EXPERIMENTAL
A crystal is a solid with well defined faces, indicating that the constituent atoms, molecules or ions are arranged in an orderly repeating pattern extending in all three spatial dimensions. X-ray Crystallography is a method of determining the arrangement of atoms in a crystal, in which a beam of X-rays strikes a crystal and diffracts into many specific directions. From the angles and intensities of these diffracted beams, and using software like WinGX (Farrugia, 1999) one should get the picture of the density of electrons within the crystal.

X-ray crystallography is widely used as fundamental study in the development of many scientific fields. In its first decades of use, this method determined the size of atoms, the lengths and types of chemical bonds, and the atomic-scale differences among various materials, especially minerals and alloys. The method also revealed the structure and functioning of many biological molecules, including vitamins, drugs, proteins and nucleic acids such as DNA. X-ray crystallography is still the chief method for characterizing the atomic structure of new materials and in discerning materials that appear similar by other experiments. In this thesis we present the structure determination of nine different molecules which are of biological interest. This section provides a brief account of basic theoretical concepts and experimental methods used to determine the structure of a molecule.

CRYSTALLIZATION AND CRYSTAL SELECTION

The accuracy of the results of crystal structure analysis of a compound depends on the quality of the intensity data used and this in turn, is determined to a very great extent by the quality of the crystal used for data collection. Hence the selection of good quality crystal of proper size is very much essential for obtaining accurate results from structure analysis. Crystallization is nothing but the creation of a dynamic equilibrium between the particles in fluid phase and solid phase from saturated solutions. Several
methods like slow evaporation method, vapour diffusion method, solute diffusion method etc are available for crystallizing a compound (Stout and Jensen, 1989). Of which, the slow evaporation method is the simplest technique adopted to obtain good diffracting quality crystals. Single crystals of all the compounds studied in this thesis were obtained by the slow evaporation technique. After careful examination under a polarizing microscope, good quality crystals were chosen for diffraction studies.

**INTENSITY DATA COLLECTION**

In this thesis, the data for some compounds have been collected either using a 4-Circle diffractometer equipped with Cu Kα / MoKα radiation or SMART CCD area detector equipped with the MoKα radiation. A crystal of suitable size was mounted and accurate values of the unit cell parameters were obtained by least squares analysis of the θ values for several Bragg reflections. Intensity data were collected at room temperature. Two or three standard reflections were monitored periodically and they showed no significant drift during the data collection period for diffractometer data. All the intensities were collected for variable scan speed, background and attenuation using the relation

\[ I_{raw} = f \left[ N_c 2(Lb+Rb) \right] \text{NPI} \]

Where \( I_{raw} \) is the relative intensity

\( N_c \) is the peak count; \( Lb \) and \( Rb \) are the left and right background counts respectively.

\( \text{NPI} \) is the scan speed parameter and \( f \) is the attenuation factor.

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

\[ F_0 = k \left[ I_{raw}/Lp \right]^{1/2} \]
Where $k$ is the scaling factor relating the arbitrary intensity counts to the absolute value of the structure amplitude.

$L$ is the Lorentz factor $= 1/\sin \theta$

And $p$ is the polarization factor $= [\cos 2\theta_m + \cos^2 \theta]/[1 + \cos 2\theta_m]$

Where $\theta$ is the Bragg angle of reflection and $\theta_m$ is the monochromator setting angle.

**CCD AREA DETECTOR**

Over the past decade significant advances in macromolecular data collection technology have been achieved. Scintillation detectors which measured single reflection have been replaced by multi-wire, fast television and imaging plate are detectors capable of measuring many reflections at a time. The latest development in data collection technology is the Charged Coupled Device (CCD) based area detector. CCD based detector systems were initially developed for synchrotron applications where the high dynamic range and fast data readout of the CCD were ideally suited for synchrotron experiments.

Siemens SMART CCD area detector three-circle diffractometer (Siemens, 1996) was equipped with graphite monochromated MoKα radiation. The data collection covered over a hemisphere of reciprocal space by a combination of three sets of exposures and each set had different $\Phi$ angle ($0^\circ$ and $180^\circ$) for the crystal and each exposure of 30s covered 0.3° in $\omega$. The crystal-to-detector distance was 4 cm and the detector swing angle was $-35^\circ$. Coverage of the unique set was over 99% complete. Crystal decay was monitored by repeating thirty initial frames at the end of the data collection and analyzing the duplicate reflections. The decay was found to be negligible for the crystals studied.
STRUCTURE SOLUTION

Let $p(x,y,z)$ be the overall electron density function for a unit cell as a whole. This is related to structure factor $F_{hkl}$, the amplitude of X-ray corresponds to diffraction maxima. Let $r$ be the position vector within a unit cell given by $r = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$ where $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ are basis vectors for the crystal. The electron density function $p(x,y,z)$ and the structure factor $F_{hkl}$ are related as follows.

$$F_{hkl} = V \int \int \int \rho(x,y,z) \exp[2\pi i (hx + ky + lz)] \, dx \, dy \, dz$$

(1)

and

$$\rho(x,y,z) = \frac{1}{V} \sum_i \sum_j \sum_k F_{hkl} \exp[-2\pi i (hx + ky + lz)]$$

(2)

where $h,k,l$ are the miller indices.

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

Where $f_j$ is the scattering amplitude of jth atom in the unit cell.

In general, $F_{hkl}$ is a complex number. Let us write $F_{hkl}$ as $F_{hkl} = |F_{hkl}| \exp[i\alpha_{hkl}]$, where $\alpha_{hkl}$ is the phase angle for the Bragg reflection. So to determine $p(x,y,z)$, we should know both $|F_{hkl}|$ and $\alpha_{hkl}$ for all hkl. However, only $|F_{hkl}|$ can be determined from the intensity measurements. This leaves one part, namely, the phase $\alpha_{hkl}$ undetermined from experiments. Determination of $\alpha_{hkl}$ using only $|F_{hkl}|$ is generally known as phase problem. There are many methods to extract information about $\alpha_{hkl}$ from the available data using additional theoretical inputs. For instance presence of centrosymmetry demands that $F_{hkl}$ should be $+|F_{hkl}|$ or $-|F_{hkl}|$ accordingly as the phase $\alpha_{hkl}$ is either 0 or $\pi$.

Broadly there are two approaches to extract the phases. They are chemical modification methods and Direct methods. In the first method, the molecule is
chemically modified by replacing one atom and then the diffractions of the native molecule and chemically modified molecule are compared. The heavy metal method, the isomorphous addition method, the isomorphous replacement method and the anomalous scattering method fall under this category.

In Direct methods information about the phases of structure factor is extracted without comparing the diffraction pattern between two crystals. This method depends on the possible relationships between the various phases. The relations may be in the form of mathematical equalities or inequalities among different phases. These relations are based on certain theoretical requirement like symmetry. For instance the extracted phases $\alpha_{hkl}$ should be such that when substituted in RHS of equation (2) it should always result in positive quantity since electron density can never be negative.

DIRECT METHODS

As stated earlier the informations about the phase factor $\alpha_{hkl}$ have to be determined from the data available in $|F_{hkl}|$. In general, the amplitude $|F_{hkl}|$ and the phase factor $\alpha_{hkl}$ are independent. However there are certain constraints on $\rho(x,y,z)$ which can be effectively used to estimate $\alpha_{hkl}$. Two important theoretical inputs are

i) The density function $\rho(x,y,z)$ has to be positive.

ii) The density function $\rho(x,y,z)$ should represent discrete atoms.

Direct method uses two important quantities namely unitary structure factor and normalized structure factor $E_{hkl}$. The unitary structure factor $U_{hkl}$ is defined as

$$ U_{hkl} = F_{hkl} / \sum_j I_j $$

The normalized structure factor $E_{hkl}$ is defined as

$$ E_{hkl} = U_{hkl} / \langle U_{hkl}^2 \rangle^{1/2} $$

6
Where \[
\langle |U_{hkl}|^2 \rangle = \frac{1}{N} \sum_{hkl} |U_{hkl}|^2
\]

\(E_{hkl}\) gives the structure factor of a reflection where atoms are replaced as points which do not suffer from thermal motion. The direct method depends on the relationship between phases. Two kinds of relationships occur. They are i) equalities ii) inequalities. Many such relationships are used whose validity can be ascertained only in terms of probability. The relationship may be probably true and the important task in direct methods is to derive such relations with the estimation of probability of the relation being true.

1. When symmetry elements are present, the origin cannot be chosen arbitrarily. For example when a two fold rotation axis is present, the origin has to lie on this axis; when there is a 2\(_1\) screw axis, the origin has to lie on this axis; when two 2\(_1\) screw axes intersect each other, the origin has to be at that intersection point and so on. The phases of the normalized structure factors change with respect to shifts in origin. Although individual phase changes with respect to a shift in origin, there are some special linear combinations of phases whose values do not change with respect to shift in origin and these are called the structure invariants. A simple example is the triplet,\(^{(or a three phase structure invariant)}\) \(\varphi_H + \varphi_K + \varphi_L\) where \(H + K + L = 0\). Probabilistic estimates for these types of invariants are already available in the literature. Those linear combinations of phases whose values remain unchanged when the origins are moved within those permitted by the space group symmetry are called the structure semi invariants.

2. The structure invariants, structure semi invariants and the origin problem on every space group are all well documented in the Tables published by the International Union of Crystallography (Hu et al., 1995). These structure invariants and semi invariants play major roles in direct methods procedures for phase determinations. The structure invariants serve to link the observed magnitudes \(|E|\) with the desired phases \(\varphi\). This is the fundamental principle of direct methods.
In SHELXS, triplets are followed by negative quartets estimation. No group scattering factors are used, and the quartets are employed as figure of merit only. An option exists to generate the two phase semi invariants from the quartets, but these relationships are not automatically used in subsequent phasing steps. Sigma-1 type of formulae are always used to estimate the one phase semi invariants.

3. Phase relations are set up using three phase or four phase invariants. This is called convergence procedure. In this procedure only reflections with strong \(|E|\) values are chosen as the reliability of the probabilistic estimates has a direct relation to the product of the normalized structure factor magnitudes of the reflections entering into the triplets or quartets. The bottom of the convergence list has few reflections which are common in many of the invariants. They are called the well-linked reflections.

4. The phases of the above small set of reflections (well-linked ones) are now assumed (for the choice of origin and enantiomorph) and knowing the values of two phases and the value of the sum of three phases, the value of the unknown phase can be found. This is called **Phase propagation or phase extension**. Magic integers are used in this process.

5. The above procedure for phase extension is well documented (Karle & Karle, 1966) and the phase refinement is carried out using the tangent formula (Karle & Hauptman, 1950)

6. Depending upon the choices of the phase values of the reflections chosen for the origin and enantiomorph, the direct methods procedures became multisolution in nature. Before doing an E-map (a Fourier synthesis with E’s as coefficients), the set with the lowest value for the combined figure of merit (as in SHELX program) is selected as the correct one.
7. Since $|E_h|$ and $\phi_h$ are now known, an $E$-map could now be calculated for this set. Mostly the entire structure will be revealed by this map. In case of incompleteness, the existing model is refined for a few cycles (isotropic refinement) and a difference Fourier will reveal the rest of the atoms.

**STRUCTURE SOLUTION**

In this thesis the structure solution of all the compounds were carried out using SHELXS programs written by G.M. Sheldrick (SHELXS97). SHELXS is primarily designed for the solution of ‘small moiety’ (1-200 atoms in the unit cell) structures from single crystal at atomic resolution. The program is a general one and is efficient for all space groups in all settings and there are effectively no limits on the number of the reflection data, atoms, phases refined in direct methods, scattering factor types etc. The direct methods routine of SHELXS is based on the random start multi solution technique. The program is more accurately described as a multiple permutation single-solution procedure, since it tries hard to identify the correct solution, which is then ‘improved’ by the E-Fourier or partial structure extension procedure. The program starts by sorting all reflections with $E>E_{mn}$ into groups. This program uses some selected one-phase seminvariants with known phases. The initial phase refinement is normally performed using reflection subsets (which include the fixed phases). The best 10% of the subset phase permutation as judged by $R_o$ (defined below) is then refined using the full set of reflections. The best set with lowest combined figure of merit (defined below) is taken as the correct structure. For the best set the residual $R$ factor for the $E$-values is calculated by the program and for a correct solution the $R$ value ranges from 20-30%. The Combined Figure of Merit (CFOM) is defined as.

$$CFOM = R_o + [0 \text{ or } (\text{NQUAL}-\text{wn})], \text{ whichever is larger}$$

Where $\text{wn}$ is a structure dependent constant which should be about 0.1 more negative than the expected value of NQUAL. CFOM should be minimum for the best solution and only this solution is retained for calculated the $E$-map.
$R_a$ is defined as

$$R_a = \frac{\Sigma w (\alpha - \alpha_{\text{est}})^2}{\Sigma w (\alpha_{\text{est}})^2}$$

Where the weight w is $1/(a_{\text{est}} + 5)$ (to avoid largest $\alpha$'s dominating) and $\alpha$ is the reliability coefficient. NQUAL is defined as:

$$\text{NQUAL} = \frac{\Sigma |\Sigma (E_1 \cdot E_2) \cdot \Sigma (E_3 \cdot E_4 \cdot E_5)|}{\Sigma ||\Sigma (E_1 \cdot E_2)\| \cdot ||\Sigma (E_3 \cdot E_4 \cdot E_5)||}$$

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 -1 for the correct solution.

**STRUCTURE REFINEMENT**

The data obtained from an X-ray diffraction experiment are a set of structure factor magnitudes $|F_o|$ and with the proposed model the structure factor magnitudes $|F_c|$ should be calculated. The objective is to get the best possible fit between the observed and calculated structure factors. Specifically it is necessary to find the atomic coordinates $(x_i, y_i, z_i)$ and factors that give maximal agreement with the observed structure factors. The ultimate purpose is therefore to refine three positional coordinates and up to six thermal parameters for each atom to give the best fit with the experimental data. The refinement stage of a structure analysis begins with a completed trial structure containing all the atoms. A structure can be refined smoothly beginning with $R$ (reliability index) in the range 0.15 to 0.35. Full matrix least-squares refinement techniques are the conventional one which is widely used in small molecular structure refinement. SHELXL-97 (Sheldrick, 2008) was used for full-matrix refinement. The least squares refinement uses the squares of the differences between the observed and calculated values as measures of their disagreement, and adjusts the parameters so that the total disagreement is a minimum. The refinement is based on $F_o^2$ because it is
impossible to refine on F using all the data which would involve taking the square root of an negative number for reflections with negative \(F_o^2\) (i.e., background higher than the peak as a result of statistical foundation). The refinement \(F_o^2\) using all the data provides a good result for weakly diffracting crystals and in particular for pseudo symmetry problems.

The residual factor or reliability index, \(R_1\) defining the correctness of the model is given by

\[
R_1 = \frac{\sum |F_o| - |F_e|}{\sum |F_o|}
\]

where the summation is made over all observed reflection \([F_o > 4\sigma(F_o)]\). Lower the \(R\) value, greater is the accuracy of the molecular model. A suitable weighting scheme is applied at the end of the refinement procedure and the weighted \(R\) factor \(wR_2\) (intensity based) is given as

\[
wR_2 = \frac{\sum w_i (|F_o|^2 - |F_e|^2)^2}{\sum w_i (|F_o|^2)}
\]

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

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

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

\[
W = 1/[\sigma^2(F_o^2) + (aP)^2 + bP] \text{ a and b are the constants and}
\]

\[
P = [2 F_e^2 + \text{Max} (F_o^2, 0)/3].
\]
CALCULATION OF GEOMETRICAL PARAMETERS

Crystal and Molecular structure determination provides us 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 in 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

$$L = \sqrt{((\Delta x)\Delta x)^2 + ((\Delta y)\Delta y)^2 - 2(\Delta x\Delta y)\cos\gamma - 2(\Delta x\Delta z)\cos\beta - 2(\Delta y\Delta z)\cos\alpha}$$

Where $a$, $b$, $c$, $\alpha$, $\beta$, $\gamma$ are the unit cell parameters $\Delta x = x_1 - x_2$, $\Delta y = y_1 - y_2$, $\Delta z = z_1 - z_2$. 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 bond (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

$$\theta = \cos^{-1} \frac{(AB)^2 + (AC)^2 - (BC)^2}{2(AB)(AC)}$$

Bond angles are useful to find the type of hybridization of a particular atom. Torsion angle is the angle inclined between the two planes designed as XYZ and YZU of the four atoms X, Y, Z and U.

Torsion angles are calculated as

$$\chi = \frac{N_1 X N_2}{|N_1| |N_2|}$$
where $N_1$ and $N_2$ are normals to the XYZ and YZU planes respectively. The magnitude of the torsion angle may also be written as

$$|\phi| = \cos^{-1} \left| \frac{e_1 \cdot e_2}{|e_1| |e_2|} \right|$$

**INTER AND INTRAMOLECULAR INTERACTIONS**

In the crystalline state, the molecules are stabilized by intermolecular interactions like hydrogen bonds and van der Waals forces. Hydrogen bonding is the specific type of non-bonded interaction between two electronegative atoms (donor and acceptor) where the hydrogen atom is bonded to them. The attractive interaction between this hydrogen with another electronegative atom of the neighboring molecules will stabilize the molecule. The usual convention for the representation of the hydrogen bond is $D-H...A$ where $D$ is the donor and $A$ is the acceptor. The hydrogen bonds are highly directional and $D-H...A$ angle should be $180^\circ$ for ideal one.

The crystal structures presented in this thesis are found to have N-H..O, C-H..O, O-H..O C-H..N, C-H..F, C-H..Cl and C-H..Br hydrogen bonds. The existence of C-H..O bonds in crystals is evident from the study of Taylor and Kennard (1982) and Desigaju (1989/1996). The most important geometrical characteristics of hydrogen bonds is that the distance between the hydrogen and acceptor atom is shorter than the sum of their van der Waals radii (Taylor & Kennard, 1982). The van der Waals radii used were as follows: $C=1.75$, $H=1.20$, $N=1.55$, $O=1.50$, $P=1.80$ and $S=1.80\text{Å}$. The ability of a C-H group to act as a hydrogen donor depends on the hybridization $[C(sp)-H>C(sp^3)-H>C(sp^3)-H]$ and increases with the number of adjacent withdrawing groups (Stenier, 1996).

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.

GRAPH SET DEFINITIONS

The graph set approach to the analysis of hydrogen-bond patterns is the fact that even complicated networks can be reduced to combinations of four simple patterns, each specified by a designator: chains (C), rings (R), intramolecular hydrogen-bonded patterns (S) and other finite patterns (D). Specification of a pattern is augmented by a subscript designating the number of hydrogen-bond donors $d$ (in the most common case covalently bonded hydrogens, but certainly not limited to them), and a superscript giving the number of hydrogen-bond acceptors $a$. In addition, the number of atoms $n$ in the pattern is called the degree of the pattern and is specified in parenthesis. The graph set descriptor is then given as $G_{da}^n$, where $G$ represents one of the four possible designators (Bernstein et al., 1995).

The graphic plots were drawn using programs PLUTON (Spek, 2009) and ZORTEP (Zsolnai, 1998). All the calculations were performed using Pentium-III PC computing systems available in the department.