by simply adding f' to the normal scattering factor of the anomalous scatterers.

The effects of the imaginary component of the dispersion correction are, however, more complex. These effects could lead to serious errors in positional parameters when the space group is polar, if data in the entire diffraction sphere are not used (Ueki *et al.*, 1966; Cruickshank & McDonald, 1967). For example, accessible data in a hemisphere are normally used for X-ray analysis when the space group is $P1$. If the hemisphere has say h positive, the x coordinates of all the atoms would be in error when the structure contains anomalous scatterers. The situation in other polar space groups has been discussed by Cruickshank & McDonald (1967). In general, in the presence of anomalous scattering, it is desirable to collect data for the complete sphere, if accurate structural parameters are required (Srinivasan, 1972).

Methods have been derived to correct for dispersion effects in observed data from centrosymmetric and noncentrosymmetric crystals (Patterson, 1963). The methods are empirical and depend upon the refined parameters at the stage at which corrections are applied. This is obviously an unsatisfactory situation and it has been suggested that the measured structure factors of Bijvoet equivalents should instead be treated as independent observations in structure refinement (Ibers & Hamilton, 1964). The effect of dispersion corrections needs to be taken into account to arrive at the correct scale and temperature factors also (Wilson, 1975; Gilli & Cruickshank, 1973).

2.4.4. Isomorphous replacement and anomalous scattering in protein crystallography

2.4.4.1. *Protein heavy-atom derivatives*

Perhaps the most spectacular applications of isomorphous replacement and anomalous-scattering methods have been in the structure solution of large biological macromolecules, primarily proteins. Since its first successful application on myoglobin and haemoglobin, the isomorphous replacement method, which is often used in conjunction with the anomalous-scattering method, has been employed in the solution of scores of proteins. The application of this method involves the preparation of protein heavy-atom derivatives, *i.e.* the attachment of heavy atoms like mercury, uranium and lead, or chemical groups containing them, to protein crystals in a coherent manner without changing the conformation of the molecules and their crystal packing. 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 prepared by replacing the disordered solvent molecules by heavy-atom-containing groups without disturbing the original arrangement of protein molecules.

2.4.4.2. *Determination of heavy-atom parameters*

For any given reflection, the structure factor of the native protein crystal ($\mathbf{F}_N$), that of a heavy-atom derivative ($\mathbf{F}_{NH}$), and the contribution of the heavy atoms in that derivative ($\mathbf{F}_H$) are related by the equation

$$\mathbf{F}_{NH} = \mathbf{F}_N + \mathbf{F}_H. \qquad (2.4.4.1)$$

The value of $\mathbf{F}_H$ depends not only on the positional and thermal parameters of the heavy atoms, but also on their occupancy factors, because, at a given position, the heavy atom may not often be present in all the unit cells. For example, if the heavy atom is present at a given position in only half the unit cells in the crystal, then the occupancy factor of the site is said to be 0.5.

For the successful determination of the heavy-atom parameters, as also for the subsequent phase determination, the data sets from the native and the derivative crystals should have the same relative scale. The different data sets should also have the same overall temperature factor. Different scaling procedures have been suggested (Blundell & Johnson, 1976) and, among them, the following procedure, based on Wilson's (1942) statistics, appears to be the most feasible in the early stages of structure analysis.

Assuming that the data from the native and the derivative crystals obey Wilson's statistics, we have, for any range of $\sin^2\theta/\lambda^2$,

$$\ln\left\{\frac{\sum f_{Nj}^2}{\langle F_N^2\rangle}\right\} = \ln K_N + 2B_N\frac{\sin^2\theta}{\lambda^2} \qquad (2.4.4.2)$$

and

$$\ln\left\{\frac{\sum f_{Nj}^2 + \sum f_{Hj}^2}{\langle F_{NH}^2\rangle}\right\} = \ln K_{NH} + 2B_{NH}\frac{\sin^2\theta}{\lambda^2}, \qquad (2.4.4.3)$$

where f_{Nj} and f_{Hj} refer to the atomic scattering factors of protein atoms and heavy atoms, respectively. K_N and K_{NH} are the scale factors to be applied to the intensities from the native and the derivative crystals, respectively, and B_N and B_{NH} the temperature factors of the respective structure factors. Normally one would be able to derive the absolute scale factor and the temperature factor for both the data sets from (2.4.4.2) and (2.4.4.3) using the well known Wilson plot. The data from protein crystals, however, do not follow Wilson's statistics as protein molecules contain highly non-random features. Therefore, in practice, it is difficult to fit a straight line through the points in a Wilson plot, thus rendering the parameters derived from it unreliable. (2.4.4.2) and (2.4.4.3) can, however, be used in a different way. From the two equations we obtain

$$\ln\left\{\frac{\sum f_{Nj}^2 + \sum f_{Hj}^2}{\sum f_{Nj}^2}\cdot\frac{\langle F_N^2\rangle}{\langle F_{NH}^2\rangle}\right\} = \ln\left(\frac{K_{NH}}{K_N}\right) + 2(B_{NH} - B_N)\frac{\sin^2\theta}{\lambda^2}. \qquad (2.4.4.4)$$

The effects of structural non-randomness in the crystals obviously cancel out in (2.4.4.4). When the left-hand side of (2.4.4.4) is plotted against $(\sin^2\theta)/\lambda^2$, it is called a comparison or difference Wilson plot. Such plots yield the ratio between the scales of the derivative and the native data, and the additional temperature factor of the derivative data. Initially, the number and the occupancy factors of heavy-atom sites are unknown, and are roughly estimated from intensity differences to evaluate $\sum f_{Hj}^2$. These estimates usually undergo considerable revision in the course of the determination and the refinement of heavy-atom parameters.

At first, heavy-atom positions are most often determined by Patterson syntheses of one type or another. Such syntheses are discussed in some detail elsewhere in Chapter 2.3. They are therefore discussed here only briefly.

Equation (2.4.2.6) holds when the data are centric. F_H is usually small compared to F_N and F_{NH}, and the minus sign is then relevant on the left-hand side of (2.4.2.6). Thus the difference between the magnitudes of $\mathbf{F}_{NH}$ and $\mathbf{F}_N$, which can be obtained experimentally, normally gives a correct estimate of the magnitude of $\mathbf{F}_H$ for most reflections. Then a Patterson synthesis with $(F_{NH} - F_N)^2$ as coefficients corresponds to the distribution of vectors between heavy atoms, when the data are centric. But proteins are made up of L-amino acids and hence cannot crystallize in centrosymmetric space groups. However, many proteins crystallize in space groups with centrosymmetric projections. The centric data corresponding to these projections can then be used for determining heavy-atom positions through a Patterson synthesis of the type outlined above.

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

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

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

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

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

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

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

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

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),

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

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

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

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

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

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

parameters HA_i, HB_i and HC_i. Then,

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

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)

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

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

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

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

$$R = \frac{\sum |F_{HLE} - F_H|}{\sum F_{HLE}}. \tag{2.4.4.15}$$

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_{HUE} 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_{HUE} 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)

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

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

$$F_H = F_{NH} \pm F_N. \tag{2.4.4.17}$$

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

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

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

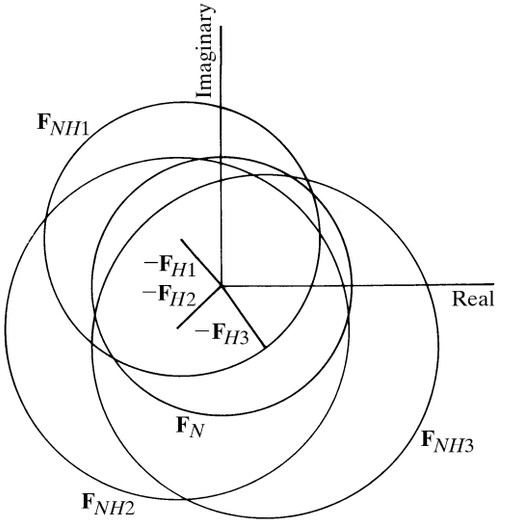

Fig. 2.4.4.1. Distribution of intersections in the Harker construction under non-ideal conditions.

Gaussian distribution. With these simplifying assumptions, the statistical procedure for phase determination could be derived in the following manner.

Consider the vector diagram, shown in Fig. 2.4.4.2, for a reflection from the ith derivative for an arbitrary value α for the protein phase angle. Then,

$$D_{Hi}(\alpha) = [F_N^2 + F_{Hi}^2 + 2F_N F_{Hi}\cos(\alpha_{Hi} - \alpha)]^{1/2}. \qquad (2.4.4.19)$$

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

$$\xi_{Hi}(\alpha) = F_{NHi} - D_{Hi}(\alpha), \qquad (2.4.4.20)$$

is a measure of the departure of α from α_N. ξ is called the lack of closure. The probability for α being the correct protein phase angle could now be defined as

$$P_i(\alpha) = N_i \exp[-\xi_{Hi}^2(\alpha)/2E_i^2], \qquad (2.4.4.21)$$

where N_i is the normalization constant and E_i is the estimated r.m.s. error. The methods for estimating E_i will be outlined later. When several derivatives are used for phase determination, the total probability of the phase angle α being the protein phase angle would be

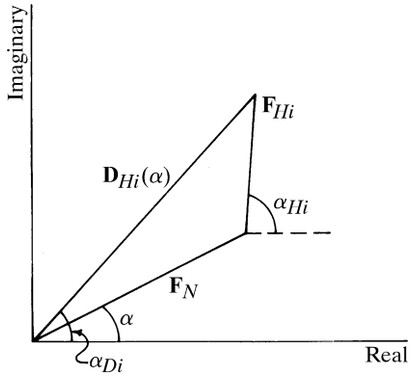

Fig. 2.4.4.2. Vector diagram indicating the calculated structure factor, $\mathbf{D}_{Hi}(\alpha)$, of the ith heavy-atom derivative for an arbitrary value α for the phase angle of the structure factor of the native protein.

$$P(\alpha) = \prod P_i(\alpha) = N \exp\left\{-\sum_i [\xi_{Hi}^2(\alpha)/2E_i^2]\right\}, \qquad (2.4.4.22)$$

where the summation is over all the derivatives. A typical distribution of $P(\alpha)$ plotted around a circle of unit radius is shown in Fig. 2.4.4.3. The phase angle corresponding to the highest value of $P(\alpha)$ would obviously be the most probable protein phase, α_M, of the given reflection. The most probable electron-density distribution is obtained if each F_N is associated with the corresponding α_M in a Fourier synthesis.

Blow and Crick suggested a different way of using the probability distribution. In Fig. 2.4.4.3, the centroid of the probability distribution is denoted by 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 the 'figure of merit' and the 'best phase', respectively. One can then compute a 'best Fourier' with coefficients

$$mF_N \exp(i\alpha_B).$$

The best Fourier is expected to provide an electron-density distribution with the lowest r.m.s. error. The figure of merit and the best phase are usually calculated using the equations

$$\begin{aligned} m\cos\alpha_B &= \sum_i P(\alpha_i)\cos(\alpha_i)/\sum_i P(\alpha_i) \\ m\sin\alpha_B &= \sum_i P(\alpha_i)\sin(\alpha_i)/\sum_i P(\alpha_i), \end{aligned} \qquad (2.4.4.23)$$

where $P(\alpha_i)$ are calculated, say, at 5° intervals (Dickerson *et al.*, 1961). The figure of merit is statistically interpreted as the cosine of the expected error in the calculated phase angle and it is obviously a measure of the precision of phase determination. In general, m is high when α_M and α_B are close to each other and low when they are far apart.

2.4.4.5. *Use of anomalous scattering in phase evaluation*

When anomalous-scattering data have been collected from derivative crystals, $F_{NH}(+)$ and $F_{NH}(-)$ can be formally treated as arising from two independent derivatives. The corresponding Harker diagram is shown in Fig. 2.4.4.4. Thus, in principle, protein phase angles can be determined using a single derivative when anomalous-scattering effects are also made use of. It is interesting to note that the information obtained from isomorphous differences, $F_{NH} - F_N$, and that obtained from anomalous differences,

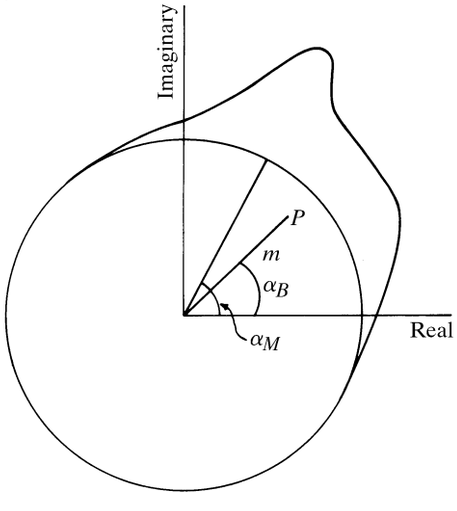

Fig. 2.4.4.3. The probability distribution of the protein phase angle. The point P is the centroid of the distribution.

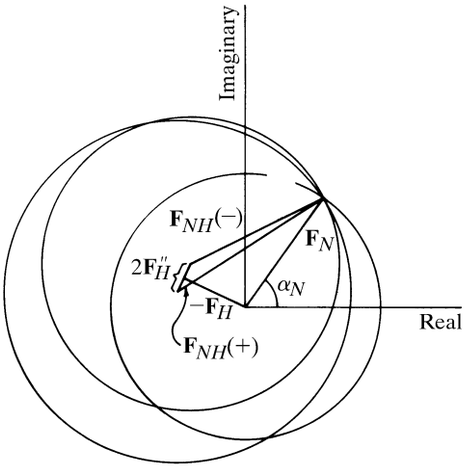

Fig. 2.4.4.4. Harker construction using anomalous-scattering data from a single derivative.

$F_{NH}(+) - F_{NH}(-)$, are complementary. The isomorphous difference for any given reflection is a maximum when F_N and F_H are parallel or antiparallel. The anomalous difference is then zero, if all the anomalous scatterers are of the same type, and α_N is determined uniquely on the basis of the isomorphous difference. The isomorphous difference decreases and the anomalous difference increases as the inclination between $\mathbf{F}_N$ and $\mathbf{F}_H$ increases. The isomorphous difference tends to be small and the anomalous difference tends to have the maximum possible value when $\mathbf{F}_N$ and $\mathbf{F}_H$ are perpendicular to each other. The anomalous difference then has the predominant influence in determining the phase angle.

Although isomorphous and anomalous differences have a complementary role in phase determination, their magnitudes are obviously unequal. Therefore, when $F_{NH}(+)$ and $F_{NH}(-)$ are treated as arising from two derivatives, the effect of anomalous differences on phase determination would be only marginal as, for any given reflection, $F_{NH}(+) - F_{NH}(-)$ is usually much smaller than $F_{NH} - F_N$. However, the magnitude of the error in the anomalous difference would normally be much smaller than that in the corresponding isomorphous difference. Firstly, the former is obviously free from the effects of imperfect isomorphism. Secondly, $F_{NH}(+)$ and $F_{NH}(-)$ are expected to have the same systematic errors as they are measured from the same crystal. These errors are eliminated in the difference between the two quantities. Therefore, as pointed out by North (1965), the r.m.s. error used for anomalous differences should be much smaller than that used for isomorphous differences. Denoting the r.m.s. error in anomalous differences by E', the new expression for the probability distribution of protein phase angle may be written as

$$P_i(\alpha) = N_i \exp[-\xi^2_{Hi}(\alpha)/2E_i^2] \times \exp\{-[\Delta H_i - \Delta H_{i\text{cal}}(\alpha)]^2/2E_i'^2\}, \qquad (2.4.4.24)$$

where

$$\Delta H_i = F_{NHi}(+) - F_{NHi}(-)$$

and

$$\Delta H_{i\text{cal}}(\alpha) = 2F''_{Hi}\sin(\alpha_{Di} - \alpha_{Hi}).$$

Here α_{Di} is the phase angle of $D_{Hi}(\alpha)$ [see (2.4.4.19) and Fig. 2.4.4.2]. $\Delta H_{i\text{cal}}(\alpha)$ is the anomalous difference calculated for the assumed protein phase angle α. F_{NHi} may be taken as the average of $F_{NHi}(+)$ and $F_{NHi}(-)$ for calculating $\xi^2_{Hi}(\alpha)$ using (2.4.4.20).

2.4.4.6. *Estimation of r.m.s. error*

Perhaps the most important parameters that control the reliability of phase evaluation using the Blow and Crick formulation are the isomorphous r.m.s. error E_i and the anomalous r.m.s. error E'_i. For a given derivative, the sharpness of the peak in the phase probability distribution obviously depends upon the value of E and that of E' when anomalous-scattering data have also been used. When several derivatives are used, an overall underestimation of r.m.s. errors leads to artifically sharper peaks, the movement of α_B towards α_M, and deceptively high figures of merit. Opposite effects result when E's are overestimated. Underestimation or overestimation of the r.m.s. error in the data from a particular derivative leads to distortions in the relative contribution of that derivative to the overall phase probability distributions. It is therefore important that the r.m.s. error in each derivative is correctly estimated.

Centric reflections, when present, obviously provide the best means for evaluating E using the expression

$$E^2 = \sum_n (|F_{NH} \pm F_N| - F_N)^2/n. \qquad (2.4.4.25)$$

As suggested by Blow & Crick (1959), values of E thus estimated can be used for acentric reflections as well. Once a set of approximate protein phase angles is available, E_i can be calculated as the r.m.s. lack of closure corresponding to α_B [*i.e.* $\alpha = \alpha_B$ in (2.4.4.20)] (Kartha, 1976). E'_i can be similarly evaluated as the r.m.s. difference between the observed anomalous difference and the anomalous difference calculated for α_B [see (2.4.4.24)]. Normally, the value of E'_i is about a third of that of E_i (North, 1965).

A different method, outlined below, can also be used to evaluate E and E' when anomalous scattering is present (Vijayan, 1981; Adams, 1968). From Fig. 2.4.2.2, we have

$$\cos\psi = (F^2_{NH} + F^2_H - F^2_N)/2F_{NH}F_H \qquad (2.4.4.26)$$

and

$$F^2_N = F^2_{NH} + F^2_H - 2F_{NH}F_H\cos\psi, \qquad (2.4.4.27)$$

where $\psi = \alpha_{NH} - \alpha_H$. Using arguments similar to those used in deriving (2.4.3.5), we obtain

$$\sin\psi = [F^2_{NH}(+) - F^2_{NH}(-)]/4F_{NH}F''_H. \qquad (2.4.4.28)$$

If F_{NH} is considered to be equal to $[F_{NH}(+) + F_{NH}(-)]/2$, we obtain from (2.4.4.28)

$$F_{NH}(+) - F_{NH}(-) = 2F''_H\sin\psi. \qquad (2.4.4.29)$$

We obtain what may be called ψ_{iso} if the magnitude of ψ is determined from (2.4.4.26) and the quadrant from (2.4.4.28). Similarly, we obtain ψ_{ano} if the magnitude of ψ is determined from (2.4.4.28) and the quadrant from (2.4.4.26). Ideally, ψ_{iso} and ψ_{ano} should have the same value and the difference between them is a measure of the errors in the data. F_N obtained from (2.4.4.27) using ψ_{ano} may be considered as its calculated value $(F_{N\text{cal}})$. Then, assuming all errors to lie in F_N, we may write

$$E^2 = \sum_n (F_N - F_{N\text{cal}})^2/n. \qquad (2.4.4.30)$$

Similarly, the calculated anomalous difference (ΔH_{cal}) may be evaluated from (2.4.4.29) using ψ_{iso}. Then

$$E'^2 = \sum_n [|F_{NH}(+) - F_{NH}(-)| - \Delta H_{\text{cal}}]^2/n. \qquad (2.4.4.31)$$

If all errors are assumed to reside in F_H, E can be evaluated in yet another way using the expression

$$E^2 = \sum_n (F_{HLE} - F_H)^2/n. \qquad (2.4.4.32)$$

2.4.4.7. *Suggested modifications to Blow and Crick formulation and the inclusion of phase information from other sources*

Modifications to the Blow and Crick procedure of phase evaluation have been suggested by several workers, although none represent a fundamental departure from the essential features of their formulation. In one of the modifications (Cullis *et al.*, 1961*a*; Ashida, 1976), all E_i's are assumed to be the same, but the lack-of-closure error ξ_{Hi} for the ith derivative is measured as the distance from the mean of all intersections between phase circles to the point of intersection of the phase circle of that derivative with the phase circle of the native protein. Alternatively, individual values of E_i are retained, but the lack of closure is measured from the weighted mean of all intersections (Ashida, 1976). This is obviously designed to undo the effects of the unduly high weight given to F_N in the Blow and Crick formulation. In another modification (Raiz & Andreeva, 1970; Einstein, 1977), suggested for the same purpose, the F_N and F_{NHi} circles are treated as circular bands, the width of each band being related to the error in the appropriate structure factor. A comprehensive set of modifications suggested by Green (1979) treats different types of errors separately. In particular, errors arising from imperfect isomorphism are treated in a comprehensive manner.

Although the isomorphous replacement method still remains the method of choice for the *ab initio* determination of protein structures, additional items of phase information from other sources are increasingly being used to replace, supplement, or extend the information obtained through the application of the isomorphous replacement. Methods have been developed for the routine refinement of protein structures (Watenpaugh *et al.*, 1973; Huber *et al.*, 1974; Sussman *et al.*, 1977; Jack & Levitt, 1978; Isaacs & Agarwal, 1978; Hendrickson & Konnert, 1980) and they provide a rich source of phase information. However, the nature of the problem and the inherent limitations of the Fourier technique are such that the possibility of refinement yielding misleading results exists (Vijayan, 1980*a*,*b*). It is therefore sometimes desirable to combine the phases obtained during refinement with the original isomorphous replacement phases. The other sources of phase information include molecular replacement (see Chapter 2.3), direct methods (Hendrickson & Karle, 1973; Sayre, 1974; de Rango *et al.*, 1975; see also Chapter 2.2) and different types of electron-density modifications (Hoppe & Gassmann, 1968; Collins, 1975; Schevitz *et al.*, 1981; Bhat & Blow, 1982; Agard & Stroud, 1982; Cannillo *et al.*, 1983; Raghavan & Tulinsky, 1979; Wang, 1985).

The problem of combining isomorphous replacement phases with those obtained by other methods was first addressed by Rossmann & Blow (1961). The problem was subsequently examined by Hendrickson & Lattman (1970) and their method, which involves a modification of the Blow and Crick formulation, is perhaps the most widely used for combining phase information from different sources.

The Blow and Crick procedure is based on an assumed Gaussian 'lumped' error in F_{NHi} which leads to a lack of closure, $\xi_{Hi}(\alpha)$, in F_{NHi} defined by (2.4.4.20). Hendrickson and Lattman make an equally legitimate assumption that the lumped error, again assumed to be Gaussian, is associated with F^2_{NHi}. Then, as in (2.4.4.20), we have

$$\xi''_{Hi}(\alpha) = F^2_{NHi} - D^2_{Hi}(\alpha), \qquad (2.4.4.33)$$

where $\xi''_{Hi}(\alpha)$ is the lack of closure associated with F^2_{NHi} for an assumed protein phase angle α. Then the probability for α being the correct phase angle can be expressed as

$$P_i(\alpha) = N_i \exp[-\xi''^2_{Hi}(\alpha)/2E''^2_i], \qquad (2.4.4.34)$$

where E''_i is the r.m.s. error in F^2_{NHi}, which can be evaluated using methods similar to those employed for evaluating E_i. Hendrickson and Lattman have shown that the exponent in the probability expression (2.4.4.34) can be readily expressed as a linear combination of five terms in the following manner.

$$-\xi''^2_{Hi}(\alpha)/2E''^2_i = K_i + A_i \cos\alpha + B_i \sin\alpha + C_i \cos 2\alpha + D_i \sin 2\alpha, \qquad (2.4.4.35)$$

where K_i, A_i, B_i, C_i and D_i are constants dependent on F_N, F_{Hi}, F_{NHi} and E''_i. Thus, five constants are enough to store the complete probability distribution of any reflection. Expressions for the five constants have been derived for phase information from anomalous scattering, tangent formula, partial structure and molecular replacement. The combination of the phase information from all sources can then be achieved by simply taking the total value of each constant. Thus, the total probability of the protein phase angle being α is given by

$$P(\alpha) = \prod P_s(\alpha) = N \exp\left(\sum_s K_s + \sum_s A_s \cos\alpha + \sum_s B_s \sin\alpha + \sum_s C_s \cos 2\alpha + \sum_s D_s \sin 2\alpha\right), \qquad (2.4.4.36)$$

where K_s, A_s *etc.* are the constants appropriate for the sth source and N is the normalization constant.

2.4.4.8. *Fourier representation of anomalous scatterers*

It is often useful to have a Fourier representation of only the anomalous scatterers in a protein. The imaginary component of the electron-density distribution obviously provides such a representation. When the structure is known and $F_N(+)$ and $F_N(-)$ have been experimentally determined, Chacko & Srinivasan (1970) have shown that this representation is obtained in a Fourier synthesis with $i[\mathbf{F}_N(+) + \mathbf{F}^*_N(-)]/2$ as coefficients, where $\mathbf{F}^*_N(-)$, whose magnitude is $F_N(-)$, is the complex conjugate of $\mathbf{F}_N(+)$. They also indicated a method for calculating the phase angles of $\mathbf{F}_N(+)$ and $\mathbf{F}^*_N(-)$. It has been shown (Hendrickson & Sheriff, 1987) that the Bijvoet-difference Fourier synthesis proposed earlier by Kraut (1968) is an approximation of the true imaginary component of the electron density. The imaginary synthesis can be useful in identifying minor anomalous-scattering centres when the major centres are known and also in providing an independent check on the locations of anomalous scatterers and in distinguishing between anomalous scatterers with nearly equal atomic numbers (Sheriff & Hendrickson, 1987; Kitagawa *et al.*, 1987).

2.4.5. Anomalous scattering of neutrons and synchrotron radiation. The multiwavelength method

The multiwavelength anomalous-scattering method (Ramaseshan, 1982) relies on the variation of dispersion-correction terms as a function of the wavelength used. The success of the method therefore depends upon the size of the correction terms and the availability of incident beams of comparable intensities at different appropriate wavelengths. Thus, although this method was used as early as 1957 (Ramaseshan *et al.*, 1957) as an aid to structure solution employing characteristic X-rays, it is, as outlined below, ideally suited in structural work employing neutrons and synchrotron radiation. In principle, γ-radiation can also be used for phase

references