CHAPTER 1

AN INSIGHT INTO THE ZONE OF X-RAY CRYSTALLOGRAPHY AND MOLECULAR DOCKING: AN OVERVIEW

1.1 GENERAL INTRODUCTION

The X-ray diffraction method depends on an interference pattern generated by passing through the atoms within a crystal. X-ray crystallography is a great technique used to find out the three-dimensional structure of molecules. It is an important system implemented to represent the three-dimensional crystal structure, their geometrical parameters, conformation of the ring and supramolecular network created by the molecules due to various intermolecular forces and thereby explaining the stability of crystal packing. The crystal structure and conformational analysis of compounds develop into vital to recognize their role. The crystal structure investigations give us important information regarding the necessary factors that maintain the molecules in the needed conformation. In drug designing, the conformation of medicinal compounds will perform a significant job. Because the thesis explains the 3D crystal structures of few biologically important molecules, a concise introduction regarding the theory and experiment such as collection of intensity data, solving the crystal structure, refinement, structural parameters and conformations of the ring and packing of molecules are discussed.
The pharmacological parameters such as absorption, distribution, metabolism, excretion and toxic (ADMET) properties, drug ability and Lipinski rule of five (molecular properties), are investigated by pre-ADMET online servers, OSIRIS properties explorer and molinspiration online server, respectively. This pharmacological result will give an idea to the scientists whether to do further wet lab studies with the compound or reject it. From the obtained pharmacological outcomes, insilico binding (docking) studies were executed for the molecules.

The recognizing of right poses of molecules in the binding pocket of a protein and calculation of affinity between the molecule and the protein were accomplished by docking method. We can also explain docking as a procedure by which two compounds interlock in a 3D space. Docking studies have been performed by GLIDE (Gride-based Ligand Docking with Energetics) software v5.5 (Friesner et al. 2006, Halgren et al. 2004 and Glide 2009) created by Schrodinger. Docking studies can be used to find effective drugs for treating deadly diseases. It is possible for us to create many conformations for the ligand (or drug) with their target protein to identify whether the drug is healthy or not. The efficient drug can be determined from the binding affinity of the compound for the target. Hence, docking result provides in determining potent drugs for treating deadly diseases.

1.2 X-RAY CRYSTAL STRUCTURE DETERMINATION

The research work consists of several important stages, such as synthesis, intensity data collection, data reduction, structure solution, refinement and calculation of geometrical parameters. Conformational features of the molecules along with intra and intermolecular forces present in the crystal structure are also discussed.
1.2.1 Synthesis

This part of the chapter provides the experimental techniques adopted in the structural analysis of small molecules in detail. The crystals under investigation were prepared or synthesized adopting various techniques 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.

1.2.2 Crystal Mounting

Specimens suitable for X-ray diffraction are selected by examining the (transparent) crystals under a polarizing microscope. The crystal which is free from any imperfection such as cracks, voids, twining etc., is selected and is fixed to a thin glassed fiber with suitable glue. This fiber is inserted into a goniometer head through a brass bin, which holds crystals in place on a diffractometer and allows it to be oriented in the X-ray beam by means of translation and angular motions for data collection.

1.2.3 Intensity Data Collection and Data Reduction

Two general techniques are available for measuring the intensities of diffracted beams. The beams can be detected either by photographic techniques or counter techniques. During the evolution of X-ray crystallography as a structural instrument, most of the work was supported with techniques method based detectors. There are two types of counters in the part of detectors. In current days, area detector and scintillation counter are used to capture the diffracted rays. Now a day, X-ray diffractometers equipped with area detectors, which can concurrently record all the beams falling in their detector area in terms of position and intensity. There are different types of area detectors based on different technologies, each with
particular advantages and disadvantages of sensitivity, size, spatial resolution, speed read-out and cost (Stout & Jensen 1989).

In the current study, intensity data for all the crystals offered in this thesis was collected utilizing a Bruker AXS Kappa APEX II single crystal CCD diffractometer employed by graphite monochromated Moka ($\lambda=0.71073 \text{Å}$) radiation (Bruker 2004) at room temperature and a Charge-coupled device. Usually, a crystal specimen of suitable size (say, 0.30 x 0.25 x 0.25 mm) is cut and mounted on a glass fiber using cyanoacrylate. The unit cell parameters were determined by reflections collected from 36 frames measured in three various crystallographic zones by using the method of difference vectors. The intensity data is collected by an average four-fold redundancy for each reflection and optimum resolution (0.75Å).

In four-circle diffractometer, Bruker AXS Kappa APEX II single crystal CCD diffractometer mounted crystal is rotated about different axes ($\varphi, \psi, \omega$) and diffracted beam is recorded by a detector (2θ). The intensity of each reflection is measured by a quantum detector with any type of scan modes like $\omega$ or $\omega/2$. In the SMART APEX-II system, goniometer has a Eulerian Cradle (three circle geometry). The crystals were mounted in the goniometer and centered. The distance of crystal set to the detector was at 50 mm. At first, a selected number of frames were scanned and integrated for an appropriate range of values. Prior to collecting the data, the strategy of the data collection was set from the unit cell parameters and direction matrix. With the help of $\omega$-scan, $\phi$-scan or both $\omega$ and $\phi$-scan methods, the data can be collected. In the $\omega$-scan, $\phi$ is fixed and in the $\phi$-scan $\omega$ is fixed, with the range of $\omega$ and $\phi$ angle as 0 to 79º and 0 to 360º, respectively. A number of frames recorded rely on the range of $\omega$ and $\phi$ angles and in addition to the frame width 0.5º. The intensity data collection, LP correction, frames integration and decay correction are completed by SAINT (Bruker 2004) program. Empirical
absorption correction (multi-scan) was completed using SADABS (Sheldrick 1996) program.

All the intensities were corrected for variable scan speed, background and attenuation by the relation

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

where

- \( I_{raw} \) - 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

\[ |F_{hkl}| = \left( K I_{raw} f L p \right)^{1/2} \]

where,

- \( K \) - Scaling factor
- \( L \) - Lorentz factor (\( L = 1/\sin \theta \))
- \( p \) - Polarization factor = \( \frac{1 + \cos^2 2\theta}{2} \)
where $\theta$ is the Bragg angle of reflection

The raw data collected from the diffractometer undergo from physical and geometrical error factors and thus cannot be used for structure elucidation straight away. Hence, the intensity data have to be corrected for Lorentz, polarization and absorption effects. The Lorentz and polarization corrections are a must for all cases because the reflection efficiency varies with the reflection angle, as 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.

The space group of the crystal is determined using the systematic absences of the reflections and by intensity statistics. If space group ambiguity arises then the contents of the unit cell, the number of molecules present in the cell, the allocation of intensity and other relevant details are analyzed in depth.

1.2.4 Structure Solution and Phase Problem

In order to get atomic positions of the molecule, intensities are converted into structure factors. Structure factor is the resulting of N waves scattered with N atoms in the unit cell. The structure factor expression 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.

Since $F_{hkl}$ is a complex quantity, it can be written as
\[ F_{hkl} = |F_{hkl}| e^{i\phi_{hkl}} \]

where, \( \phi_{hkl} \) is a phase of the reflection hkl.

The structure factor can be expressed in terms of the integral of electron-density. The Fourier transform of the structure factor yields electron-density.

**ELECTRON DENSITY ⇔ STRUCTURE FACTOR**

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

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

In the above equation, we have to simply map \( \rho(x, y, z) \) and situate the maxima in it, in order to obtain the molecular structure. But the process is not so straightforward to sum the series of \( \rho(x, y, z) \). i.e,

**SUMMATION OF ELECTRON-DENSITY = ELECTRON-DENSITY**

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

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

where \( V \) is the volume of the unit cell and \( x, y \) and \( z \) are the fractional coordinates of any point in the unit cell. So to locate an atom, one has to find a point at which electron-density is maximum and to do this one has to compute the above synthesis for which one has to know not only the magnitude \( |F_{hkl}| \) which is obtainable from experiment but also the phase \( \phi_{hkl} \) which is missing from the experiments. Hence the phase \( \phi_{hkl} \) is necessary if one wants to calculate electron-density to set up the position of atoms. This is called the ‘PHASE PROBLEM’ in crystallography.

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

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

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

1.2.5 Direct Methods

Basic assumptions

The basis of direct methods is that there are two imposed conditions that restrict the relative values of phase angles.
1. Electron-density function calculated with deduced phases should never be negative.

2. The electron-density maps should have high values at and near atomic positions and have nearly zero values everywhere else.

Possible phase angles are constrained by these two conditions so that relative phase angles mainly depend on the mathematical expressions for Fourier series. In the initial stage, atoms are assumed to be of point atom type [As fall-off of intensity at high scattering angles is due to atomic size and atomic vibrations, \(|F_{hkl}|\) is now replaced by \(|E_{hkl}|\) which do not vary with \(\sin\theta/\lambda\)].

**Direct methods Procedures**

1. Normalized structure factors \(|E_{hkl}|\) are determined from the observed magnitudes \(|F_{hkl}|\) of the structure factors. Normalized structure factor \(|E_{hkl}|\), is given by

\[
|E_{hkl}|^2 = \frac{|F_{hkl}|^2}{\epsilon \sum_j f_j^2}, \quad \text{where } \epsilon \text{ is a constant.}
\]

Only high \(|E_{hkl}|\) values (those greater than 1.5 are used) signify greater validity of the probabilistic estimate.

2. The magnitudes \(|E_{hkl}|\) of the normalized structure factor are uniquely calculated by crystal structure and are independent of the choice of origin but the values of phase’s \(\phi_{hkl}\) depend on the choice of the origin. There exists a certain linear combination of phases, which are called structure invariants, whose values are determined by the structure alone and are
independent of the choice of origin. In shifting the origin by a vector, the phases turn out to be origin dependent while the amplitudes are not. But there are certain specific phase relations of the form given below which do not change with a shift in origin.

Triplet (Three phase structure invariant) relation is given by

$$\phi_H + \phi_K + \phi_L$$ such that \(H+K+L = 0\)

Quartet (Four phase structure invariant) is given by

$$\phi_L + \phi_M + \phi_N + \phi_P$$ such that \(L+M+N+P = 0\)

3. Convergence procedure is utilized to determine the starting set for the phase relations. Only strong \(|E_{hkl}|\) values are chosen in the generation of invariants so that the reliability of the probabilistic estimates has a direct relation to the normalized structure-factor magnitudes entering in the triplets or quartets. Sets of three Bragg reflections are selected with indices that satisfy the triple-product sign relationship \(\Sigma_2\) formula.

$$\phi(H) \approx \langle \phi(K) + \phi(H-K) \rangle_K$$

where \(\phi(H)\) means the relative phase angle of the Bragg reflection \(hkl\).

4. The phases of the above small set of reflections are now assumed to be known and knowing the values of two phases and sum of the three phases, the value of the unknown phase can be found. This is called the phase extension or phase
propagation. This propagation and phase refinement are carried out using tangent formula

\[
\tan \theta_k = \frac{\sum_{k,j} |E_{k, h-k, j}| \sin(\phi_{k, j} + \phi_{h-k, j})}{\sum_{k,j} |E_{k, h-k, j}| \cos(\phi_{k, j} + \phi_{h-k, j})}
\]

5. Depending on the choices of the phase values for the reflections chosen for the origin and enantiomorph, the direct methods procedures become multi-solution in nature from which the correct solution can be picked up. Before doing an E-map, the set with the lowest value for the combined figure of merit is choosed as the correct one.

\[
\text{CFOM} = R_\alpha + (0 \text{ or } (NQUAL - \text{wn}), \text{ whichever is larger})^2
\]

where \(\text{wn}\) is a structure dependant constant which should be about 0.1 more negative than the expected value of NQUAL.

\(R_\alpha\) is defined as

\[
R_\alpha = \sum w[\alpha - \alpha_{\text{est}}/\sum w[\alpha_{\text{est}}]^2
\]

where the weight \(w\) is \(1/(\alpha_{\text{est}}+5)\) (to avoid the largest \(\alpha_s\) dominating) and \(\alpha\) is the reliability coefficient.

NQUAL is defined as

\[
\text{NQUAL} = \frac{\left[ \sum (E1 \ast E2) \ast \sum (E3 \ast E4 \ast E5) \right]}{\sum \left[ \sum (E1 \ast E2) \ast \sum (E3 \ast E4 \ast E5) \right]^n}
\]
where the outer summations are performed over all refined reflections and inner summations are over the triplets and negative quartet reflections involving a given reflection. NQUAL approached to -1 for the right solution.

6. Since $|E_h|$ and $\phi_h$ are now known, an E-map could now be determined for this set. Mostly the entire structure will be revealed by this map. In the case of incompleteness, the existing model is refined for few cycles (i.e. isotropic refinement) and a difference Fourier will reveal the rest of the atoms.

### 1.2.6 Structure Refinement

The structures derived from various methods like direct methods or Patterson techniques are only estimated in the sense that the positions of the atoms are slightly away from the correct ones. This is because phases which decide the accuracy of the positions are in error due to various reasons.

Differences between the observed and calculated values can arise from random errors in the observations and defects in the model. The trial structure obtained from the structure solution is refined in order to get the accurate atomic positions and the associated thermal parameters.

As the structure determination means determining both the positions of the atoms and their individual strengths accurately in the lattice, one should refine the structure to achieve this. This can be achieved by various structure refinement processes:

- Full-Matrix Least-Squares Method
- Block Diagonal Least-Squares Method
- Rigid Body Refinement
- Energy Minimization
- Simulated Annealing
- Maximum Entropy Method
- Maximum Likelihood

Among them, full matrix least-squares refinement method is the conventional one and broadly used one in small molecular structure determination. In the least-squares method, one minimizes the sum of the squares of the errors, that is, the squares of the difference between the observed and the calculated structure amplitude. Thus, the refinement involves refining the positions \((x_j, y_j, z_j)\) and their corresponding thermal parameters so that the best fit would be obtained from the experimental data. The refinement is based on \(F_o^2\) since it is impossible to refine on \(F\) by all the data which would occupy collecting the square root of a negative number for reflections with negative \(F_o^2\) (after data reduction). The refinement based on \(F_o^2\) by all data provides a better result even for weakly diffracting crystals.

The residual factor or reliability index defining correctness of the model

\[
R_i = \frac{\sum |F_o| - |F_c|}{\sum |F_o|}
\]

where

\(|F_o|\) – 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

\[
wr^2 = \frac{\sum w_i \left( |F_o|^2 - |F_c|^2 \right)^2}{\sum w_i |F_o|^2}
\]

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

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

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

\[
w = 1/ [\sigma^2 (F_o^2) + (aP)^2 + bP]
\]

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

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

1.2.7 Computation of Geometrical Parameters

Crystal structure determination gives the unit cell constants and fractional atomic coordinates of all the atoms and their related thermal displacement parameters. The geometrical parameters such as bond lengths, bond angles and torsion angles can be intended from the coordinates of the relevant atoms.
In 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 specified by the law of cosines in three dimensions as

\[
L = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - 2ab \Delta x \Delta y \cos \gamma - 2ac \Delta x \Delta z \cos \beta - 2bc \Delta y \Delta z \cos \alpha}
\]

where \(a, b, c, \alpha, \beta\) and \(\gamma\) are the unit cell parameters and \(\Delta x = x_1-x_2\), \(\Delta y = y_1-y_2\), \(\Delta z = z_1-z_2\). The above expression can be applied for any crystal system to determine the bond lengths. Bond length values are helpful to identify the nature of chemical bonds (triple, double, partial double or single bond) present in the molecule.

Bond angle formed by the three atoms A, B and C (where the angle is subtended between the bonds AB and AC) can be calculated by 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 specified by ABC and BCD formed by four atoms A, B, C and D is given by

\[
\cos \chi = \frac{\mathbf{N}_1 \cdot \mathbf{N}_2}{\|\mathbf{N}_1\| \|\mathbf{N}_2\|}
\]

where \(\mathbf{N}_1\) and \(\mathbf{N}_2\) are vectors normal to ABC and BCD planes, respectively.
1.2.8  **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 present in the molecule. The probable conformation of five, six and seven membered rings are shown in Figures 1.1, 1.2 and 1.3, respectively. Five-membered rings are found to have only two probable conformations i.e. half-chair or envelope as shown in Figure 1.1. In 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 probable conformations in addition to sofa and half-chair conformations. The sofa has only one mirror and half-chair gives two-fold symmetry as shown in Figure 1.2. Seven member rings posses chair, twisted chair, boat and twisted boat conformations as shown in Figure 1.3. Several rings are found to be distorted, without having any defined conformations. In fact, conformations are explained from the puckering and asymmetry parameters which provide the extent of the ring from the ideal conformations (Cremer & Pople 1975, Duax et al. 1976). Platon (Spek 2009) program is utilized for the estimated of geometrical parameters.

The asymmetry parameters have been proposed by Duax, Weeks & Rohrer (1976) to provide a quantitative estimation of how a ring of any size deviates from ideal symmetry and to use in describing its conformation.

The conformation of the ring can also be understood from the torsion angles about the bonds forming the ring. In the case of five member rings, there will be five torsion angles about the five bonds of the ring. When all the five torsion-angles are small, then the ring will be nearly planar and the algebraic sum of torsion angles will be zero. If one of the dihedral angles is
close to zero and others are relatively large i.e. four atoms are in one plane and remaining atom is out of the plane, then the conformation is known as the envelope. The situation in which torsion angles about two consecutive bonds are nearly equal (both in magnitude and sign) and others relatively large and different, then the conformation is called as half-chair conformation. In the case of the conformation of six-member rings, the torsion angles will be alternating between $+60^\circ$ and $-60^\circ$ in chair conformation (in actual, the modulus of torsion angles can be in the range of $50^\circ$ to $70^\circ$). For a boat conformation (Rao et al. 1967) two of the non-consecutive torsion angles will be zero and will be separated by two torsion angles around $+60^\circ$ and $-60^\circ$. In half chair form, two consecutive torsion angles will be around zero. In other words, five of six atoms will be in a plane. These factors can be effectively used to describe and compare the conformation of the molecules occurring in different compounds, in a quantitative way. In seven-membered rings, asymmetry parameters cannot be used to distinguish all main conformations. Inspection of the signs of the individual torsion angles is necessary for a complete identification. Computation of these asymmetry parameters is included as an algorithm in the program PLATON (Spek 2009).
Figure 1.1  The three most symmetric conformations observed in the five-membered rings. Symmetries are indicated on the right.
Figure 1.2 The most commonly observed conformations of six-membered rings. The symmetries are indicated on the right.
Figure 1.3 The most commonly observed conformations of seven-membered rings. The symmetries are indicated on the right.

1.2.9 Forces Stabilizing the Crystal Structure

The packing energy of crystal, named as the lattice energy, is the summation of a large number of relatively weak intermolecular interactions (0.5-2.0 KJ per mol), relatively strong molecular interactions (30 KJ per mol) and particularly strong intramolecular and inter-ionic interactions (150 KJ per mol) (Perlstein 1994). The intra molecular interactions are accountable for the
bonding of atoms to form molecule while the inter-molecular forces minimize the energy of the molecules in crystal and are primarily responsible for the formation of crystals (Buckingham 1999). The inter-molecular forces, which can be attractive or repulsive in nature, consist of non-bonded, electrostatic and ionic interactions. Both hydrogen bond and Van der Walls interactions are attractive interactions falling in the category of non-bonded (i.e. non-covalent) interactions (Desiraju 2002) are the major inter-molecular forces in most pharmaceutical crystals (Desiraju & Steiner 1999). Though, if the molecule is polar and charged, ion-ion contributions can importantly affect the overall crystal packing energy (Smith 1981). Hence, the forces stabilizing the crystal structure are generally of a non-covalent type such as ion-ion interactions, dipole-dipole interactions, ion-dipole interactions, hydrogen bonding interactions, cation…π interactions, hydrophobic interactions, aromatic π-π stacking interactions and van der Walls interactions etc.

The intra-molecular and inter-molecular contacts perform an important role in medical field. Scientists with different areas carry out their research in this field for more than 40 years.

Molecular recognition caused due to weak, reversible and selective binding between two molecules or within a molecule, is of central significance in biology and chemistry; depends on non-covalent interactions. The weak non-covalent interactions between aromatic units are known to play significant roles in various fields of chemistry, biology, and material designs. In designing and synthesizing the crystals with non-covalent interactions play major roles in chemical engineering. Weak non-covalent interactions are important for making a proper fit of a molecule to a macromolecule (receptor) in order to provide its desired biological function (Drug-design), thus has gained utmost significance in last decades. X-ray crystallography takes a crucial role in understanding these interactions with determining the
structures of small molecules (ligand), receptors (macromolecules) as well as receptor-ligand complexes.

(i) **Hydrogen Bond**

A hydrogen bond is a particular type of non-bonded interaction between two electronegative atoms (donor and acceptor) wherever the hydrogen atom is bonded to them. The attractive interaction between this hydrogen with another electronegative atom of the adjacent molecules provides extra stability to the molecule. The common convention of the representation of the hydrogen bond is D-H… A, Where D is the donor and A is acceptor atom. The hydrogen bonds are very directional and the D-H…A angle should be 180° for an ideal one. Bifurcated hydrogen bond formed due to hydrogen bond through two acceptor atoms or two donor atoms.

The most significant geometrical characteristics of a hydrogen bond are that the distance between the proton and the acceptor atom is shorter than the sum of their van der Waals radii (Taylor & Kennard 1982). The van der Waals radii are:

\[
C=1.75 \text{ Å}, \ H=1.20 \text{ Å}, \ N=1.55 \text{ Å}, \ O=1.50 \text{ Å}, \ P=1.80 \text{ Å} \text{ and } S=1.80 \text{ Å}.
\]

The capability of a C-H group to act as a proton donor depends on the hybridization \([C (sp)-H > C (sp^3)-H > C (sp^3)-H]\) and enhance with the number of adjacent withdrawing groups (Steiner 1996).

(ii) **C-H…π interaction**

The attraction between the C–H bond with the \(\pi\)-system is regarded as the C-H… \(\pi\) interaction. It was first proposed by Nishio and coworkers to explain the preference of conformations in which bulky alkyl and phenyl
groups had close contact (Kodama 1977). Such close contact has been observed in the stable conformations of a number of molecules (Tsuzuki et al. 2000). Statistical analysis of the crystal structure database by Nishio and co-workers have revealed that more than 75% of organic crystals have a short contact of the C-H with $\pi$–systems (Umezawa et al. 1998, 1999). The C-H–π interaction is observed as the crucial driving force of crystal packing. Such contacts have also been examined in crystals of proteins (Quiocho & Vyas 1984, Umezawa & Nishio 1998) and the significance of this interaction in molecular recognition has been appreciated.

In the features of the growing importance of the C-H–π interaction, very small is known about the physical origin and scope of the interaction. This is very much preferred for the understanding of conformational preferences, crystal packing and communicated aspects of molecular recognition. Quantitative approximations on the interaction are desired with those who complete force field simulations, as the energy of this type of interaction is very small (< 2 Kcal/mol). The confirmation of C-H–π interaction came from studies on conformational aspects of a series of compounds having an aliphatic group on one side of the molecule and a phenyl group sited on the other side.

The mean distance H–π decreases with increase in acidity of the hydrogen atom and that the electrostatic force gives at least in interactions relating activated C-H group. This trend is the characteristics of hydrogen bonding interaction and has been seen in other weak interactions, such as the C-H–O contact (Steiner 1997). The authors have evaluated the C-H–π distances as observed in X-ray studies with those from neutron diffraction. The distances do not vary very significantly.
It is broadly believed that the C-H… π interaction mostly originates from a charge transfer process from the π system to the anti bonding orbital of C–H bond. However, dispersion forces also contribute, but the contribution from coulomb forces is insignificant. Thus the C–H… π interaction can play its role in polar media and also in a non-polar atmosphere, unlike normal H–bonding. This makes it important in biological systems.

(iii) Aromatic π-π Interaction

The attraction between two aromatic residues occurs in the same or different molecules are known as arene–arene interaction or π–π interaction. The π–π interactions are weak in strength with the energy ranging from 0-50 KJ/mol (Steed and Atwood 2000). These interactions have always been present in nature (e.g. in protein and DNA) but are less studied because of their weakness and complex nature of mechanism.

There are four types of π–π interaction geometries, edge-to-face, face-to-face, parallel displaced and Y-shaped (Headen et al. 2010, Wheeler et al. 2010). Parallel-displaced and edge-to-face geometries are the most general geometries found in structural chemistry. The parallel-displaced geometry is most general in DNA base pair stacking. Amongst the different intermolecular interactions that exist between the molecules, the π–π stacking interactions between the planar aromatic molecules are least understood (Grimme et al. 2008). Computational studies on benzene dimer confirmed that the parallel-displaced and edge-to-face geometries are almost iso-energetic with a binding energy of about 2 Kcal/mol (Tsuzuki et al. 2000). Face-to-face stacked geometry is usually observed with donor–acceptor pairs. The benzene per fluorobenzene interaction is an admirable example of this type of aromatic interaction and has been calculated to provide -15.5 KJ /mol stability (West et al. 1997). Also, there are also continuums of intermediate geometries (Blundell et al.
1986, Tsuzuki et al. 2002). Various computational techniques used for studying non-covalent interactions, as well as $\pi-\pi$ interactions, have been reviewed newly (Riley et al. 2010).

(iv) **Vander Waal’s forces**

Weak attractive forces between uncharged atoms or molecules are together referred to as Vander Waal’s forces. These forces occur from the electrostatic attraction of the nuclei of one molecule in the electrons of various molecules. The repulsion arising between the electrons of two molecules with the nuclei of two molecules counteract the electrostatic attractions. The vander Waal’s forces are short range forces i.e., they are important only when the molecules are very close to one another.

1.3 **PHARMACOLOGICAL INVESTIGATIONS**

1.3.1 **Molinspiration**

Lipinski's rule (Lipinski et al. 2001& 2004) of five also named as the Pfizer's rule of five or basically the Rule of five (RO5) is a rule of thumb to estimate drug-likeness or find out if a chemical compound with an assured biological activity or pharmacological has properties that would create it a possible orally active drug in humans. The rule was invented by Christopher A (Lipinski et al. 1997).

Lipinski's rule states:

- Not exceed 5 hydrogen bond donors
- Not exceed 10 hydrogen bond acceptors
- A molecular mass lower than 500 daltons
- An octanal-water partition coefficient log P not greater than 5
- No exceed one number of violations.

Molinspiration, the web-based software was utilized to obtain parameters such as miLogP, volume, TPSA, etc., calculated by the methodology developed by Molinspiration as an addition of fragment-based contributions and rectification factors (Verma 2012 and Molinspiration 2015). The miLogP parameter is used to check better permeability through the cell membrane. TPSA is associated with hydrogen bonding potential of compound. Calculation of volume developed at molinspiration is based on group contributors. A number of rotatable bonds determine molecular flexibility. It is a good descriptor of absorption and bioactivity of drugs.

All the parameters were checked with the help of software molinspiration online. The obtained molecular properties result score of each compound are compared with a standard drug.

1.3.2 Osiris

The computer programmer OSIRIS (Osiris property explorer 2014) is used to predict drug score, drug-likeness, mutagenic, turmeric, irritant and reproductive effect. This programmer gives overall drug score values to consider the compound as a drug. It is used to screen compounds that are most likely to have high binding affinities represented as drug score. Prediction results are valued and color coded. Orange and red indicates moderate and severe toxicity of the compounds, respectively. The green color represents that the compounds are not toxic and indicates drug-conform behavior.
1.3.3 ADMET Predictions

To find out the drug ability of the compounds, Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) properties were executed by Pre-ADMET (ADMET predictors 2015) online server. This pharmacological analysis of compounds will assist pharmaceutical scientists to go for the top candidates for improvement in addition to discarding those with a low probability of success.

Using Pre-ADMET online server, CaCo-2 cell permeability, plasma proteins binding, blood-brain penetration, skin permeability, human intestinal absorption and MDCK cell permeability properties of all compounds were calculated. Toxicity parameters such as carcinogenicity prediction in mouse and rat were done by the pre-ADME online server.

a) MDCK cell permeability provides an experimental and computational screening model for the calculation of intestinal drug absorption. The ranges of MDCK cell permeability predictions are shown in Table 1.1.

b) Human Intestinal Absorption (HIA): Calculating human intestinal absorption of drugs is significant for recognizing potential drug candidate. Pre-ADMET can predict percent human intestinal absorption (%HIA). The ranges of HIA predictions are shown in Table 1.1.

c) Blood Brain Barrier Penetration: Calculating BBB penetration means predicting whether compounds pass across the blood-brain barrier. Pre-ADMET can calculate in vivo data on rates of BBB penetration. The ranges of blood brain barrier predictions are shown in Table 1.1.
d) Plasma Protein Binding (PPB): Generally, unbound drug is offered for transport or across diffusion the cell membranes, and also for interaction with a pharmacological target. As a result, a degree of plasma protein binding of a drug manipulates not only the drug’s action but also its nature and efficacy. The ranges of PPB are shown in Table 1.1.

e) Skin permeability: In the pharmaceutical, cosmetics and agrochemical fields, it is significant to predict the skin permeability rate for crucial parameters for the transdermal delivery of drugs and for the risk assessment of all chemicals that come into contact with the skin either fortuitously or by design. The Pre ADMET can predict in vitro data on a human for skin permeability. Pre ADMET predicts in vitro skin permeability and the result value is specified as log Kp, Kp (cm/hours) is defined as

\[ Kp = \frac{Km \cdot D}{h} \]

where Km is distribution coefficient between vehicle and stratum corneum, D is average diffusion coefficient (cm²/h) and h is the thickness of skin (cm) (Singh et al. 1993).

f) CaCo-2 cell permeability: Caco-2 cells, a well-differentiated intestinal cell line resultant from human colorectal carcinoma, exhibit several of the morphological and functional properties of the in-vivo intestinal epithelial cell barrier. The ranges used are shown in Table 1.1.
Table 1.1 Upper and lower limits of ADME predictions

<table>
<thead>
<tr>
<th>Human intestinal absorption (HIA %)</th>
<th>CaCo-2 cell permeability (nm/sec)</th>
<th>MDCK cell permeability (nm/sec)</th>
<th>Plasma Protein Binding (%)</th>
<th>Blood-Brain Barrier penetration(c.brain/c.blood)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorly 0 ~ 20 %</td>
<td>Low less than 4</td>
<td>Low less than 25</td>
<td>Chemicals strongly bound More than 90%</td>
<td>CNS active compounds (+) More than 1</td>
</tr>
<tr>
<td>Moderate 20 ~ 70 %</td>
<td>Moderate 4 ~ 70</td>
<td>Moderate 25 ~ 500</td>
<td>Chemicals weakly bound Less than 90%</td>
<td>CNS inactive compound Less than 1</td>
</tr>
<tr>
<td>Well 70-100 %</td>
<td>High more than 70</td>
<td>High more than 500</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1.3.4 Carcinogenicity Prediction

Carcinogenicity is a toxicity that produces cancer in the body. Usually carcinogenicity test involves long time (usually 2 years), presently only in vivo test methods are established. Generally, the test uses mice or rats, exposing them to a compound and the variable to be observed is the existence of cancer. Pre ADMET predicts the result from its model which is built from the data of NTP (National Toxicology program) and USFDA, which are the results of the in vivo carcinogenicity tests of mice and rats for 2 years. Carcinogenicity prediction results are shown in Table 1.2.

Table 1.2 Carcinogenicity prediction results using Pre-ADMET online server

<table>
<thead>
<tr>
<th>Type</th>
<th>NTP Definition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Clear evidence of carcinogenic activity (carcinogen)</td>
<td>Negative prediction</td>
</tr>
<tr>
<td>Positive</td>
<td>No evidence of carcinogenic activity (non carcinogen)</td>
<td>Positive prediction</td>
</tr>
</tbody>
</table>
1.4 MOLECULAR DOCKING

Docking method aims to exact orientation of ligands inside the active site of the protein and to calculate the attraction between the protein and compound.

Molecular docking can be classified into two major kinds. **Rigid-flexible docking** (or rigid docking) is the molecular docking which permits only ligand (small molecule or donor) to change its orientation during docking computation. **Flexible-flexible docking** (or flexible docking, induced fit docking) change its orientation, especially around the active site.

We can determine perfect or efficient drugs for curing diseases by docking. It is possible to creat various poses to find whether the drug is potent or not. Thus, insilico binding results are useful in discovery of drugs which are efficient towards curing certain diseases.

1.4.1 Ligand Preparation

Ligprep module of v 2.3 of Schrodinger suite 2009 is utilized for the generation of ligands. The entire molecules investigated in this thesis are selected as ligand molecules. If the asymmetric unit contains many molecules, docking study was calculated for each separate molecule. At first, the 3D structure of the ligand was converted into PDB file and then saved as mol file.

1.4.2 Protein Preparation

The protein structure drawn form protein data bank cannot be used directly for docking studies because they contain heavy atoms, metal ions and water. Hence the protein structure does not have any data about topologies and bond orders. Therefore, for docking purpose, the protein structure drawn from protein data bank can be generated in an appropriate manner. The
GLIDE software was employed for the generation of protein structure. This software generates protein for insilico binding studies. Also, this software uses OPLS-AA fields for minimization of energy.

### 1.4.3 Induced Fit Docking

GLIDE software is used to carryout induced fit docking (Friesner et al. 2004 and 2006, Halgren et al. 2004, Glide 2009). In induced fit docking, the ligand fits into the target forming different conformations of the ligand-protein complex. For each conformation the ranking is provided by GLIDE score. The conformation with top rank is the best docked complex. The Discovery studio molecular graphics software v4.5 (Discovery studio visualizer 2015) was used to analyze the interactions and to create of high-resolution images.

### 1.5 AIM OF THE WORK

The literature review identifies that benzothiazepine, triazole, thiazolidine, acrylate, benzothiazole and coumarine derivatives are efficient medicinally energetic derivatives. Since the medicinal activity of entire molecules depends on their 3D structure, it is noteworthy to find them. Based on this, crystal structure analysis, pharmacological evaluations and docking analysis of ten molecules consisting of two benzothiazepine derivatives, two triazole derivatives, two thiazolidine derivatives, two acrylate derivatives, one benzothiazole derivative and one coumarine derivative have been carried out in the present work.