CHAPTER – I

INTRODUCTION
1.1. X-ray Crystallography

The Crystalline state of the matter is characterized by a structure in which component atoms, molecules or ions are arranged such that they are in close proximity to one another, there are relatively strong interactions between one another and there is a long range, well defined, three dimensional order. Max von Laue, in 1912, discovered that crystalline substances act as three-dimensional diffraction gratings for X-ray wavelengths similar to the spacing of planes in a crystal lattice. X-rays have the proper wavelength (in the Angstrom range, ~10⁻⁸ cm) to be scattered by the electron cloud of an atom of comparable size. Based on the diffraction pattern obtained from X-ray scattering of the periodic assembly of molecules or atoms in the crystal, the electron density can be reconstructed. Additional phase information must be extracted either from the diffraction data or from supplementing diffraction experiments to complete the reconstruction (the phase problem in crystallography). A model is then progressively built into the experimental electron density, refined against the data and the result is a quite accurate molecular structure.

The knowledge of accurate molecular structures is a pre-requisite for rational drug design and for structure based functional studies to aid the development of effective therapeutic agents and drugs. Crystallography can
reliably provide the answer to many structure related questions, from global folds to atomic details of bonding. In this thesis we present the structural, conformational and molecular modelling studies of ten different synthesised molecules and natural molecules from Aagle marmoles (Bael) and Calophyllum inophyllum (Punna) leaf extract.

**Crystallization and Crystal Selection**

The process of crystallization is one of the ordering, wherein randomly arranged ions, atoms or molecules take up regular positions in the solid state. It involves the phenomenon of nucleation and it may be considered to be in dynamic equilibrium between particles in the fluid phase and solid phase from saturated solutions. Several techniques are available for crystallization of small molecules such as slow evaporation, slow cooling, diffusion methods etc.

Slow evaporation technique yielded good quality single crystals for all the compounds presented in this thesis. Crystals were examined under a polarizing microscope before mounting in the goniometer head for data collection.

**Intensity Data Collection**

In the present study, reflection data from the crystals were measured by Bruker SMART APEXII CCD diffractometer (Bruker, 2008) system. For intensity data collection, it is equipped with graphite mono-chromated MoK₂\(\alpha\) (\(\lambda=0.7107\text{Å}\)) radiation. Once the crystal is centered, a preliminary rotational image is often collected to screen the sample quality and to select parameters for later steps. An automatic collection routine can then be used to collect a
preliminary set of frames for determination of the unit cell. Reflections from these frames are auto-indexed to select the reduced primitive cell and calculate the orientation matrix (which relates the unit cell to the actual crystal position within the beam). The primitive unit cell is refined using least-squares and then converted to the appropriate crystal system and Bravias lattice. This new cell is also refined using least-squares to determine the final orientation matrix for the sample.

After the refined cell and orientation matrix have been determined, intensity data is collected. Generally this is done by collecting a sphere or hemisphere of data using an incremental scan method, collecting frames in 0.1° to 0.3° increments (over certain angles while others are held constant). For highly symmetric materials, collection can be constrained symmetrically to reduce the collection time. Data is typically collected between 4° and 60° 2θ for molybdenum radiation. A complete data collection may require anywhere between 6-24 hours, depending on the specimen and the diffractometer. Exposure times of 10-30 seconds per frame for a hemisphere of data will require total run times of 6-13 hours.

The intensity data collection, frames integration, Lorentz polarization (LP) correction and decay correction were done using SAINT (Bruker, 2008) software. Empirical absorption correction (multi-scan) was performed using SADABS (Bruker, 2008) program. The Bruker AXS SMART APEX-II single crystal diffractometer is shown in (Fig. 1.1).
Figure 1.1. Bruker AXS SMART APEX-II single crystal diffractometer
All the intensities were corrected for variable scan speed, background and all the intensities were corrected for variable scan speed, background and attenuation using the relation.

\[ I_{\text{raw}} = f [N_c - 2(L_b+R_b)] NPI \]

where,

- \( I_{\text{raw}} \) - Relative intensity
- \( N_c \) - Peak count
- \( L_b \) and \( R_b \) - Left and right background counts, respectively.
- \( NPI \) - Scan speed parameter
- \( f \) - Attenuation factor

The observed structure factor for each reflection is obtained using the equation.

\[ |F_{hkl}| = k \left( \frac{I_{\text{raw}}}{L_p} \right)^{1/2} \]

where,

- \( k \) - Scaling factor
- \( L \) - Lorentz factor \( \left( L = \frac{1}{\sin^2 \theta} \right) \)
- \( p \) - Polarization factor = \( \frac{1 + \cos^2 \theta}{2} \)

where \( \theta \) is the Bragg angle of reflection
Data Reduction

The raw data collected from the diffractometer suffers from physical and geometrical error factors and hence cannot be used for structure elucidation straight away. Therefore the intensity data have to be corrected for Lorentz, polarization and absorption effects.

The Lorentz and polarization corrections are must for every case, since the reflection efficiency varies with the reflection angle, whereas the absorption correction has to be applied depending upon the nature of the compound and the radiation used, i.e. depending on the linear absorption coefficient value. If space group ambiguity arise then the contents of the unit cell, the number of molecules present in the cell, the distribution of intensity and other relevant details are analyzed in depth.

Structure Solution

In order to obtain atomic positions of the molecule, intensities are converted into structure factors. Structure factor is the resultant of N waves scattered by N atoms in the unit cell. The general expression for the structure factor is

\[ F_{hkl} = \sum_{j=1}^{N} f_j e^{2\pi i (hx_j + ky_j + lz_j)} \]

where,

- \( x_j, y_j, z_j \) are fractional co-ordinates of \( j^{th} \) atom
- \( N \) - total number of atoms in the Unit Cell.
- \( f_j \) – Atomic scattering factor of the \( j^{th} \) atom
Since $F_{hkl}$ is a complex quantity, it can be written as

$$F_{hkl} = |F_{hkl}| e^{i\phi_{hkl}}$$

where, $|F_{hkl}|$ is the Structure Factor Amplitude $\alpha \sqrt{I_{hkl}}$ of a Bragg reflection $h, k, l$ and $\phi_{hkl}$ is phase of the reflection $h, k, l$. The structure factor can be expressed in terms of the integral of electron-density. The Fourier transform of the structure factor yields electron-density function.

The general expression for electron-density function $\rho(x, y, z)$ is given by

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

If the structure factor magnitude and phases are known, the electron-density distribution of the unit cell can be calculated with peaks revealing atomic positions. If the electron-density is known, three dimensional structure of the molecule can be elucidated. But the process is not straightforward; for to sum the series $\rho(x, y, z)$, we have to find out the complex structure factor $F_{hkl}$

Therefore, $\rho(x, y, z)$ becomes

$$\rho(x, y, z) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |F_{hkl}| e^{-2\pi i (hx + ky + lz) + i\phi_{hkl}}$$
In the equation we replaced $I_{hkl}$ of the previous equation by $|F_{hkl}| e^{i\phi_{hkl}}$ and

$$|F_{hkl}| \propto \sqrt{I_{hkl}}$$

In order to locate the electron density maxima from the left hand side of the above equation, one should be able to compute the right hand side of the equation, in which the structure factor $F_{hkl}$ is the coefficient (this equation is also called a Fourier synthesis or electron density equation). Thus if we know the structure factors, we can calculate the actual real structure (the density of the electrons in real space). The electron density is calculated for each point $(x, y, z)$ in space. Since, $F_{hkl}$ involves two components, $|F_{hkl}|$ and $\phi_{hkl}$, the computation of electron density is possible only if both the structure factor magnitudes and phases of the Bragg’s reflections are known. But unfortunately experimental measurement gives only structure factor amplitude $|F_{hkl}|$ thus $\phi_{hkl}$ remains unknown. This is known as Phase problem in crystallography, namely, the non-availability of phases of Bragg’s reflections, which are required along with the known structure factor magnitudes, to locate an atomic site. Thus unless the phases are known, atomic sites could not be located.
Methods of Solving the Phase Problem

Several methods are available to solve the phase problem and some of them are:

(i) Direct methods
(ii) Patterson method
(iii) Isomorphous replacement method
(iv) Anomalous Dispersion method

These methods can be effectively useful to locate the approximate positions of the atoms of trial structure of a molecule in the unit cell. As direct methods have been used to solve the crystal structures presented in this thesis, a brief summary of the assumptions and stepwise procedure adopted in this method are given below:

Direct Methods

Direct methods are used to calculate the phases directly by simple mathematical procedures from a single set of X-ray intensities. The basic postulates of direct methods are the positivity (the electron density is positive everywhere) and the atomicity (the atoms are spherically symmetric). 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.
This constraint however does not hamper the evaluation of $\mathfrak{a}_{h\kappa l}$. The various steps involved in the direct methods are:

i) Conversion of observed structure factors $|F_{h\kappa l}|$ to normalized structure factors $|E_{h\kappa l}|$ which are independent of $\theta$.

ii) Setting up of phase relations using triple phase relations (triplets) and four phases relations (quartets).

iii) Selection of a few reflections, the phases of which are assigned apriori.

iv) Phase propagation and refinement using tangent formula (Karle and Hauptman, 1950).

v) Calculation of best phase sets and expressing the reliability of the phases in terms of Combined Figure of Merit (CFOM).

vi) Calculation of electron density map (E-map) with $|E_{h\kappa l}|$ as the Fourier Coefficient.

The compounds used in the thesis were solved by direct methods using SHELXS97 program (Sheldrick, 2008) of WinGX software package.

**Structure Refinement**

Structure refinement consists of obtaining the best fit between a set of observed measurements and the quantities calculated from a model postulated to explain them. Differences between the observed and the calculated values can
arise due to random errors (statistical fluctuations) in the observations and defects in the model (systematic errors). The trial structure obtained from the structure solution is refined in order to get the accurate atomic positions and the associated thermal parameters. Though several structure refinement process are in vogue, the full-matrix least-squares refinement technique is the conventional one and more widely used in small molecular structure determination. \textit{SHELXL97} (Sheldrick, 2008) computer program was used for the refinement. The least-squares refinement consists of using the squares of the differences between the observed and calculated values as a measure of their disagreement, and adjusting the parameters so that the total disagreement tends to a minimum. The refinement is based on $F_o$ which would involve taking the square root of a negative number of reflections with negative $F_o$ (i.e. background higher than the peak as a result of statistical fluctuation). The refinement on $F_o^2$ using all the data provides a good result for weakly diffracting crystals and in particular for pseudosymmetry problems.

The residual factor or reliability index $R_1$

$$R_1 = \frac{\sum |F_o| - |F_c|}{\sum |F_o|}$$

$|F_o|$– Observed structure factor amplitude

$|F_c|$– Calculated structure factor amplitude

The summation is taken over all the observed reflections. $R$ value should be a minimum for the accurate model. A suitable weighting scheme is applied at the end of refinement procedure and the weighted $R$-factor is given by
The Goodness of Fit is always based on $F^2$

$$wR_2 = \frac{\Sigma w_i \sqrt{(|F_o|^2 - |F_c|^2)^2}}{\Sigma w_i (|F_o|^2)^2}$$

$$GoF = S = \left[ \sum (w(F_o^2 - F_c^2))/(n-p) \right]^{1/2}$$

where, $n$ is the number of reflections and $p$ is the total number of parameters refined.

$$w = 1/[a^2(F_o^2) + (aP)^2 + bP]$$

where, $a$ and $b$ are the constants and $P = [2F_c^2 + Max(F_o^2, 0)/3]$

**Calculation of Geometrical Parameters**

Crystal structure determination provides the unit cell constants and fractional atomic coordinates of all the atoms and their associated thermal displacement parameters. The geometrical parameters such as bond lengths, bond angles and torsion angles can be calculated from the coordinates of the relevant atoms.

For a triclinic lattice, the distance between the two points with fractional atomic coordinates $(x_1, y_1, z_1)$ and $(x_2, y_2, z_2)$ is given by the law of cosines in three dimensions as

$$L = \{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - 2ab\Delta x\Delta y \cos \gamma - 2ac\Delta x\Delta z \cos \beta - 2bc \Delta y\Delta z \cos \alpha \}^{1/2}$$
where \(a, \ b, \ c, \ \alpha, \ \beta, \ \gamma\) are the unit cell parameters \(\Delta x = x_1 - x_2, \ \Delta y = y_1 - y_2 \text{ and } \Delta z = z_1 - z_2\).

The above equation can be applied for any crystal system to calculate the bond lengths. Bond length values are useful to identify the nature of chemical bonds (triple, double, partially double or single bond) present in the molecule.

Bond angle formed by the three atoms \(A, B, C\) where the angle is subtended by the bonds \(AB\) and \(AC\) can be calculated using the formula

\[
\cos \theta = \frac{(AB)^2 + (AC)^2 + (BC)^2}{2(AB)(AC)}
\]

Bond angles are useful to find the type of hybridization of a particular atom.

Torsion angle is the angle of inclination between the two planes given by \(ABC\) and \(BCD\) formed by four atoms \(A, B, C\) and \(D\) is given by

\[
\cos \chi = \frac{N1 \cdot N2}{|N1| \cdot |N2|}
\]

where \(N1\) and \(N2\) are vectors normal to \(ABC\) and \(BCD\) planes respectively.

Torsion angles are useful to understand the orientation/conformation of one plane/ring with another plane/ring formed by the various groups of atoms in a molecule.
Ring Conformations

Ring conformation can be predicted with the help of a mirror plane lying perpendicular to the ring plane and the two-fold symmetry lying in the ring plane as depicted in Fig.1.2. For a six membered ring three mirrors and three 2-fold symmetries (for a chair), two mirror symmetries (for a boat) and two 2-fold symmetries (for a twist boat) are the possible conformations in addition to sofa and half-chair conformations. The sofa has only one mirror and half-chair possesses one two-fold symmetry. Many rings are found to be distorted, without having any defined conformations. Five-membered rings are found to have only two possible conformations i.e. half-chair or envelope Fig.1.3. In practice, conformations are described from the puckering and asymmetry parameters which give the extent of deviation of the ring from the ideal conformations (Cremer and Pople, 1975; Duax et al., 1976; Nardelli, 1983). PLATON (Spek, 2009) program is used for the calculation of geometrical parameters.

The asymmetry parameters have been proposed (Duax et al., 1976) to give a quantitative evaluation of how a ring of any size deviates from ideal symmetry and to help in describing its conformation. Calculations of these asymmetry parameters are included as an algorithm PLATON (Spek, 2009).

Molecular Interactions

In crystalline state, the molecules are stabilized by intramolecular and intermolecular interactions like hydrogen bonds, van der Waal’s forces and
possibly some short contacts between the two atoms. Hydrogen bonding is the

\[ \Delta C_S = \left[ \frac{\sum_{i=1}^{m} (\phi_i + \phi'_i)}{m} \right]^{\frac{1}{2}} \]

Mirror asymmetry

\[ \Delta C_2 = \left[ \frac{\sum_{i=1}^{m} (\phi_i - \phi'_i)^2}{m} \right]^{\frac{1}{2}} \]

Two-fold asymmetry

where, \( \phi_i, \phi'_i \) are symmetry related torsions
m is the number of individual comparisons

Figure 1.2. Possible conformations of the six-membered rings
Figure 1.3. Possible conformations of the five-membered rings
specific type of non-bonded interaction between two electronegative atoms (donor and acceptor), where the hydrogen atom is bonded to them. The schematic representation of the hydrogen bond is D—H…A, where D is the donor and A is the acceptor.

The crystal structures presented in this thesis are found to have C—H…O, O—H…O, O—H…N, O—H…Cl, C—H…N and N—H…O types of hydrogen bonds. The existence of C—H…O bonds in crystals is evident from the study of Taylor and Kennard (1982) and Desiraju (1991; 1996). The ability of a C—H group to act as a proton donor depends on the hybridization \([C(sp)―H]\) > \(sp^2\) → \(H\) > \(C(sp^3)―H\), and increases with the number of adjacent withdrawing groups (Steiner, 1997).

Weak attractive forces between uncharged atoms or molecules are collectively referred to as van der Waals forces. These forces arise from the electrostatic attraction of the nuclei of one molecule by the electrons of a different molecule. The repulsion arising between the electrons of two molecules as well as the nuclei of two molecules counteract the electrostatic attractions, but there is always a small net attractive force. The van der Waals forces are short range forces i.e., they are significant only when the molecules are very close to one another. The molecular graphics of the molecules and their packings are obtained using ORTEP-3 (Farrugia, 1997) and PLATON (Spek, 2009).
1.2 Extraction and Characterisation of Bioactive Compounds from Plants Leaves.

Natural products from medicinal plants, either as pure compounds or as standardized extracts, provide unlimited opportunities for new drug leads because of the unmatched availability of chemical diversity. According to the World Health Organization (WHO), nearly 20,000 medicinal plants exist in 91 countries including 12 mega biodiversity countries. Plants used for traditional medicine contain a wide range of substances that can be used to treat chronic as well as infectious diseases. Due to the development of adverse effects and microbial resistance to the chemically synthesized drugs, men turned to ethnopharmacognosy. They found literally thousands of photochemicals from plants as safe and broadly effective alternatives with less adverse effect. Clinical trials directed towards understanding the pharmacokinetics, bioavailability, efficacy, safety and drug interactions of newly developed bioactive compounds and their formulations (extracts) require a careful evaluation. The premier steps to utilize the biologically active compound from plant resources are extraction, isolation and characterization of bioactive compound.

Extraction

Extraction is the crucial first step in the analysis of medicinal plants, because it is necessary to extract the desired chemical components from the plant materials for further separation and characterization. The basic operation included steps, such as pre-washing, drying of plant materials or freeze drying, grinding to obtain a homogenous sample and often improving the kinetics of analytic
extraction and also increasing the contact of sample surface with the solvent system. Proper actions must be taken to assure that potential active constituents are not lost, distorted or destroyed during the preparation of the extract from plant samples.

The selection of solvent system largely depends on the specific nature of the bioactive compound being targeted. The extraction of hydrophilic compounds uses polar solvents such as methanol, ethanol or ethyl-acetate. As the target compounds may be non-polar to polar, the suitability of the methods of extraction must be considered. The most common chemical procedure for obtaining organic constituents from dried plant tissues is to continuously extract powdered material in a Soxhlet apparatus with a range of solvents, starting with petroleum ether and chloroform (to separated lipids and terpenoids) and then using alcohol and ethyl acetate (for more polar compounds) (Raaman, 2006).

**Soxhlet Extractor**

The Soxhlet extractor (Fig.1.4) was developed by Franz Von Soxhlet, a German agricultural chemist. In Soxhlet extractor, within an enclosed flask there is an inverted condenser pointing down into the flask from the top. Just below that condenser will be suspended either what’s called a soxhlet basket or a recovery vessel depending on whether extracting or recovering solvent. The condenser will have cold liquid circulating through it to keep the condenser cold. In the bottom of the main flask, solvent is placed. To do an extraction, the powdered plant material is placed in the soxhlet basket, a vessel with perforated sides and bottom so that
Figure 1.4. A typical Soxhlet extractor loaded with sample
liquid can fall through it. When gentle heat is applied to the main flask, the solvent begins to evaporate and the solvent vapours reach the cold condenser at the top of the flask and begin to liquify on the sides of the condenser. The re-condensed solvent on the sides of the condenser begins flowing down the sides and dripping off on the end of the condenser. This solvent drips into the top of the soxhlet basket where it saturates the herb being extracted. The solvent flows through the basket and out of the holes in the bottom of the basket carrying the extract with it into the bottom of the flask. The extract laden solvent falling from the soxhlet basket will be dark in colour and as it becomes clearer one can know that the plant material has leached out and the process is finished.

1.3 Characterisation of Lead Chemical Compounds in the Crude Extract by Gas Chromatography Mass Spectrometry (GC–MS)

Mass Spectrometry

Mass spectrometry is the science of displaying the spectra of the masses of the molecules comprising a sample of material. It is used for determining the elemental composition of a sample, the masses of particles and of molecules, and for elucidating the chemical structures of molecules such as peptides and other chemical compounds. Mass spectrometry works by ionizing chemical compounds to generate charged molecules or molecular fragments and measuring their mass-to-charge ratio (Sparkman, 2000).

In a MS procedure, a sample, which may be solid, liquid or gas, is ionized. The ions are separated according to their mass-to-charge ratio. The ions are detected by a mechanism capable of detecting charged particles. The signal is
processed into the spectra of the relative abundance of ions as a function of the mass-to-charge ratio. The atoms or molecules can be identified by correlating known masses by the identified masses or through a characteristic fragmentation pattern. The following laws govern the dynamics of charged particles in electric and magnetic fields in vacuum:

\[ F = Q (E + v \times B) \rightarrow \text{Lorentz force law} \]

\[ F = ma \rightarrow \text{Newton’s second law of motion in non-relativistic case} \]

Here, \( F \) is the force applied to the ion, \( m \) is the mass of the ion, \( a \) is the acceleration, \( Q \) is the ion charge, \( E \) is the electric field, and \( v \times B \) is the vector cross product of the ion velocity and the magnetic field.

Equating the above expressions for the force applied to the ion yields:

\[ (m/Q)a = E + v \times B \]

This is the classic equation of motion for charged particles. Together with the particle’s initial conditions, it completely determines the particle’s motion in space and time in terms of \( m/Q \). Thus mass spectrometers could be thought of as “mass-to-charge spectrometers”. When presenting the data, it is common to use the dimensionless \( m/z \), where \( z \) is the number of elementary charges (\( e \)) on the ion (\( z = Q/e \)). This quantity, although informally called the mass-to-charge ratio, represents the ratio of the mass number and charge number, \( z \).

A mass spectrometer consists of three components: ion source, mass analyzer, and detector (Chhabil Dass, 2007). Fig.1.5. The ionizer converts some
Figure 1.5. Schematic representation of Mass Spectroscopy
Figure 1.6. Schematic representation of GC-MS and the instrument.
portion of the sample into ions. There are a wide variety of ionization techniques, depending on the phase (solid, liquid, gas) of the sample, and the efficiency of various ionization mechanisms for the target species. An extraction system removes ions from the sample and gives them a trajectory which allows the mass analyzer to sort the ions mass-to-charge. The detector measures the value of an indicator quantity and thus provides data for calculating the abundance of each ion present.

**Gas Chromatography Mass Spectrometry (GC –MS)**

GC – MS (Fig.1.6) analysis was carried out in the Indian Institute of Crop Processing Technology (IICPT) Thanjavur. GC –MS plays a key role in the analysis of unknown component of natural orgin. GC-MS ionises compounds and measures their mass numbers. Ionization method includes EI (electron ionization) and CI(chemical ionization). Typically, the CI method is used. The EI method produces ions by colliding thermal electrons emitted from a filament with sample gas molecules. This method provides with high stability in ionization and the obtained mass spectra, how good reproducibility. The EI method provides with good results for quantitative analysis as well. Quantification with GC-MS, in which only ions specific to the compounds are measured, is a highly selective method without interfering components. GC techniques involves the separation of volatile components in a test sample using suitable capillary column coated with polar or non-polar or intermediate polar chemicals. The following equipment and running parameters are maintained.
**Gas Chromatograph Programme**

- Column: Elite – 1 (100% Dimethyl polysiloxane) 30 * 0.25mmx
- Equipment: GC clarus 500 Perkin Elmer
- Carrier Gas: Helium, 1ml per min, split 10:1 mass detector turbo mass gold
- Detector: Mass detector: Turbo mass gold-Perkin Elmer.
- Software: Turbomass 5.2

**Oven temperature programme**

- 110°C - 2min hold
- Upto 200°C at the rate of 5°C/min - 9 min hold
- Injecter temperature - 250°C

**MS programme**

- Library/year: NIST-version/2005
- Inlet temperature: 200°C
- Source temperature: 200°C
- Electron energy: 70eV

**1.4 Molecular Modelling (docking)**

The application of computational methods to study the formation of intermolecular complexes has been the subject of intensive research during the last decade. Molecular docking is a key tool in structural molecular biology and computer-assisted drug design. The goal of ligand-protein docking is to predict the predominant binding mode(s) of a ligand with a protein of known three-
dimensional structure. Successful docking methods search high-dimensional spaces effectively and use a scoring function that correctly ranks candidate dockings. Docking can be used to perform virtual screening on large libraries of compounds, rank the results, and propose structural hypotheses of how the ligands inhibit the target, which is invaluable in lead optimization.

**Molecular Docking Approaches**

Molecular docking can be categorized into two main kinds as following:

- **Rigid-flexible docking** (also rigid docking) is the molecular docking which allows only ligand (donor or small molecule) to change its orientation (depending on its torsional degree of freedom) during docking calculation.

- **Flexible-flexible docking** (also flexible docking, induced fit docking), in addition to rigid docking, macromolecule (protein enzyme) also change its orientation, especially around active site.

**ADME Property**

**ADME** is an abbreviation in pharmacokinetics and pharmacology for absorption, distribution, metabolism, and excretion, and describes the disposition of a pharmaceutical compound within an organism. The four criteria all influence the drug levels and kinetics of drug exposure to the tissues and hence influence the performance and pharmacological activity of the compound as a drug.
• Lipinski’s rule of five

Lipinski’s rule of five also known as the Pfizer's rule of five or simply the Rule of five (RO5) is a rule of thumb to evaluate druglikeness or determine if a chemical compound with a certain pharmacological or biological activity has properties that would make it a likely orally active drug in humans. The rule was formulated by Christopher A. Lipinski in 1997, based on the observation that most medication drugs are relatively small and lipophilic molecules.

The rule is important to keep in mind during drug discovery when a pharmacologically active lead structure is optimized step-wise to increase the activity and selectivity of the compound as well as to insure drug-like physicochemical properties are maintained as described by Lipinski's rule.

Lipinski’s rule states that, in general, an orally active drug has no more than one violation of the following criteria:

• Not more than 5 hydrogen bond donors (nitrogen or oxygen atoms with one or more hydrogen atoms)
• Not more than 10 hydrogen bond acceptors (nitrogen or oxygen atoms)
• A molecular mass less than 500 daltons
• An octanol-water partition coefficient $\log P$ not greater than 5

The molecular docking studies presented in the thesis follows Grid-based Ligand Docking – Energetics and Induced fit Docking protocol using Maestro Graphical
User Interface of the *Glide* software purchased from Schrödinger (Glide version 4.0, Schrödinger, LLC, New York, NY, 2005).

**Grid-based Ligand Docking – Energetics (GLIDE)**

Glide is a ligand docking program for predicting protein-ligand binding modes and ranking ligands via high-throughput virtual screening. Glide utilizes two different scoring functions, SP and XP GlideScore, to rank-order compounds. Three modes of sampling ligand conformational and positional degrees of freedom are available to determine the optimal ligand orientation relative to rigid protein receptor geometry.

The Schrödinger GLIDE for docking one or more ligands using the following steps:

1. Glide searches for favorable interactions between one or more ligand molecules and a receptor molecule, usually a protein. Each ligand must be a single molecule, while the receptor may include more than one molecule, e.g., a protein and a cofactor. Glide can be run in rigid or flexible docking modes; the latter automatically generates conformations for each input ligand. The combination of position and orientation of a ligand relative to the receptor, along with its conformation in flexible docking, is referred to as a *ligand pose*.

2. The ligand poses that Glide generates pass through a series of hierarchical filters that evaluate the ligand’s interaction with the receptor.
3. The initial filters test the spatial fit of the ligand to the defined active site, and examine the complementarity of ligand-receptor interactions using a grid-based method patterned after the empirical ChemScore function.

4. Poses that pass these initial screens enter the final stage of the algorithm, which involves evaluation and minimization of a grid approximation to the OPLS-AA nonbonded ligand-receptor interaction energy. Final scoring is then carried out on the energy-minimized poses.

5. By default, Schrödinger’s proprietary GlideScore multi-ligand scoring function is used to score the poses. If GlideScore was selected as the scoring function, a composite Emodel score is then used to rank the poses of each ligand and to select the poses to be reported to the user. Emodel combines GlideScore, the nonbonded interaction energy, and, for flexible docking, the excess internal energy of the generated ligand conformation.

**Receptor Grid Generation**

Glide searches for favourable interactions between one or more ligand molecules and a receptor molecule, usually a protein. The shape and properties of the receptor are represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand poses. For receptors that adopt more than one conformation on binding, you might want to prepare grids for each conformation, to ensure that possible activities are not missed. Glide can, however, handle different hydroxyl conformations with a single grid generation. The options in each tab of the **Receptor Grid Generation** panel allow you to define the receptor structure by excluding any cocrystallized ligand that
may be present, determine the position and size of the active site as it will be represented by receptor grids, and set up Glide constraints. Ligand docking jobs cannot be performed until the receptor grids have been generated.

**Receptor Ligand Docking**

- **HTVS (High Throughput Virtual Screening)**
- **SP (Standard precision)**
- **XP (Extra precision)**

**HTVS (High Throughput Virtual Screening)**

High throughput virtual screening (HTVS) is intended for the rapid screening of very large numbers of ligands. HTVS has much more restricted conformation sampling than SP docking, and cannot be used with constraints, core-in-place, or rigid docking. Advanced settings are not available HTVS, but are fixed at predetermined values.

**SP (Standard precision)**

SP mode techniques cannot represent larger adjustments to the receptor, like conformational changes. The explicit-water technology and descriptors are also used in Glide SP scoring.

**XP (Extra precision)**

Extra-precision mode is a refinement tool designed for use only on good ligand poses. The more extensive XP docking method and specialized XP scoring method are strongly coupled: the more precise poses produced by XP docking are necessary for the more demanding XP scoring method. Because XP docking mode
requires considerably more CPU time, one should screen large sets of ligands first in standard-precision (SP) mode or in high-throughput virtual screening (HTVS) mode. Only the top-scoring ligands should be docked using XP mode. The chief purposes of the XP method are to weed out false positives and to provide a better correlation between good poses and good scores.

**Induced Fit Docking (IFD)**

Schrödinger has developed and validated an Induced Fit Docking protocol based on Glide and the Refinement module in Prime that accurately predicts ligand binding modes and concomitant structural changes in the receptor. In IFD, both the ligand and the receptor are flexible which enable to dock the ligand at the receptor’s binding site to generate multiple poses of the receptor-ligand complex, each including unique structural conformations of the receptor to fit the ligand pose and ranks them by Glide score (G-score) to find the best structure of the docked complex. G-score takes into account a number of parameters like hydrogen bonds (H-bond), hydrophobic contacts (Lipo), vander-Waals (vdW), columnic (Coul), polar interactions in the binding site (Site), metalbinding term (Metal) and penalty for buried polar group (BuryP) and freezing rotatable bonds (RotB) G-score = H bond + Lipo + Metal + Site + 0.130 Coul + 0.065 vdW – BuryP – RotB.

The Schrödinger IFD protocol models induced fit docking of one or more ligands using the following steps:

1. Constrained minimization of the receptor (Glide protein preparation, refinement only) with an RMSD cutoff of 0.18 Å.
2. Initial Glide docking of each ligand using a softened potential (van der Waals radii scaling). By default, a maximum 20 poses per ligand are retained, and by default poses to be retained must have a Coulomb-vdW score less than 100 and an H-bond score less than – 0.05.

3. One round of Prime side-chain prediction for each protein/ligand complex, on residues within a given distance of any ligand pose (default 5 Å).

4. Prime minimization of the same set of residues and the ligand for each protein/ligand complex pose. The receptor structure in each pose now reflects an induced fit to the ligand structure and conformation.

5. Glide redocking of each protein/ligand complex structure within specified energy of the lowest-energy structure (default 30 kcal/mol). The ligand is now rigorously docked, using default Glide settings, into the induced-fit receptor structure.

6. Estimation of the binding energy (IFDScore) for each output pose.

The protein structures downloaded from protein data bank (PDB) and the ligand structures drawn cannot be used as such for docking. Hence the protein and ligand(s) preparation must precede the use of the protocol.

**Protein Preparation**

A typical PDB structure file consists only of heavy atoms, can contain waters, cofactors, and metal ions, and can be multimeric. The structure generally has no information on bond orders, topologies, or formal atomic charges.
Terminal amide groups can also be misaligned, because the X-ray structure analysis cannot usually distinguish between O and NH2. Ionization and tautomeric states are also generally unassigned. Glide calculations use an all-atom force field for accurate energy evaluation. Thus, Glide requires bond orders and ionization states to be properly assigned and performs better when side chains are reoriented when necessary and steric clashes are relieved. These can be achieved by the "Protein Preparation Wizard" of the Glide software. This Wizard allows one to properly prepare a protein for docking studies. This also follows the Optimized Potential for Liquid Simulations-All Atoms (OPLS-AA) force fields for energy minimization.

**Ligand Preparation**

The structures that are docked must be good representations of the actual ligand structures as they would appear in a protein-ligand complex. This means that the structures supplied to Glide must meet the following conditions:

1. They must be three-dimensional (3D).
2. They must have realistic bond lengths and bond angles. Glide only modifies the torsional internal coordinates of the ligand during docking, so the rest of the geometric parameters must be optimized beforehand.
3. They must each consist of a single molecule that has no covalent bonds to the receptor, with no accompanying fragments, such as counter ions and solvent molecules.
4. They must have all their hydrogens (filled valences).
5. They must have an appropriate protonation state for physiological pH values (around 7). For example, carboxylic acids should be deprotonated and aliphatic amines should be protonated. Otherwise a neutral aliphatic amine could improperly act as a hydrogen-bond acceptor in the docking calculations, or could occupy a hydrophobic region without incurring the large desolvation penalty.

6. They must be supplied in Maestro, SD, Mol2, or PDB format.

All of the above conditions can be met by using LigPrep module of the Glide software.

**Visualization and Analysis**

The *PyMol* Molecular Graphics System was used to analyze the hydrogen bond interactions and in the preparation of high resolution images. The hydrophobic interactions were obtained as *Ligplot* diagram by submitting the docked complexes to the online PDBsum server (http://www.ebi.ac.uk/pdbsum).

In this thesis, the compounds were subjected to property analysis of Lipinski’s rule of five, they all found to obey. Docking was carried out using Schrodinger’s GLIDE (Grid based Ligand Docking with Energetic). The prepared protein and the ligand are loaded on the Maestro GUI. IFD can be carried out by picking the co-crystallized compound in the protein structure or by mentioning the active site residues, where the test ligands should be docked. In case of more ligand, XP-mode of GLIDE virtual screening was performed. The obtained docking results were screened by the following parameters: i) Glide Energy ii) Docking score and iii) the number of h-bond with distance less than or equal to
3Å. From these best 10 ligands with top docking parameters (i and ii) were taken and subjected to IFD protocol with the respective target protein. A comparison of the IFD results of the actual drug molecule with that of the drug like compounds through the structure–function relation was carried out. The inhibitory properties of the compounds against their target proteins with which they were docked are explained in detail.