CHAPTER 3

CRYSTAL STRUCTURE DETERMINATION

3.1 X-RAY CRYSTALLOGRAPHY

X-ray crystallography is a standard technique for solving crystal structures. Today X-ray crystallography is widely used in materials and biological research. Structures of very large biological machinery (e.g. protein, DNA complexes, and virus particles) have been solved using this method. 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 off 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 three-dimensional molecular structure, their geometrical description in terms of bond lengths, bond angles, torsion angles, non-bonded distances, conformation adopted by the molecule, intra-molecular forces and molecular packing governed by intermolecular forces are the outcomes of crystal structure determination.
3.2 THE METHODOLOGY

This section provides an overview of the X-ray crystallographic structure analysis procedures of small organic molecules and hence it is essential to give a brief summary of the crystal structure determination.

3.2.1 Preparation, Selection and Mounting of Samples

The sample to be used for structure determination must be a single crystal, in which all the unit cells are identical and should be aligned in the same orientation, so that they can scatter to give a clear diffraction pattern consisting of individual X-ray beams, each in definite direction.

3.2.1.1 Crystal growth

To perform X-ray crystallography, it is necessary to grow crystals with edges around 0.1-0.3mm. Among the various methods of growing single crystals from a saturated solution; slow evaporation method occupies a predominant place owing to its versatility and simplicity.

All the samples presented in this thesis were obtained by slow evaporation method at room temperature. In this method generally the compound (solute) is dissolved in a suitable solvent or some time with mixed solvents. In almost all cases, the vapour pressure of the solvent above the solution is higher than the vapour pressure of the solute and therefore, the solvent evaporate more rapidly and the solution become supersaturated. Therefore the supersaturating solution is maintained under suitable conditions in such a way that the rate of evaporation of the solvent should be minimized. Best crystals are usually produced when the solution is free from mechanical
vibration and are allowed to evaporate without disturbance (Stout and Jensen 1989).

3.2.1.2 Crystal selection

The crystals were examined with a polarizing microscope under crossed polaroids to check for imperfection such as cracks, voids, twining or for microscopic sub-crystals also to make sure that complete extinction of light at certain angles of rotation of the stage. If the crystal is not single it will cause problems when diffraction data are measured. The crystals are cut with a sharp scalpel, in such a way that the size of the crystal should not exceed the diameter of the X-ray collimator used in the diffractometer. This is to ensure that the sample has to be completely bathed in X-ray at any orientation.

3.2.1.3 Crystal mounting

The selected crystals were glued on to a thin glassed fiber with suitable glue. The fibre through a brass bin is inserted into a goniometer head, 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.

3.2.2 Determination of unit cell parameters and intensity data collection.

The unit cell dimensions \((a,b,c,\alpha,\beta,\gamma)\) can be determined from accurately measured \(2\theta\) values (the angular deviation from the direct undeviated beam) of about 25 reflections. The crystal is rotated at various axes \((\phi, \chi, \omega)\) and the diffracted beam is then recorded with the detectors \((2\theta)\). The intensity of each reflection is measured with a quantum detector by any
type of scan modes like $\omega$ or $\omega/2\theta$. The data collection procedure depends on the type of diffractometer used for the experiment. In the present study the intensity data were collected using a four-circle diffractometer (Enraf-Nonius CAD-4) and Bruker AXS APEX-II CCD area detector with $\omega$-2$\theta$ scan mode or $\omega$ and $\varphi$ scans and with monochromatic radiation (CuK$\alpha$ & MoK$\alpha$).

### 3.2.3 Data Reduction

The raw data collected from diffractometer suffers from physical and geometrical factors and hence could not be used for structure elucidation immediately. The collected intensity data is to be corrected for Lorentz and polarization and absorption effects and is given by the equation,

$$I_{hkl} = \left(\frac{KI_{hkl}}{LP}\right)$$

(3.1)

Where $P$ is the Polarization factor, given by

$$P = \left(\frac{1 + \cos^2 2\theta}{2}\right)$$

(3.2)

$L$ is the Lorentz factor, depends on the measurement technique used and is given by

$$L = \frac{1}{\sin 2\theta}$$

(3.3)

The absorption of X-rays by the crystal is governed by the relation

$$I = I_0 \exp(-\mu t)$$

(3.4)

where $\mu$ is the linear absorption coefficient and the X-ray beam travels an average distance $t$ inside the crystal. Absorption correction becomes vital,
when the crystal has more absorbing elements for the incident X-ray wavelength.

### 3.2.4 Space Group Determination

The space group of the crystal can be determined from the systematic absences of the $hkl$ reflections. If the space group ambiguity arises, then the content of the unit cell is analysed by measuring the density. In some cases, intensity statistics are used to finalise the space group, particularly, to distinguish between centrosymmetric and non-centrosymmetric cases.

### 3.2.5 Structure Factor

The X-ray radiation scattered by one unit cell of a structure in any direction in which the diffraction maximum has particular combination of amplitude and phase is known as structure factor.

The general expression for the structure factor is given by

$$F_{hkl} = \sum_{j=1}^{N} f_j \exp\left[2\pi i (hx_j + ky_j + lz_j)\right]$$  (3.5)

Where $f_j$ is the atomic scattering factor for the $j^{th}$ atom, $x_j, y_j, z_j$ are the fractional coordinates of the $j^{th}$ atom and $N$ denotes the total number of atoms in the unit cell.

Also, $F_{hkl}$ is a complex quantity, written as

$$F_{hkl} = |F_{hkl}| e^{i\phi_{hkl}}$$  (3.6)

$|F_{hkl}|$ is the structure amplitude and $\phi_{hkl}$ is the phase angle of the reflection $hkl$. 
3.2.6 Electron Density

The intensities of the diffraction pattern and the arrangement of atoms in the unit cell of the crystal structure are related to each other by Fourier transform. The diffraction pattern is the Fourier transform of the electron density and the electron density itself is the Fourier transform of the diffraction pattern. Thus the atomic positions in the molecular structure is determined by noting the electron density maxima in the unit cell and it is given by the equation,

$$\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F_{hkl} \exp\left[-2\pi i (hx + ky + lz)\right]$$  \hspace{1cm} (3.7)

Where $\rho(x, y, z)$ is the electron density at position $(x, y, z)$, $V$ is the volume of the unit cell and $F_{hkl}$ is the structure factor for a reflection $hkl$.

3.2.7 The Phase Problem

The above equation (3.7) requires the exact values of the complex quantity $F_{hkl}$ in terms of magnitude and phases. However, it is possible to obtain only the structure amplitudes $|F|$ directly from the observed intensity as its square root. The phase angles cannot be directly measured from the experimental conditions.

The unavailability of phase angles of the diffracted beams leads to the central difficulty in structure determination using X-ray crystallography. This is referred to as the "phase problem", which arises from the fact that the diffraction data contains information only on the amplitude but not the phases.
3.2.8 Structure Solution

Several methods are used to solve the phase problem. Some of them are given below

- Direct methods
- Heavy atom methods
- Anomalous dispersion methods
- Isomorphous replacement method

The above methods can be successfully used to trace out the approximate positions of all the atoms (trial structure of a molecule) in a unit cell. This process is known as structure solution. If the molecule consists of limited number of light atoms, then direct methods can be used for the structure determination.

3.2.9 Direct Methods

Direct methods are used to calculate the phases directly by simple mathematical means from a single set of X-ray intensities. The basic postulates of direct methods are positivity (i.e., the electron density is positive everywhere) and atomicity (the atoms are spherically symmetric). The structure amplitudes and phases are linked through knowledge of electron density by Fourier transformation. A mathematical constraint on the function $\rho(x,y,z)$ imposes corresponding constraint on the structure factor. This constraint is sufficient to evaluate $\phi_{hkl}$ directly. The steps involved are given below:
(1) Conversion of observed structure factors $|F_{hkh}|$ to normalised structure factors $|E_{hkh}|$ which are independent of $\theta$.

(2) Setting up of phase relations using triple phase relation (triplets) and four phase relations (quartets).

(3) Selection of a few reflections, the phases of which are assigned aprior.

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

(5) The correct solution are then to be selected by using combined figures of merit (CFOM).

(6) Calculation of E-map and getting the trial structure.

3.2.10 Structure Refinement

Structure refinement is the process of improving the parameters for all atoms in an approximate (trial) structure, until the best fit of calculated structure factor amplitudes to those observed is obtained. Differences between the observed and the calculated values can arise from the random errors (statistical fluctuations) in the observations and defects in the model (systematic errors). The trial structure obtained from the structure solution is refined to get the accurate atomic positions and the associated thermal parameters. This process usually requires many successive stages. Full matrix least-squares refinement technique is the conventional one and widely used in small molecular structure determination.

3.2.11 The SHELX programs

Now a days structure solution and structure refinement is done using SHELX97 program package developed by Professor George Sheldrick of University of Gottingen, Germany. It is a set of programs for crystal
structure determination from single crystal diffraction data. These programs are used in well over 50% of small molecule structure determinations. Now computing improvements led to increased use of SHELX for macromolecules also. SHELX97 contains six executable programs and are given below,

SHELXS – structure solution by Patterson and Direct methods
SHELXL – structure refinement
SHELXPRO – user interface to SHELX and other programs
SHELXH – for refinement of very large structures.
SHELXWAT – automatic water devising for macromolecules.
SHELXA – to make absorption correction.

Direct methods are very successful for small molecule structures with upto 100 (non-H) atoms, where they have become the standard way to solve the phase problem. Usually direct methods will be initiated with the single SHELXS command TREF (e.g. TREF 5000).

The structure solution by direct methods includes various initial processing R-values. They are $R_{int}$ and $R_{\sigma}$ which are usually represented by,

$$ R_{int} = \frac{\sum |F_o^2 - F_\sigma^2 (mean)|}{\sum F_o^2} \tag{3.8} $$

Where both summations involve all input reflections, and

$$ R_{\sigma} = \frac{\sum |\sigma F_o^2|}{\sum F_o^2} \tag{3.9} $$

summation over all reflections.

$R_{int}$ gives an estimate of the correctness of the symmetry of the crystal system whereas $R_{\sigma}$ gives the quality of collected data.
3.2.12 The R-factor

The agreement between calculated and observed structure factors is usually represented by a residual index called R-factor, which describes the correctness of the model structure, given by

$$R = \frac{\sum |F_o| - |F_i|}{\sum |F_o|}$$  \hspace{1cm} (3.10)

Lower the R-value, greater the accuracy of the molecular model. When $R = 0\%$, there is perfect agreement between observed and calculated intensities. But this cannot be achieved due to the systematic and random errors present in the data collection and refinement procedures. In general, a good structural model will give a final $R$ value of $\sim 5\%$.

3.2.13 Weighted R-factor

The residual factor used very widely in any computer program for crystal structure determination is weighted R-factor and is given by,

$$wR^2 = \left(\frac{\sum w(F_o^2 - F_c^2)^2}{\sum w(F_o^2)^2}\right)^{1/2}$$  \hspace{1cm} (3.11)

where each reflection has its own weight $w$.

$$w = \frac{1}{\sigma^2 F_o^2 + (K1P)^2 + K2P}$$  \hspace{1cm} (3.12)

$K1$ and $K2$ are constants and $P = (F_o^2 + 2F_c^2)/3$

Values of $wR^2$ and other residual factors based on $F^2$ are generally higher than those based on $F$ values.
3.2.14 Goodness of fit

The Goodness of Fit (GOOF), $S$, is a measure of how good the model is.

$S$ is defined as

$$S = \left( \sum w(F_o^2 - F_c^2)^2 \right)^{1/2} \quad (3.13)$$

where $n$ is the number of reflections used in refinement and $p$ is the total number of parameters refined including the overall scale factor. The goodness of fit is always based on $F^2$.

3.3 INTERPRETATION OF RESULTS

The structure thus determined is represented as atoms joined together by chemical bonds. Thus from the atomic coordinates, unit cell geometry and symmetry, many geometrical results can be derived. These include bond lengths, bond angles, torsion angles, shapes and conformations of rings, the planarity (or) otherwise of a group of atoms, degree of association and intermolecular geometry such as hydrogen bonding, van der Waals contacts, π-interactions, stacking of planar aromatic groups.

3.3.1 Bond Length

The interest in determining the structure of a molecule often relates to quantities that can be derived from the structural parameters. Prominent among these are bond lengths and bond angles. The equation by which the bond length is evaluated is given below:
For a triclinic lattice the distance between two points in fractional coordinates \((x_1, y_1, z_1)\) and \((x_2, y_2, z_2)\) is given by the law of cosines in three dimensions,

\[
l = \left\{ (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - 2ab\Delta x\Delta y\cos \gamma \\
- 2ac\Delta x\Delta z\cos \beta - 2bc\Delta y\Delta z\cos \alpha \right\}^{1/2}
\]

(3.14)

where \(a, b, c, \alpha, \beta\) and \(\gamma\) are the unit-cell parameters.

The standard deviation in the bond length is given by the equation,

\[
\sigma_l = \left\{ (\sigma_{x_1}^2 + \sigma_{x_2}^2) \left( \frac{\Delta x - \delta x}{l} \right)^2 + (\sigma_{y_1}^2 + \sigma_{y_2}^2) \left( \frac{\Delta y - \delta y}{l} \right)^2 + (\sigma_{z_1}^2 + \sigma_{z_2}^2) \left( \frac{\Delta z - \delta z}{l} \right)^2 \right\}^{1/2}
\]

(3.15)

where \(\sigma_{x_1}\) and \(\sigma_{x_2}\) are the standard deviations in the positions of atoms 1 and 2 in the \(x\) direction, with similar meaning for \(\sigma_{y_1}, \sigma_{y_2}, \sigma_{z_1}, \sigma_{z_2}\). \(\Delta x\) is \((x_2 - x_1)\) and so on for \(y\) and \(z\) and \(l\) is the bond length.

### 3.3.2 Bond Angle

If the lengths \(AB, AC\) and \(BC\) are known, then the law of cosines provides a direct means of computing the angle

\[
\theta = \cos^{-1} \left[ \frac{(AB)^2 + (AC)^2 - (BC)^2}{2(AB)(AC)} \right]
\]

(3.16)

Also the standard deviation in the bond angle is given by the equation,

\[
\sigma_{\theta} = \left[ \frac{\sigma_A^2}{(AB)^2} + \frac{\sigma_B^2}{(AC)^2} + \frac{\sigma_C^2}{(AC)^2} \right]^{1/2}
\]

(3.17)

where \(\sigma_A, \sigma_B, \sigma_C\) are the standard deviations in the positions of the atoms \(A, B, C\).
3.3.3 Torsion Angle

For an arrangement of four atoms 1,2,3,4, the torsion angle \( \chi(1,2,3,4) \) is defined by the angle between the planes 1,2,3 and 2,3,4 and its value lies between in the range \(-180^\circ < \chi \leq 180^\circ\). If \( \chi = 0^\circ \) the configuration is called cis, for \( \chi = 180^\circ \) it is trans and \( \chi = \pm 60^\circ \) it is \( \pm \text{gauche} \).

3.3.4 Least Squares Planes

While discussing the results of the crystal structure determination, it is necessary to discuss whether sets of four or more atoms are planar within experimental error or not. The planarity is described in terms of the least squares plane through the set of atoms, that is, the plane that minimizes \( \sum_m d_m^2 \), where \( d_m \) are the perpendicular distances of the \( m \) atoms from the plane.

3.3.5 Ring Conformation

As far as the closed rings are concerned, the conformations of the ring differ due to the hybridization nature of the atoms. Commonly occurring closed rings are five and six membered ones. Apart from these seven and eight membered also occur rarely. The six membered rings possess various conformations namely planar, chair, boat, sofa, twist and half chair. The five membered rings adopt planar, envelope, half chair or twist conformations. The possible conformations in which the eight membered rings could adopt are crown, saddle, boat-boat, boat, twist-chair-chair, chair, chair-chair, twist-chair, twist-boat-chair and boat-chair and is shown in Figure 3.2. The condition for ideal conformations of any type is given by the endo cyclic torsion angles of the ring. Any deviation from the ideal ring conformations is described by the puckering parameters suggested by
The conformation of a n-membered ring is uniquely defined by its n-3 puckering parameters. For a six membered ring there are three puckering degrees of freedom. These are described by the single amplitude phase pair \( (q_2, \phi_2) \) and a single puckering coordinate \( q_3 \).

The ring conformation can also be calculated from the asymmetry parameters (Duax et al. 1976; Nardelli 1983). These parameters are calculated as root mean squares of the sum of mirror-related torsion angles \( \Delta C_s \) or root mean squares of the differences of two fold-axis related torsion angles \( \Delta C_2 \).

The two equations used to calculate the asymmetry parameters are

\[
\Delta C_s = \sum_{i=1}^{m} \left( \frac{\phi_i + \phi_i'}{m} \right)^2 \quad (3.18)
\]

\[
\Delta C_2 = \sum_{i=1}^{m} \left( \frac{\phi_i - \phi_i'}{m} \right)^2 \quad (3.19)
\]

where \( m \) is the number of individual comparisons and \( \phi_i \) and \( \phi_i' \) are the symmetry related torsion angles. The ring conformation also may be defined with the help of symmetry elements. Two types of symmetry that needed to be considered in order to define ring conformation are mirror planes perpendicular to the dominant ring plane and twofold axes lying in the ring plane (Figure 3.1).

In the case of an eight membered ring the standard puckering analysis yields three amplitudes \( (q_2, q_3, q_4) \) and two phase angles \( (\phi_2, \phi_3) \). The five parameters can be mapped by reducing the number of amplitudes through the introduction of an angular variable \( \theta \) \( (0<\theta<\pi) \) such that \( q_4 = Q \cos \theta \). A set of values describing various conformations are given by Evans and Boeyns (1988). The possible symmetry elements with respect to the mean plane through a puckered eight membered ring include vertical axis and
planes, and horizontal two fold axes. The possible vertical axes are $C_4$, $S_4$ and $C_2$.

### 3.3.6 Crystal Packing

A crystal is packed together by non-covalent or weak forces like van der Waals forces, ionic forces, hydrogen bonding and so on. Molecules have the tendency to develop weak interactions that are so important, because they join together and thus generate very strong molecular conformations. For example, the secondary, tertiary and quaternary structure of proteins, the double helix of the DNA, the membrane structures are all maintained by weak interactions. The more weak interactions, the more stable is the resulting conformation.
Figure 3.1 The possible conformations of the five and six membered rings
Figure 3.2 The possible conformations of eight membered ring
3.3.7 Hydrogen bonds

A hydrogen bond is the attractive force that arises between the donor covalent pair D-H in which a hydrogen atom H is bound to a more electronegative atom D (donor), and other non-covalently bound nearest neighbour electronegative (acceptor) atom A. Hydrogen bonds play a crucial role in determining the structure of water, the folding of proteins and the pairing of base in DNA etc. For this reason, crystal-packing studies are essential to understand the laws governing the intramolecular and intermolecular H-bonding in a molecular crystal. An important feature of hydrogen bond is that they are highly directional. The strongest hydrogen bonds are those in which donor, hydrogen and acceptor atoms are collinear. Strong hydrogen bonds of 20-40 kcal/mol, generally formed between charged donors and acceptors, are nearly as strong as covalent bonds. Weak hydrogen bonds of 1-5 kcal/mol, sometimes formed with carbon as the proton donor, are no stronger than van der Waals interactions. Moderate hydrogen bonds, which are the most common are formed between neutral donors and acceptors, are from 5-15 kcal/mol.

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 geometrical parameters for characterizing a hydrogen bond is shown in Figure 3.3

![Figure 3.3](image)

**Figure 3.3** Definition of the geometrical parameters $D_0$, $d$ and $\theta$ for characterizing a hydrogen bond D-H…A. Donor and acceptor are denoted by D and A respectively.
The most important geometrical characteristics of hydrogen bonds is that the distance between the hydrogen and the acceptor atom is shorter than the sum of their van der Waals radii (Taylor and Kennard 1982).

In general, crystal structures are governed by O-H…O, N-H…O, N-H…N and C-H…O types of hydrogen bonds, which are briefly discussed below, in addition to short contacts and halogen atoms involved contacts. An O-H…O interaction is a hydrogen bond, if the H…O distance is significantly less than the sum of their van der Waals radii (2.6Å) and the angle O-H…O>170° (Olovsson and Jonsson 1976). At present, even the long range electrostatic interactions are considered to be weak hydrogen bonds with H…O distance<3.0Å and the angle O-H…O>90° (Steiner and Saenger 1993).

N-H…O and N-H…N hydrogen bonds are another important class of hydrogen bonds. The formation of base pairs in DNA and folding of proteins are mainly due to the presence of N-H…O and N-H…N hydrogen bonds. The ideal H…O distance is 2.05Å and N-H…O distance is 2.9Å with an angle of 170° (Legros and Kvick 1980; Berkovitch-Yellin and Leiserowitz 1984). Regarding N-H…N hydrogen bonds, the distance criteria are 3.10Å.

The concept of C-H…O hydrogen bond was not accepted in earlier days. After the conclusive evidence of the existence of C-H…O bonds in crystals provided by Taylor and Kennard (1982), the C-H…O hydrogen bond was paid more attention. It is also given as an electrostatic interaction with C…O distance 3.0-4.0Å and angle 90-180° (Desiraju 1991). The ability of a C-H group to act as proton donor depends on the hybridization [C(sp)-H>C(sp²)-H > C(sp³) – H], and increases with the number of adjacent electron withdrawing groups (Steiner 1996).
In addition to the above hydrogen bonds, there are various other hydrogen bonds like C-H…N, C-H…Cl, C-H…S and N-H…S, which are pronounced in crystal structural chemistry. Apart from them, there is also a possibility of interactions like C-H…π interactions in molecular crystals, where the delocalised electrons in an unsaturated terminal alkynes (C≡C) are ready to interact with the C-H group (Steiner 1996). The π…π interactions are also present in planar molecules, where the planar rings stack one over the other.

### 3.3.8 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 parentheses. The graph set descriptor is then given as $G_d^a(n)$, where $G$ represents one of the four possible designators (Bernstein et al 1995).

### 3.3.9 van der Waals forces

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 van der Waals forces are used for non-specific attractions between two atoms that are close to each other. These interactions depend on the distance between the respective atoms (or atom groups or molecules). At too close distances, repulsive forces are dominating (overlapping of electron shells). The energy of van der Waals attractions is only slightly higher than that of thermal molecular movements: -0.7 to -1 kcal/mol (-3 to -4 kJ/mol). Consequently, van der Waals attractions are under physiological conditions only of importance, if as many atoms of a molecule as possible are engaged. They are strongest, if the involved molecular structures complement each other. Van der Waals attractions are additive and have thus a much greater impact on macro than on small molecules.