CHAPTER 1

An overview of Alzheimer disease, X-ray Crystallography, Molecular docking, Intermolecular interactions, Molecular dynamics and QM/MM based Charge density analysis

1.1 Introduction of Alzheimer disease

In a neurology conference held on 1907, the German psychiatrist Alois Alzheimer was presented new disorder which was diagnosed for an old woman and defined a rare pre-senile dementia occurring among aged people, named as “Alzheimer disease (AD); it initially affects a person’s ability to carry out daily functions and leads to death [Parihar 2004; Small 2006; Stelzmann 1995]. The AD patient has cognition, memory impairment and unpredictable behavior. And, AD is an irreversible, complex, multifactorial, progressive deterioration, chronic, common cause of dementia and neurodegenerative disorder. Further, the current treatments with cholinesterase inhibitors can temporarily alleviate the cognitive symptoms of the illness and slow down the neuronal cell death [Scarpini 2003; Mattson 2004; Masters 2015]. In order to provide exact or suitable therapy for AD, many scientists and pharmaceutical companies are optimistically working to identify the exact disease mechanism. Till date there is no medical treatment available to cure the disease completely. Globally, ~50 million people are living with AD and every year ~8 million people are suffered newly [Alzheimer's Association 2018]. Due to that, the CEO of Alzheimer’s Disease International (ADI) addressed a global plan (2017-2022) on dementia at the 70th session of the World Health Assembly in Geneva on 29th May 2017 and mentioned that “the most devasting disease of the 21st century is Dementia and the future treatment depends on research. Therefore, the ADI proposed a global plan that the nationally 1% of the cost should be devoted to dementia research”. The present study focused onto identify a potential inhibitor to control the level of Alzheimer disease. Basically, many diverse factors are leading to neuronal cell death such as low level of acetylcholine, deposition of amyloid beta, aggregation of tau proteins, oxidative stress, neuro-inflammation, excitotoxicity, genetic and heredity factors. Some of the recent reports [Rashid 2014; Choudhary 2015] also outline the about AD and it was characterized by different hypothesis and are named as cholinergic hypothesis, amyloid hypothesis, tau hypothesis and others (oxidative stress and genetic). Hence, the present study is targeted to inhibit the
important enzymes acetylcholinesterase (AChE), β-secretase (BACE1) and Glycogen Synthase Kinase 3 (GSK3β).

1.1.1 Acetylcholinesterase

The cholinergic hypothesis proposes that acetylcholinesterase (AChE) is a serine protease, which plays a vital role in the central and peripheral nervous system, where catalyzes the hydrolysis of the neurotransmitter acetylcholine (ACh) in cholinergic synapses and subsequently, it can affect a number of pathogenic processes [Kumar 2015]. AChE is the most significant enzyme in memory and cognition [Bartus 1982]. Due to the abnormal activity of AChE in the AD patients, there are five AChE inhibitors were approved by FDA as anti-AD drugs. The plant derived drugs such as galantamine, huperzine A, rivastigmine and the synthetic drugs Donepezil and Tacrine are being used for the treatment of AD [Sun 2011; Yiannopoulou 2013]. However, all these drugs are not completely cure the disease [Casey 2010]. Hence, the potential drugs are essential to restrain the Alzheimer disease [Murray 2013; Anekonda 2005]. For the design of new AChE inhibitors, the detailed structural knowledge of AChE is essential [Dvir 2010; Wiesner 2007]; so far, a considerable number of AChE crystal structures were determined from different organisms, such as Homo sapiens, Tropedo Californica, Drosophila melanogaster, Mus musculus and Electrophorus electricus. The molecular structure of recombinant human acetylcholinesterase (rhAChE) was achieved from the Homo sapiens, which reveals the exact size and shape of active site of this enzyme [Dvir 2010]. The active site of AChE contains different types of active sub-sites namely; acyl binding pocket, anionic sub-site, esteratic subsite or catalytic triad, omega loop, oxyanion hole and peripheral anionic subsite. This active site occupies 20 Å of narrow deep gorge at the bottom of the AChE. The molecular mechanism of AChE is that the catalytic site contains serine amino acid, the hydroxyl group of serine directly interacts with acyl carbon on acetylcholine to break the choline. Then, the acetylated serine is hydrolyzed by water to form acetate; now, the AChE is capable for next function.

1.1.2 Beta (β) Secretase

According to amyloid hypothesis [Selkoe 2016; Hardy 2002], the accumulation of amyloid beta (Aβ) triggers physiological changes in the brain and eventually provoking cognitive dysfunction. The insoluble and neurotoxic Aβ42 is generated from the large transmembrane amyloid precursor protein (APP) by the
sequential action of β-Secretase (BACE1) followed by γ-secretase and then accumulated in senile plaques; it leads to neuronal cell loss [Bayer 2010]. Moreover, the abnormal deposition of Aβ exists in the clinical report of AD patients [LaFerla 2007]. To reduce the abnormal aggregation, potential inhibitor is required to control the BACE1 activity. Further, the structural part of BACE1 is essential to design a good drug molecule. BACE1 is the aspartate protease with N-terminal domain, transmembrane region, cytosolic domain and strand residues [Barman 2013]. Its active site has different potential region such as catalytic dyed, flap region and 10s loop [Hong 2002]. Several research groups are independently focused on the development of BACE1 inhibitors. The discovery of verubecestat and acylguanidine molecules are the successful challenges of structure-based design, medicinal chemistry optimization, in vivo screening in the improvement of BACE1 inhibitors [Keränen 2017; Thaisrivongs 2016]. This evidence gives a strong support that the research has concentrated on the development of new drugs for the treatment of AD. Additionally, the clinical results of phase 1 and 2 reveal that the proposed molecules have positive effect against BACE1 and it also reduces the Aβ accumulation in the animal models [Kennedy 2016].

1.1.3 Glycogen synthase kinase 3

According to tau hypothesis [Iqbal 2005; Williams 2006], the tau protein is a microtubule-associated protein which plays an important role in the stabilization of neuronal cytoskeleton. The three-dimensional structure of tau protein is divided into different regions such as acidic region, N and C-terminals, proline-rich domain and tau binding region (repeated R1, R2, R3, R4 domains). The level of phosphorylation is required to the routine function of tau. In the Alzheimer disease, these repeating domains are highly phosphorylated and lead to loss the stabilization. Hence, this abnormal phosphorylation of tau protein is confirmed due to the over expression of Glycogen synthase kinase 3 (GSK3β) enzyme and could be toxic. The animal model shows that the increases of GSK3β activity in tau hyperphosphorylation induce the memory deficit, increase the amyloid β production and reduce the acetylcholine synthesis. Therefore, a potential inhibitor is required to control this GSK3β activity [Cohen 2001; Bhat 2003; Calcul 2012].

To understand the binding mechanism of inhibitors of AChE, BACE1 and GSK3β, the present study employed different experimental and computational
techniques such as X-ray crystallography, charge density analysis, QM/MM, molecular docking and molecular dynamics.

1.2 X-ray Crystallography and Charge density analysis

To determine the three-dimensional structure of molecules, often X-rays are being used due to its wavelength, which is similar to the atomic spacing in a crystal lattice. The structure and conformational studies of molecules are very much essential to understand the structure-activity relation of biomolecules and it plays an important role in the rational drug design. Nowadays, the X-ray crystallography is a common technique to investigate the crystal structure of materials.

![Diagram](image-url)

**Figure 1.1:** Flow chart shows the steps involved in experimental charge density analysis
X-ray Crystallography and Charge density analysis

Chapter 1

X-rays are generated by a cathode ray tube and further filtered to produce monochromatic radiation, collimated to concentrate and directed towards the sample. These diffracted X-rays are satisfy Bragg's Law \((2d\sin\theta = n\lambda)\) and then detected, processed and counted. All possible diffraction directions of the lattice should be attained while changing the geometry of the incident rays, the orientation of the centered crystal and the detector. Figure 1.1 shows the steps involved in crystal structure and topological analysis such as crystallization, crystal selection, intensity data collection, data reduction, structure solution and structure refinement [Massa 2004; Zachariasen 1945; Stout 1989; Velmurugan 2008].

1.2.1 Crystallization

Crystallization is the process involves the nucleation to arrange ions, atoms or molecules randomly in the solid state from the fluid phase of saturated solutions [Mersmann 2001; Glynn 1990]. There are several techniques are available for crystallization of small molecules such as slow evaporation, slow cooling, diffusion methods, etc. In the present study, the crystals were grown from slow evaporation method.

1.2.2 Selection of single crystal

The selection of single crystal with good quality, morphology leads to obtain an excellent X-ray intensity data. The selected crystal should be a reasonable size, shape without any defects like cracks, twinning, etc. The scattering power and the intensity are weak when the selected crystal is either too small or large. The modeling of the structure is affected by the variation in the scale factor and large absorption.

1.2.3 Data collection

This is an important step in the X-ray diffraction intensity measurement of crystal. To perform the experimental charge density analysis, require high resolution X-ray diffraction intensity to the resolution limit at least \(\sin\frac{\theta}{\lambda} = 1.0 \text{ Å}\). During the data collection, the crystal must be cooled to reduce the thermal motions of atoms in the crystal. This facility used to measure the high-resolution diffraction intensity. The requirement of good redundancy and completeness up to high resolution data set is governed by multiple measurements of the same reflection and the atomic thermal motions are controlled by the low temperature.
1.2.4 Data reduction

The data reduction is the process which converts the integrated intensity to structure factor amplitudes after correcting the spot shape, subtraction of the background intensity, crystal decay, averaging and merging of symmetry equivalent data. The intensity data might be corrected by Lorentz, polarization and absorption effects in all cases \( I = I_0 e^{-\mu t} \), \( I_0 \) is the incident intensity, \( \mu \) is the absorption coefficient and \( t \) is the thickness of the crystal. This developed procedure helps to retain the accurate reproducible data [Blessing 1997].

1.2.5 Structure solution

The structure solution is playing a key role in the crystal structure determination to locate the exact atomic position in the unit cell and electron density. However, it requires both magnitudes of structure factor and phases of each atom in the unit cell.

\[
F_{hkl} = |F_{hkl}| e^{i\phi_{hkl}}
\]  

(1.1)

The electron density at \( x, y, z \) position in the unit cell can be expressed as

\[
\rho_{x,y,z} = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F_{hkl} e^{-2\pi i(hx+ky+lz)}
\]

(1.2)

The phase information of diffraction measurements cannot be directly obtained from the diffraction pattern. Hence, the electron density is not possible to determine from the Fourier Transform. This fact is known as the “Phase problem” in crystallography [Cowtan 2003]. Over the years, this problem was tackled successfully with the help of mathematical methods such as

1. Trial & Error Method
2. Patterson Method
3. Direct Method
4. Heavy Atom Method
5. Isomorphous Replacement Method

These methods are effectively used to locate the approximate positions of the atoms of trial structure of a molecule in the unit cell. In this thesis, the direct methods have been used to solve the crystal structures.

1.2.6 Direct method: This is used directly to calculate the phases by simple mathematical procedures from a single set of X-ray intensities. The basic postulates of direct method is that the electron density is positive everywhere (positivity) and
the atoms are considered spherically symmetric (atomicity). The structure factor amplitudes and phases are linked with electron density through Fourier transformation. A mathematical constraint on the function \( \rho(x) \) imposes a corresponding constraint on the structure factor. The evaluation of phase \( \phi_{hkl} \) is not determined directly from the constraint, several steps are involved in the direct method such as

- Convert the observed structure factors \( |F_{hkl}| \) into normalized structure factors \( |E_{hkl}| \).
- The triplets (triple phase relations) and quartets phase relations (four phase relations) were considered.
- To assign initial phases to be selected from the origin.
- The phase propagation and refinement were carried out using tangent formula.
- The reliability of the phases was confirmed from Combined Figure of Merit.

In the calculation of phases, the structure invariants and structure semi invariants are used by Direct methods [Hauptman 1986; Velmurugan 2008].

1.2.7 Structure refinement

Refinement is a major role in the modelling of electron density to refine and improve the atomic co-ordinates of the structure obtained from the direct methods or any other phase determining methods to get the actual model of the structure. This crystallographic refinement model supports to reduce the errors and the refinement will take until the model difference between the observed and calculated should be less. The residual disagreement for the exact refinement is 0.0 and the total disagreement is ~ 0.59. The end of the refinement process is called reliability factor or R-factor [Sheldrick 2008, 2015].

\[
R_1 = \frac{\sum_{hkl} |F_o| - |F_c|}{\sum_{hkl} |F_o|} \tag{1.3}
\]

\( |F_o| \) is the magnitude of the observed structure factors and \( |F_c| \) is calculated structure factors. The summation was taken over all the observed reflections and the R value should be minimum at the converging point of the refinement which shows that the model is closure to the actual structure. The suitable weighting scheme can be applied at the end of refinement procedure and the weighted R-factor is nearly 3 to 4 times that of \( R_1 \).
\[ wR_2 = \left( \frac{\sum W_i (|F_0| - |F_c|)^2}{\sum W_i |F_0|^2} \right)^{1/2} \]  \hfill (1.4) 

The Goodness of Fit is always based on \( F^2 \)

\[ \text{GooF} = S = \left[ \frac{\sum (w(F_o^2 - F_c^2)/ (n-p))}{n-p} \right]^{1/2} \]  \hfill (1.5) 

where, the \( n \) is the number of reflections, \( p \) is the total number of parameters refined.

\[ w = \frac{1}{\sigma^2 (F_o^2) + (aP)^2 + bP} \]  \hfill (1.6) 

where, \( a \) and \( b \) are the constants

\[ P = [2F_c^2 + \text{Max} (F_o^2, 0)/ 3] \]  \hfill (1.7) 

1.2.8 Spherical atom model

This is known as classical structure determination or independent atom model (IAM) which is commonly used to determine the three-dimensional structure and geometrical parameters of molecule. The negative charges are highly localized around the nuclei due to the dominance of the core scattering in the total scattering of an atom. Therefore, the atoms are considered as a superposition of spherical atomic densities. The IAM modeled structure is generally used as initial step for the multipole refinement as well as for the wave function determination with the help of quantum mechanics.

1.2.9 Aspherical atom model

This is an advanced crystallography method which allows to determine the topological and electrostatic properties of the molecule. The IAM model is easy to locate the nuclear position of heavy atom than the small like hydrogen due to the lack of core shell electrons. Therefore, the charge density analysis is aimed to analyze the electron distribution and carry out the detailed information of chemical bonding in the crystal environment. Indeed, the topology of electron density gives deeper insights about the electron density at the bond critical point (bcp) and the Laplacian of electron density is second derivative of electron density which provides the charge concentration and depletion at the bcp. The electrostatic properties are exploring the chemical reactivity region of the molecule. The sum of core and valence electron density is taken as a total electron density of an atom in the molecule. In which, the valence electron densities are deformed in the molecule due to the bonding nature. The electron density portion of an atom is

\[ \rho_{\text{atom}}(r) = \rho_{\text{core}}(r) + \rho_{\text{valence}}(r) + \rho_{\text{deformation}}(r, \theta, \phi) \]  \hfill (1.8)
1.2.10 Kappa formalism

The kappa formalism is an improved method of IAM which allows the charge transfer between atoms due to electron distribution in the bonding region [Coppens 1997; Su 1998; Hansen 1879].

\[ \rho_{\text{atom}}(r) = \rho_{\text{core}}(r) + P_{\text{valence}} \kappa^3 \rho_{\text{valence}}(\kappa r) \]  
(1.9)

Where, the \( P_{\text{valence}} \) shows the valence population and the \( \kappa \) indicates the coefficient of the spherical expansion/contraction (the atom is contacted means the \( \kappa > \) unity and the atom is expanded means the \( \kappa < \) unity).

1.2.11 Aspherical atom formalism

According to Hansen-Coppens Multipolar Model [Stewart 1973; Hansen 1879], the electron density of an atom can be divided into three components such as spherical core density \( \rho_{c}(r) \), spherical valence density \( \rho_{v}(\kappa r) \) and aspherical valence density \( \rho_{d}(\kappa'r) \).

\[ \rho_{\text{atom}}(r) = P_{c}\rho_{c}(r) + P_{v}\kappa^3 \rho_{v}(\kappa r) + \sum_{l=0}^{l_{\text{max}}} \kappa'^3 R_{l}(\kappa' r) \sum_{m=0}^{l} P_{lm} Y_{lm}(\theta, \phi) \]  
(1.10)

Where, \( \kappa' \) is the aspherical valence density, \( R_{l} \) is the radial functions, \( P_{lm} \) is the multipole population and \( Y_{lm}(\theta, \phi) \) is the density normalized real spherical harmonic function. The \( \rho_{c}(r) \), and \( \rho_{v}(\kappa r) \) are calculated from Hartree-Fock (HF) atomic wave functions. In the radial function \( (R_{l}) \), the \( l \) indicates the different level of multipole functions [dipoles \( (l=1) \), quadrupoles \( (l=2) \), octupole \( (l=3) \) and hexadecapoles \( (l=4) \)].

To obtain the total population of an atom, the set of monopole terms are composed to multipole expansion. Therefore, this Hansen-Coppens multipolar model was proposed that the electron density is expanded with respect to nuclear motion [Gatti 2012].

1.2.12 Deformation density

The deformation density is defined as the difference between the total electron density and IAM modeled promolecule density [Spackman 1986]. The deformation density is strongly based on the function and population of multipole refinement, the deformation density map has been plotted to understand the recognition of bonding features, diagnostic purposes and to visualize the charge accumulation and lone-pair regions in the molecule.

\[ \Delta \rho_{\text{def}}(r) = \rho(r) - \rho_{\text{pro}}(r) \]  
(1.11)
\[
\Delta \rho_{\text{def}}(r) = \frac{1}{V_{\text{cell}}} \sum_{H} [F_{\text{OBS}}(H) - F_{\text{IAM}}(H)] e^{-2\pi H r_j}
\]

\[
\Delta \rho_{\text{exp}}(r) = \frac{1}{V_{\text{cell}}} \sum_{H} |F_{\text{OBS}}(H)| e^{i\phi_{\text{mult}}} - |F_{\text{IAM}}(H)| e^{i\phi_{\text{IAM}}} e^{-2\pi H r_j}
\]

### 1.3 Molecular mechanics

The molecular mechanics is a classical computational method, which is used to determine the potential energy surface of the molecular system with the help of potential functions [Leach 2001]. The potential functions explain the force constants and equilibrium values. There are some regulations that the molecular mechanics treats the electrons and the nucleus as a perfect sphere, the bonds are treated as springs. The sum of bond stretching, angle bending, torsional energies, and non-bonding interactions are known as “potential energy function” and it is easily determined computationally by using the force field. There are huge number of force fields are available and the popular force fields are classical force field (AMBER, CHARMM, CVFF, GROMOS, OPLS and etc) and polarizable force field (PFF, CFF, PIPE and etc) [MacKerell 2001; Krishnan 2008].

\[
E = E_{\text{covalent}} + E_{\text{non-covalent}}
\]

\[
E_{\text{covalent}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}}
\]

\[
E_{\text{non-covalent}} = E_{\text{electrostatic}} + E_{\text{van der Waals}}
\]

The main application of this method is to calculate the energy, motion and interactions of the small (<100 atoms) as well as macromolecules (>10000 atom) with the help of molecular dynamics.

### 1.4 Molecular docking

In the molecular modeling, the molecular docking is one of the most developed computational approach used to determine the ligand binding mode in the active site of protein. The basic definition of the molecular docking is “the ability of protein interacts with ligand molecule to form a supramolecular complex, which may enhance or inhibit its biological function” [Roy 2015]. Based on the binding pose and the binding affinity, the molecular docking can be classified into protein-ligand docking, protein-nucleic acid docking, protein-protein docking [Halperin 2002; Elokely 2013]. Figure 1.2 shows the steps involved in the molecular docking.

Due to the importance of structure-based drug design, the ligand-protein docking is being considered as an important research area.
There are several software’s available to perform molecular docking and to predict potential molecules to inhibit proteins [Roy 2015; Halperin 2002; Elokely 2013; Doss 2014]. In general, when the ligand present in the active site of protein, it includes the conformational modification in protein. It worth to note that the protein flexibility, size of the cavity and degrees of freedom of protein are the important components, require to explore the binding of ligands in the active site.

**Figure 1.2:** The flowchart clearly explains in details of the molecular docking.

According to this concern, it is categorized into soft docking (both ligand and protein flexible), side-chain flexibility, molecular relaxation (both side-chain and backbones are flexible in the active site region) and protein-ensemble docking (homology modeling). Further, the ligand sampling is an another important
component in protein-ligand docking which helps to generate the possible ligand orientation in the binding site of desired protein. The sampling algorithms are applied to identify the exact molecular surface of the binding site based on the shape matching (conformationally fix the ligand into protein), systematic search (blind and rank) and stochastic approach (orientation and conformation randomly).

The accuracy of the protein-ligand docking is confirmed from the scoring function; in which, the swiftness and precision are the important aspects of the scoring function and also would be the proficient as well as consistent computationally. There are several scoring functions are proposed according to their methods of derivation.

1. **Force filed scoring function**: the binding energy is separated into vdW, electrostatic, bond bending/stretching/torsional energies

2. **Empirical scoring function**: The set of weighted empirical energy are adding to binding energy (vdW, electrostatic, hydrogen-bonding, desolvation, hydrophobicity and entropy).

3. **Knowledge-based scoring function**: The binding energy is determined from the structural information’s.

4. **Clustering and entropy-based scoring function**: The binding energy is calculated from the binding modes and different clusters

**Methodology of docking**

- In the docking analysis, the protein structure of our interest is essential that can be determined either from X-ray crystallography or nuclear magnetic resonance (NMR). Before the selection of protein structure, the important criteria are that the protein structure is derived with high resolution (<2 Å), low B-factor and complex (protein with cofactor). Moreover, the protein structure can be predicted by homology modeling technique if the structure is not available in the protein data bank (PDB). The sequence identification
Molecular docking Chapter-1

of the target is >50%, seems to be correct homology model. The interest of the ligand selection is fully based on the binding site character, size and intermolecular interactions. Therefore, the protein structure of interest is highly important to perform docking study.

- The **ligand preparation** is the initial step in the docking process which requires three dimensional coordinates of the ligands. The ionization and tautomeric states should be assigned. In the **virtual screening** method, the ligands are selected based on testing, scientific expertise, chemical intuition and blindly by compound ranking (score and pose). The important thing is that the ligands should obey Lipinski’s “rule of five” which gives that the ligand molecules of interest is never exceed 5 hydrogen bond donor, 10 hydrogen bond acceptor, molecular mass is <500 D, the lipophilicity is >5 and exhibit pharmacokinetic properties (ADMET).

- In the **protein preparation** step, the retrieved protein structure must be minimized by using software-specific force fields after adding hydrogen atoms, formal bond order, charges, side chains, missing residues as well as removal of water molecules beyond 5 Å of active site of the protein.

- After ensuring that the protein and ligands are present in the correct form to docking, the receptor grids are going to generate using grid generation program (if need); generally, the grid box is created at the center of the active site of protein. In the induced-fit method, the grid is generated automatically in the cocrystal region.

- The ligand-protein complex is ready after the ligands re-docked into the active site of the protein. Then, the interaction energies and intermolecular interactions are calculated to compare with those of protein-ligand and the crystal structure to confirm whether the ligand molecule bind with the active site of protein or not. The docking pose of each conformation is ranked based on their score value and interactions. The main application of the docking is in the drug design and drug discovery [Roy 2015; Halperin 2002; Elokely 2013; Doss 2014].
1.5 Intermolecular interactions

According to Richard Feynman, the intermolecular interaction was defined as “The world is built by atoms that repel each other at short distances as well as attract at longer ones.” The world would disintegrate within femtosecond if the intermolecular interactions were suddenly switched off [Piela 2014]. The intermolecular interaction plays an important role in the molecular recognition and it acts between the molecules like protein-carbohydrates/fats/nucleic acids and also within these molecules to form secondary structure due to attractive or repulsive force between non-bonded atoms [Hunter 2001; Babine 1997; Williams 2003; Banner 2003]. There are several types of interactions exist between protein and ligand such as hydrogen bonding, van der Waals, hydrophobic/hydrophilic, electrostatic, halogen bond, etc.

**Hydrogen bond:** The most important and essential interaction in biology is hydrogen bonding interaction which is formed by the attraction of lone pair of electronegative atom (nitrogen, oxygen and fluorine) with hydrogen [Hubbard et al. 2010; Williams 2003; Banner 2003; Scheiner, 2001]. The hydrogen bond is generally known as weak bond and also considered as partially covalent or electrostatic in nature. The hydrogen bond is stronger and more directional than the van der Waals interaction. The protein structures have huge number of hydrogen bonds due the amide and carbonyl groups present in the backbone of protein. The strength of hydrogen bond is estimated from both experimental and theoretical reports exhibit the interaction energy greater than 25 kcal/mol, bond length below 2.5 Å and the bond angle less than <120˚ [Pace 2014].

**Van der Waals interaction:** The weak attractive force exists between atoms or molecules. The van der Waals forces are the short-range forces at the distance is proportional to $1/r^7$ and it is observed in the condensation of gas to liquid, liquid to solid. The dipole-dipole force, inductive force and dispersion forces are considered as van der Waals type of interactions. These interactions are weaker than hydrogen bond and not directional.

**Hydrophobic/hydrophilic interactions:** The hydrophobic (water hates) interaction is defined as the tendency of water to exclude non-polar molecules. This is not spontaneous while mixing hydrophobes (non-polar molecules) and water molecules; however, it is highly possible between hydrophobes due to high enthalpy. Whereas,
the hydrophilic (water loves) type of interaction is defined as the charged part of the molecules are attracted to water; hence, the sum of all the intermolecular forces leads to hydration process [Du 2016; Barratt 2005].

**Electrostatic interaction:** the electrostatic interactions are observed between cationic and anionic character of atoms, it is either attractive or repulsive depends on sign of charge. The electrostatic interactions are sometime called as ionic bond, strong interaction, act as long rage (due to the distance between ions are $1/r^2$) [Williams 2017].

**Halogen bonding:** The attractive non-covalent interactions are formed between an electrophilic region of halogen atom to nucleophilic region of another molecule. A high effective nuclear charge of halogen atom leads to the possibility of halogen bonding and it is directional. The potential significance of this interaction is responsible for ligand binding, folding and recognition [Sirimulla, 2013; Zhou 2009; Auffinger 2004].

**Other interactions:** The $\pi-\pi$ stacking is another attractive non-covalent interaction occurred between aromatic rings which present in the nucleobase staking of DNA, RNA and protein folding. The $\pi-\pi$ interaction is divided into different types based on the substituent effects, i.e. sandwich and T-shaped interactions. Moreover, the alkyl and aryl groups are forming the interaction to stabilizing the protein structures [Ribas 2002; Sinnokrot 2002, 2004; Meyer 2003; Hunter 2004; Ringer 2006; Baron 2010].

### 1.6 Molecular dynamics

The molecular dynamics (MD) is one of the first simulation methods to understand the dynamics of the liquids using the computer simulation, which is widely used to investigate the dynamical behavior of the structure of small and macromolecules at the atomic level [Leach 2001; Tamar 2002; Frenkel 2002; McCammon 1987]. The MD allows to study the microscopic behavior of the chemical process with the help of high-speed computer. The MD methods are classified into two groups based on mathematical formalism of the modeling, i.e. classical MD simulation (using the classical mechanics laws) and quantum MD simulation (using the quantum equations). The quantum MD simulation gives more new insights at electronic level and solving several biological problems than the classical one. The classical MD is comprising thousands of atoms over time scales.
Molecular Dynamics

Chapter 1

of nanoseconds. Nowadays, more researchers are using the classical MD than the quantum MD due to the requirement of high computation. Particularly, this method of MD simulations is used computationally to design the new drug molecules and to predict the chemical, biological properties faster than the experimental process like synthesis, characterization, etc.

MD simulation is used to solve the classical equations of motion numerically for a group of atoms. According to Newton’s 2\textsuperscript{nd} law of motions is that “the applied force is equal to the rate of change of momentum”. In the molecular dynamics, the force \( F_i \) can be determined while \( i \) atom is treated with mass \( m_i \) and charge \( q_i \).

\[
F_i = m_i a_i; \ a_i = \frac{d^2 x_i}{dt^2}; \ F_i = \frac{d^2 x_i}{dt^2}
\]

(1.17)
The initial set of position and velocities of the atoms can be evaluated with the subsequent time in the computer; in which, the atoms or molecules are bumping each other, vibrating, wandering around in the fluid, oscillating, evaporating (if there is a free surface) and so on. Therefore, the Feynman is quoted that “the living things of everything do can be understood in terms of the jiggling and wiggling of atoms” [Vlachakis 2014; Miller 2008]. The MD simulation can be performed on system (thousands of atoms) from the time range of few picoseconds to hundreds of nanoseconds. In this regard, the time integration of the equation of motion is expressed from the Verlet algorithm to define the position of atom \( r_i \) at time \( t \). This time integration algorithms are based on finite difference methods, where the time is discretized on a finite grid, the time step \( \Delta t \) being the distance between consecutive points on the grid [Salomon-Ferrer 2013; Tuckerman 2000; Van Gunsteren 2006; Deng 2009; Meller 2001]. The time evaluation of the system can follow,

\[
r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + \frac{F_i}{m_i}\Delta t^2
\]

(1.18)

1.6.1 Methodology of the simulation

During the MD simulation, there are several steps are involved. In which, the MD simulation can start with initial positional coordinates to apply the force on individual atoms in the system. The following steps may be required to understand the behavior of the atoms or molecules at the given time.

**Preparation:** This is the crucial step and many potential problems can arise. The required data file should be in the PDB file format including the atom names, atom numbers, residues names, residue numbers, chain identifier (if more than one chain)
and the coordinates of all atoms. The non-protein structures like ligand, co-factors, water molecules, monoatomic ions may cause problem unless the loading of extra libraries. Further, some of the residues name are not rendered in PDB due to the tautomeric and protonation states; hence, the special names are available in the softwares according the position and character of the amino acids. Once clean up these problems, the PDB files are ready to generate the parameter files and it requires the force field, the Hamiltonian of the simplest Amber force field is,

$$E_{\text{total}} = \sum_{bonds} k_b (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} V_n [1 + \cos(n\phi - \gamma)]$$

$$+ \left[ \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + q_i q_j \right]$$

(1.19)

The non-bonded terms are parameterized by Lennard-Jones potential.

$$V_{i,j} = 4\varepsilon_{i,j} \left[ \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right]$$

(1.20)

$$V_{i,j} = \varepsilon_{i,j} \left[ \left( \frac{R_{\text{min}}}{r_{i,j}} \right)^{12} - 2 \left( \frac{R_{\text{min}}}{r_{i,j}} \right)^{6} \right]$$

(1.21)

The above-mentioned parameters must be specified in the parameter files.

**Solvation:** Water is the fundamental solvent, used to treat the biological molecules during the MD simulation. The solvent creates the electrostatic interactions both locally and globally. This is the very important step in the MD simulation. The relation between partial atomic charges ($q_i, q_j$), dielectric constant ($\varepsilon_{ij}$) and particle distance ($r_{ij}$)

$$V_{\text{elec}} = \frac{q_i q_j}{\varepsilon_{\text{elec}} r_{ij}}$$

(1.22)

There are two solvation methods, that is implicit solvent model and explicit solvent model. Implicit solvent model is used to treat the solvent as a continuous medium and explicit solvent model rely on discrete solvent model.

**Minimization:** The aim of the minimization is to find out the local energy minimum of the system. Generally, the molecules can exist energetically unfavorable with relative atom placements and the optimal atom placement requires an input energy. The computational process essential to put minimum energy (ME) conformation of the system using force fields. There are several algorithms used to minimize the complex system. In which, the steepest descent and conjugated gradient methods are
well known algorithms used in the minimization process. Both methods have a linear rate of convergence and easy to implement. Although, the steepest descent is slow that the conjugated gradient method. The rate of convergence in steepest descent method is,

\[ f(x) = \frac{1}{2} x^T Q x - b^T x \]  

(1.23)

The rate of convergence in conjugated gradient method is,

\[ \phi(x) = \frac{1}{2} x^T A x - b_x \]  

(1.24)

There are two stages of minimization of the complex system using steepest descent and conjugated gradient methods; first state, the complex is restrained with native position and solvents are relaxed. Then, they will continue without any restrains.  

**Heating process:** The initial velocities (0 K) are assigned to the atoms in the system during the minimization. In this phase, new velocities are given due to that the system will move to the higher temperature. The canonical ensemble (NVT) was used normally to the system with fixed number of particles N, volume V, and temperature T. The Langevin and Berendsen thermostats are used in the heating process to balance the temperature.  

**Equilibration:** This process deals to equilibrate the kinetic and potential energies which is an essential step in the MD simulation and it also helps to stabilize the complex, pressure, temperature and energy with respect time. There are several equilibration methods are proposed as coupling schemes and gradual heating and quenching of the system.  

**Production phase:** This is the final step of the MD methodology; in which, all constraints of the complex might be removed. Then, the production phase can be performed to the desired time scale (picoseconds to nanoseconds or more). During the production running, the thermodynamic properties (pressure, temperature, volume) are calculated to confirm the good simulation.  

**Analysis of MD simulation:** The desired time step has been completed after saving the coordinates and velocities. Over the certain period of time, the MD trajectories are used to calculate the structural, dynamical, kinetic and thermodynamic properties of the system. It also helps to visualize the conformational changes of the system at atomic level. Among the huge properties, the energy, root mean square deviation
(RMSD), root mean square fluctuation (RMSF), β-factor, radius of gyration are the important parameters to investigate the behavior of the complex.

\[
\text{Energy} = \langle E \rangle = \frac{1}{N} \sum_{i=1}^{N} E_i
\]  
\[ (1.25) \]

\[
\text{RMSD} = \left\langle \left( r_i^\alpha - r_i^\beta \right)^2 \right\rangle^{\frac{1}{2}} = \sqrt{\frac{1}{N_i} \sum_{i} (r_i^\alpha - r_i^\beta)^2}
\]  
\[ (1.26) \]

\[
\text{RMSF} = \left( \frac{1}{N_f} \sum_f (r_i^f - r_i^{ave})^2 \right)
\]  
\[ (1.27) \]

\[
B_i = \frac{8}{3} \pi^2 (\text{RMSF})^2
\]  
\[ (1.28) \]

\[
\text{Radius of Gyration} = \sqrt{\frac{1}{N_i} \sum_{i} (r_i - r_{cm})^2}
\]  
\[ (1.29) \]

1.6.2 Binding free energy calculation

The molecular mechanics energies combined with the Poisson–Boltzmann or Generalized Born and Surface Area continuum solvation (MM/PBSA and MM/GBSA) methods are widely used to predict the free energy of protein-ligand complex. The MD trajectories are separated as complex, receptor and ligand to understand the conformational changes of receptor while ligand enters into the active site of receptors [Sun 2018, 2014; Woo 2005; Genheden 2015]. The following equations were used to determine the energies of those complexes with Poisson-Boltzmann and generalized Born surface area continuum solvent model.

\[
\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \{ \Delta G_{\text{receptor}} - \Delta G_{\text{ligand}} \}
\]  
\[ (1.30) \]

\[
\Delta G_{\text{bind}} = \Delta H - T \Delta S \approx \Delta E_{\text{MM}} - \Delta G_{\text{sol}} - T \Delta S
\]  
\[ (1.31) \]

\[
\Delta E_{\text{MM}} = \Delta E_{\text{internal}} - \Delta E_{\text{electrostatic}} + \Delta E_{\text{vdw}}
\]  
\[ (1.32) \]

\[
\Delta G_{\text{sol}} = \Delta G_{\text{PB/GB}} + \Delta G_{\text{SA}}
\]  
\[ (1.33) \]

Where, the binding free energy is divided into three terms as molecular mechanical energy (\(\Delta E_{\text{MM}}\)), solvation energy (\(\Delta G_{\text{sol}}\)) and the entropy contribution (\(T \Delta S\)). The molecular mechanical energy is the sum of intermolecular energy of bond, angle, dihedral angle (\(\Delta E_{\text{internal}}\)), electrostatic energy (\(\Delta E_{\text{electrostatic}}\)) and van der Waals energy (\(\Delta E_{\text{vdw}}\)). The solvation energy is derived from the contribution of polar (\(\Delta G_{\text{PB/GB}}\)) and non-polar (\(\Delta G_{\text{SA}}\)) energies. The MM/PBSA is more accurate than the MM/GBSA due to more physically sound. Further to understand the nature of
complex and ligand binding, the decomposition energy of each interaction was determined using MM/GBSA without entropy contribution [Woods 2014; Gohlke 2003].

\[
\Delta G_{\text{residue-inhibitor}} = \Delta G_{\text{vdw}} - \Delta G_{\text{electrostatic}} + \Delta G_{\text{solvation}}
\]  

(1.34)

The van der Waals and electrostatic interaction energies of intermolecular interactions present between ligand and each residue were calculated.

1.7 QM/MM calculation

The Schrodinger equations are not easy to apply for biomolecules like enzymes, receptors, protein-ligand complex, DNA and RNA due to that the molecules contain thousands of atoms. In this regard, the biomolecular systems are investigated in explicit solvent model using the QM/MM approaches [Senn 2009; Menikarachchi 2010; Van Der Kamp 2013; Groenhof 2013]. The complex is divided into two parts such as QM region (the ligand molecule with active site amino acids) and MM region (entire system except QM region). The QM region is dealt by basis set to acquire the electronic properties and the MM region handled by force filed to obtain the energy of bonding and non-bonding interactions. The linked atoms are taken to treat in both QM and MM level. The QM/MM Hamiltonian is derived as,

\[
H = H_{\text{QM}} + H_{\text{MM}} + H_{\text{QM-MM}}
\]  

(1.35)

These three components were generally used to earn the energy of the total system. The PM3 Hamiltonian of semi-empirical method was used to perform the QM/MM calculation on protein-ligand complex.

1.7.1 Quantum mechanics

This method is used to understand the nature of molecular system at electronic level such as chemical reactions, electron density distribution, bond dissociation energy, dipole moment, etc. The Gaussian software [Frisch 2005] is widely used to carry out the electronic properties of the molecules [Richard 2002].

Schrodinger equation

The Schrodinger equation is used to determine the quantum state of physical system and it defines the relation between Hamiltonian operator (H) and energy of the particle (E)

\[
H\psi(\vec{r}) = E\psi(\vec{r})
\]  

(1.36)

\[
H = \frac{-\hbar^2}{8\pi^2 m} \nabla^2 + V
\]  

(1.37)
The time-dependent Schrödinger equation is

\[
\left\{ -\frac{\hbar^2}{8\pi^2m} \nabla^2 + V \right\} \psi(\hat{r}, t) = \frac{i\hbar}{2\pi} \frac{\partial \psi(\hat{r}, t)}{\partial t}
\] (1.38)

Where, \( \psi \) denotes the wave function, \( m \) is mass of the particle, \( \hbar \) is Planck’s constant and \( V \) is the potential energy of the moving particle [Griffiths 2018].

### 1.7.2 Parameterized model number 3 (PM3)

It is one of the semiempirical method to perform the quantum chemical calculation in computational chemistry which was developed by JJP Stewart. The important feature of this method is the methodology of parameterization. The PM3 method applies 3\textsuperscript{rd} MNDO parameterization automatically, whereas AM1 is 2\textsuperscript{nd} MNDO. However, the accuracy of structural and thermodynamical parameters of the both AM1 and PM3 is almost similar; although, the PM3 is more accurate to predict the energies than the other semiempirical methods. The empirical torsional potential was used to fix the low rotational barriers of amide group obtained by PM3 method [Stewart 2004].

### 1.7.3 Hartree-Fock method

The HF method is one the simplest and important approximation theory to solve the many-body Hamiltonian [Parr 1989; Levine 1991]. The Hartree product is defined that the wave function of \( N \)-electron molecule considered as a one electron wave function which means the probability of the finding electron in a small volume is independent. Thus determined approximate wave function known as “Hartree-Fock” method or self-consistent field method (SCF). The Schrödinger equation is converted into the set of Hartree-Fock equation,

\[
\sum_i \psi_i(r) = \left( -\frac{1}{2} \nabla^2 + V_{\text{ion}}(r) \right) \psi_i(r) + \sum_j \int dr' \frac{|\psi_j(r')|^2}{|r-r'|} \psi_i(r) \\
- \sum_j \delta_{\sigma_i \sigma_j} \int dr' \frac{\psi_j^*(r') \psi_i(r') \psi_i(r)}{|r-r'|}
\] (1.39)

### 1.7.4 Density functional theory

The DFT method is also used to determine the electronic structure of many-body system with the help of another function. The interesting point of DFT is that the relations between total electronic energy and electron density [Argaman 1998; Parr 1989; Levine 1991]. The total energy of the system is expressed in terms of electron density rather than the wave function. In the Hohenberg-Kohn theorem, the
electron density is defined as the unique function which express the ground-state energy and density. The DFT method is more accurate than the HF due to the combined treatment of approximation and electron correlation methods. Therefore, the DFT is derived from the Thomas-Fermi and Slater’s models in the quantum chemistry. In DFT method, the electronic energy is defined as the summation of kinetic energy ($E^T$), potential energy ($E^V$) of the nuclear-electron and nuclear-nuclear repulsion, the electron-electron repulsion ($E^J$), exchange correlation ($E^{XC}$) and electron-electron interaction [$E=E^T+E^V+E^J+E^{XC}$]. In this calculation, the derived wave function of the electronic state is satisfied Schrodinger equation.

$$\hat{H}\psi = \left[ \hat{T} + \hat{V} + \hat{U} \right] \psi = \left[ \sum_i^N \left( -\frac{\hbar^2}{2m_i} \nabla_i^2 \right) + \sum_i^N V(\vec{r}_i) + \sum_{i<j}^N U(\vec{r}_i, \vec{r}_j) \right] \psi = E\psi \quad (1.40)$$

1.7.5 Basis sets

The basis sets are the collection of atomic functions, used to determine the wave function of electronic structure of molecular system in HF and DFT methods and it determines the atomic orbitals mathematically within a system [Davidson 1986]. In the quantum chemical calculation, the basis sets are applied to determine the atomic orbitals computationally which helps to investigate the chemical properties of the molecular system. The location of the electrons in the system are accurately determined while using larger basis sets; whereas the standard basis sets are used to calculate the linear combination of atomic functions to form the orbitals. Nowadays, there are several basis sets are available that is minimal basis set, split valance basis set, polarized basis set, correlation-consistent basis set, and completeness-optimized basis set [Jensen 2001; Dunning 1989; Lehtola, 2015; Manninen 2006]. In which, the few available basis sets are given here.

<table>
<thead>
<tr>
<th>Minimal basis set</th>
<th>STO-3G, STO-4G, STO-6G, STO-3G*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarized basis set</td>
<td>def2-SV(P), def2-SVP, def2-SVPD, def2-TZVP, def2-TZVPD, def2-TZVPP, def2-TZVPPD, def2-QZVP, def2-QZVPD, def2-QZVPP, def2-QZVPPD</td>
</tr>
</tbody>
</table>
Correlation-consistent basis set: cc-pVDZ, cc-pVTZ, cc-pVQZ, cc-pV5Z, aug-cc-pVDZ, cc-pCVTZ

1.8 Quantum theory of atoms in molecules (QTAIM)

The professor Richard F. W. Bader and his co-workers developed the QTAIM method to determine the electronic structure of molecules quantum mechanically [Bader 1991, 1994, 2005]. This method has been widely used to understand the reactivity, strength of covalent and non-covalent interactions at electronic level in molecules such as small molecules, macromolecules, protein-ligand complex, DNA and RNA. The electron density of N electron contained molecular system is defined as the probability of finding electrons in the small volume.

\[ \rho(r, X) = N \int \psi^*(x, X)\psi(x, X) \, dt' \]  

(1.41)

The \( N \) denotes the number of electron present in the \( dt' \) volume of the molecular system, the \( x \) and \( X \) are electronic and nuclear coordinates. The topological properties of electron density of molecules in both experimental and theoretical, which leads to understand the new insights of chemical bonding and its derivatives are extensively used to describe the exact notion of chemical bonding from bond paths and bond critical points (bcp). Further, the Popelier and coworkers were demonstrated to predict the structure-activity relationship with the help of QTAIM method [Kumar 2016]. The topological properties are interrelated with the geometrical parameters, which can be derived from the wave function obtained from the quantum chemical calculation.

Electron density \( \rho(r) \): The \( \rho(r) \) is a physical quantity which plays an important role in the atoms in molecules (AIM) and it insights the topological properties of electron density of atoms in molecules [Matta 2007; Popelier 2000; Bader 1991; Gatti 2012]. The topology of electron density can be described in terms of gradient vector fields \( (\nabla \rho) \) and its critical points \( (cp) \). The collection of gradient path is called as gradient vector field that the curves are perpendicular to the \( \rho(r) \) and follow the direction of steepest ascent in \( \rho(r) \). The \( \nabla \rho \) is zero means that the gradient path either originates from a minimum or terminates at saddle point; this point is defined as critical points \( (cp) \) and the point denote as,
Quantum Theory of Atoms in Molecules  Chapter-I

\[ \nabla \rho = i \frac{\partial \rho}{\partial x} + j \frac{\partial \rho}{\partial y} + k \frac{\partial \rho}{\partial z} = 0 \]  (1.42)

Where, \( \hat{i}, \hat{j}, \hat{k} \) are unit vectors. A function is an either maximum or minimum, it is determined by the sign of its second derivative, or curvature, at the stationary point. In general, the arbitrary choice of coordinate axes can be assigned nine second derivatives of the curvatures of \( \rho \) at a point in space. The Hessian matrix of the charge density is an ordered 3x3 array and it can be diagonalized since it is real and symmetric.

The diagonalized Hessian is denoted by \( A \) and given by,

\[ A = \begin{bmatrix}
\frac{\partial^2 \rho}{\partial x^2} & \frac{\partial^2 \rho}{\partial x \partial y} & \frac{\partial^2 \rho}{\partial x \partial z} \\
\frac{\partial^2 \rho}{\partial y \partial x} & \frac{\partial^2 \rho}{\partial y^2} & \frac{\partial^2 \rho}{\partial y \partial z} \\
\frac{\partial^2 \rho}{\partial z \partial x} & \frac{\partial^2 \rho}{\partial z \partial y} & \frac{\partial^2 \rho}{\partial z^2}
\end{bmatrix} = \begin{bmatrix}
\frac{\partial^2 \rho}{\partial x^2} & 0 & 0 \\
0 & \frac{\partial^2 \rho}{\partial y^2} & 0 \\
0 & 0 & \frac{\partial^2 \rho}{\partial z^2}
\end{bmatrix} = \begin{bmatrix}
\lambda_1 & 0 & 0 \\
0 & \lambda_2 & 0 \\
0 & 0 & \lambda_3
\end{bmatrix} \]  (1.43)

In the diagonalized form Hessian matrix, the three eigenvalues \( \lambda_1, \lambda_2, \lambda_3 \) and these are the curvatures of density along the main curvature axes at the point \( r_c \) as flowing order \( \lambda_1 < \lambda_2 < \lambda_3 \). In \( \rho(r) \), the critical point is classified by the rank \( (\omega) \) of non-zero eigenvalues \( \lambda_i \) and the signature \( (\sigma) \) of the sum of the signs of the eigenvalues \( \lambda_i \) of the Hessian matrix. The critical point with \( \omega < 3 \) are not found in the equilibrium charge distribution, it always finds \( \omega=3 \); these three curvatures are sharing \( \pm 1 \) of \( \sigma \) depends on the positive or negative curvature. Hence, there are only four possible types of CP’s with signature that is nuclear critical point \( (3, -3) \), bond critical point \( (3, -1) \), ring critical point \( (3, +1) \) and cage critical point \( (3, +3) \). These four critical points are only existed in a molecule and the topological critical point of this relationship is known as Poincare-Hopf (PH) relationship, i.e the set of \( n_{NCP} - n_{BCP} + n_{RCP} - n_{CCP} \) is 1 indicates the isolated molecule and the sum of \( n_{NCP} - n_{BCP} + n_{RCP} - n_{CCP} \) is 0 means infinite crystals. This set is known as “characteristic set” of the system. The details of the critical points are

- \( (3, -3) \): Peak – all curvatures are negative and \( \rho \) is a local maximum at \( r_c \).
- \( (3, -1) \): passes or saddle points – two curvatures are negative and \( \rho \) is a maximum at \( r_c \) in the plane defined by their corresponding axes. \( \rho \) is a minimum at \( r_c \) along the third axis, perpendicular to this plane.
- \( (3, +1) \): Pales – two curvatures are negative and \( \rho \) is a minimum at \( r_c \) in the plane
• defined by their corresponding axes. \( \rho \) is a maximum at \( r_c \) along the third axis, perpendicular to this plane.

• (3, +3): pits – all curvatures are positive and \( \rho \) is a local minimum at \( r_c \).

**Laplacian of electron density:** The contributions of paired electrons from bonding or lone pairs are difficult to detect due to that the contributions of the core electrons are dominating the topology of electron density. However, the amplification of small changes in the topology of the electron density distribution is determined from the second order derivatives, obtained in the Hessian matrix [Matta 2007; Popelier 2000]. The Laplacian of electron density equation as

\[
\nabla^2 \rho = \frac{\partial^2 \rho}{\partial x^2} + \frac{\partial^2 \rho}{\partial y^2} + \frac{\partial^2 \rho}{\partial z^2}
\]

Where, the Laplacian carries the negative value \( \nabla^2 \rho < 0 \) means that the scalar field \( \rho \) is locally concentrated and it is positive \( \nabla^2 \rho > 0 \) indicates that the \( \rho \) is locally depleted. The covalent bonds are characterized by an overlapping of the valence shells and it refers to the valence shell regions of the bonding patterns due to the accumulation of charge density \( \nabla^2 \rho < 0 \) in the bonding region; therefore, this pattern is known as *bond critical point (bcp)*. This interaction of the valence shells in the covalent bonds are called *open shell or shared interactions*. Further, the charge depletion \( \nabla^2 \rho > 0 \) at the electropositive atom indicates that the *bcp* is moved towards the electronegative atom. The interactions can occur between the atoms or ions in a closed shell, it displays the *closed-shell interactions*. Therefore, the Laplacian of electron density plays an important role throughout the AIM theory and provides strong insight into the structural and reactive properties.

**Ellipticity:** The two negative curvatures of \( \rho \) at the *bcp* is known as the bond ellipticity (\( \varepsilon \)),

\[
\varepsilon = \frac{\lambda_1}{\lambda_2} - 1 \text{ where } (|\lambda_1| \geq |\lambda_2|)
\]

If \( \lambda_1 = \lambda_2 \) means \( \varepsilon = 0 \), it indicates that the bond is cylindrically symmetrical. The \( \varepsilon \) is used to measure the \( \pi \) - character, reaches up to maximum ellipticity which is attributed in the double bond whereas in the triple bond, the ellipticity get decreases due to the increasing bond order [Matta 2007]. The strength of chemical bond is defined by bond order \( [e^{A(\rho_b - B)}] \); where A and B are constants, it depends on the nature of bonded atoms. The \( \rho_b \) is an amount of electron density at the *bcp* and it
correlates with binding energy, the $\rho_b < 0.20$ au means *shared or covalent bonding*; the $\rho_b > 0.10$ au is *closed-shell interaction*.

**Bond path:** The vector that points in the direction of greatest increase in the electron density is called a gradient vector and the trajectory of tiny segments of gradient vectors is known as gradient path [Matta 2007]. The pairs of gradient path which originate at each (3,-1) critical point and terminate at the nuclei through the charge distribution linking line of the neighbouring nuclei. This atomic interaction line is called as bond path (BP). The bond path length is equal to bond length.

**Energy densities:** This is used to determine the mechanics of bonding interactions from the information contained one electron density matrix [Abramov 1997; Espinosa 1998]. According to Bader, the atomic interactions are characterized by the values of electron density $\rho(r)$, Laplacian of electron density $\nabla^2 \rho(r)$ and also total energy density $H(r)$ at the bcp of (3, -1). The virial theorem expresses the relationship of virial field, kinetic energy density $G(r)$ and Laplacian of electron density,

$$\left(\frac{\hbar}{4m}\right)\nabla^2 \rho(r) = 2G(r) + V(r) \quad (1.46)$$

The quantity $G(r)$ is the kinetic energy density of electrons and the value is positive everywhere; the $V(r)$ is the potential energy density and it is negative everywhere. Therefore, this equation determines that the lowering of the potential energy dominates the total energy where electronic charge is concentrated [$\nabla^2 \rho(r) < 0$]; similarly, the kinetic energy is dominant in regions [$\nabla^2 \rho(r) > 0$]. The local statement of the total energy is the sum of kinetic $G(r)$ and potential $V(r)$ energy density [$H_r = G_r + V_r$]. The equation of the $G(r)$ is

$$G(r) = \frac{3}{10} (3\pi^2)^{\frac{2}{3}} \rho_{\infty}^{\frac{5}{3}} + \frac{1}{6} \nabla^2 \rho(r) \quad (1.47)$$

$$V(r) = \frac{1}{4} \nabla^2 \rho(r) - 2G(r) \quad (1.48)$$

$$D_c = -\frac{1}{2} V(r) \quad (1.49)$$

Where, the $D_c$ denotes the bond dissociation energy of the bonds.

**1.9 Electrostatic potential**

The electrostatic potential (ESP) is another important application of electron density analysis which is used to generate the area of the charge distribution in the molecules. It also provides the calculation of intermolecular interaction energies,
crystal dipole moment, electronegativity and partial charges [Koster 1993; Gadre 1999]. The molecular electrostatic potential is the potential energy of the protons at particular locations in the molecule and the equation is defined as,

$$ ESP(r) = ESP_{\text{nuc}}(r) + ESP_{\text{elec}}(r) = \sum_j \frac{Z_j}{|r - R_j|} - \int \frac{\rho(r)}{|r - r'|} dr' \quad (1.50) $$

The ESP is defined as the energy required to move a positive unit charge from infinity to the point in space defined by \( r \). The ESP contribution is derived from the both nuclei and electrons present in the molecule. In ESP equation, the first term indicates the electrostatic potential of the nuclei and second term obtained from the electron density. It can be calculated independently from the crystal environment, applying the formalism of Su, Coppens and Macchi [Coppens 1997]. The electrostatic forces are generating the path of a reactant to reach the reactive site of a molecule due to long range forces. Interestingly, the ESP is used to predict the possible reactive sites of the molecule where the nucleophilic (attracted to regions with positive potential) and electrophilic (attracted towards the regions with negative potential) regions are present in the molecule. Recently, the researchers are extensively used this ESP map to understand the molecular properties as well as identification of the drug binding regions in the molecule. It is also used to predict the strength of intermolecular interactions and molecular recognition [Tasi 1993; Popelier 2004; Politzer 2001].