

Chapter - 6

THEORY OF STRUCTURE ANALYSIS

6.1 INTRODUCTION

The discovery of x-ray and Von Laue's experiment of x-ray diffraction by a single crystal brought forth a momentous change in the concept of crystalline solids. The idea of external regular geometrical shape to be regarded as the basic feature and characteristic of crystallinity of solids was then replaced by the regularity in the arrangement of atoms and molecules in space. This new idea of crystallinity established that x-ray is capable of diffraction and bringing the inside structures of the crystalline materials. As liquid crystals show anisotropic properties like solid, to elucidate the formation and stability of mesophases, a systematic study had been undertaken by many scientists to find relations between mesophase characteristics and molecular interactions revealed in crystal structure by x-ray diffraction analysis^[121,140-146].

6.2 X-RAY DIFFRACTION BY CRYSTALS

A crystal with its periodic structure was found to act as a three dimensional diffraction grating for x-rays. If the periodicity of the structure is defined by three non-coplanar unit translations a , b , c then for radiation of wavelength λ with unit vector S_0 , diffracted maxima would occur along directions with unit vectors S , only when the following three Laue^[147] equations are satisfied simultaneously:

$$\text{a. } (S - S_0) = h\lambda; \quad \text{b. } (S - S_0) = k\lambda; \quad \text{c. } (S - S_0) = l\lambda;$$

h, k, l being integer.

Lawrence Bragg^[148,149] extended Laue's idea by showing that each diffracted beam may be considered as 'reflection' by an array of parallel lattice planes with inter planar spacing d_{HKL} , where HKL are Millar indices of the possible crystal face parallel to the array, provided that the x-ray beam is incident at a particular glancing angle θ_{HKL} to be 'reflected', naturally, at the same angle. Thus the direction of the observed maxima are given, by what we know as Bragg equation, $2d_{\text{HKL}}\sin\theta_{\text{HKL}} = \lambda$. Here HKL are same integers as in Laue's equation.

X-rays are scattered by the electrons of an atom, the nucleus has no part in it. Electrons in the atom, scattering coherently, give rise to a resultant amplitude f depending on the direction of observation. This quantity f is known as the atomic

scattering factor or form factor. Considering now all the N atoms in the unit cell, the resultant diffracted amplitude for the reflection hkl , known as the structure factors $F(hkl)$, is given by the expression

$$F(hkl) = \sum_{j=1}^N f_j \exp 2\pi i (hx_j + ky_j + lz_j) \quad 6.1$$

where f_j is the form factor of the j^{th} atom, with fractional co-ordinate x_j , y_j , and z_j .

Equation 6.1 may be written as

$$F_{\mathbf{H}} = \sum_{j=1}^N f_j \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_j) \quad 6.2$$

The vector \mathbf{r}_j represents the position of the j^{th} atom in the unit cell and \mathbf{H} the position in reciprocal space of the point (hkl) . Evidently, $F_{\mathbf{H}}$ is a complex number which may be written in terms of its modulus and phase:

$$F_{\mathbf{H}} = |F_{\mathbf{H}}| \exp(i\phi_{\mathbf{H}})$$

Now as far as x-ray diffraction is concerned, the crystal is just a periodic distribution of electrons in three dimensions. The electron density $\rho(\mathbf{r})$ at location \mathbf{r} in the crystal may then be represented by a Fourier series in three dimensions:

$$\left. \begin{aligned} \rho(\mathbf{r}) &= \frac{1}{V} \int_{\mathbf{S}} F_{\mathbf{H}} \exp(-2\pi i \mathbf{H} \cdot \mathbf{r}) d\mathbf{s} \\ \text{or} \\ \rho(\mathbf{r}) &= \frac{1}{V} \sum_{\mathbf{h}} \sum_{\mathbf{k}} \sum_{\mathbf{l}} F_{\mathbf{H}} \exp(-2\pi i \mathbf{H} \cdot \mathbf{r}) \end{aligned} \right\} \quad 6.3$$

So, if one sums up the Fourier series for a fine enough grid of points spanning the entire unit cell, the structure is obtained. But measurement of intensities of the reflections provides us with $|F_{\mathbf{H}}|^2$ and hence $|F_{\mathbf{H}}|$, the phases $\phi_{\mathbf{H}}$ are lost.

This is the well known **Phase Problem** in the crystallography. To overcome this problem, we generally take help of four methods viz., ① *Patterson function*, ② *Direct methods*, ③ *Isomorphous replacement technique* and ④ *Anomalous scattering method*. Patterson map is very useful when heavy atoms are present in the unit cell. In organic crystals where there are no heavy atoms, Patterson method is quite inadequate, and usually direct methods are used.

6.3 DIRECT METHODS

Direct methods try to evaluate phases ϕ_H directly from the measured intensities I_{obs} through purely mathematical techniques. Roughly one can say that, since the crystal structure can be described by a limited numbers of parameters (the positions of the atoms) and since many more intensities can be measured, relationship among the structure factor F_H and thus among the phases ϕ_H , will exist. Therefore, the goal of the Direct methods is to identify as many of these relationship as possible. Thus if phases of two reflection were known for Σ_2 phase relations, described latter in this chapter, that of a third one related to these two can also be determined. But how can direct methods calculate phases from observed structure factor magnitudes when the amplitude and phase of a wave are physically independent quantities? $|F_H|$ and ϕ_H are linked through a knowledge of the electron density $\rho(\mathbf{r})$. Since $\rho(\mathbf{r})$ is related to structure factors by a Fourier transformation, constraints on the electron density impose corresponding constraints on the structure factors. Because the structure amplitudes are known, most constraints restrict the values of structure factor phase or, more precisely, the phases of structure invariants and seminvariants. In favourable cases, these constraints are sufficient to determine phase value directly.

6.4 STRUCTURE INVARIANTS AND SEMINVARIANTS

A structure invariant is a quantity, the value of which remains unchanged for any shift of the origin of the unit cell. A simple example is the intensity of a reflection, or F^2 because it can be measured and therefore is independent of any shift of the origin. The structure factor F is not a structure invariant since for any shift in the origin by say $\Delta\mathbf{r}$ the phase of 'F' changes by $2\pi \mathbf{H} \cdot \Delta\mathbf{r}$ radians. However, if the sum of the indices of the structure factors equals zero then it can easily be shown that the product of these structure factors is a structure invariant. Thus $F_{-H}F_KF_{H-K}$ is an example of a structure invariant and hence sum of their phases viz.,

$$\phi_{-H} + \phi_K + \phi_{H-K} = \phi(\mathbf{H}, \mathbf{K}) = \phi_3 \quad (\text{say}) \quad 6.4$$

is also an invariant. The value of structure invariant, however, is not always known, even though its values can only be a function of other structure invariants, e.g., intensities.

The structure seminvariants^[150] are those linear combinations of the phase whose values are uniquely determined by the crystal structure alone, when the choice of origin is restricted within permissible values. It originates from space group symmetry. For each space group they have to be derived separately. In any space group any structure invariant is also a structure seminvariant, but reverse is always not true. A complete theory concerning this subject is given in a series of papers by Hauptman and Karle^[151-153] and by Schenk^[154].

6.5 SOFTWARE PACKAGES BASED ON DIRECT METHODS

Although a substantial work was done by Ott^[155], Banerjee^[156], Avrami^[157] and Goedkoop^[158], the magnum offers of crystal structure determination was the work of H. Hauptmann and J. Karle^[159] of achieving it by direct methods. Different computer programs viz., MULTAN, SIMPLE^[160], SHELX^[161], MDM^[162], MAGIC^[163], RANTAN^[164], MAGEX^[165], YZRAC^[166], MITHRIL^[167] etc. are available based on direct methods — of which MULTAN (MULTiple TANGent) developed by Germain, Main and Woolfson^[168] was used by me.

A systematic account of the development of the direct methods is beyond the scope of this dissertation. Only I shall discuss the basic principle and some working formulae.

6.6 STEPS TO SOLVE THE STRUCTURE

Several steps have to be taken to solve the crystal structure after intensity data collection. Steps are as follows:

- I) Data reduction due to Lorentz factor and Polarisation factor
- II) Absorption correction, absolute scaling and temperature factor
- III) Estimation of $|E|$'s from F_{obs} values
- IV) Set up phase relationships
- V) Starting phase determination

VI) Phase extension and refinement

VII) Calculation of figure of merit

VIII) E-map interpretation

IX) Refinement of structures via Fourier synthesis, Difference Fourier synthesis and Least-square refinement.

Once the intensity data are collected in an automatic single crystal diffractometer (like ENRAF – Nonius, Phillips, Seifert, Rigaku) using monochromated x-ray beam, next task is to correct them for some geometric factors such as Lorentz factor, Polarisation factor and for some physical factors such as, absorption and temperature factor.

Step – I

Lorentz factor

The relation between intensity I_{hkl} and $|F_{hkl}|$ as used in diffractometric study is

$$|F_{hkl}| = \left[\frac{KI_{hkl}}{L_{hkl} P_{hkl}} \right]^{\frac{1}{2}}$$

where K is a constant for the experiment, but L_{hkl} , the Lorentz factor, and P_{hkl} , the polarisation factor, differ from reflection to reflection and depend on the geometry of the experimental set-up.

Lorentz factor is composed of two parts:

$$L_{hkl} = \frac{1}{\sin 2\theta_{hkl}} = \left(\frac{1}{2} \frac{1}{\sin \theta_{hkl}} \right) \left(\frac{1}{\cos \theta_{hkl}} \right)$$

the first factor is a measure of the relative intensity reflected by a unit volume at glancing angle θ and the second factor is an inverse measure of the relative rate of change, with glancing angle, of the path length. L_{hkl} can also be interpreted as the relative time opportunity for the various planes of the crystal to reflect.

Polarisation factor

This factor arises due to the nature of x-ray beam and the dependence of scattering amplitude on the orientation of electric vector E of the beam. For a

single electron scattering amplitude is proportional to $\sin\phi$ where ϕ is the angle between direction of the reflected beam and \mathbf{E} . For parallel and perpendicular component of \mathbf{E} , ϕ will be $(90^\circ - 2\theta_{hkl})$ and 90° respectively. Thus the ratio of the amplitude is $\cos 2\theta_{hkl}:1$. The mean intensity for these two states of polarisation is defined as the polarisation factor.

$$P_{hkl} = \frac{1}{2}(1 + \cos^2 2\theta_{hkl}).$$

In our case the incident beam is partially polarised during monochromatisation by reflection from the basal plane of a graphite crystal and P_{hkl} takes the form,

$$P_{hkl} = P_{\text{erf}} \frac{\cos^2 2\theta_m + \cos^2 2\theta_{hkl}}{1 + \cos^2 2\theta_m} + (1 - P_{\text{erf}}) \frac{\cos^2 2\theta_m + \cos^2 2\theta_{hkl}}{1 + \cos 2\theta_m}$$

where $P_{\text{erf}} =$ a constant depending on the crystal used in the monochromator (0.5 in our experiment).

$\theta_m =$ Bragg angle of reflection from the monochromator crystal.

Step – II

Absorption

To obtain the absorption correction for a reflection it is necessary to calculate the absorption of intensity for the actual path length travelled within the crystal by the beam reflecting from each infinitesimal portion of the crystal and then to integrate these results over the entire volume of the crystal¹. This problem cannot be solved explicitly for a crystal of general shape. In general, the practical approach has been to try to minimise the effects of absorption as much possible by crystal shaping and/or the use of more penetrating radiation, e.g., MoK_α .

Temperature factor and absolute scale

In increasing temperature, electron clouds about the nucleus spread over a large volume and thus to cause the scattering power of the real atom to fall off

more rapidly than that of the ideal, stationary model. The proper scattering factor for a real atom is, thus, not simply f_0 , rather given by the expression

$$f = f_0 e^{-B(\sin^2 \theta)/\lambda^2} \quad 6.5$$

where B is called Debye Waller Temperature factor and is related to the mean-square amplitude $\overline{u^2}$ of atomic vibration by

$$B = 8\pi^2 \overline{u^2}$$

It is convenient to have an estimate of the average value of B for the whole structure before beginning the actual analysis. Wilson plot method^[170] provides a realistic working value of B and at the same time calculates the scale factor k to place all the observed intensities on an approximately absolute basis. Expression used for the purpose is

$$\ln \left[\frac{\langle |F_{\text{obs}}|^2 \rangle}{\sum_{i=1}^N f_{oi}^2} \right] = \ln k - 2B(\sin^2 \theta)/\lambda^2 \quad 6.6$$

Thus if the left side of equation 6.6 is evaluated for each of the small shells of constant f to cover the entire reciprocal space and the values are plotted against $(\sin^2 \theta)/\lambda^2$, the result should be a straight line in which the slope is $-2B$ which gives the thermal parameter and the extrapolated intercept at $(\sin^2 \theta)/\lambda^2 = 0$ is $\ln k$ giving the scale factor k needed to convert $|F_{\text{obs}}|$ to $|F_{\text{abs}}|$ by

$$|F_{\text{obs}}|^2 = k |F_{\text{abs}}|^2 \quad 6.7$$

Step III

Estimation of $|E|$'s from $|F_{\text{obs}}|$ values

Since in the direct methods phases of the structure factors are estimated directly from the structure amplitudes it becomes necessary that the structure amplitudes be judged on their intrinsic merit where allowance is made for the decrease of the atomic scattering factor with increasing scattering angle. Ordinarily, the amplitudes of the different structure factors, F_H , cannot be compared directly, since the scattering factor decreases with increasing reflection angle θ . The observed $|F_H|$ is therefore modified so that they correspond to the hypothetical diffracted wave which would be obtained if atoms were stationary

point atoms. Thus the modified structure factor, called 'Normalised structure factor', E_H is defined by,

$$|E_H|^2 = \frac{|F_H|^2}{\varepsilon \sum_{j=1}^N f_j^2} = \frac{I_h}{\langle I \rangle} \quad \text{and} \quad \langle |E_H|^2 \rangle = 1$$

where ε is an integer characteristic of the space group symmetry elements of the crystal. $I_h = |F_H|^2$ and $\langle I \rangle$ is the expected intensity at that value of $\sin\theta/\lambda$. Among several ways the best method to obtain expected intensity is to use the K-curve, that is plot of equation 6.6.

Extrapolating the K-curve at low angle is very important. It is always a good idea to check the scaling of E's by examining $\langle |E|^2 \rangle$ as a function of $\sin\theta/\lambda$. Values of $\langle |E|^2 \rangle$ well above unity may give rise to difficulties in phase determination.

Step IV

Set up phase relationship, phase extension and refinement

For this, phases of only the strongest reflections are required. In practice a suitable number of reflections ($4 \times$ no. of independent atoms + 100) are chosen. If the crystal is triclinic or centrosymmetric, more reflections may have to be used.

The most commonly used phase relation is a three phase structure invariant based on positivity of electron density criterion, as proposed by Karle and Karle^[153].

$$\phi_H \approx \phi_K + \phi_{H-K} \quad 6.8$$

which for centrosymmetric structure is expressed by signs as

$$S(H) \approx S(K) S(H-K) \quad 6.9$$

These relationships are probability relations and the probability is high when the reflections have large $|E|$ values in addition to satisfying the criterion $H + K + L = 0$. These are called Σ_2 phase relations. Relation 6.8 is used to generate phases ϕ_H when the values of the phases on the right-hand side are known and it is used in a cyclic manner to propagate the phases to all the selected reflections.

The question now arises of finding the overall estimate if there are several

pairs of known phases, the estimate from each of which might well be different.

The answer to this important problem was given by Karle and Hauptman^[171] (1956) when they introduced the tangent formula

$$\tan \phi_{\mathbf{H}} \approx \frac{\sum_{\mathbf{K}} k(\mathbf{H}, \mathbf{K}) \sin(\phi_{\mathbf{K}} + \phi_{\mathbf{H}-\mathbf{K}})}{\sum_{\mathbf{K}} k(\mathbf{H}, \mathbf{K}) \cos(\phi_{\mathbf{K}} + \phi_{\mathbf{H}-\mathbf{K}})} = \frac{B(\mathbf{H})}{A(\mathbf{H})} \quad 6.10$$

where $k(\mathbf{H}, \mathbf{K}) = 2\sigma_3\sigma_2^{-\frac{3}{2}} |E_{\mathbf{H}}| |E_{\mathbf{K}}| |E_{\mathbf{H}-\mathbf{K}}|$

$$\sigma_n = \sum_{j=1}^N Z_j^n$$

Z_j being the atomic number of the j^{th} atom in a unit cell containing a total of N atoms. For identical atoms $\sigma_3\sigma_2^{-\frac{3}{2}} = N^{-\frac{1}{2}}$.

In order to use the tangent formula to obtain a new phase, the value of some phases have to be known and put into the right-hand side of the tangent formula. The set of the known phases is called a starting set from which the tangent formula derives more and more new phases and refines them to self-consistency.

When $\phi_{\mathbf{K}}$ and $\phi_{\mathbf{H}-\mathbf{K}}$ are known then the distribution proposed by Cochran and Woolfson^[172] for centrosymmetric structure and by Cochran^[173] for non-centrosymmetric structure are as follows:

Centrosymmetric case:

$$P_+(\mathbf{H}, \mathbf{K}) = \frac{1}{2} + \frac{1}{2} \tanh \left[\frac{1}{2} k(\mathbf{H}, \mathbf{K}) \right] \quad 6.11$$

Non-centrosymmetric case:

$$P[\phi(\mathbf{H}, \mathbf{K})] = \frac{\exp \{k(\mathbf{H}, \mathbf{K}) \cos[\phi(\mathbf{H}, \mathbf{K})]\}}{2\pi I_0 \{k(\mathbf{H}, \mathbf{K})\}} \quad 6.12$$

where I_0 is a zero-order modified Bessel function of the first kind.

But in this way all phases cannot be determined with acceptable reliability. It is therefore useful at this stage to eliminate about 10% of these reflections whose phases are most poorly defined by the tangent formula 6.10. An estimate of the reliability of each phase is obtained from $\alpha(\mathbf{H})$:

$$\alpha(\mathbf{H}) = \{A(\mathbf{H})^2 + B(\mathbf{H})^2\}^{\frac{1}{2}} \quad 6.13$$

when 6.13 contains only one term, as it may in the initial stages of the phase determination, then $\alpha(\mathbf{H}) = k(\mathbf{H}, \mathbf{K})$.

The larger the value of $\alpha(\mathbf{H})$, the more reliable is the phase estimate. The relation between $\alpha(\mathbf{H})$ and the variance is given by Karle and Karle^[153] (1966) as

$$\sigma^2(\mathbf{H}) = \frac{\pi^2}{3} + 4 \sum_{t=1}^{\infty} \frac{(-1)^t}{t^2} \frac{I_t \{\alpha(\mathbf{H})\}}{I_0 \{\alpha(\mathbf{H})\}}$$

From 6.13 it can be seen that $\alpha(\mathbf{H})$ can only be calculated when the phases are known. However, an estimate of $\alpha(\mathbf{H})$ can be obtained from the known distribution of three phase structure invariants^[173]. The estimated $\alpha(\mathbf{H})$ at the initial stage is given approximately by

$$\alpha_{\text{est}}(\mathbf{H}) = \sum_{\mathbf{K}} k(\mathbf{H}, \mathbf{K}) \frac{I_1 \{k(\mathbf{H}, \mathbf{K})\}}{I_0 \{k(\mathbf{H}, \mathbf{K})\}} \quad 6.14$$

Step V

Starting phase determination

As the tangent phasing process is usually initiated with a few 'known' phases so to fix the origin and enantiomorph is the first step in phase extension. This is done imposing the condition in terms of structure factor seminvariant phases. The selection of starting phases which are involved in reliable structure invariant relationship is critical to the success of 'multisolution' methods. The generator reflections are sorted by a convergence-type process by Germain, Main and Woolfson^[168] which maximises the connection between starting phases. At the end of the convergence procedure there is obtained a number of reflections sufficient to fix the origin and enantiomorphs whose phases are known and a number of few other reflections to which different phase values are assigned to create different starting points for phase extension through Σ_2 relations. The strength of convergence is that it ensures, as far as possible, that the initial phases will develop through strong and reliable phase relationships. For each starting phase set phases of all the selected strong reflections are generated and refined as explained in earlier section. Thus we get multiple phase set.

Step VII

Calculation of figure of merit

When a number of sets of phases have been developed by MULTAN it is necessary to rank them according to some Figure-of-Merit (FOM), prior to computing a Fourier map (i.e., E-map). Combining all weights from various FOM viz., Absolute Figure-of-Merit (ABSFOM), Relative Figure-of-Merit (RFOM), R-factor Figure-of-Merit (RFAC), Psi (zero) Figure-of-Merit (PSIO) etc. Combined Figure-of-Merit (CFOM) are calculated for each set. The most likely correct sets of phases are those with the highest value of CFOMs.

Step VIII

E-map interpretation

E-map are calculated using the best set of phases as indicated by the FOM. The complete interpretation of the maps is done in three stages:

- a) peak search
- b) separation of peaks into potentially bonded clusters
- c) application of simple stereochemical criteria to identify possible molecular fragments.

The molecular fragments obtained in this way can be compared with the expected molecular structure. The computer can thus present the user with a list of peaks and their interpretation in terms of the expected molecular structure quite automatically. It is also common practice to output a picture of molecule as an easy check on the structure the computer has found.

Step IX

Refinement of structures

Generally for refinement of a model structure obtained from E-map we use following three methods, e.g., 1) Fourier synthesis, 2) Difference Fourier synthesis and 3) Least square refinement.

The Fourier synthesis gives the refined co-ordinates of the atoms and also tends to reveal the position of any atom which is not included in computing the

structure factors. The Difference Fourier map is very useful for correcting the position of an atom used in structure factor calculation. This is also very useful in locating H-atoms towards the final stages of refinement procedure. Difference Fourier synthesis is free from series termination effect which arises due to limitation in the amount of data being collected.

An analytical method of refinement of great power and generality is that based on the principle of Least Squares. In brief, Least-Square refinement consists in using the square of the difference between observed and calculated values as a measure of their disagreement and adjusting the parameters so that the total disagreement is a minimum.

An agreement between the calculated structures factors F_c and those observed, F_o , indicates the degree of refinement. The most common method of assessing the agreement is calculating the residual or reliability index of the form

$$R = \frac{\sum (|F_o| - |F_c|)}{\sum |F_o|} \quad 6.15$$

the summation being over all the reflections. Evidently, the lower the value of R , the better is the agreement. Another form of the residual of common use is

$$R_w = \left[\frac{\sum w (|F_o - F_c|^2)}{\sum w |F_o|^2} \right]^{\frac{1}{2}} \quad 6.16$$

where the frequently used weight is,

$$w = \frac{1}{\sigma^2(F_o)}$$

$\sigma(F_o)$ being the standard deviation of F_o .