

14. ANOMALOUS DISPERSION

14.1. Heavy-atom location and phase determination with single-wavelength diffraction data

BY B. W. MATTHEWS

14.1.1. Introduction

As is well known, the successful introduction of the method of isomorphous replacement by Green *et al.* (1954) was the turning point in the subsequent development of protein crystallography as we now know it.

The idea that the phases of X-ray reflections from a protein crystal could be obtained by the introduction of heavy atoms into the crystal was not new, having been suggested by J. D. Bernal in 1939 (Bernal, 1939). The isomorphous-replacement method was used as early as 1927 by Cork (1927) in studying the alums. Bokhoven *et al.* (1951) subsequently extended the method to the study of a noncentrosymmetric projection of strychnine sulfate, using what would now be termed the method of single isomorphous replacement. They also suggested that by using a double isomorphous replacement, a unique phase determination could be obtained, even for noncentrosymmetric reflections. The details of the double (or multiple) isomorphous-replacement method were worked out by Harker (1956), who introduced the very useful concept of phase circles. Another contribution which was of great practical value, and which will provide the basis for much of the subsequent discussion, is the method introduced by Blow & Crick (1959) for the treatment of errors in the isomorphous-replacement method. In addition to the determination of protein phases by the method of substitution with heavy atoms, it is now routine to supplement this information by utilizing the anomalous scattering of the substituted atoms. The underlying principles trace back to articles by Bijvoet (1954), Ramachandran & Raman (1956), and Okaya & Pepinsky (1960). The first application of the anomalous-scattering method to protein crystallography was by Blow (1958), who used the anomalous scattering of the iron atoms to determine phase information for a noncentrosymmetric projection of horse oxyhaemoglobin.

In the following discussion, we first review the classical method of phase determination by isomorphous replacement, then discuss the inclusion of single-wavelength anomalous-scattering data, and conclude by discussing the use of such data for heavy-atom location. Part of the review is based on Matthews (1970).

14.1.2. The isomorphous-replacement method

Consider a protein crystal with an isomorphous heavy-atom derivative, *i.e.* a modified crystal in which heavy atoms occupy specific sites throughout the crystal, but which is in all other respects identical to the unsubstituted 'parent' crystal. Let the structure factors of the protein crystal be $\mathbf{F}_P(\mathbf{h})$, of the isomorph be $\mathbf{F}_{PH}(\mathbf{h})$, and of the heavy atoms $\mathbf{F}_H(\mathbf{h})$. (Note: Structure *amplitudes* are indicated by italic type, *e.g.* F_P , and *vectors* by bold-face type, *e.g.* \mathbf{F}_P .) In practice, one can measure the structure amplitudes F_P and F_{PH} , and it is desired to obtain from these observable quantities the value of the phase angle of $\mathbf{F}_P(\mathbf{h})$ so that a Fourier synthesis showing the electron density of the protein structure may be calculated. It will be assumed, for the moment, that the positions and occupancy of the sites of heavy-atom binding have been determined as accurately as possible.

From the heavy-atom parameters, the corresponding structure factor $\mathbf{F}_H(\mathbf{h})$ is calculated. To determine φ , the phase of $\mathbf{F}_P(\mathbf{h})$, we

construct a set of phase circles, as proposed by Harker (1956). From a chosen origin O (Fig. 14.1.2.1a), the vector OA is drawn equal to $-\mathbf{F}_H$. Circles of radius F_P and F_{PH} are then drawn about O and A , respectively. The intersections of the phase circles at B and B' define two possible phase angles for F_P . Note that the angles are

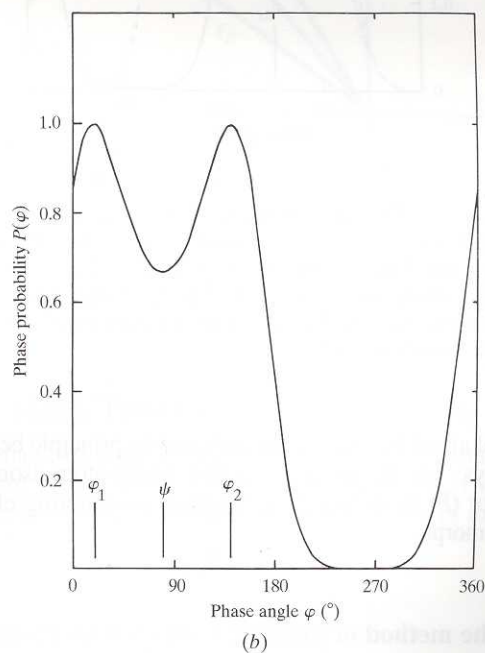
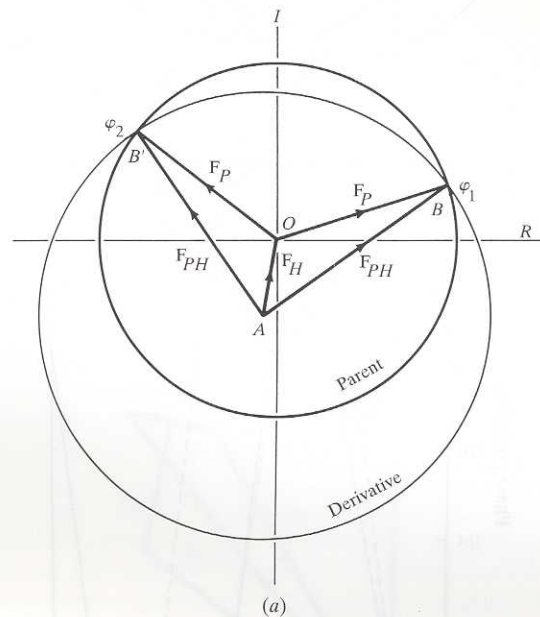


Fig. 14.1.2.1. (a) Harker construction for a single isomorphous replacement. φ_1 and φ_2 are the 'most probable' phases for \mathbf{F}_P . (b) Phase probability distribution for a single isomorphous replacement. This and subsequent probabilities are unnormalized. [All figures in this chapter are reproduced with permission from Matthews (1970). Copyright (1970) International Union of Crystallography.]

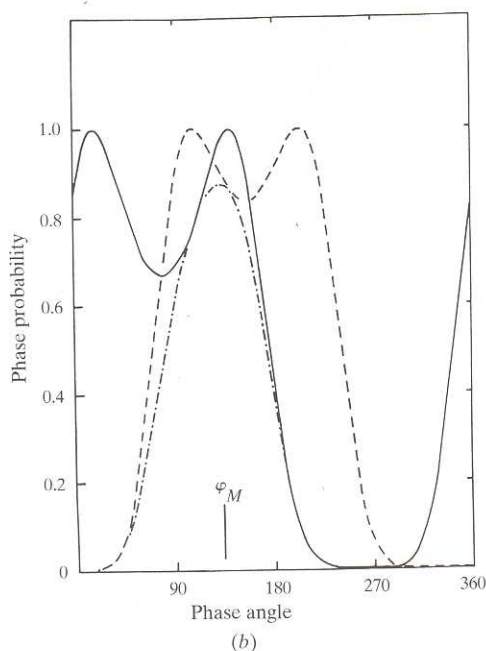
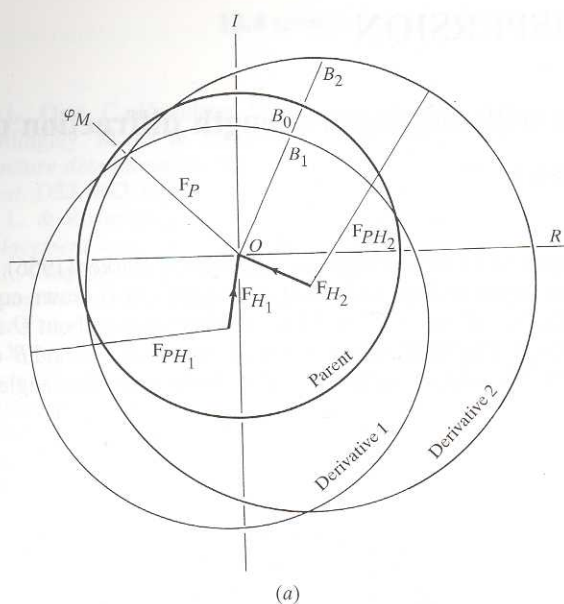


Fig. 14.1.3.1. (a) Harker construction for a double isomorphous replacement. φ_M is the 'most probable' phase for \mathbf{F}_P . (b) Phase probability distribution corresponding to the double isomorphous replacement shown in part (a). The curve for derivative 1 is solid, that for derivative 2 is dashed, and that for the combined distribution is drawn as a dotted-and-dashed line.

symmetrical about \mathbf{F}_H . This ambiguity may in principle be resolved in two ways: (a) by using a second heavy-atom isomorphous derivative or (b) by utilizing the anomalous-scattering effects for the first isomorph.

14.1.3. The method of multiple isomorphous replacement

The phase information provided by a second isomorph is illustrated in Fig. 14.1.3.1(a). In theory, the three phase circles will intersect at a point and the phase ambiguity will be resolved. In practice, there will be errors in the observed amplitudes F_P and F_{PH} and in the heavy-atom parameters (and thus in \mathbf{F}_H). Also, the isomorphism may be imperfect. As a result, the intersections of the three phase

circles may not coincide. Another complication arises from the fact that for reflections where \mathbf{F}_H is small, the circles will be essentially concentric and will not have well defined points of intersection. In other words, the phase determination will become indeterminate. The method of Blow & Crick (1959) was introduced as a way to take all these factors into account. It has had an extraordinary impact, not only as a practical method for protein phase determination, but also in influencing all subsequent thinking in this area.

14.1.4. The method of Blow & Crick

Blow & Crick pointed out that in practice the phase angle φ can never be determined with complete certainty. Rather, there is a finite probability that any arbitrary phase angle may be the correct one. Consider the vector diagram shown in Fig. 14.1.4.1, in which \mathbf{F}_H is known and we wish to determine the probability $P(\varphi)$ that the arbitrary phase angle φ is the correct phase of \mathbf{F}_P . Strictly, one should allow for the possibility of errors in \mathbf{F}_H , F_P and F_{PH} , and should consider the probability that the vector \mathbf{F}_P occupies all possible positions in the Argand diagram. However, Blow & Crick suggested that the analysis might be considerably simplified by assuming that F_P and \mathbf{F}_H are known accurately and that all the error lies in the observation of F_{PH} . In other words, it was assumed that the vector \mathbf{F}_P must lie on the circle of radius F_P , and the probability distribution of F_P could be evaluated as a function of φ only.

For an arbitrary phase angle φ , the phase triangle (Fig. 14.1.4.1) will not close exactly. If we define F_C to be the vector sum of \mathbf{F}_H and $F_P \exp(i\varphi)$, then the lack of closure of the phase triangle is given by

$$\varepsilon = F_C - F_{PH}. \quad (14.1.4.1)$$

Following Blow & Crick, if E is the r.m.s. error associated with the measurements, and the distribution of error is assumed to be Gaussian, then the probability $P(\varphi)$ of the phase φ being the true phase is

$$P(\varphi) = N \exp(-\varepsilon^2/2E^2), \quad (14.1.4.2)$$

where N is a normalizing factor such that the sum of all probabilities is unity. The un-normalized probability distribution corresponding to Fig. 14.1.4.1 (and Fig. 14.1.2.1a) is shown in Fig. 14.1.2.1(b). The two most probable phase angles ($\varphi = \varphi_1$ and $\varphi = \varphi_2$) are the alternative phases of F_P for which the phase triangle is closed.

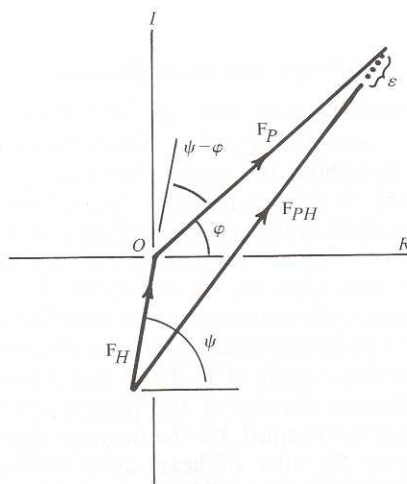


Fig. 14.1.4.1. Vector diagram illustrating the lack of closure, ε , of an isomorphous-replacement phase triangle.

Individual probability distributions for the additional heavy-atom derivatives are derived in an analogous manner and may be multiplied together to give an overall probability distribution. The joint probability distribution corresponding to Fig. 14.1.3.1(a) is shown in Fig. 14.1.3.1(b), and in this case the most probable phase is that which simultaneously fits best the observed data for the two isomorphous derivatives.

The main objection which may be made to the Blow & Crick treatment is that it assumes that there is no error in F_P . In practice, however, this is not a serious limitation.

14.1.5. The best Fourier

A protein crystallographer desires to obtain a Fourier synthesis that can most readily be interpreted in terms of an atomic model of the structure. One synthesis which could be calculated is the 'most probable Fourier', obtained by choosing the value of $F_P(\mathbf{h})$ for each reflection which corresponds to the highest value of $P(\varphi)$. Blow & Crick pointed out that although this Fourier is the most likely to be correct, it has certain disadvantages. In the first place, it might tend to give too much weight to uncertain or unreliable phases, and, in the second place, for cases where $P(\varphi)$ is bimodal, there is a strong chance of making a large error in the phase angle. Blow & Crick suggested that in cases such as this, a compromise is needed, and that the centroid of the phase probability distribution provides just the required compromise. They showed that the corresponding synthesis is the 'best Fourier', which is defined to be *that Fourier transform which is expected to have the minimum mean-square difference from the Fourier transform of the true F 's when averaged over the whole unit cell.*

The centroid of the phase probability distribution may be defined as a point on the phase diagram with polar coordinates (mF_P, φ_B) , where φ_B is the 'best' phase angle. The quantity m , which acts as a weighting factor for F_P , is called the 'figure of merit' of the phase determination. Its magnitude, between 0 and 1, is a measure of the reliability of the phase determination.

14.1.6. Anomalous scattering

All atoms, particularly those used in preparing heavy-atom isomorphs, give rise to anomalous scattering, especially if the energy of the scattered X-rays is close to an absorption edge. The atomic scattering factor of the atom in question can be expressed as

$$\mathbf{f} = f_0 + \Delta f' + if'' = f' + if'', \quad (14.1.6.1)$$

where f_0 is the normal scattering factor far from an absorption edge, and $\Delta f'$ and f'' are the correction terms which arise due to dispersion effects. The quantity $\Delta f'$, in phase with f_0 , is usually negative, and f'' , the imaginary part, is always $\pi/2$ ahead of the phase of the real part ($f_0 + \Delta f'$). It may be noted that by using different wavelengths, the term $\Delta f'$ is equivalent to a change in scattering power of the heavy atom and produces intensity differences similar to a normal isomorphous replacement, except that in this case the isomorphism is *exact* (Ramaseshan, 1964). This is the basis of the multiwavelength-anomalous-dispersion (MAD) method (Hendrickson, 1991) discussed in Chapter 14.2. Here we focus on measurements based on a single wavelength, traditionally referred to as the 'anomalous-scattering method'.

The anomalous scattering of a heavy atom is always considerably less than the normal scattering (for Cu $K\alpha$ radiation, $2f''/f'$ ranges from about 0.24 to 0.36), but there are several factors which tend to offset this reduction in magnitude (e.g. see Blow, 1958; North, 1965).

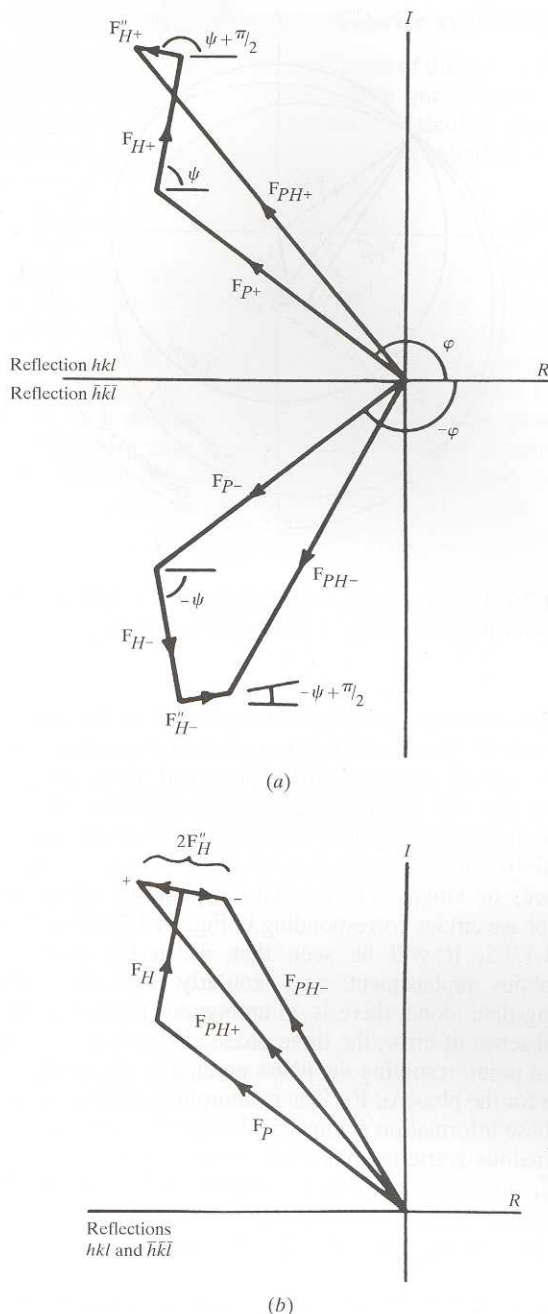


Fig. 14.1.7.1. (a) Vector diagrams illustrating anomalous scattering for the reflections hkl and $\bar{h}\bar{k}\bar{l}$. (b) Combined vector diagram for reflections hkl and $\bar{h}\bar{k}\bar{l}$.

14.1.7. Theory of anomalous scattering

Suppose that two isomorphous crystals are differentiated by N heavy atoms of position \mathbf{r}_n and scattering factor $(f'_n + if''_n)$. Then, for the reflection hkl , the calculated structure factor of the N atoms is

$$F_H(\mathbf{h}) + iF''_H(\mathbf{h}) = \sum_{n=1}^N f'_n(\mathbf{h}) \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_n) + i \sum_{n=1}^N f''_n(\mathbf{h}) \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_n). \quad (14.1.7.1)$$

If the heavy atoms are all of the same type, i.e. they all have the same ratio of $f''_n/f'_n (=k)$, then F_H and F''_H are orthogonal, and $F''_H = F_H/k$.

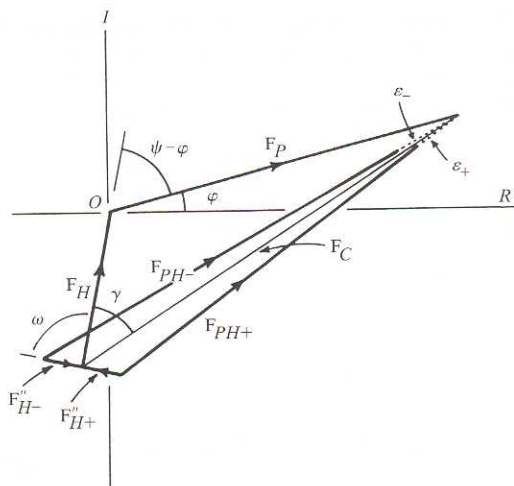
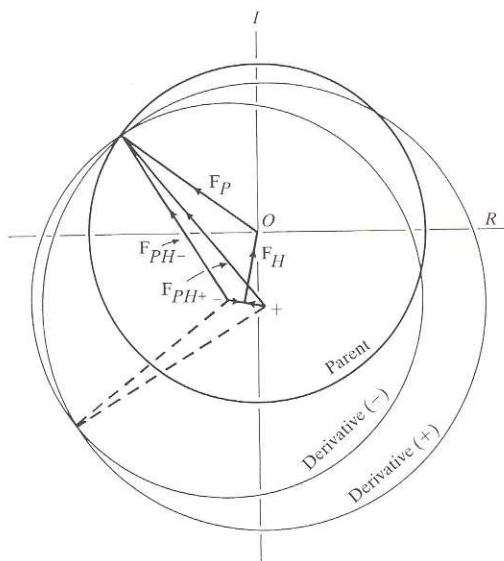


Fig. 14.1.8.1. Vector diagrams illustrating lack of closure in the anomalous-scattering method.

Fig. 14.1.7.2. Harker construction for a single isomorphous replacement with anomalous scattering, in the absence of errors.

The relation between the structure factors of the reflection hkl and its Friedel mate $\bar{h}\bar{k}\bar{l}$ is illustrated in Fig. 14.1.7.1(a). The situation can be conveniently represented (Fig. 14.1.7.1b) by reflecting the $\bar{h}\bar{k}\bar{l}$ diagram through the real axis onto the hkl diagram. In cases such as this, where Friedel's law breaks down, we shall refer to the difference $\Delta_{PH} = (F_{PH+} - F_{PH-})$ as the Bijvoet difference, or simply the anomalous-scattering difference. The Harker phase circles corresponding to Fig. 14.1.7.1(b) are shown in Fig. 14.1.7.2. It will be seen that, as in the case of single isomorphous replacement, and similarly with the anomalous-scattering data alone, there is an ambiguous phase determination. In the absence of error, the three phase circles (Fig. 14.1.7.2) will meet at a point, resolving the phase ambiguity and giving a unique solution for the phase of F_P . The isomorphous-replacement method gives phase information symmetrical about the vector F_H , whereas the anomalous-scattering phase information for F_{PH} is symmetrical about F_H'' , which, for heavy atoms of the same type, is at right angles to F_H . In other words, the two methods complement each other, one method providing exactly that information which is not given by the other.

On average, the experimentally measured isomorphous-replacement difference, $(F_{PH} - F_P)$, will be larger than the anomalous-scattering difference, $(F_{PH+} - F_{PH-})$. The former, however, relies on measurements from different crystals and is also susceptible to errors due to non-isomorphism between the parent and derivative crystals. The latter can be obtained from measurements on the same crystal, under closely similar experimental conditions, and is not affected by non-isomorphism. Therefore, it is desirable to use methods that take into account the different sources of error in the respective measurements (Blow & Rossman, 1961; North, 1965; Matthews, 1966b). One method is as follows.

14.1.8. The phase probability distribution for anomalous scattering

From Fig. 14.1.8.1, it can be seen that the most probable phase angle will be the one for which $\epsilon_+ = \epsilon_-$. At any other phase angle, there will be an 'anomalous-scattering lack of closure' which we define to be $(\epsilon_+ - \epsilon_-)$. The value of $(\epsilon_+ - \epsilon_-)$ can readily be calculated as a function of φ (Matthews, 1966b; Hendrickson, 1979). Thus, if the r.m.s. error in $(\epsilon_+ - \epsilon_-)$ is E' , and the distribution of error is

assumed to be Gaussian, then from measurements of anomalous scattering, the probability $P_{ano}(\varphi)$ of phase φ being the true phase of F_P can be estimated using an equation exactly analogous to equation (14.1.4.2).

An example of an anomalous-scattering phase probability distribution is shown by the dotted curve in Fig. 14.1.8.2. The asymmetry of the distribution arises from the fact that $P_{ano}(\varphi)$ is the phase probability distribution for F_P rather than that of F_{PH} , which would be symmetrical about the phase of F_H'' . The overall probability distribution obtained by combining the anomalous-scattering data with the previous isomorphous-replacement data (Fig. 14.1.2.1b) is given by

$$P(\varphi) = NP_{iso}(\varphi)P_{ano}(\varphi) \tag{14.1.8.1}$$

and is illustrated in Fig. 14.1.8.2.

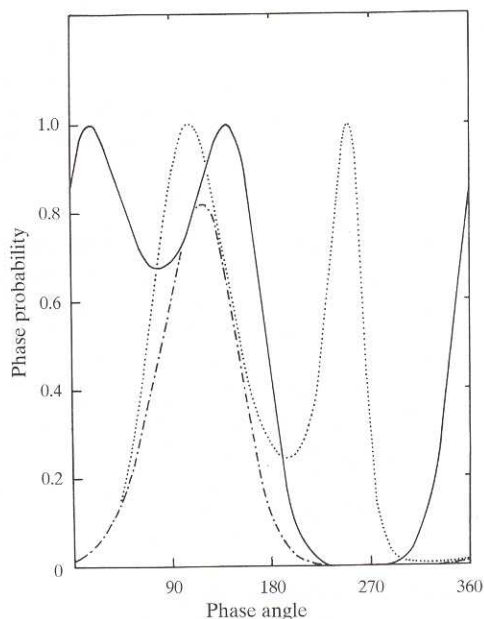


Fig. 14.1.8.2. Combination of isomorphous replacement and anomalous-scattering phase probabilities for a single isomorphous replacement. $P_{iso}(\varphi)$ is drawn as a solid line, $P_{ano}(\varphi)$ as a dotted line, and the combined probability distribution is drawn as a dotted-and-dashed line.

14.1.9. Anomalous scattering without isomorphous replacement

The treatment outlined above of phase determination by anomalous scattering assumed that data were available for a parent crystal devoid of anomalous scatterers and an anomalously scattering isomorphous heavy-atom derivative. It is not uncommon that the native protein itself contains atoms which scatter anomalously or has been engineered to contain such scatterers. In such cases, measurements will usually be made at multiple wavelengths in order to exploit MAD phasing (Hendrickson, 1991). If, however, measurements are available only at a single wavelength, they can be utilized to obtain some phase information (e.g. Matthews, 1970).

14.1.10. Location of heavy-atom sites

During the development of protein crystallography, it was understood that heavy-atom sites might be located from difference Patterson functions, but there was substantial debate as to the type of function that was preferable (Perutz, 1956).

Blow (1958), and also Rossmann (1960), advocated a Patterson function with amplitudes $(F_{PH} - F_P)^2$. It relies on the admittedly crude assumption that the desired scattering amplitude of the heavy atoms, $|F_H|$, can be approximated by

$$|F_H| \simeq |F_{PH} - F_P|. \quad (14.1.10.1)$$

The approximation does have one very helpful characteristic, namely, that it tends to be most accurate when $|F_{PH} - F_P|$ is large, i.e. when F_H is parallel or antiparallel to F_P (cf. Fig. 14.1.4.1). Thus, the numerically largest coefficients in the Patterson function tend to represent $|F_H|^2$ correctly. Given a well behaved isomorphous heavy-atom derivative, and accurately measured data, experience has shown that a map with coefficients $(F_{PH} - F_P)^2$ can give an excellent representation of the desired heavy-atom-heavy-atom vector peaks.

14.1.11. Use of anomalous-scattering data in heavy-atom location

A relation exactly analogous to equation (14.1.10.1) can be used to approximate the anomalous heavy-atom scattering amplitude, namely,

$$|F_H''| \simeq \frac{1}{2} |F_{PH+} - F_{PH-}| \quad (14.1.11.1)$$

(see Fig. 14.1.7.1b). As noted above, if all the heavy atoms are the same, $F_H = kF_H''$. Thus, a Patterson function with coefficients $(F_{PH+} - F_{PH-})^2$ should also show the desired heavy-atom-heavy-atom vector peaks (Blow, 1957; Rossmann, 1961).

For each individual reflection, however, and as is also the case for phase determination, the information that is provided by the isomorphous-replacement difference $(|F_{PH}| - |F_P|)$ is exactly complementary to that provided by the anomalous-scattering measurement $(|F_{PH+}| - |F_{PH-}|)$. By combining both sets of experimental measurements, it is possible to obtain a much better estimate of the heavy-atom scattering, $|F_H|$, for every reflection (Kantha & Parthasarathy, 1965a,b; Matthews, 1966a; Singh & Ramaseshan, 1966). One formulation (Matthews, 1966a) can be written as

$$F_H^2 = F_P^2 + F_{PH}^2 - 2F_P F_{PH} \{1 - [wk(F_{PH+} - F_{PH-})/2F_P^2]^{1/2}\}, \quad (14.1.11.2)$$

where $F_{PH} = (F_{PH+} + F_{PH-})/2$ and w is a weighting factor (from 0 to 1) that is an estimate of the relative reliability of the measurements of $(F_{PH+} - F_{PH-})$ compared with $(F_{PH} - F_P)$.

14.1.12. Use of difference Fourier syntheses

The discussion above has focused on the use of difference Patterson functions to locate heavy-atom sites. Once one or more putative sites have been located, they can be used to calculate approximate protein phases, which, in turn, can be used to calculate difference Fourier series with coefficients in the form

$$m(F_{PH} - F_H) \exp(-i\varphi_B), \quad (14.1.12.1)$$

where m is the figure of merit and φ_B is the 'best', albeit approximate, phase of the protein structure factor. Putting aside errors due to inaccuracies in φ_B , such maps do not give the true heavy-atom vector, F_H . Rather, they give, essentially, the projection of F_H along F_P (cf. Fig. 14.1.4.1). Nevertheless, subject to certain limitations, such difference maps are extraordinarily powerful in locating secondary sites in a given heavy-atom derivative, or in using approximate phases from one derivative to search for heavy-atom sites in other putative derivatives. It is in this context, however, that certain limitations of the single-isomorphous-replacement (SIR) method have to be kept in mind. These are noted in the next section.

14.1.13. Single isomorphous replacement

Although phase determination from a single heavy-atom derivative in the absence of anomalous-scattering data is, in principle, ambiguous, it was realized early on that useful phase information can still be obtained (Blow & Rossmann, 1961). As shown in Fig. 14.1.2.1(a), the two possible phases for the protein are φ_1 or φ_2 . In terms of the analysis of Blow & Crick (1959), the 'best' phase to use for the protein is the average of φ_1 and φ_2 . This is also equivalent to using both φ_1 and φ_2 . With this in mind, a situation that is of special concern is one in which the heavy-atom distribution used to determine the phases happens to have a centre of symmetry. One common way in which this can occur is when one has a heavy-atom derivative with a single site in space group $P2_1$. A related situation occurs when there are multiple sites in space group $P2_1$, but all have

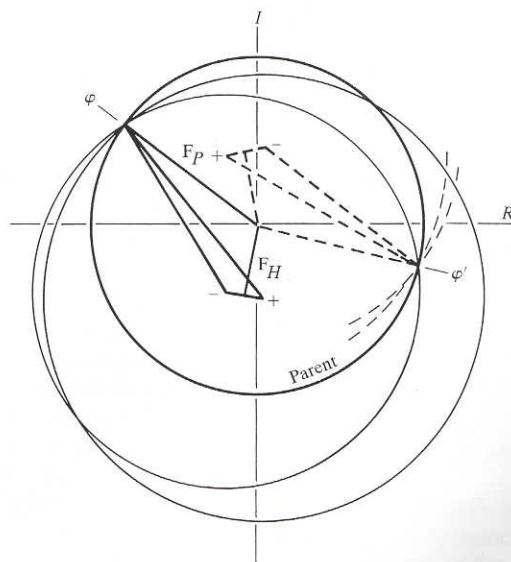


Fig. 14.1.13.1. The consequences of using the incorrect hand for the heavy-atom arrangement in phase determination. The correct heavy-atom arrangement gives the construction drawn in solid lines with φ uniquely determined as the correct phase for F_P . If the enantiomeric arrangement of heavy atoms is used, the dashed construction will result, leading to the incorrect phase φ' . Even though φ is correct and φ' is incorrect, both phase determinations have an identical figure of merit.

14. ANOMALOUS DISPERSION

the same y coordinate. If the origin of coordinates is considered to be at the site of centrosymmetry, then all of the heavy-atom vectors F_H (Fig. 14.1.2.1a) will necessarily have phases of 0 or π . If such phases are used, for example, to try to identify heavy-atom-binding sites in a second derivative, the map will show the correct sites, but will also show spurious peaks of equal height related by the centre of symmetry. Faced with this choice, one must arbitrarily choose one of the alternative peaks which, in turn, will define an overall handedness for the heavy-atom arrangement. In the absence of any anomalous-scattering data, one can proceed with the structure determination in the standard way, but it must be kept in mind that either the correct electron-density map or its mirror image will ultimately be obtained.

An alternative approach is to include anomalous-scattering data in the initial phase determination, *i.e.* to use single isomorphous replacement with anomalous scattering (SIRAS). It must be remembered that in calculating phases from anomalous-scattering data, it is first necessary to determine the coordinates of the heavy atoms *in their absolute configuration*. If the wrong hand is used in

the SIRAS method (illustrated in Fig. 14.1.13.1), the resultant electron-density map will generally bear no relation to the correct electron density.

The recommended procedure, therefore, is as follows. One arbitrarily chooses one possible heavy-atom arrangement for heavy-atom derivative 1, calculates SIRAS phases and calculates a difference-electron-density map for derivative 2. The handedness of the derivative 1 coordinates are then inverted and the overall calculation repeated. The calculation based on the correct heavy-atom arrangement should show peaks at the heavy-atom sites of the second derivative. The calculation based on the incorrect arrangement shows noise (Matthews, 1966a). This procedure determines the absolute configuration of the heavy-atom arrangement and, at the same time, shows the derived sites for the second and subsequent derivatives.

Acknowledgements

This work was supported in part by NIH grant GM21967.