

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

The situation is more complex for three-dimensional acentric data. It has been shown (Rossmann, 1961) that

$$(F_{NH} - F_N)^2 \simeq F_H^2 \cos^2(\alpha_{NH} - \alpha_H) \quad (2.4.4.5)$$

when F_H is small compared to F_{NH} and F_N . Patterson synthesis with $(F_{NH} - F_N)^2$ as coefficients would, therefore, give an approximation to the heavy-atom vector distribution. An isomorphous difference Patterson synthesis of this type has been used extensively in protein crystallography to determine heavy-atom positions. The properties of this synthesis have been extensively studied (Ramachandran & Srinivasan, 1970; Rossmann, 1960; Phillips, 1966; Dodson & Vijayan, 1971) and it has been shown that this Patterson synthesis would provide a good approximation to the heavy-atom vector distribution even when F_H is large compared to F_N (Dodson & Vijayan, 1971).

As indicated earlier (see Section 2.4.3.1), heavy atoms are always anomalous scatterers, and the structure factors of any given reflection and its Friedel equivalent from a heavy-atom derivative have unequal magnitudes. If these structure factors are denoted by $\mathbf{F}_{NH}(+)$ and $\mathbf{F}_{NH}(-)$ and the real component of the heavy-atom contributions (including the real component of the dispersion correction) by \mathbf{F}_H , then it can be shown (Kartha & Parthasarathy, 1965) that

$$\left(\frac{k}{2}\right)^2 [F_{NH}(+) - F_{NH}(-)]^2 = F_H^2 \sin^2(\alpha_{NH} - \alpha_H), \quad (2.4.4.6)$$

where $k = (f_H + f_H')/f_H''$. Here it has been assumed that all the anomalous scatterers are of the same type with atomic scattering factor f_H and dispersion-correction terms f_H' and f_H'' . A Patterson synthesis with the left-hand side of (2.4.4.6) as coefficients would also yield the vector distribution corresponding to the heavy-atom positions (Rossmann, 1961; Kartha & Parthasarathy, 1965). However, $F_{NH}(+) - F_{NH}(-)$ is a small difference between two large quantities and is liable to be in considerable error. Patterson syntheses of this type are therefore rarely used to determine heavy-atom positions.

It is interesting to note (Kartha & Parthasarathy, 1965) that addition of (2.4.4.5) and (2.4.4.6) readily leads to

$$(F_{NH} - F_N)^2 + \left(\frac{k}{2}\right)^2 [F_{NH}(+) - F_{NH}(-)]^2 \simeq F_H^2. \quad (2.4.4.7)$$

Thus, the magnitude of the heavy-atom contribution can be estimated if intensities of Friedel equivalents have been measured from the derivative crystal. F_{NH} is then not readily available, but to a good approximation

$$F_{NH} = [F_{NH}(+) + F_{NH}(-)]/2. \quad (2.4.4.8)$$

A different and more accurate expression for estimating F_H^2 from isomorphous and anomalous differences was derived by Matthews (1966). According to a still more accurate expression derived by Singh & Ramaseshan (1966),

$$\begin{aligned} F_H^2 &= F_{NH}^2 + F_N^2 - 2F_{NH}F_N \cos(\alpha_N - \alpha_{NH}) \\ &= F_{NH}^2 + F_N^2 \pm 2F_{NH}F_N \\ &\quad \times (1 - \{k[F_{NH}(+) - F_{NH}(-)]/2F_N\}^2)^{1/2}. \end{aligned} \quad (2.4.4.9)$$

The lower estimate in (2.4.4.9) is relevant when $|\alpha_N - \alpha_{NH}| < 90^\circ$ and the upper estimate is relevant when $|\alpha_N - \alpha_{NH}| > 90^\circ$. The lower and the upper estimates may be referred to as F_{HLE} and F_{HUE} , respectively. It can be readily shown (Dodson & Vijayan, 1971) that the lower estimate would represent the correct value of F_H for a vast majority of reflections. Thus, a Patterson synthesis with F_{HLE}^2 as coefficients would yield the vector distribution of heavy atoms in

the derivative. Such a synthesis would normally be superior to those with the left-hand sides of (2.4.4.5) and (2.4.4.6) as coefficients. However, when the level of heavy-atom substitution is low, the anomalous differences are also low and susceptible to large percentage errors. In such a situation, a synthesis with $(F_{NH} - F_N)^2$ as coefficients is likely to yield better results than that with F_{HLE}^2 as coefficients (Vijayan, 1981).

Direct methods employing different methodologies have also been used successfully for the determination of heavy-atom positions (Navia & Sigler, 1974). These methods, developed primarily for the analysis of smaller structures, have not yet been successful in *a priori* analysis of protein structures. The very size of protein structures makes the probability relations used in these methods weak. In addition, data from protein crystals do not normally extend to high enough angles to permit resolution of individual atoms in the structure and the feasibility of using many of the currently popular direct-method procedures in such a situation has been a topic of much discussion. The heavy atoms in protein derivative crystals, however, are small in number and are normally situated far apart from one another. They are thus expected to be resolved even when low-resolution X-ray data are used. In most applications, the magnitudes of the differences between F_{NH} and F_N are formally considered as the 'observed structure factors' of the heavy-atom distribution and conventional direct-method procedures are then applied to them.

Once the heavy-atom parameters in one or more derivatives have been determined, approximate protein phase angles, α_N 's, can be derived using methods described later. These phase angles can then be readily used to determine the heavy-atom parameters in a new derivative employing a difference Fourier synthesis with coefficients

$$(F_{NH} - F_N) \exp(i\alpha_N). \quad (2.4.4.10)$$

Such syntheses are also used to confirm and to improve upon the information on heavy-atom parameters obtained through Patterson or direct methods. They are obviously very powerful when centric data corresponding to centrosymmetric projections are used. The synthesis yields satisfactory results even when the data are acentric although the difference Fourier technique becomes progressively less powerful as the level of heavy-atom substitution increases (Dodson & Vijayan, 1971).

While the positional parameters of heavy atoms can be determined with a reasonable degree of confidence using the above-mentioned methods, the corresponding temperature and occupancy factors cannot. Rough estimates of the latter are usually made from the strength and the size of appropriate peaks in difference syntheses. The estimated values are then refined, along with the positional parameters, using the techniques outlined below.

2.4.4.3. Refinement of heavy-atom parameters

The least-squares method with different types of minimization functions is used for refining the heavy-atom parameters, including the occupancy factors. The most widely used method (Dickerson *et al.*, 1961; Muirhead *et al.*, 1967; Dickerson *et al.*, 1968) involves the minimization of the function

$$\varphi = \sum w(F_{NH} - |\mathbf{F}_N + \mathbf{F}_H|)^2, \quad (2.4.4.11)$$

where the summation is over all the reflections and w is the weight factor associated with each reflection. Here F_{NH} is the observed magnitude of the structure factor for the particular derivative and $\mathbf{F}_N + \mathbf{F}_H$ is the calculated structure factor. The latter obviously depends upon the protein phase angle α_N , and the magnitude and the phase angle of \mathbf{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

2.4. ISOMORPHOUS REPLACEMENT AND ANOMALOUS SCATTERING

parameters HA_i , HB_i and HC_i . Then,

$$\begin{aligned} \mathbf{F}_{HA} &= \mathbf{F}_{HA}(HA_i) \\ \mathbf{F}_{HB} &= \mathbf{F}_{HB}(HB_i) \\ \mathbf{F}_{HC} &= \mathbf{F}_{HC}(HC_i). \end{aligned} \quad (2.4.4.12)$$

A set of approximate protein phase angles is first calculated, employing methods described later, making use of the unrefined heavy-atom parameters. These phase angles are used to construct $\mathbf{F}_N + \mathbf{F}_H$ for each derivative. (2.4.4.11) is then minimized, separately for each derivative, by varying HA_i for derivative A, HB_i for derivative B, and HC_i for derivative C. The refined values of HA_i , HB_i and HC_i are subsequently used to calculate a new set of protein phase angles. Alternate cycles of parameter refinement and phase-angle calculation are carried out until convergence is reached. The progress of refinement may be monitored by computing an R factor defined as (Kraut *et al.*, 1962)

$$R_K = \frac{\sum |F_{NH} - |\mathbf{F}_N + \mathbf{F}_H||}{F_{NH}}. \quad (2.4.4.13)$$

The above method has been successfully used for the refinement of heavy-atom parameters in the X-ray analysis of many proteins. However, it has one major drawback in that the refined parameters in one derivative are dependent on those in other derivatives through the calculation of protein phase angles. Therefore, it is important to ensure that the derivative, the heavy-atom parameters of which are being refined, is omitted from the phase-angle calculation (Blow & Matthews, 1973). Even when this is done, serious problems might arise when different derivatives are related by common sites. In practice, the occupancy factors of the common sites tend to be overestimated compared to those of the others (Vijayan, 1981; Dodson & Vijayan, 1971). Yet another factor which affects the occupancy factors is the accuracy of the phase angles. The inclusion of poorly phased reflections tends to result in the underestimation of occupancy factors. It is therefore advisable to omit from refinement cycles reflections with figures of merit less than a minimum threshold value or to assign a weight proportional to the figure of merit (as defined later) to each term in the minimization function (Dodson & Vijayan, 1971; Blow & Matthews, 1973).

If anomalous-scattering data from derivative crystals are available, the values of F_H can be estimated using (2.4.4.7) or (2.4.4.9) and these can be used as the 'observed' magnitudes of the heavy-atom contributions for the refinement of heavy-atom parameters, as has been done by many workers (Watenpaugh *et al.*, 1975; Vijayan, 1981; Kartha, 1965). If (2.4.4.9) is used for estimating F_H , the minimization function has the form

$$\varphi = \sum w(F_{HLE} - F_H)^2. \quad (2.4.4.14)$$

The progress of refinement may be monitored using a reliability index defined as

$$R = \frac{\sum |F_{HLE} - F_H|}{\sum F_{HLE}}. \quad (2.4.4.15)$$

The major advantage of using F_{HLE} 's in refinement is that the heavy-atom parameters in each derivative can now be refined independently of all other derivatives. Care should, however, be taken to omit from calculations all reflections for which F_{HLE} is likely to be the correct estimate of F_H . This can be achieved in practice by excluding from least-squares calculations all reflections for which F_{HLE} has a value less than the maximum expected value of F_H for the given derivative (Vijayan, 1981; Dodson & Vijayan, 1971).

A major problem associated with this refinement method is concerned with the effect of experimental errors on refined

parameters. The values of $F_{NH}(+) - F_{NH}(-)$ are often comparable to the experimental errors associated with $F_{NH}(+)$ and $F_{NH}(-)$. In such a situation, even random errors in $F_{NH}(+)$ and $F_{NH}(-)$ tend to increase systematically the observed difference between them (Dodson & Vijayan, 1971). In (2.4.4.7) and (2.4.4.9), this difference is multiplied by k or $k/2$, a quantity much greater than unity, and then squared. This could lead to the systematic overestimation of F_{HLE} 's and the consequent overestimation of occupancy factors. The situation can be improved by employing empirical values of k , evaluated using the relation (Kartha & Parthasarathy, 1965; Matthews, 1966)

$$k = \frac{2 \sum |F_{NH} - F_N|}{\sum |F_{NH}(+) - F_{NH}(-)|}, \quad (2.4.4.16)$$

for estimating F_{HLE} or by judiciously choosing the weighting factors in (2.4.4.14) (Dodson & Vijayan, 1971). The use of a modified form of F_{HLE} , arrived at through statistical considerations, along with appropriate weighting factors, has also been advocated (Dodson *et al.*, 1975).

When the data are centric, (2.4.4.9) reduces to

$$F_H = F_{NH} \pm F_N. \quad (2.4.4.17)$$

Here, again, the lower estimate most often corresponds to the correct value of F_H . (2.4.4.17) does not involve $F_{NH}(+) - F_{NH}(-)$ which, as indicated earlier, is prone to substantial error. Therefore, F_H 's estimated using centric data are more reliable than those estimated using acentric data. Consequently, centric reflections, when available, are extensively used for the refinement of heavy-atom parameters. It may also be noted that in conditions under which F_{HLE} corresponds to the correct estimate of F_H , minimization functions (2.4.4.11) and (2.4.4.14) are identical for centric data.

A Patterson function correlation method with a minimization function of the type

$$\varphi = \sum w[(F_{NH} - F_N)^2 - F_H^2]^2 \quad (2.4.4.18)$$

was among the earliest procedures suggested for heavy-atom-parameter refinement (Rossmann, 1960). This procedure would obviously work well when centric reflections are used. A modified version of this procedure, in which the origins of the Patterson functions are removed from the correlation, and centric and acentric data are treated separately, has been proposed (Terwilliger & Eisenberg, 1983).

2.4.4.4. Treatment of errors in phase evaluation: Blow and Crick formulation

As shown in Section 2.4.2.3, ideally, protein phase angles can be evaluated if two isomorphous heavy-atom derivatives are available. However, in practice, conditions are far from ideal on account of several factors such as imperfect isomorphism, errors in the estimation of heavy-atom parameters, and the experimental errors in the measurement of intensity from the native and the derivative crystals. It is therefore desirable to use as many derivatives as are available for phase determination. When isomorphism is imperfect and errors exist in data and heavy-atom parameters, all the circles in a Harker diagram would not intersect at a single point; instead, there would be a distribution of intersections, such as that illustrated in Fig. 2.4.4.1. Consequently, a unique solution for the phase angle cannot be deduced.

The statistical procedure for computing protein phase angles using multiple isomorphous replacement (MIR) was derived by Blow & Crick (1959). In their treatment, Blow and Crick assume, for mathematical convenience, that all errors, including those arising from imperfect isomorphism, could be considered as residing in the magnitudes of the derivative structure factors only. They further assume that these errors could be described by a