MACROMOLECULAR CRYSTALLOGRAPHY*†

M. VIJAYAN

(*Molecular Biophysics Unit, Indian Institute of Science, Bangalore* 560012, *India*)

Received on July 28, 1975

Key words: Protein crystallography, isomorphous replacement, anomalous dispersion.

INTRODUCTION

The most fruitful application of x-ray crystallography in recent years has been in the area of the structure analysis of biological macromolecules like proteins. In the last one and half decades the structures of several proteins have been analysed and our current understanding of the structure of and structure-activity relationships in proteins are largely based on the results of protein crystallography. These results are indeed very illuminating but we shall not be concerned with them here. This article would be concerned exclusively with the technique of x-ray structure analysis as applied to biological macromolecules. Until recently, the only class of macromolecules analysed by single crystal x-ray technqiues consisted of globular proteins. However, recently the structure of a transfer RNA molecule has also been determined using protein crystallographic techniques. But we would often use, for the sake of convenience, the term "protein" to refer to macromolecules in general.

What is presented here is only a bird's-eye view of macromolecular crystallography with more emphasis on theoretical results than on experimental details. For more exhaustive treatments of the subject reference may be made to Holmes and Blow [13] and North and Phillips [20]. A non-mathematical introduction to the subject is given in the article by Eisenberg [9].

---

* Adapted from a talk given in the Workshop on Crystallography, held along with the National Conference on Crystallography, 26-29 December 1974, Bangalore.

† Contribution No. 65 from the Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India.

CRYSTALLIZATION

The obvious first step in the structure analysis of a protein is its crystallization. This is a difficult step for two reasons. First, proteins appear to be intrinsically more resistant to crystallization and, secondly, they are usually available only in small quantities. There is no unique method of crystallization applicable to all proteins. Crystallization would depend upon the concentration of protein in the solution, the pH and the ionic strength of the medium, the presence of other substances in the solution and several other factors. Each of these parameters has to be varied and the optimum conditions for crystallization has to be worked out by trial and error for each protein. There have been some instances when proteins have been crystallized without any particular effort on the part of the investigator, but in most cases it takes several months, or years, to standardise the conditions of crystallization. Some of the different techniques for the crystallization of macromolecules have been described by Zeppezauer, Eklund and Zeppezauer [27] and Hample *et al.* [10].

It should be mentioned that there is one important difference between ordinary crystals and protein crystals. In ordinary crystals, whether they are organic or inorganic, the atoms, ions or molecules are closely packed. In other words, the crystal structure is usually such as to achieve the close packing of the contents. However, 30 to 50% of a protein crystal is usually filled with the solvent consisting mostly of water. This is a fundamental difference. Protein crystals, like protein molecules themselves, are stable only when they are in equilibrium with the aqueous solvent. Therefore, when x-ray photographs are taken, crystals are mounted in capillary tubes along with the solvent which is usually referred to as the mother liquor.

PREPARATION OF HEAVY ATOM DERIVATIVES

The only method that has been successfully used in the x-ray analysis of biological macromolecules is the isomorphous replacement method which is often used in conjunction with anomalous dispersion method to derive phase angles. Therefore, having obtained suitable crystals, the next step in the analysis is the preparation of heavy atom derivatives. This involves the attachment of "heavy atoms" like mercury, lead and uranium or chemical groups containing such heavy atoms to the protein crystals in a coherent manner without changing the conformation of the molecules and their crystal packing. Thus, ideally the structures of a protein crystal and a derivative crystal should be identical, except for the presence of heavy atoms

or heavy atom containing groups in the latter. This is only rarely possible in ordinary crystals as the molecules in them are closely packed. Protein crystals, however, contain large solvent regions and isomorphous derivatives can be obtained by replacing the disordered solvent molecules by heavy atom containing groups without disturbing the original arrangement of protein molecules. Thus the preparation of isomorphous heavy atom derivatives and hence the x-ray analysis of proteins are possible only because of the presence of the large solvent regions in their crystals.

Blake [2] has given a detailed review on the preparation of protein heavy atom derivatives. Some further information can be obtained from Eisenberg [9] also.

Collection of Intensity Data

The x-ray analysis of a protein involves the collection of intensity data from the crystals of the native protein and the derivatives. As the unit cell dimensions of the crystals are large, the number of reflections to be collected is also large. For preliminary studies, precession photographs are invariably used for determing the unit cell dimensions, screening heavy atom derivatives and so on. In the early years of protein crystallography, intensity data were also collected on precession photographs. Subsequently, diffractometers largely replaced precession cameras as instruments for data collection. Collection of intensity data using a diffractometer is generally more accurate and more rapid than that by conventional precession photography. The second factor is of considerable importance as crystals are damaged on long exposure to x-rays. Each crystal can be used for data collection only for 40 to 100 hours.

One particular aspect merits special consideration when collecting data from a protein crystal. As is well known, a Bragg reflection occurs when a reciprocal lattice point cuts the surface of the Ewald's sphere [5]. In protein crystals, the direct cell dimensions are large and consequently the reciprocal lattice dimensions are small. In other words, the reciprocal space is densely populated with lattice points. Thus, in practice, several reciprocal lattice points would cut the surface of the Ewald's sphere at the same time and hence several reflections would occur simultaneously. Conventional diffractomers are attached with only one counter and hence only one reflection can be measured at a time. Diffractometers with multiple counters, have been developed to take advantage of the simultaneous occurrence of reflections [21]. However, the number of counters that can be

attached to a diffractometer is often too small to allow the simultaneous measurement of all the reflections occurring at any given angular setting of the crystal.

The possibility of measuring several reflections simultaneously has also encouraged workers in the area of crystallographic instrumentation to have a fresh look at photographic techniques. Photographic techniques are superior to diffractometer techniques in two ways. First, in the former, one has a permanent record of the diffraction data. Secondly, a large number of reflections can be simultaneously recorded on a photograph. If an oscillation photograph or a precession photograph without layer line screen is taken, all the reflections in the given angular range are recorded on a single photograph. The time taken to measure all these reflections one by one on a diffractometer would, in many cases, be greater than that required to record them on an oscillation photograph or a precession photograph wthout layer line screen. But such photographs would be too complicated to be indexed manually by conventional methods. However, sophisticated computer controlled macrodensitometers, which have been developed recently, can be employed for indexing the photographs and measuring the intensities [1].

Determination of Heavy Atom Positions

As we shall see later, the phase angles of the structure factors from protein crystals can be calculated only if the structural parameters of the heavy atoms in the derivative crystals are known. We have to determine not only the positions and the temperature factors of the heavy atoms, but also the occupancy factors because often at a given position the heavy atom may not be present in all the unit cells. For example, if the heavy atom is present at a given position only in half the unit cells, then the occupancy factor of the site is said to be 0·5. If the heavy atom is present at a particular position in three quarters of the unit cells, then the occupancy factor of that site is 0·75 and so on. The occupancy factors cannot be inferred from chemical analysis. They have to be determined by x-ray methods.

Isomorphous difference Patterson synthesis

The argand diagram corresponding to a structure factor from a heavy atom derivative is shown in Figure 1. In the diagram, the magnitudes of the structure factor of the heavy atom derivative $\vec{F}_{PH}$, the structure factor of the protein $\vec{F}_P$ and the heavy atom contribution $\vec{F}_H$ are denoted by F_{PH},

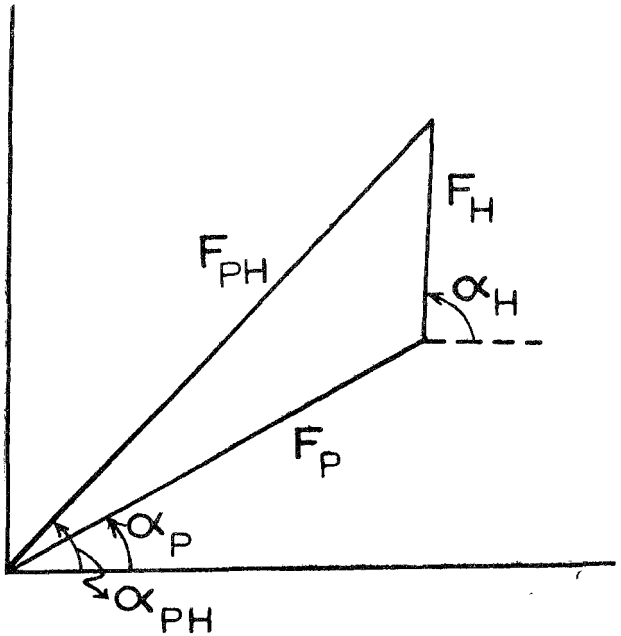

FIG. 1. Argand diagram corresponding to the structure factor of a heavy atom derivative

F_P and F_H respectively. The corresponding phase angles are α_{PH}, α_P and α_H. Of these, only F_{PH} and F_P can be obtained directly from experimental data. What we are interested initially is in the value of the vector $\vec{F}_H$. The Fourier transform of the complete set of $\vec{F}_H$ vectors, obviously corresponds to the distribution of heavy atoms. Therefore, in order to determine the heavy atom distribution, one should get an approximation to this vector or at least an approximation to its magnitude. This can de done in a simple manner if the data is centric.

Proteins are made up of L-amino acids and therefore cannot crystallize in centrosymmetric space groups. However, many proteins crystallize in

space groups with centrosymmetric projections. For centric data, the phase angles are 0 or 180° and hence we have

$$F_H = | F_{PH} \pm F_P |.$$

For most of the reflections the minus sign would be relevant. Thus, a Patterson synthesis with $(F_{PH} - F_P)^2$ as coefficients would give the vector distribution of heavy atoms in the derivative. Also, $| F_{PH} - F_P |$ can be used as the observed value of the heavy atom contribution in standard least sqaures procedures to refine the heavy atom parameters. On account of these advantages, centric data, when present, are extensively used in the initial stages of the analysis.

The situation is more complicated in the case of the general non-centric data. For non-centric data, it has been shown by Kartha and Parthasarathy [15] that

$$(F_{PH} - F_P)^2 = F_H^2 \cos^2 (\alpha_{PH} - \alpha_H) \qquad (1)$$

when F_H is small compared to F_{PH} and F_P. The Patterson synthesis with $(F_{PH} - F_P)^2$ as coefficients would, therefore, give an approximation to the heavy atom vector distribution. Patterson synthesis of this type has been extensively used in protein crystallography to determine heavy atom positions. The properties of this synthesis have been theoretically analysed by Ramachandran [23], Kartha and Parthasarathy [15] and Phillips [22]. Also, Dodson and Vijayan [8] have shown that this Patterson synthesis would provide a good approximation to the heavy atom vector distribution even when F_H is large.

Anomalous Difference Patterson Synthesis

Atomic scattering factors are normally calculated on the assumption that the binding energy of the electrons in an atom is negligible compared to the energy of the incident x-rays and that the distribution of electrons is spherically symmetric. When this assumption is valid, the atomic scattering factor is a real positive number and its value decreases as the scattering angle increases on account of the finite size of the atoms. When the binding energy of the electrons is appreciable, the atomic scattering factor at any given scattering angle can be written as

$$f = f_0 + f' + if'' \qquad (2)$$

where f_0 is a real positive number and corresponds to the atomic scattering factor for a spherically symmetrical collection of free electrons in the atom. The second and the third terms are referred to as the real and the imaginary components of the "dispersion correction" respectively. f' and f''

have appreciable values when an absorption edge of the atom concerned is in the neighbourhood of the wavelength of the incident radiation. Atoms with high atomic numbers have several absorption edges and their scattering factors always contain appreciable correction terms. For example, for Cuk_α radiation, the real and imaginary components of the dispersion correction for mercury are $-4{\cdot}80$ and $7{\cdot}04$ respectively; the corresponding values for uranium are $-4{\cdot}85$ and $13{\cdot}06$ [6]. If the dispersion correction term in the scattering factor of an atom have appreciable values for the given radiation, that atom is usually referred to as an anamalous scatteer. The effects on the structure factors or the intensities of Bragg reflections results ing from dispersion corrections are referred to as anomalous dispersion effects.

The argand diagram corresponding to a reflection (hkl) from a derivative and its Friedel partner ($\bar{h}\bar{k}\bar{l}$) is given in figure 2. Here it is assumed that the heavy atoms in the derivative are the only anomalous scatterers. It is also assumed that all the heavy atoms are of the same type. The phase angle of the imaginary component of the heavy atom contribution is then 90° ahead of that of the real component for all the reflections. The magnitude of the structure factor of the derivative for the reflection (hkl) is denoted by $F_{PH}(+)$ in the figure. The corresponding structure factor for reflection ($\bar{h}\bar{k}\bar{l}$) is denoted by $F_{PH}(-)$. In the absence of anomalous dispersion $F_{PH}(+) = F_{PH}(-)$ and the Friedel's law is obeyed. When anomalous dispersion is present, Friedel's law is violated and, as can be seen from the figure, $F_{PH}(+)$ is no longer equal to $F_{PH}(-)$. A composite view of the vector relationships for reflections (hkl) and ($\bar{h}\bar{k}\bar{l}$) can be obtained by reflecting the vectors corresponding to reflection ($\bar{h}\bar{k}\bar{l}$) about the real axis in the argand diagram (*see* Figure 3). The $\vec{F}_P$ and $\vec{F}_H$ vectors corresponding to the two reflections then superpose exactly. $\vec{F}_{PH}(+)$ and $\vec{F}_{PH}(-)$, however, have different magnitudes and phases.

The difference between the magnitudes of $F_{PH}(+)$ and $F_{PH}(-)$ is obviously related to the magnitude of the heavy atom contribution. Kartha and Parthasarathy [15] have shown that

$$\left(\frac{k}{2}\right)^2 (F_{PH}(+) - F_{PH}(-))^2 = F_H^2 \sin^2 (\alpha_{PH} - \alpha_H) \qquad (3)$$

where $k = (f_H + f_H')/f_H''$. A Patterson synthesis with the left hand side of (3) as coefficients would also yield the vector distribution corresponding to the heavy atom positions. However, $(F_{PH}(+) - F_{PH}(-))$ is a small difference between two large quantities and is liable to be in considerable

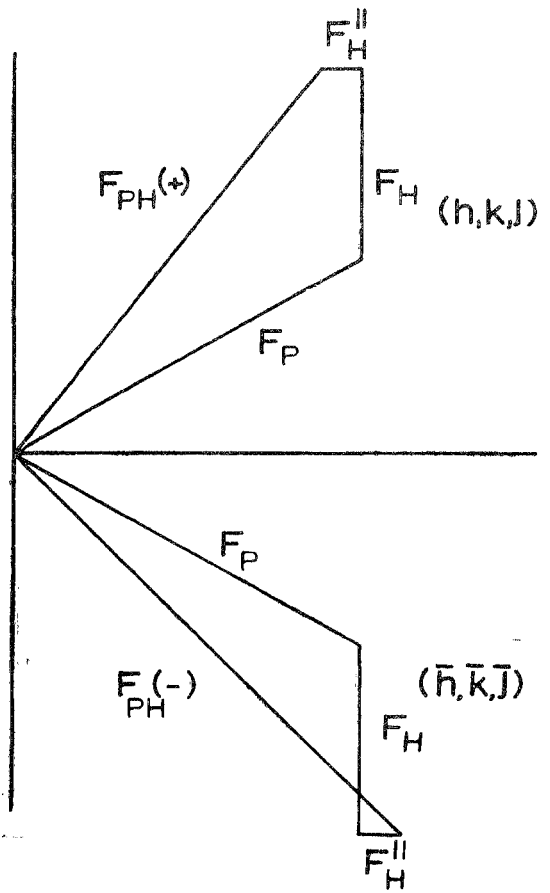

FIG. 2. Argand diagram for reflections (hkl) and $(\bar{h}\bar{k}\bar{l})$ from a heavy atom derivative in the presence of anomalous dispersion. F_H is the real part of the heavy atom contribution including that due to the real component of the dispersion correction. F_H'' is the imaginary component of the heavy atom contribution.

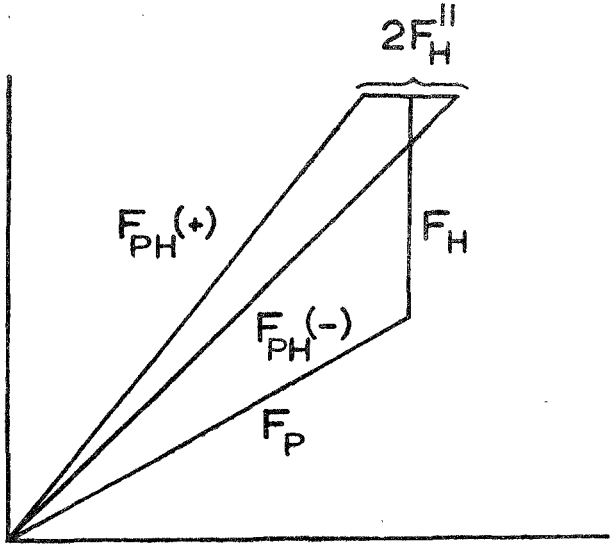

FIG. 3. Composite argand diagram for reflections (hkl) and $(\bar{h}\bar{k}\bar{l})$ from a heavy atom derivative in the presence of anomalous dispersion.

rror. Therefore, in practice, Patterson synthesis with the left hand side of (3) as coefficients are rarely used to determine heavy atom positions.

Combination of isomorphous and anomalous differences

So far we have considered isomorphous and anomalous differences eparately. But if we add equations (1) and (3), we get

$$(F_{PH} - F_P)^2 + \left(\frac{k}{2}\right)^2 (F_{PH}(+) - F_{PH}(-))^2 = F_H^2 \qquad [15] \qquad (4)$$

Here, F_{PH} can be approximated to $(F_{PH}(+) + F_{PH}(-))/2$. A different xpression for F_H^2 in terms of isomorphous and anomalous differences was lerived independently by Mathews [16]. According to a more accurate xpression given by Singh and Ramaseshan [26]

$$F_H^2 = F_{PH}^2 + F_P^2 - 2F_{PH}F_P \cos(\alpha_P - \alpha_{PH})$$
$$= F_{PH}^2 + F_P^2 \pm 2F_{PH}F_P \{1 - [k(F_{PH}(+) - F_{PH}(-))/2F_P]^2\}^{1/2} \quad (5)$$

In (5), the lower estimate is relevant when $|\alpha_P - \alpha_{PH}| < 90°$ and the upper estimate is relevant when $|\alpha_P - \alpha_{PH}| > 90°$. The lower and upper estimates of F_H are referred to as F_{HLE} and F_{HUE} respectively. In most of the cases, F_{HLE} would represent the correct value of F_H. Therefore, a Patterson synthesis with F_{HLE}^2 as coefficient would yield the vector distribution of heavy atoms in the derivative. A Patterson synthesis of this type would obviously be superior to that with $(F_{PH} - F_P)^2$ or $[F_{PH}(+) - F_{PH}(-)]^2$ as cofficients.

Difference Fourier Synthesis

Once a set of approximate phase angles has been calculated, by methods to be described later, using one or more heavy atom derivatives, the heavy atom positions in yet another derivative can be more easily determined by a difference Fourier synthesis with coefficients

$$(F_{PH} - F_P)\exp(i\alpha_P).$$

Difference syntheses of this type can also be used for improving the information obtained from different Patterson syntheses regarding the heavy atom distribution in the derivatives which have been used to calculate the phase angles. The difference Fourier technique is obviously very powerful when the data are centric. However, some limitations on its usefulness are introduced when the data are non-centric. The nature and extent of these limitations have been analysed, theoretically as well as through model calculations by Dodson and Vijayan [8]. They have shown that, when the data are non-centric, the difference Fourier syntheses should be interpreted with caution especially when the level of substitution in the derivative is high.

Refinement of Heavy Atom Parameters

After determining the approximate heavy atom positions, temperature factors and occupancies, the next step in the analysis is the refinement of these parameters. Least squares methods with different minimisation functions have been used by different workers. In the procedures most commonly used [7], the function

$$\sum w(F_{PH} - |\vec{F}_P + \vec{F}_H|)^2 \quad (6)$$

is minimised. Here F_{PH} is the observed magnitude of the structure factor for the particular derivative and $\vec{F}_P + \vec{F}_H$ is the calculated structure factor. The latter/obviously depends upon the phase angle α_P and the magnitude and the phase of $\vec{F}_H$ which are in turn dependent on the heavy atom parameters. Let us assume that we have three derivatives *A*, *B* and *C* and that we have already determined the heavy atom parameters *HAi*, *HBi* and *HCi*. Then,

$$\vec{F}_{HA} = \vec{F}_{HA}\ (HAi)$$

$$\vec{F}_{HB} = \vec{F}_{HB}\ (HBi)$$

$$\vec{F}_{HC} = \vec{F}_{HC}\ (HCi)$$

A set of approximate protein phase angles are first calculated making use of the unrefined heavy atom parameters. These phase angles are used to construct $\vec{F}_P + \vec{F}_H$ for each derivative. Then function (6) is minimised by varying *HAi* for derivative *A*, *HBi* for derivative *B* and *HCi* for derivative *C*. The refined values of *HAi*, *HBi* and *HCi* are then used to calculate a new set of protein phases. Alternate cycles of refinement of heavy atom parameters and phase angle calculations are carried out until convergence is reached. One significant feature of this refinement procedure is the dependence of the refined heavy atom parameters in one derivative on those in the other derivatives through the calculated protein phases. This dependence sometimes leads to different kinds of spurious effects [8].

Another refinement method employed by many workers [4, 14] makes use of the F_{HLE} values obtained by combining isomorphous and anomalous data. In this method, a minimisation function

$$\Sigma\, w\,(F_{HLE} - F_H)^2 \qquad (7)$$

is used. Care should be taken to omit from the calculations all reflections for which F_{HUE} is likely to be the correct estimate of F_H. F_H is a function of the heavy atom parameters *Hi* which are varied to minimise (7). The main advantage of this method is that the heavy atom parameters in each derivative can be refined independently of the information available on those in other derivatives. The refined parameters, however, are considerably affected by the statistical errors in the data. Usually, the anomalous differences are of the same order of magnitude as the errors present in the data. This leads to a systematic overestimation of F_{HLE} values and hence to the overestimation of the occupancy factors. This difficulty can be overcome

either by properly adjusting the weighting factors in the minimisation function or by using empirically estimated values of k in (5) [8]. The empirical values of k for different ranges of Bragg angles can be calculated using the expression

$$k_{empirical} = 2\,\Sigma \mid F_{PH} - F_P \mid / \Sigma \mid F_{PH}(+) - F_{PH}(-) \mid$$

given by Mathews [17].

Isomorphous Replacement and Anomalous Dispersion Methods

Before going on to describe the procedure for calculating protein phase angles using isomorphous replacement and anomalous dispersion methods, let us briefly outline these methods as applied to protein crystallography. For any given reflection, the structure factors of the derivative and the protein, and the heavy atom contribution are related as shown in Fig. 1. Here F_{PH} and F_P are experimentally observed quantities whereas F_H and α_H can be calculated once the heavy atom parameters are known. The phase angle of the protein structure factor is then given by

$$\alpha_P = \alpha_H \pm \phi \tag{8}$$

$$\text{where } \cos(\pi - \phi) = \frac{F_{PH}^2 - F_P^2 - F_H^2}{2F_P F_H}.$$

Thus we get two possible values for the protein phase angle symmetrically placed below and above the phase angle of the heavy atom contribution. One of these values would correspond to the correct protein phase angle. This ambiguity can be resolved if data from two independent derivatives are available. Two equations like (8) would then be available.

$$\alpha_P = \alpha_{H1} \pm \phi_1$$

and

$$\alpha_P = \alpha_{H2} \pm \phi_2$$

where subscripts 1 and 2 refer to derivative 1 and 2 respectively. Thus there are two possible sets of values. That value which is common to both the sets would obviously correspond to the correct protein phase angle. The situation can be graphically demonstrated with the aid of the so-called Harker construction [11] shown in Fig. 4. Here one draws a circle with F_P as radius and the origin of the argand diagram as the centre. Two more circles are drawn with F_{PH1} and F_{PH2} as the radii and with the ends of vectors $\vec{F}_{H1}$ and $\vec{F}_{H2}$ as the centres. Both these circles interesect the F_P circle at two points each. One of the points of intersection is common. The angle

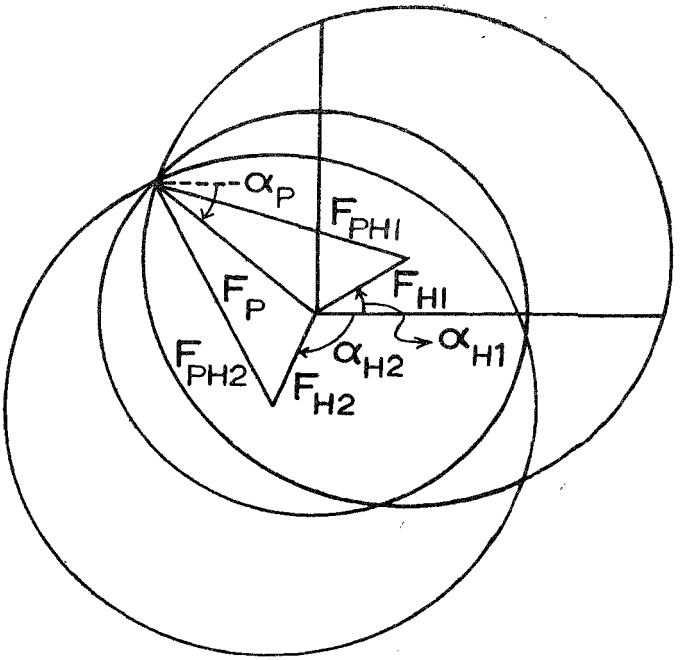

FIG. 4. Harker construction when two heavy atom derivatives are available.

between the horizontal line passing through the point of intersection of the three circles and the line/connecting the point of intersection with the origin is the protein phase angle. Therefore, protein phase angles can be determined if a minimum of two independent heavy atom derivatives are available.

In the presence of anomalous dispersion effects in the diffraction pattern from a derivative crystal, F_{PH} (+) and F_{PH} (−) can be formally considered as arising from two independent derivatives. A Harker diagram can then

be constructed as shown in Fig. 5. Therefore, when anamalous dispersion effects are present, in principle, data from a single derivative is sufficient tc determine the protein phase angles.

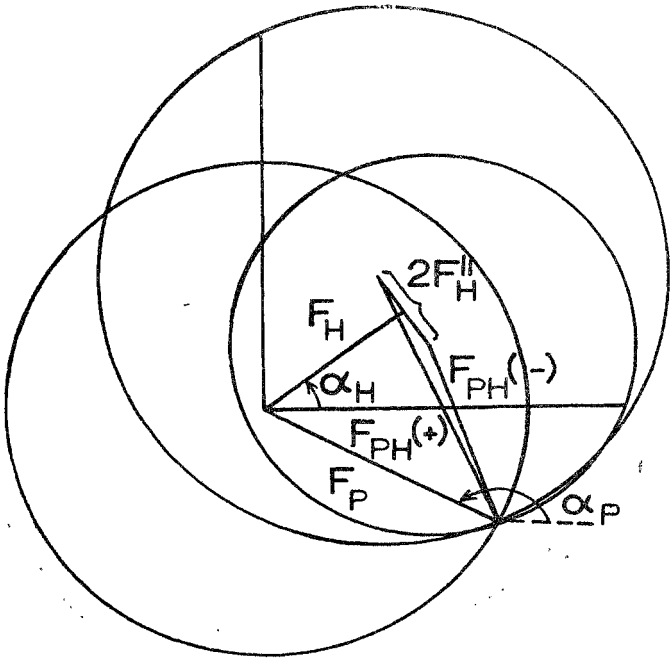

FIG. 5. Harker construction using anomalous difference.

CALCULATION OF PHASE ANGLES

As we saw earlier, if all the data sets are entirely error-free and the derivatives are strictly isomorphous, the protein phases can be determined using a minimum of two derivatives. In the presence of anomalous dispersion effects, in principle, only one derivative is required to calculate

phase angles provided the above conditions hold good. However, the data always contain considerable errors and the derivatives are not often strictly isomorphous. Therefore, in practice, data from several derivatives are used in the phase angle calculations. On account of the errors in the data and the departures from strict isomorphism, all the circles would not intersect at a single point in the Harker diagram; instead there would be a distribution of intersections. Therefore, what one obtains from calculations is not a unique phase angle, but a probability distribution for the phase angle.

Most Probable Phase, "Best Phase" and the Figure of Merit

The procedure currently used for determining protein phase angles based on probability considerations was developed by Blow and Crick [3]. Consider the argand diagram for a reflection from a particular derivative with an arbitrary value α for the protein phase angle (Fig. 6). Referring to Fig 6,

$$D_H(\alpha) = \{F_P^2 + F_H^2 + 2F_P F_H \cos(\alpha_H - \alpha)\}^{1/2}. \quad (9)$$

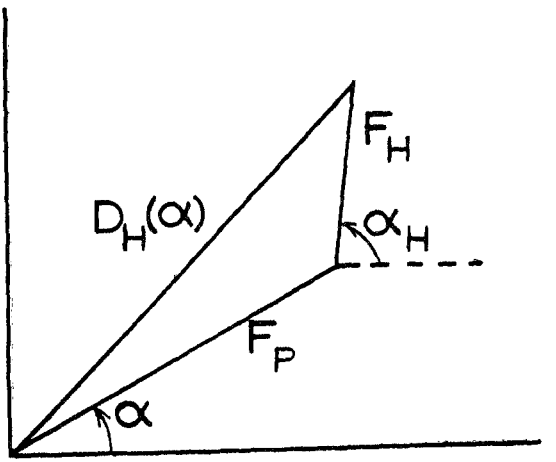

FIG. 6. Argand diagram corresponding to the structure factor of a heavy atom derivative with an arbitrary protein phase angle.

If α corresponds to the true protein phase angle α_P, then $D_H(\alpha)$ coincides with F_{PH}. The amount by which $D_H(\alpha)$ differs from F_{PH}, namely,

$$\xi_H(\alpha) = F_{PH} - D_H(\alpha), \tag{10}$$

is a measure of incorrectness of the phase angle α. This quantity is called the lack of closure. Then the probability for α being the correct protein phase angle is defined as

$$P(\alpha) = N \exp - \{\xi_H^2(\alpha)/2E^2\} \tag{11}$$

where N is the normalisation constant and E is an error estimate which includes the effects of non-isomorphism as well as the errors in the data. When centric data are available E can be estimated using the relation

$$E^2 = \sum_{hkl} \{\vec{F}_{PH} - |\vec{F}_P + F_H|\}^2/n \tag{12}$$

where n is the total number of reflections included in the summation. When the entire data are non-centric, the expression

$$E^2 = \sum_{hkl} (F_{HLE} - F_H)^2/n \tag{13}$$

can be used instead.

When a number of heavy atom derivatives are available, the total probability of a phase α being correct would be

$$P(\alpha) = N \exp - \sum_i \{\xi_{Hi}^2(\alpha)/2E_i^2\} \tag{14}$$

where the summation is over all the derivatives.

When $P(\alpha)$ for any particular reflection is plotted around a circle of unit radius, as shown in Fig. 7, the phase corresponding to the highest peak in the probability distribution would give the most probable protein phase α_M of the reflection. Then, a Fourier synthesis with

$$F_P \exp(i\alpha_M)$$

as coefficients would give the most probable electron-density distribution in the protein.

Blow and Crick have described a different way of using the probability distribution. In Fig. 7, the centroid of the probability distribution is at point P. The polar coordinates of P are m and α_B where m, a fractional positive number with a maximum value of unity, and α_B are referred to as

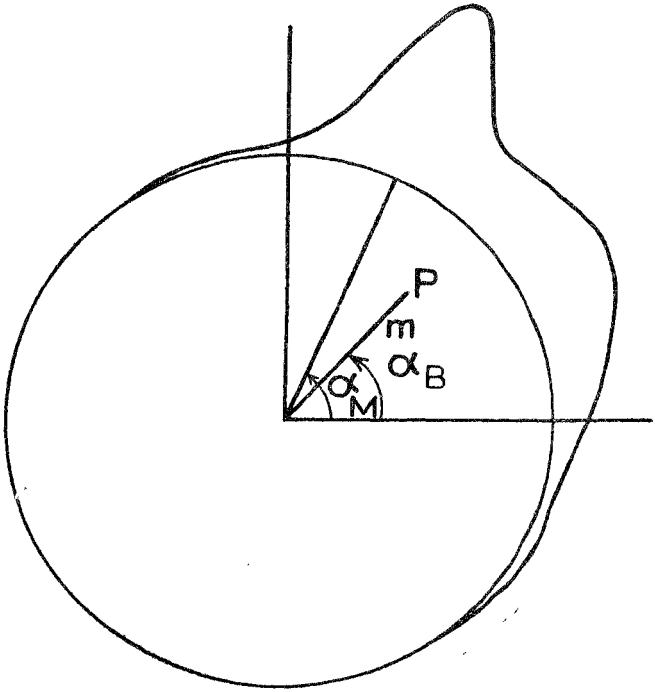

FIG. 7. Probability distribution of the protein phase angle.

the "figure of merit" and the "best phase" respectively. A Fourier synthesis with

$$mF_p \exp(i\alpha_B)$$

as coefficients is called the "best Fourier". Defined in this manner the best Fourier would give the electron density distribution with the lowest root mean-square error. The best Fourier synthesis rather than the most probable Fourier synthesis is usually employed in the structure analysis

of proteins. In practice, the figure of merit and the best phase are calculated using the expressions,

$$m \cos \alpha_B = \sum_i P(\alpha_i) \cos \alpha_i / \sum_i P(\alpha_i) \quad (15)$$

and

$$m \sin \alpha_B = \sum_i P(\alpha_i) \sin \alpha_i / \sum_i P(\alpha_i)$$

where $P(\alpha_i)$ are calculated, say, at 5° intervals [7]. Obviously the figure of merit gives an estimate of the precision of the calculated phase angle. It will have a high value when the best phase and the most probable phase are close to each other. The figure of merit can be statistically interpreted as the cosine of the expected error in the calculated phase angle.

Inclusion of Anomalous Scattering Data

As pointed out earlier, when the intensities of the Friedel equivalent reflections have also been measured from the crystals of a derivative, $F_{PH}(+)$ and $F_{PH}(-)$ can be formally treated as arising from two different derivatives. Treated in this manner, the effect of anomalous differences on phase determination would only be marginal as, for any given reflection, the difference between $F_{PH}(+)$ and $F_{PH}(-)$ is usually small. North [19] has, however pointed out that the error in $|F_{PH}(+) - F_{PH}(-)|$ would normally be much smaller than that in $|F_{PH} - F_P|$. First, the former is obviously free from the effects of non-isomorphism. Secondly, as $F_{PH}(+)$ and $F_{PH}(-)$ are measured from the same crystal, both these quantities can be expected to contain the same systematic errors. These errors are eliminated in the difference between the two quantities. Thus, the mean square error in anomalous differences is usually much smaller than that in isomorphous differences. Therefore, different estimates of the root mean square error E should be used for isomorphous and anomalous differences. Then, for any given derivative, the new expression for the probability distribution of the protein phase angle in the presence of anomalous dispersion data would be,

$$P(\alpha) = N \exp - (\xi_H^2(\alpha)/2E^2) \exp - \{(\triangle H_{obs} - \triangle H_{cal})^2/2E'^2\} \quad (16)$$

where $\triangle H_{OBS} = |F_{PH}(+) - F_{PH}(-)|$, $\triangle H_{CAL}$ is the corresponding value of anomalous difference calculated for the phase angle α and E' is the root mean square error in anomalous differences [19], [16]. For evaluating $\xi_H(\alpha)$ and E, F_{PH} can be taken as the average value of $F_{PH}(+)$ and $F_{PH}(-)$; similarly F_H can be approximated to be the magnitude of the real

part of the heavy atom contribution. Methods are available for estimating the value of E', but satisfactory results are obtained if the value of E' is taken as a third of corresponding E-value.

The phase determination procedures discussed so far are based on the Blow and Crick formalism. There are several objections to this formalism but it is the only one available and attempts to improve upon it has not so far met with success. In all the protein structure determinations todate, phase angles were calculated using this formalism.

Electron Density Map

After calculating the phase angles, the "best" Fourier synthesis is computed which gives the electron-density distribution in the protein crystals. Apart from other factors, the information that can be obtained from the electron density map depends also on the "resolution" of the Fourier map. Theoretically, the resolution of a Fourier map is approximately equal to $0{\cdot}7\ d_m$ where d_m is the minimum interplanar spacing of the Bragg reflections included in the Fourier series. However, the series are never complete and the coefficients normally contain different kinds of errors. Therefore, the nominal resolution of the map is taken to be d_m. As is well known, for a given wavelength of radiation, the interplanar spacing is inversely proportional to the sine of the Bragg angle. The minimum interplanar spacing therefore corresponds to the maximum Bragg angle upto which data have been collected. Thus, if the data have been collected to a maximum Bragg angle corresponding to a spacing of 6 Å, the resulting map is said to have 6 Å resolution; if the minimum spacing is 3 Å, we have a 3 Å resolution map and so on.

Unlike in the case of ordinary crystals, atomic resolution is rarely achieved in protein electron-density maps. Even though protein molecules have finite three-dimensional structures, they have a certain amount of flexibility even in crystals. Also, unlike in ordinary crystals, protein molecules are not closely packed in crystals. They exist in equilibrium with the solvent and, therefore, the molecules can wobble slightly. Consequently, protein crystals are always slightly disordered. This results in the data fading out at higher Bragg angles and the resolution would be consequently poor. For most protein crystals, the data would go only upto, say, 2·5 Å resolution. In some cases, they may go upto 2 Å resolution and very rarely upto 1·7 or 1·5 Å resolution.

Even when fairly high resolution (better than about 3 Å) data are present, the preliminary studies are usually carried out using low resolution data. The number of intensities to be measured is proportional to the cube of the sine of the maximum Bragg angle, and the number of reflections to be recorded and processed increases very rapidly with increasing resolution. Therefore, data at higher resolution are usually collected only after promising results have been obtained from low resolution studies. Valuable information like the overall shape and the sub-unit structure of the molecule can be obtained even from a low resolution electron-density map. The map usually becomes interpretable in terms of the structure and conformation of the molecule when the resolution is 3Å or better.

The complete three-dimensional structure of the protein can be deduced from a good high resolution electron-density map. The interpretation of the map usually consists of constructing a molecular model of the protein which could be superposed on the electron-density map. A molecular model of the protein, so constructed, would contain a wealth of information. In fact, from a biological point of view, the most interesting part of the work starts after the model is obtained. However, a discussion on the results of protein crystallography does not come under the purview of this article.

CONCLUSION

Macromolecular crystallography has come to stay as a well established method, in fact the only method, for determining the detailed three-dimensional structure of globular proteins and transfer RNA. This method is now being applied to the study of even larger molecular aggegates like viruses, and it could perhaps be expected that the ultrastructure of even larger and more complex biological particles would be unravelled by protein crystallographic techniques in the foreseeable future.

Macromolecular crystallography has now developed into a vast and well organised subject. In this article, we have not discussed many important aspects of this subject such as the numerous applications of rotation and translation functions [24], the application of direct methods to the determination of heavy atom positions [18] and to the refienment and the extension of protein phases [12]; Sayre, [25], and the use of the difference Fourier technique in studying the interactions of proteins with small molecules. What we have attempted to give here is only a brief outline of the subject.

ACKNOWLEDGEMENT

The author thanks Professor G. N. Ramachandran for his encouragement in preparing this article.

REFERENCES

[1] Arndt, U. W., Champness, J. N., Phyzackerley, R. P. and Wonacott, A. J. *J. Appl. Cryst*., 1973, **6**, 457.

[2] Blake, C. C. F. .. *Advances in Protein Chemistry*, Vol 23, Edited by C. B. Anfinsen *et al*., Chapter 2. New York, Academic Press, 1968.

[3] Blow, D. M. and Crick, F. H. C. *Acta Cryst*., 1959, **12**, 794.

[4] Blundell, T. L., Dodson, E., Dodson, G. and Vijayan, M. *Contemp. Phys*., 1971, **12**, 209.

[5] Buerger, M. J. .. *X-ray Crystallography*, New York, John Wiley, 1942.

[6] Cromer, D. T. .. *Acta Cryst*., 1965, **18**, 17.

[7] Dickerson, R. E., Kendrew, J. C. and Strandberg, B. E. *Acta Cryst*., 1961, **14**, 1188.

[8] Dodson, E. and Vijayan, M. .. *Acta Cryst*., 1971, **B27**, 2402.

[9] Eisenberg, D. .. *The Enzymes*, Vol. 1, Edited by P. D. Boyer, Chapter 1, New York, Academic Press, 1970.

[10] Hample, A., Labanonskas, M. Connors, P. G., Kirkergard, L., Raj Bhandary, U. L., Sigler, P. B. and Bock, R. M. *Science*, 1968, **162**, 1384.

[11] Harker, D. .. *Acta Cryst*., 1956, **9**, 1.

[12] Hendrickson, W. A. and Karle, J. *J. Biol. Chem*., 1973, **248**, 3327

[13] Holmes, K. C. and Blow, D. M. *The Use of X-ray Diffraction in the Study of Protein and Nucleic Acid Structure*, New York, Interscience, 1966.

[14] Kartha, G. .. *Acta Cryst*., 1965, **19**, 883.

[15] Kartha, G. and Parthasarathy, R. *Acta Cryst*., 1965, **18**, 745.

[16] Mathews, B. W. . *Acta Cryst.*, 1966, **20**, 230

[17] Mathews, B. W. .. *Acta Cryst.*, 1966, **20**, 82.

[18] Navia, M. A. and Sigler, P. B. *Acta Cryst.*, 1974, **A30**, 706.

[19] North, A. C. T. . *Acta Cryst.*, 1965, **18**, 212.

[20] North, A. C. T. and Phillips, D. C. *Progress in Biophysics*, Vol. 19, Part I, Edited by J. A. V Butler and D. Noble, Chapter 1. Oxford, Pergamon Press, 1969.

[21] Phillips, D. C. .. *J. Sci. Instrum.*, 1964, **41**, 123.

[22] Phillips, D. C. .. *Advances in Structure Research by Diffraction Methods*, Vol. 2, Edited by R. Brill and A. Mason, New York and London, Interscience, 1966, p. 75.

[23] Ramachandran, G. N. .. *Advanced Methods in Crystallography*, Edited by G. N Ramachandran, Chapter 2, New York, Academic Press, 1964.

[24] Rossmann, M. G. .. *Molecular Replacement Method*, Edited by M. G. Rossmann, New York, Gordon and Breach, 1972.

[25] Sayre, D. .. *Acta Cryst.*, 1974, **A30**, 180.

[26] Singh, A. K. and Ramaseshan, S. *Acta Cryst.*, 1966, **21**, 279.

[27] Zeppezauer, M., Eklund, H. and Zeppezauer, E. S. *Arch. Biochem. Biophys.*, 1968, **126**, 564.