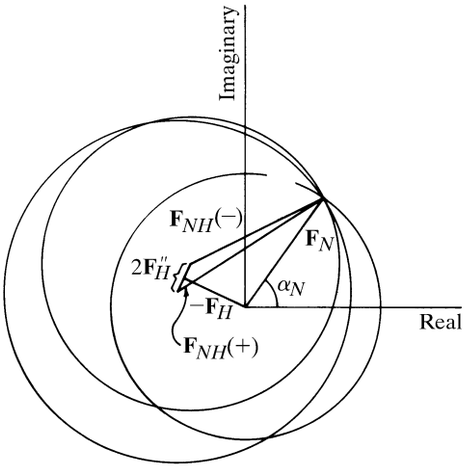


Fig. 2.4.4.4. Harker construction using anomalous-scattering data from a single derivative.

$F_{NH}(+) - F_{NH}(-)$, 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 $\alpha_N$ 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 $\alpha_{Di}$ 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 $\alpha$. $F_{NHi}$ 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 $\alpha_B$ towards $\alpha_M$, 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 $\alpha_B$ [*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 $\alpha_B$ [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 $\psi_{\text{iso}}$ if the magnitude of $\psi$ is determined from (2.4.4.26) and the quadrant from (2.4.4.28). Similarly, we obtain $\psi_{\text{ano}}$ if the magnitude of $\psi$ is determined from (2.4.4.28) and the quadrant from (2.4.4.26). Ideally, $\psi_{\text{iso}}$ and $\psi_{\text{ano}}$ 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 $\psi_{\text{ano}}$ 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 $(\Delta H_{\text{cal}})$ may be evaluated from (2.4.4.29) using $\psi_{\text{iso}}$. 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 $\xi_{Hi}$ for the $i$th 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 $\alpha$. Then the probability for $\alpha$ 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 $\alpha$ 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 $s$th 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, $\gamma$-radiation can also be used for phase



references