make the simulations easier to understand, semi-empirical approaches use parameters obtained from experimental data and is referred to as a "semi-empirical" because it is developed using both theory and experimental ("empirical") data [48-50]. The method is an intermediate between molecular mechanics and ab-initio approaches which refines the molecular energy and wave function by diagonalizing the Fock matrix. The methods correctness is dependent on the calculations parameters which yield rather reasonable results when used with suitable parameters when determining the energy and structure of the chemical system. However, results obtained from this method are not trustworthy until they are compared to either the high-level ab-initio calculations or the experimental collection of data.

Semiempirical approaches make use of three approximations: (a) Removing core electrons from the computations; (b) Using the fewest possible basis sets; and (c) Using fewer two-electron integrals. CNDO, PPP, NDDO and INDO are only a few of the

processes that fall under the semiempirical method. The differential of overlap integral is set to zero in all four of these processes, which make use of the Zero Differential Overlap (ZDO) approximation.

SCF semiempirical methods PPP, CNDO, INDO, and NDDO have been developed in an increasing order of sophistication. While CNDO, INDO, and NDDO employ all of the valence electrons, the PPP technique is restricted to p electrons. The differential of the overlap integral is set to zero in all four because they make use of the ZDO approximation, which significantly lowers the number of integrals that must be calculated. These techniques were previously mostly parameterized using experimental values (ionization energies, electron affinities), with the exception of PPP and CNDO, which also occasionally used the outcomes of minimal-basis-set, or low-level, ab initio computations. Of these early techniques, only INDO/S and its variation ZINDO/S, which are parameterized to replicate experimental UV spectra, are still widely used today.

These days, NDDO-based AM1 and PM3 semiempirical SCF approaches are the most widely used. They are meticulously parameterized to replicate experimental variables. Similar in performance, AM1 and PM3 often produce quite acceptable geometries but less than ideal formation heats and relative energies. Relatively little-used SAM1 (semi-ab initio method 1) is a variant of AM1 that is thought to be superior to AM1. In the near future, extensions of AM1 (RM1) and PM3 (PM6) are probably going to become the norm for general-purpose semiempirical approaches since they appear to be significant advancements.

### 2.2.6. Density Functional Theory (DFT)

The solution to the 3N-dimensional Schrödinger equation for the wave function ( $\psi$ ) typically evades the question of how to handle a system of N-interacting electrons in an external potential  $v(r)$ . This problem is recast using DFT in terms of a universal functional of the density  $E_{xc}[n(r)]$  and the electronic-density distribution  $n(r)$ . A system with N electrons would have a wave function with three coordinates for each electron and an additional coordinate for the spin, for a total of 4N coordinates. In contrast, electron density depends only on three coordinates, regardless of the number of electrons in the system. This illustrates the advantage of DFT over other approaches. Thus, compared to the ab-initio methods which are the focus of wave function-based theory, Density Functional Theory (DFT) locates electrons precisely because it works with

---

---



electron density rather than individual electrons. Because of this, regardless of system size, the electron density keeps the same number of variables constant while the wave function's complexity rises with the number of electrons. With the introduction of electron density in 1965, Kohn and Sham [51] laid the groundwork for the current use of DFT in computational chemistry. As a result of the method's full potential and wide range of applications being made apparent by advances in computing power, Walter Kohn was awarded the 1998 Nobel Prize in Chemistry for his contributions to the development of Density Functional Theory (Figure 2.3). Over the past few decades, DFT has emerged as the primary technique for simulating chemical systems using quantum mechanics. Quantum chemists have also embraced it in recent years, and it is now extensively utilized for the simulation of molecular energy surfaces.

## The Nobel Prize in Chemistry 1998



**Walter Kohn**  
Prize share: 1/2



**John A. Pople**  
Prize share: 1/2

**Figure 2.3:** Walter Kohn and John A. Pople received Nobel Prize in chemistry in 1998. (Picture source: [http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/1998/](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1998/))

The problem of solving many electrons Schrödinger equation is replaced by the problem of finding sufficiently accurate approximations to  $E_{xc}[n(r)]$  and then solving appropriate single electron equations. Finding adequate approximations to and then solving suitable single electron equations takes the role of the issue of solving the many electrons Schrödinger equation.

### 2.2.6.1 The Kohn-Sham Formalism

The Hohenberg-Kohn theorems are foundational principles in the framework of Density Functional Theory (DFT), formulated by Pierre Hohenberg and Walter Kohn in 1964 and laid the groundwork for the development of DFT. The foundation of density functional theory thus lies in the two Hohenberg-Kohn theorems, which postulates that an atom's or molecule's ground-state properties are dictated by its electron density function and that a trial electron density must provide an energy that is either larger than or equal to the true energy. But theorem is true only if the exact functional is used. With the approximate functionals in use today, DFT can give an energy below the true energy.

**First Hohenberg-Kohn Theorem:** This theorem states that the external potential ( $V_{\text{ext}}$ ) of a many-electron system uniquely determines its ground-state electron density ( $n(\mathbf{r})$ ). In simpler terms, this theorem asserts that the ground-state electron density is a unique functional of the external potential. Mathematically, it can be expressed as:  $V_{\text{ext}}(\mathbf{r}) \rightarrow n(\mathbf{r})$ . This theorem implies that the ground-state properties of a system, such as energy, electron density, and other observables, can be determined solely from the electron density, without explicit knowledge of the wave function.

**Second Hohenberg-Kohn Theorem:** This theorem establishes the existence of a universal functional of the electron density, known as the "universal functional" or "universal energy functional" ( $F[n]$ ) which uniquely determines the ground-state energy of a system. Mathematically, it can be expressed as:  $F[n] \rightarrow E_{\text{gs}}$  where  $F[n]$  the universal functional and  $E_{\text{gs}}$  is the ground-state energy. This theorem implies that although the exact form of the universal functional is unknown, it exists and uniquely determines the ground-state energy for any given electron density.

These theorems provide a rigorous foundation for the development of DFT, which has become an essential tool for predicting the properties of molecules, solids, and other materials. By treating the electron density as the central quantity of interest, DFT offers a computationally efficient approach to solving the quantum many-body problem, enabling the study of complex systems that would be intractable using traditional quantum mechanical methods.

The Kohn-Sham equations are named after Walter Kohn and Lu Sham, who introduced them in 1965, are a set of equations derived to map the many-body problem of interacting electrons onto a set of non-interacting electrons moving in an effective

potential. The KS equations are solved self-consistently to obtain the ground-state electronic structure of a system. Within the Kohn-Sham formalism, which states that the ground state energy of a non-degenerate electronic system and the corresponding electronic properties are uniquely defined by the electron density, for a set of interacting electrons, DFT calculations can be carried out very conveniently in terms of single particle orbitals. The central premise in the Kohn-Sham approach is that a system's energy functional may be divided into two parts: an exact part that treats electrons as non-interacting particles and a minor correction term that takes the electron-electron interaction into account.

The electronic energy is given by

$$E[\rho(r)] = T_{ni}[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)] + \Delta T[\rho(r)] + \Delta V_{ee}[\rho(r)] \quad (8)$$

where,  $T_{ni}[\rho(r)] =$  Kinetic energy of non-interacting electrons,  $V_{ne}[\rho(r)] =$  Nuclear-electron interaction,  $V_{ee}[\rho(r)] =$  Electron-electron repulsion,  $\Delta T[\rho(r)] =$  Correction to the kinetic energy derived from the electron-electron interaction,  $\Delta V_{ee}[\rho(r)] =$  Non-classical correction to the electron-electron repulsion energy.

The above equation can be expressed as:

$$E[\rho(r)] = -\frac{1}{2} \sum_{i=1}^n \int \psi_i^*(r_1) \nabla_i^2 \psi_i(r_1) dr_1 - \sum_{x=1}^N \int \frac{Z_x}{r_{X_i}} \rho(r_1) dr_1 + \frac{1}{2} \iint \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 + E^{xc}[\rho] \quad (9)$$

where, n is the number of electrons and N is the number of nuclei.  $\frac{1}{2}$  for the  $V_{ee}$  term is required to prevent double counting each electron-electron term. The exchange-correlation term,  $E^{xc}[\rho]$  is given as:

$$E^{xc}[\rho] = \Delta T[\rho(r)] + \Delta V_{ee}[\rho(r)] \quad (10)$$

And, the ground state electron density can be written as a set of one electron orbitals given by:

$$\rho(r) = \sum_{i=1}^n |\psi_i(r)|^2 \quad (11)$$

In the Kohn-Sham approach the energy of a system is expressed as a divergence from the energy of an idealized system with noninteracting electrons. Since the idealized system's wavefunction can be precisely represented by a Slater determinant (wavefunctions and orbitals added as a mathematical convenience to get at the electron

density), the energy of the system can be computed exactly. The sole unknown term in the expression for the DFT energy is the exchange-correlation functional, which is found in the relatively small difference between the real and idealized system's energies. The primary challenge with DFT is estimating this functional. The Kohn-Sham equations can be obtained, in an analogy to the Hartree-Fock equations, from the energy equation by reducing the energy with respect to the Kohn-Sham orbitals. The KS orbitals are a collection of molecular orbitals that are qualitatively comparable to the orbitals of wavefunction theory. The molecular orbitals of the KS equations are enlarged using basis functions, and the energy is iteratively refined using matrix methods.

The one-electron KS operator plays a crucial role in the KS equations. It is constructed to mimic the behavior of the true many-electron system while being mathematically tractable. The key point to note is that the one-electron KS operator is conceptually similar to the Fock operator used in Hartree-Fock theory, with a notable difference: the exchange operator in Hartree-Fock theory is replaced by an exchange-correlation potential, denoted as  $V_{XC}$  in DFT.

In Hartree-Fock theory, the exchange operator accounts for the antisymmetry of the wave function, ensuring that the total wave function of the system obeys the Pauli exclusion principle. Additionally, it captures a portion of the electron correlation effects. However, in DFT, the exchange-correlation potential ( $V_{XC}$ ) serves a dual purpose:

**Exchange Contribution:** Like the exchange operator in Hartree-Fock theory,  $V_{XC}$  ensures the correct antisymmetry of the electron density, thus handling the exchange effects.

**Correlation Contribution:**  $V_{XC}$  also incorporates the effects of electron correlation beyond what is captured by the exchange term alone.

By including both exchange and correlation effects in the exchange-correlation potential, the one-electron KS operator effectively addresses both the Hartree and exchange terms in the total electronic energy expression, offering a computationally efficient approach to solve for the electronic structure of systems. This understanding sets the stage for discussing the approximations employed in DFT, such as LDA and GGA, which aim to approximate the exchange-correlation potential  $V_{XC}$  to make calculations feasible for real-world systems.

### 2.2.6.1.1. Local Density Approximation (LDA)

Local density approximation (LDA) is the simplest approximation that treats the electron density as roughly equivalent to that of a homogeneous electron gas and pairs two electrons with opposite spins in each KS orbital. This approximation assumes that exchange energy ( $E_{xc}$ ) at any point in space depends on the electron density at that point [15]. The exchange correlation function  $E_{xc}[\rho(\vec{r})]$  is expressed as:

$$E_{XC}^{LDA}[\rho] = \int \rho(\vec{r}) \epsilon_{XC}[\rho(\vec{r})] d\vec{r} \quad (12)$$

where,  $\epsilon_{XC}[\rho(\vec{r})]$  is the exchange-correlation energy per particle of a homogeneous electron gas of density  $\rho(\vec{r})$ . The exchange correlation energy  $\epsilon_{XC}[\rho(\vec{r})]$  is a combination of two parts:

$$\begin{aligned} \epsilon_{XC}[\rho(\mathbf{r})] &= \epsilon_X[\rho(\mathbf{r})] + \epsilon_C[\rho(\mathbf{r})] \\ \epsilon_C[\rho(\mathbf{r})] &= -3/4(3/\pi)^{1/3}(\rho(\mathbf{r}))^{1/3} + \epsilon_c^{vwn}[\rho(\mathbf{r})] \end{aligned}$$

The first term represents the exchange energy which is attributed to the exchange of an electron in an electron gas and the second term gives the description of correlation energy that evolved from the instantaneous movement of electron as a result of which they come close to each other and are repulsed. The implementation of the exchange correlation energy in the equation brings accuracy in the energy, structure and other properties of the system than the HF approximation having higher computational cost. However, the assumption of a homogeneous electron gas in the system results a rather poor calculated energy in LDA functional. LDA has been largely replaced by methods which use gradient corrected (“nonlocal”) functionals.

### 2.2.6.1.2. Generalized gradient approximation (GGA) and Meta-generalized gradient approximations (meta-GGA)

The assumption of homogeneous electron gas in LDA is not found to be robust in calculating the ground state energy of a molecular system. Hence, GGA functional are developed to bring accuracy in the exchange-correlation energy which accounts not only the electro density but also the gradient of the electron density. GGA assign one set of spatial orbitals to a-spin electrons, and another set of orbitals to b-electrons; this latter “unrestricted” assignment of electrons constitutes the local-spin-density approximation (LSDA). In 1998, a GGA functional B88 was developed by Becke. The  $E_{XC}^{GGA}$  is

comprised of two parts exchange and correlation  $E_{XC}^{GGA} = E_X^{GGA} + E_C^{GGA}$ . Further, many correlation functionals are developed; among them most popular is LYP functional developed by Lee, Yang and Parr. The other functionals are Perdew 1986 (P86), Perdew-Wang correlation functional (PW91), combination of B88 with LYP (BLYP) and Perdew-Becke-Ernzerhof exchange-correlation functional (PBE).

A refinement in the GGA functionals was done by implementing additional semi-local information ahead of the first order density gradient that have in GGA functional. These new set of functionals are termed as the meta-GGA functional and they explicitly depend on the Laplacian of the spin density or the local kinetic energy density. Examples of meta-GGA functional are TPSS, M06-L, B95, B98 etc.

### 2.2.6.1.3. Hybrid Functionals

To bring more accuracy in the exchange-correlation energy, further development of functional has been made in the approximation. In the hybrid functional, a percentage of Hartee-Fock exchange is introduced with the correlation functional and the energy is expressed as,

$$E_{xc} = E_x^{HF} + E_{xc}^{DFT} \quad (13)$$

On the popular hybrid functional B3LYP energy is given by,

$$E_{XC}^{B3LYP} = (1-a)E_x^{LDA} + aE_x^{HF} + bE_x^{B88} + cE_C^{LYP}$$

where, the parameters a, b & c are the empirical parameters that control the contribution of each term in the functional, determined by fitting experimental data and calculated data for the atomization energies, ionization potentials and proton affinities of second and third period elements. These parameters are typically optimized to provide the best overall performance across a range of chemical systems and properties. VWN is the Vosko, Wilk and Nusair functional and LYP is Lee, Yang and Parr functional. Other popular hybrid functionals are B3P86, B3PW91, O3LYP (Becke Three-Parameter Hybrid Functionals), MN15, M11, MN12-SX, M06, M06-2X, TPSSh, PBE0 etc. Range-separated hybrids are a type of exchange-correlation functional that aims to address the long-range and short-range electron-electron interactions differently. These functionals are particularly useful for systems with large band gaps, charge transfer excitations, and other properties where the standard hybrid functionals may fail. Examples of long range hybrids include LC-wHPBE, CAM-B3LYP and short range include wB97X, wB97X-D.

### 2.2.6.2. Basis Set

The atomic or molecular orbitals within an atom or molecule are represented mathematically by a basis set, limiting every electron to a specific area of space. The atomic orbitals are combined linearly to form the molecular orbitals in the Ab Initio approach. The basis sets are classified into two categories Slater type Orbitals (STOs) and Gaussian Type orbitals (GTOs) [52-53]. STOs are shown to be the exact solutions to the problem of hydrogen atoms and they offer a rather accurate description of atomic wave functions. Additionally, it is capable of approximating the proper behavior at the nucleus. Gaussian type orbitals (GTO), however, are developed as alternatives because of the computing challenges. GTOs are made by the linear combination of Primitive Gaussian Functions, described by a radial dependence and are frequently used basis functions. There are various types of basis function as the Minimal Basis Sets, Split Valence Basis Sets, Numerical basis sets, Polarization Functions, Diffuse Functions, Effective Core Potentials.

DFT is quite versatile and finds applications in various scientific fields with prominent uses being in used to study the electronic structure of materials, solids, including semiconductors and insulators helping to predict and understand properties such as electronic band structures, crystal structures, and magnetic properties. In the realm of chemistry, DFT is employed to investigate molecular structures, bond energies, reaction mechanisms, and spectroscopic properties, particularly useful for understanding the behavior of large molecules and complex systems. DFT is valuable in the study of catalysis, providing insights into the mechanisms of chemical reactions occurring on catalyst surfaces. This is crucial for designing more efficient catalysts for industrial processes. DFT plays a role in computational drug design by providing insights into the electronic structure and properties of drug molecules. It helps in understanding interactions with target proteins and predicting molecular properties. It can be applied to study enzyme reactions, protein-ligand interactions, and other biological processes at the molecular level. DFT helps in predicting and interpreting the optical properties of materials, such as electronic excitations and optical spectra. This is crucial for designing materials for applications in optics and photonics. Overall, Density Functional Theory is a powerful tool that spans various scientific disciplines, contributing to a deeper understanding of the fundamental properties of matter and guiding the development of new materials and technologies.



## 2.3. Softwares Packages used to carry out the computational work

### 2.3.1. The Gaussian Package

Gaussian is a computing program, first made available as Gaussian 70 by John Pople and his research team in 1970 [51]. Based on the foundational principles of quantum mechanics, the Gaussian algorithm predicts molecular structures, vibrational frequencies, relative energies, structures and properties of molecules which are experimentally challenging to observe *i.e.*, the short-lived intermediates and transition structures, thermo-chemical properties and kinetics of a molecular system or a chemical reaction. With progressing time and research, the software has been updated at regular intervals de-bugging the errors and making it feasible and viable with advancing computing facilities. The most popular version of Gaussian, is Gaussian 09; however, Gaussian 16 is the most recent version employed in the quantum chemical computations. As per availability we have used Gaussian09 for the various DFT related calculations [54].

### 2.3.2. The AMBER MD Package

Molecular dynamics (MD) allows the study of biological and chemical systems at the atomistic level on timescales of femtoseconds to milliseconds. Not only does it provide a means of tracking processes that are challenging to identify using experimental methods, it also enhances experiment. Numerous software packages exist for performing MD simulations of which AMBER is one of the widest used for nucleic acid simulations [55]. AMBER is a collection of simulation programs for biomolecules. It was started in the latter part of the 1970s and is been maintained up by an active development community since then. Originally developed under the leadership of Peter Kollman, a large group of people actively collaborated in the development of Amber: Carlos Simmerling at Stony Brook; Adrian Roitberg at the University of Florida; Tom Cheatham at the University of Utah; Ray Luo at UC Irvine; Ken Merz at Michigan State University; Maria Nagan at Stony Brook; Dan Roe at NIH; Junmei Wang at the University of Pittsburgh; Darrin York at Rutgers University; and many more.

The software refers to two different things. Firstly, it is a collection of molecular mechanical force fields for biomolecule simulation; these force fields are used in many different simulation programs and are available to the public. Secondly, it provides a collection of demonstrations and source codes for molecular simulation programs. As per availability we have used AMBER18 for the various MD related calculations [56].

## Bibliography

- [1] Parrill, A. L., and Lipkowitz, K. B. *Reviews in Computational Chemistry*, 2015.
- [2] Keith, J. A., Vassilev-Galindo, V., Cheng, B., Chmiela, S., Gastegger, M., Müller, K. R., and Tkatchenko, A. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chemical reviews*, 121(16):9816-9872, 2021.
- [3] Fernbach, S. and Taub, A. H. *Computers and Their Role in the Physical Sciences*. Routledge, 1970.
- [4] Cao, Y., Romero, J., and Aspuru-Guzik, A. Potential of quantum computing for drug discovery. *IBM Journal of Research and Development*, 62(6):6-1, 2018.
- [5] Shaker, B., Ahmad, S., Lee, J., Jung, C., and Na, D. In silico methods and tools for drug discovery. *Computers in biology and medicine*, 137:104851, 2021.
- [6] Kalita, J. M., Saha, A., Nath, D., and Patangia, U. Advances in computer aided drug design. *Universal Journal of Pharmaceutical Sciences and Research*, 1:17-22, 2015.
- [7] Pandey, S., and Singh, B. K. Current advances and new mindset in computer-aided drug design: A review. *The Pharma Innovation*, 6(8, Part B):72, 2017.
- [8] Lewars, E. G. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, Springer Dordrecht Heidelberg London, New York, 2<sup>nd</sup> edition, 2011.
- [9] Berkert, U. and Allinger, N. L. *Molecular Mechanics*, American Chemical Society, Washington, D.C, 1982.
- [10] Lewars, E. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, Kluwer Academic Publishers, Dordrecht, Netherlands, 2003.
- [11] Ringsdorf, H., Schlarb, B., and Venzmer, J. Molecular architecture and function of polymeric oriented systems: models for the study of organization, surface recognition, and dynamics of biomembranes. *Angewandte Chemie International Edition in English*, 27(1):113-158, 1988.
- [12] Acharya, K. R. and Lloyd, M. D. The advantages and limitations of protein crystal structures. *Trends in Pharmacological Sciences*, 26(1):10-14, 2005.
- [13] Karplus, M., and Petsko, G. A. Molecular dynamics simulations in biology. *Nature*, 347(6294):631-639, 1990.

- [14] Karplus, M. and McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9):646-652, 2002.
- [15] Leach, A. R. Molecular modelling: principles and applications. Pearson education, 2nd edition, 2001.
- [16] Grubmüller, H., Heller, H., Windemuth, A. and Schulten, K. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, 6(1-3):121-142, 1991.
- [17] Van Gunsteren, W. F. and Berendsen, H. J. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation*, 1(3):173-185, 1988.
- [18] Itoh, S. G., Okumura, H., and Okamoto, Y. Generalized-ensemble algorithms for molecular dynamics simulations. *Molecular Simulation*, 33(1-2):47-56, 2007.
- [19] Barclay, P. L., and Zhang, D. Z. Periodic boundary conditions for arbitrary deformations in molecular dynamics simulations. *Journal of Computational Physics*, 435:110238, 2021.
- [20] Lemak, A. and Balabaev, N. On the Berendsen thermostat. *Molecular Simulation*, 13(3):177-187, 1994.
- [21] Bussi, G., Donadio, D. and Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):1-8, 2007.
- [22] Martoňák, R., Laio, A. and Parrinello, M. Predicting crystal structures: the ParrinelloRahman method revisited. *Physical Review Letters*, 90(7):075503(1-4), 2003.
- [23] Zhakhovskii, V. V., Inogamov, N. A., Petrov, Y. V., Ashitkov, S. I., and Nishihara, K. Molecular dynamics simulation of femtosecond ablation and spallation with different interatomic potentials. *Applied Surface Science*, 255(24):9592-9596. 2009.
- [24] González, M.A. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique*, 12: 169-200, 2011.
- [25] Baker, C. M. Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 5(2):241-254, 2015.
- [26] Leontyev, I.V. and Stuchebrukhov, A.A. Electronic polarizability and the effective pair potentials of water. *Journal of Chemical Theory and Computation*, 6(10):3153-3161, 2010.

- [27] Liu, Y.-P., Kim, K., Berne, B., Friesner, R. A. and Rick, S. W. Constructing ab initio force fields for molecular dynamics simulations. *The Journal of Chemical Physics*, 108(12):4739-4755, 1998.
- [28] Aina, A.A., Misquitta, A.J. and Price, S.L. From dimers to the solid-state: Distributed intermolecular force-fields for pyridine. *The Journal of Chemical Physics*, 147(16):161722(1-12), 2017.
- [29] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. and Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179-5197, 1995.
- [30] Beachy, M. D., Chasman, D., Murphy, R. B., Halgren, T. A. and Friesner, R. A. Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. *Journal of the American Chemical Society*, 119(25):5908-5920, 1997.
- [31] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. and Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577-8593, 1995.
- [32] Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. and Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *Journal of Chemical Theory and Computation*, 9(9):3878-3888, 2013.
- [33] Gotz, A.W., Williamson, M.J., Xu, D., Poole, D., Le Grand, S. and Walker, R.C., Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *Journal of Chemical Theory and Computation*, 8(5):1542-1555, 2012.
- [34] Levitt, M. Molecular dynamics of native protein. I. Computer simulation of trajectories. *Journal of Molecular Biology*, 168(3):595-617, 1983.
- [35] Schuler, L. D., Daura, X. and Van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *Journal of Computational Chemistry*, 22(11):1205-1218, 2001.
- [36] Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. and Jorgensen, W. L. Evaluation and re-parametrization of the OPLS-AA force field for proteins via comparison with

- accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28):6474-6487, 2001.
- [37] Senn, H. M. and Thiel, W. QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition*, 48(7):1198-1229, 2009.
- [38] Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O. and Shaw, D. E. Longtimescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology*, 19(2):120-127, 2009.
- [39] Hedger, G. and Sansom, M. S. Lipid interaction sites on channels, transporters and receptors: recent insights from molecular dynamics simulations. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1858(10):2390-2400, 2016.
- [40] Childers, M. C. and Daggett, V. Insights from molecular dynamics simulations for computational protein design. *Molecular Systems Design & Engineering*, 2(1):9-33, 2017.
- [41] Liu, X., Shi, D., Zhou, S., Liu, H., Liu, H., and Yao, X. Molecular dynamics simulations and novel drug discovery. *Expert opinion on drug discovery*, 13(1):23-37, 2018.
- [42] Salo-Ahen, O. M., Alanko, I., Bhadane, R., Bonvin, A. M., Honorato, R. V., Hossain, S., ... and Vanmeert, M. Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes*, 9(1):71, 2020.
- [43] Shukla, R., and Tripathi, T. Molecular dynamics simulation in drug discovery: opportunities and challenges. *Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design*, 295-316, 2021.
- [44] Chivian, D., Robertson, T., Bonneau, R., and Baker, D. Ab initio methods. *Structural bioinformatics*, 44:547-557, 2003.
- [45] Carsky, P., and Urban, M. *Ab initio calculations: methods and applications in chemistry* (Vol. 16). Springer Science & Business Media, 2012.
- [46] (a) Hartree, D. R. The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume, 24, pages, 89-110. Cambridge University Press, 1928. (b) Hartree, D. R. The wave mechanics of an atom with a non-coulomb central field. Part II. Some results and discussion. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume, 24, pages, 111-132. Cambridge University Press, 1928. (c) Hartree, D. R. The wave mechanics of an atom with a

- non-coulomb central field. Part III. Term values and intensities in series in optical spectra. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume, 24, pages, 426-437. Cambridge University Press, 1928.
- [47] Fock, V. Approximation method for solving the quantum mechanical multibody problem. *Journal of Physics*, 61:126-148, 1930.
- [48] Stewart, J. J. Optimization of parameters for semiempirical methods II. Applications. *Journal of computational chemistry*, 10(2):221-264, 1989.
- [49] Besler, B. H., Merz Jr, K. M., and Kollman, P. A. Atomic charges derived from semiempirical methods. *Journal of computational chemistry*, 11(4):431-439, 1990.
- [50] Thiel, W. Semiempirical quantum-chemical methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(2):145-157, 2014.
- [51] Levine, I. N. *Quantum chemistry*, Prentice Hall, Engelwood Cliffs, NJ, 5<sup>th</sup> edition, 2000.
- [52] Boese, A. D., Martin, J. M., and Handy, N. C. The role of the basis set: Assessing density functional theory. *The Journal of chemical physics*, 119(6):3005-3014, 2003.
- [53] Kreutzer, J., Blaha, P., & Schubert, U. Assessment of different basis sets and DFT functionals for the calculation of structural parameters, vibrational modes and ligand binding energies of Zr<sub>4</sub>O<sub>2</sub> (carboxylate) 12 clusters. *Computational and Theoretical Chemistry*, 1084:162-168, 2016.
- [54] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., ... and Fox, D. J. 2009.
- [55] <https://ambermd.org/>
- [56] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2018), AMBER 2018, University of California, San Francisco. pdf.