CHAPTER I

INTRODUCTION TO X-RAY CRYSTAL STRUCTURE DETERMINATION:

Structure analysis by X-ray diffraction technique is the most powerful method of determining the relative atomic positions in a molecular structure. In this chapter, a brief account of the principles of X-ray diffraction theory is outlined. However, for detailed studies one can refer to monographs, reviews and books (Stout & Jensen, 1968; Ladd & Palmer, 1977; Dunitz, 1979; Luger, 1980; Woolfson, 1987; Hauptman, 1988 & Giacovazzo, Monaco, Viterbo, Scordari, Gilli, Zanotti & Catti, 1992).

A crystal structure can be considered as a three-dimensional periodic packing of identical unit cells. The cell contains a number of unique atoms, a set of symmetry elements and additional atoms generated by the symmetry from the unique ones. Alternatively, the structure can be described by means of a three-dimensional periodic electron density distribution $\rho(\mathbf{r})$, the maxima of which coincide with the atomic positions. Because $\rho(\mathbf{r})$ is continuous and periodic it can be expressed as a Fourier series.

$$
\rho(\mathbf{r}) = \frac{1}{V} \sum F_H \exp[2\pi i (\mathbf{H}.\mathbf{r})]
$$

(1.1)

$V = \text{Volume of the unit cell,}$

$F_H = \text{the structure factor of the reflection } \mathbf{H}(=hkl)$

$\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}; (x,y,z) \text{ fractional coordinates with respect to the unit cell axes } \mathbf{a}, \mathbf{b} \text{ and } \mathbf{c}.$

$\mathbf{H} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*; \text{ the indices } (h,k,l) \text{ are integers relative to the reciprocal lattice axes } \mathbf{a}^*, \mathbf{b}^* \text{ and } \mathbf{c}^*.$
Typically the summation in (1.1) may involve several thousands $F_H$. Once the Fourier coefficients $F_H = |F_H| \exp(i\Phi_H)$ are known, $\rho(x)$ can be calculated and the atomic positions determined. However, an X-ray diffraction experiment leads to partial knowledge of the $F_H$ only. When a monochromatic parallel X-ray beam (wavelength $\lambda \approx 1\text{Å}$) is incident on a crystal, the electrons will scatter the radiation. Due to interference, the total scattered amplitude from the unit cell, commonly referred to as the structure factor (s.f.) will be significantly present only for the scattered contributions from the $N$ atoms in the unit cell.

\[ F_H = |F_H| \exp(i\Phi_H) = \sum_{j=1}^{N} f_j(H) \exp[2\pi i (H \cdot \mathbf{r}_j)] \]

\[ = \sum_{j=1}^{N} f_j(H) \exp[2\pi i (h x_j + k y_j + l z_j)] \]  \hspace{1cm} (1.2)

with $f_j(H)$ the atomic scattering factor and $\mathbf{r}_j = (x_j, y_j, z_j)$ the position of atom $j$ in the unit cell.

The magnitude of the scattered wave $F_H$ can be obtained from the diffraction experiment. For ideal imperfect crystals the measured intensity of the scattered radiation $I_H$ can be shown to be proportional to $|F_H|^2$,

\[ I_H \propto |F_H|^2 \]  \hspace{1cm} (1.3)

However, the phase $\Phi_H$ needed to compute the electron density (ref. eqns (1.1) & (1.2)) cannot be obtained experimentally and as a result this expression cannot be evaluated
ab initio. The problem of finding the unknown phases $\Phi_H$ is usually referred to as the "phase problem" in crystallography.

Solving the phase problem:

The phase problem may be attacked in several ways, depending on the type of structure under investigation, e.g. via the Patterson function (Patterson, 1935). This function displaying all interatomic distances $u$ can be calculated from the measured diffraction data

$$P(u) = \frac{2}{V} \sum_{H} |F_H|^2 \cos(2\pi H \cdot u) \tag{1.4}$$

From $P(u)$ the positions of a small number of relatively heavy atoms, if present, can be derived and this results in a partial phasing model. When no relatively heavy atoms are present but a model of the structure is available, then the so called Patterson search techniques can be applied. However, when no model is available the Patterson method is not useful at all since the interatomic distance peaks are multi-interpretable. For these equal-atom structures another method can be used, viz, the so called Direct Methods of phase determination.

Basis of Direct Methods:

The Direct Methods try to derive the unknown phases directly from the structure magnitudes by means of algebraic and probabilistic expressions. Its base is formed by the existence of structure invariants (s.i.'s) and structure semi-invariants (s.s.i.'s) which are quantities invariant to origin shifts and special origin shifts respectively. The phase of the $F_H$ itself depends on the choice of the origin of the unit cell and is not a s.i.
Examples of s.i.'s are the triplet invariant

\[ \Psi_3 = \Phi_{H1} + \Phi_{H2} + \Phi_{-H1-H2} \]  

the quartet invariant

\[ \Psi_4 = \Phi_{H1} + \Phi_{H2} + \Phi_{H3} + \Phi_{-H1-H2-H3} \]

and the quintet invariant

\[ \Psi_5 = \Phi_{H1} + \Phi_{H2} + \Phi_{H3} + \Phi_{H4} + \Phi_{-H1-H2-H3-H4} \]

The normalized structure factors \( E_H \) are defined as

\[ |E_H| = \frac{|F_H|}{\left( \sum_{j=1}^{N} f_j^2(H) \right)^{1/2}} \]  

The \( E_H \), which in the equal-atom case can be interpreted as the scattered amplitude from a structure of point atoms, does not depend on the scattering angle \( \Theta \) and serves as an absolute measure of the scattering power of the crystal in the direction \( H \).

In the case of centrosymmetric structures, the probability expression for the sign product \( S_H S_K S_{-H-K} \) to be positive is

\[ P(S_H S_K S_{-H-K} = +) = \frac{1}{2} + \frac{1}{2} \tanh(E_3) \]  

with

\[ E_3 = \sqrt{3} \sigma_2^{-3/2} |E_HE_K E_{-H-K}| \]  

and

\[ \sigma_n = \sum_{j=1}^{N} f_j^2(H) \]

and by cochran (1955) for non-centrosymmetric structures,

\[ P(\Psi_3) = \exp(2E_3 \cos(\Psi_3))/(2\pi I_0(2E_3)) \]
$I_0$ is a modified Bessel function.

For equal atom structures, (1.10) becomes

$$E_3 = \frac{1}{(N)^{1/2}}\left(\left|E_H E_K E_{-H-K}\right|\right)$$

(1.13)

Larger is the triplet product of the E's, sharper is the probabilistic distribution and better is the estimate. As $N$ becomes large (for larger structures) $E_3$ becomes small and the variance of the distribution becomes large and the estimate becomes poor.

These exponential based expressions can also be obtained with the j.p.d. (joint probability distribution) technique by keeping the characteristic function in exponential form while retaining only a limited number of terms (e.g. Giacovazzo, 1980) or by transforming the series-expansion into an exponential expression (Karle, 1972; Karle & Gilardi, 1973).

Direct Methods in practice:

Direct method procedures start with normalizing the $|F_H|'$s into and setting up the phase relations among the phases of reflections with the largest $|E_H|'$s. The current procedures continue with the selection of a number of starting set phases usually by means of the convergence procedure of Germain, Main & Woolfson (1970) in which successively the weakest linked reflections are identified. This leads to a starting set consisting of a small number of phases, those fixing the origin (and enantiomorph) and a number of unknowns to which as a start two phase alternatives can be used (either numerical values (multisolution approach) or symbolic ones (symbolic addition approach)). Most popular is the multisolution technique, incorporated in many direct methods program systems such as MULTAN.
in which numerical starting values are selected for e.g. with the magic-integer principle (White & Woolfson, 1970) or at random as in the RANTAN procedure (Yao Jia-Xing, 1981). However, a single set of numerical starting phases is not sufficient since the true phase values are not known apriori, so a number of equivalently likely sets of numerical starting values must be employed. The next step in the multisolution technique is extending the starting phases with the tangent expression (Karle & Hauptman, 1956; Karle & Karle, 1966),

\[
\tan(\Phi_H) = \frac{\sum_{j=1}^{N} E_{3j} W_{Kj} W_{H-Kj} \sin(\Phi_{Kj} + \Phi_{H-Kj})}{\sum_{j=1}^{N} E_{3j} W_{Kj} W_{H-Kj} \cos(\Phi_{Kj} + \Phi_{H-Kj})}
\]

This expression may be obtained in various ways, e.g. by taking the derivative of the j.p.d. of the phase $\Phi_H$ taking part in $n$ triplets, given $|E_{H}|$ and all other magnitudes and phases involved.

In practice the weights $W_{Kj}$ and $W_{H-Kj}$ are inserted in expression (1.14) as statistical weights for the known phases $\Phi_{Kj}$ and $\Phi_{H-Kj}$ respectively. Because of the $\pi$ ambiguity of the tangent, the phase will be calculated from $\sin(\Phi_H)$ and $\cos(\Phi_H)$ instead of the tangent itself. Throughout the phase determination, consistency criteria are calculated in order to reject phase sets which are unlikely to be correct. Finally, a number of numerical solutions remain ordered in likelihood according to the criteria employed and for the most likely solutions' E-map can be calculated.
The alternative to the multisolution technique is the symbolic phase determination or symbolic addition method. In this technique symbolic phase values, A, B, C etc, are assigned to the unknown starting set of phases, representing only allowable numerical phase value (Gillis, 1948; Zachariasen, 1952; Karle & Karle, 1966). This technique, which has the apparent advantage that only one phase set expressed in symbols need to be extended, has been incorporated in the direct method structure determination package SIMPEL (Overbeek & Schenk, 1978; Schenk & Kiers, 1985). Employing the triplet relations, only phases for which a unique phase indication is obtained are accepted provided the (sum of) E3's of the relations involved meet a strict reliability threshold value. After extending this single symbolic phase set consistancy criteria are calculated for sets of numerical trial values. Finally, for the most likely trial set a further numerical phase extension is performed with (1.14) to determine the still unphased reflections.

When part of a structure is known, the unknown part of the structure can be solved by using direct methods. For phase extension and for the refinement of input phases and amplitudes, the difference structure factors, phased by the partial structure are used as input to a weighted tangent-refinement process. The method is referred to as DIRDIF (Beurskens, Bosman, Doesburg, Gould, van den Hark, Prick, Noordik, Beurskens & Parthasarathi, 1983; Beurskens et al., 1990).

It is most useful if the known part is only marginally sufficient to solve the structure. It can also be used if the known atoms lie in special or pseudo-special position (origin
ambiguity), or if, for non-centro symmetric structures, the known atoms form a centrosymmetric arrangement (enantiomorph ambiguity). The observed structure amplitudes and positional parameters of the known atoms are used by the computer program DIRDIF, to produce a greatly improved electron density map.

Once a model of a structure has been found by one of the previous methods, it is necessary to improve the preliminary coordinates by some process of refinement. The common procedure is to compare the calculated structure factors, $|F_c|$, for a proposed model with the actually observed ones. Then a necessary condition for the proposed model to be correct is that the calculated values duplicate those observed. The best agreement is obtained for

$$D = \sum_{i=1}^{m} w_i (|F_o| - |F_c|)^2_i$$

being minimal; $w_i$ is a weighting factor for an observation $|F_o|$. It has become common practice to utilize a residual of the following form for this purpose

$$R = \frac{\sum_i |F_o| - |F_c|}{\sum_i |F_o|}$$

The weighted R value (wR) used by the SHELX76 (Sheldrick, 1976) program is
The weighted R value \( (wR^2) \) used by the SHELXL93 (Sheldrick, 1993) program is

\[
wr^2 = \frac{\sum w(\frac{|Fo|^2 - |Fc|^2)}{\sum w(|Fo|)^2})}{\sum w(|Fo|)^2} \quad (1.18)
\]

After completing the refinement of the structure, it can be interpreted and here again a number of computing programs may be very useful for calculating bond lengths, bond angles, torsion angles, Newman projections, stereoviews and conformation energies.

Molecular Conformations:

The internal parameters of a molecule are parameters that characterize the molecular conformation and hence they are known as conformational parameters. They are: bond length, bond angle and torsion angle. Of these three, the torsion angle is the most important one, since, a variation of the torsion angles leads to different conformations of a molecule. It can be defined as follows:

The torsional (or dihedral) angle \( \tau \) of four atoms A, B, C, D with a chemical bond between AB, BC and CD, (Fig.1.1) is defined as the angle between the two planes through A, B, C and B, C, D. As it can vary from \(-180^\circ\) to \(+180^\circ\), its sign is defined according to the convention of Klyne and Prelog(1960).
The initial conformation corresponding to \( \tau = 0 \) is that in which all the atoms are coplanar and the end atoms A and D are cis with respect to BC. In any orientation of CD, the torsion angle \( \tau \) is given by the rotation in degrees it has undergone from the \( (\tau = 0) \) original position. The torsion angle is considered positive when it is measured clockwise from the front substituent A to the rear substituent D and negative when it is measured anti-clockwise. \( \tau(D-C-B-A) \) has the same sign and magnitude as \( \tau(A-B-C-D) \) (see also Luger, 1980).

Conformations in ring structures:

The conformational analysis of interest in crystallography varies from molecule to molecule. Actually it mainly depends on the type of the molecule. If the molecule contain rings, as it happens in the structures studied by the author, the planarity or otherwise of the ring is an important conformational feature.

The nonplanar rings are described in terms of well-known objects in common use like the chair, the boat, the envelope, the sofa etc.

Five-membered rings:

The five-membered rings generally assume (i) the envelope conformation in which four atoms are on a plane while the fifth is displaced from the plane forming the flap or (ii) the half-
chair conformation in which two atoms lie on either side of the plane of the remaining three atoms (Fig 1.2(a)).

Six-membered rings:

The six-membered rings assume varieties of conformations some of which are shown in Fig 1.2(b) (Duax, Weeks & Rohrer, 1976). If one is interested in knowing the difference between these forms (without constructing a model) the simplest way is to look at the torsion angles within the ring. In the case of chair conformation, the torsion angles will be alternating between +60° and -60° (In actual examples the modulus of the torsion angle can be in the range of 50 to 70°). For a boat conformation, two of the non-consecutive torsion angles will be zero and these will be separated by two torsion angles around +60° and -60°. In the "half chair" form, two consecutive torsion angles will be around zero. In other words, 5 out of 6 atoms will be in one plane. These factors can be effectively used to describe and compare the conformations of the molecules occurring in different compounds, in a quantitative way.

In the case of ring molecules where the ring is known to be planar, the deviations of the atoms from the least-square plane is an important feature of molecular conformation and provides information about the effects of substitution on the ring planarity. Two possible representations are employed: (i) the equation to the least-square plane and the deviation of the atoms are given; (ii) the torsion angles about the various bonds of the ring are listed out. It can easily be seen that, for a completely planar molecules, all the torsion angles will be zero. Hence, the deviation of the observed torsion angles from
Fig. 1–2(a) The three most symmetric conformations observed in general for five-membered rings have the symmetries indicated on the right.
Fig. 1.2(b) The most commonly observed conformations of six-membered rings. The mirror and twofold rotational symmetries are indicated on the right.
Fig. 1-2 (c) Various conformations of cycloheptane
this ideal value gives an indication of the amount of non planarity present in the molecule. The algebraic sum of the torsion angles will, however, be zero.

Seven-membered rings:

The possible conformations of the seven-membered rings may be grouped into two families which cannot be interconverted without altering the bond angles. One family includes chair, twist-chair and all the intermediate conformations while the other consists of boat, twist-boat and the intermediate forms (Fig 1.2(c)) (Anet, 1988; Bocian & Strauss 1977; Hendrickson, 1967 & Burkert & Allinger, 1982).

Hydrogen Bonds and Conformation:

It is a well-known fact that hydrogen bonding is one of the important factors that contributes to the stability of a structure. Thus it forms a part of the molecular conformation in that the conformation and the subsequent packing of the molecules should yield the formation of as many hydrogen bonds as possible (Ramakrishnan, 1980; Desiraju, 1989).

![Geometrical H-bond parameters](image)

Fig 1.3. Geometrical H-bond parameters.

Conformationally the hydrogen bond can be represented as shown in Fig 1.3. The most natural geometrical criterion for H-bonding between the groups D-H and A would seem to be that the
the van der Waals approach (Olovsson & Jonsson, 1976; Ramanadhan & Chidambaram, 1978), namely

\[ d_{H\cdots A} < r_H + r_A \]

Where \( D-H \) is the donor group and \( A \) is the acceptor group. \( r_D \) and \( r_A \) are the van der Waals radii for the donor and acceptor atoms. \( r_H=1.20\text{Å}, \ r_C=1.75\text{Å}, \ r_N=1.55\text{Å}, \ r_Cl=1.80\text{Å} \) and \( r_O=1.50\text{Å} \) are the van der Waals radii used by the author (Taylor & Kennard, 1982), to calculate the H-bonding geometries, in this thesis. In crystal structure studies, the angle \( D-H\cdots A \) is conventionally known as the hydrogen bond angle. It has also been possible to assess in a qualitative manner the strength of the hydrogen bond as it exists in a crystal structure (for example, Nethaji & Vasantha Pattabhi, 1989). A shorter \( D\cdots A \) contact is considered to be a stronger H-bond. However, to assess the effect of H-bonding on the geometry of the molecules, \( H\cdots A \) distances (Olovsson & Jonsson, 1976) are also considered. The stronger hydrogen bonds are those in which all the three atoms (donor, proton and acceptor) are collinear. Hydrogen bonds are relatively weak interactions as the energies involved are of the order of 1 to 7 Kcal/mole and are within the range of thermal fluctuations at physiological temperatures.

**Measurement of Density:**

There are several physical properties of crystal which could be determined early in a crystallographic investigation and which could prove useful in interpreting and using the subsequent X-ray results. One of these is the density of the crystal. The
method most commonly used is the floatation method. Two liquids, one lighter and one heavier than the crystals are taken, and when mixed together, the proportions of the two liquids are adjusted until the crystal remains suspended in the medium. The density of the liquid, determined by weighing a sample of known volume, is that of a crystal (Stout & Jenson, 1968). The density thus measured was combined with the measured dimensions of the unit cell, to obtain the accurate value for the total molecular weight of its contents. It was also used to deduce the number of molecules per unit cell of the compound. In this thesis densities have been measured by the author, by using the floatation method for some of the compounds only. For rest of the compounds it was not possible to measure because of the lack of sufficient suitable samples.

Computer Programs* used by the author:

Data collection was done using SDP (Frenz, 1978) and MolEN (Fair, 1990) computer systems. Cell refinement and data reduction were done using SDP (Frenz, 1978) and MolEN (Fair, 1990) computer systems. The crystal structures of the compounds presented in the thesis were solved by the programs SHELXS86 (Sheldrick, 1985), MULTAN80 (Main et al, 1980) and DIRDIF (Beurskens et al, 1990). Least-squares refinements of these compounds were carried out by means of the programs SHELX76 (Sheldrick, 1976) and SHLXL93 (Sheldrick, 1993). The programs

*All calculations were performed on Micro VAX II and Pentium PC systems available in the Bharathidasan University.
PARST (Nardelli, 1983) and SHELXL93 (Sheldrick, 1993) were used for geometry calculations and molecular diagrams were prepared using PLUTO (Motherwell & Clegg, 1978) and DTMM (Crabbe & Appleyard, 1989). The thermal ellipsoid plots were drawn by employing the programs ORTEP (Johnson, 1976) using the LVP16 plotter.