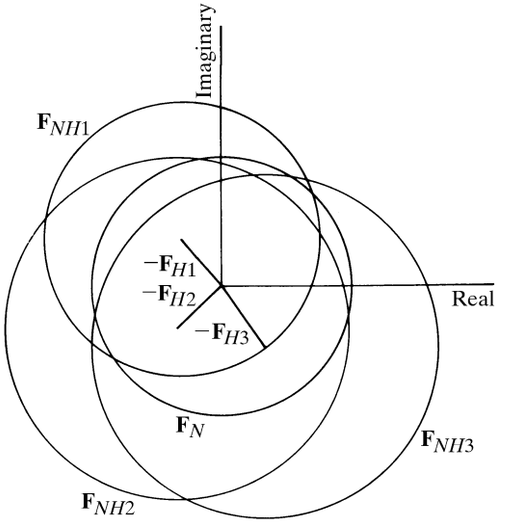

Fig. 2.4.4.1. Distribution of intersections in the Harker construction under non-ideal conditions.

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

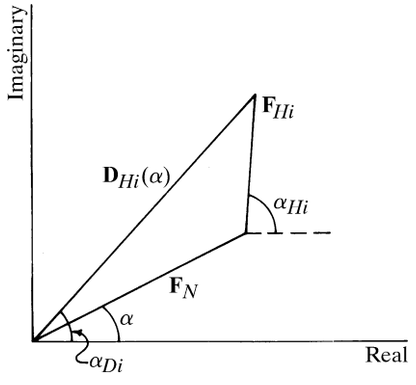

Fig. 2.4.4.2. Vector diagram indicating the calculated structure factor, $\mathbf{D}_{Hi}(\alpha)$, of the ith heavy-atom derivative for an arbitrary value α for the phase angle of the structure factor of the native protein.

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

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,

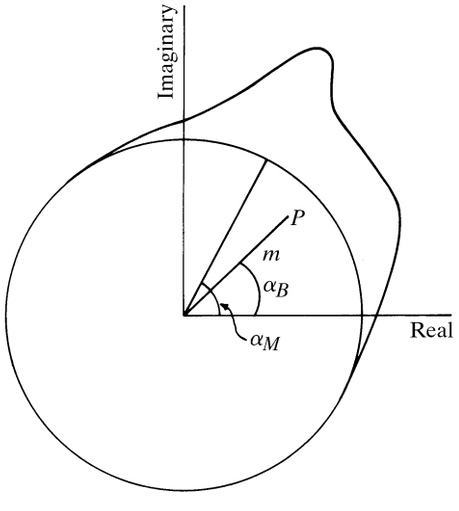

Fig. 2.4.4.3. The probability distribution of the protein phase angle. The point P is the centroid of the distribution.

references