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

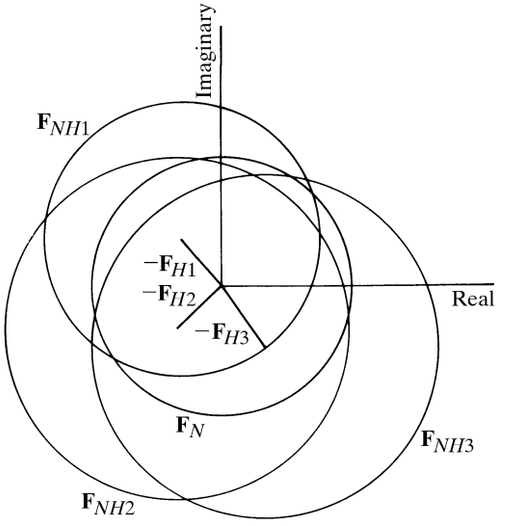

Fig. 2.4.4.1. Distribution of intersections in the Harker construction under non-ideal conditions.

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

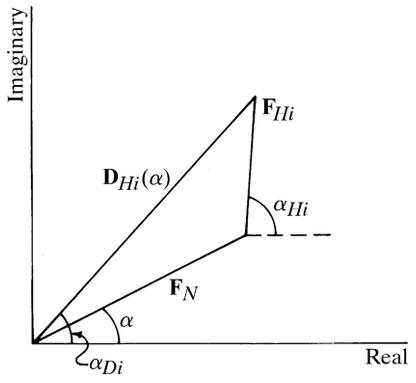

Fig. 2.4.4.2. Vector diagram indicating the calculated structure factor, $\mathbf{D}_{Hi}(\alpha)$, of the ith heavy-atom derivative for an arbitrary value α for the phase angle of the structure factor of the native protein.

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

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,

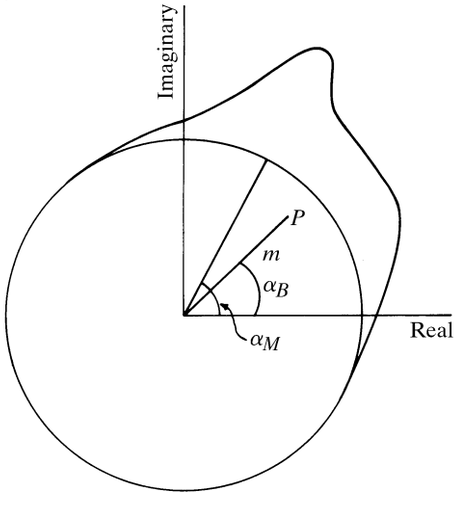

Fig. 2.4.4.3. The probability distribution of the protein phase angle. The point P is the centroid of the distribution.

references