

## 16. DIRECT METHODS

## 16.1.2.1. SIR differences

Given the individual normalized structure-factor magnitudes for native structures ( $|E_{\text{nat}}|$ ) and for structures containing heavy atoms ( $|E_{\text{der}}|$ ), as well as the atomic scattering factors  $|f_j| = |f_j^0 + f_j' + if_j''| = [(f_j^0 + f_j')^2 + (f_j'')^2]^{1/2}$  which allow for the possibility of anomalous scattering, then greatest-lower-bound estimates of SIR difference- $E$  magnitudes are

$$|E_{\Delta}| = \frac{\left( \sum_{j=1}^{N_{\text{der}}} |f_j|^2 \right)^{1/2} |E_{\text{der}}| - \left( \sum_{j=1}^{N_{\text{nat}}} |f_j|^2 \right)^{1/2} |E_{\text{nat}}|}{q \left[ \left( \sum_{j=1}^{N_{\text{der}}} |f_j|^2 \right) - \left( \sum_{j=1}^{N_{\text{nat}}} |f_j|^2 \right) \right]^{1/2}}, \quad (16.1.2.2)$$

where  $q = q_0 \exp(q_1 s^2 + q_2 s^4)$  is a least-squares-fitted empirical renormalization scaling function, dependent on  $\sin(\theta)/\lambda$ , that imposes the condition  $\langle |E_{\Delta}|^2 \rangle = 1$  and serves to define  $q_0$ ,  $q_1$  and  $q_2$ .

## 16.1.2.2. SAD differences

Given Friedel pairs of normalized structure-factor magnitudes ( $|E_{+\mathbf{H}}|$ ,  $|E_{-\mathbf{H}}|$ ) and the atomic scattering factors ( $f^0$ ,  $f_j'$  and  $f_j''$ ), then the greatest-lower-bound estimates of SAD difference  $|E|$ 's are

$$|E_{\Delta}| = \frac{\left[ \sum_{j=1}^N (f_j^0 + f_j')^2 + (f_j'')^2 \right]^{1/2} \|E_{+\mathbf{H}}\| - \|E_{-\mathbf{H}}\|}{2q \left[ \sum_{j=1}^N (f_j'')^2 \right]^{1/2}}, \quad (16.1.2.3)$$

where, again,  $q$  is an empirical renormalization scaling function that imposes the condition  $\langle |E_{\Delta}|^2 \rangle = 1$ .

In the case of MAD data, the anomalous differences for all wavelengths can be combined in the form of  $F_A$  structure factors as described in Chapter 14.2 of this volume by Smith & Hendrickson. In the *SHELX* system, the program *SHELXC* prepares the  $|F_A|$  values and the phase shifts  $\alpha$  for SAD, SIR, SIRAS, RIP (radiation-damage-induced phasing) and MAD experimental phasing (Sheldrick, 2008, 2010), and the corresponding  $|E_A|$  values are derived by *SHELXD*.

## 16.1.2.3. Difference intensities and direct methods

There are, of course, difficulties with normalized difference intensities. Direct methods normalize the data to give difference  $E$  magnitudes, and the macromolecular temperature factors incorporated in this process can be a source of error. The temperature-factor correction, often large for macromolecules, can take small, statistically dubious, amplitude differences at high resolution when working with anomalous difference data and magnify them into large normalized structure-factor differences that are then used in direct methods. These methods are very sensitive to such errors and will often fail unless a suitable resolution truncation limit is imposed. Usually this is around 0.5 Å larger than the experimental diffraction limit of the data or up to the resolution shell with  $I/\sigma(I) = 10$ . Morris *et al.* (2004) have extended their work on profiles to devise a method of normalization based on  $\langle |E|^2 \rangle$  profiles, and this should be more robust than traditional methods.

## 16.1.3. Starting the phasing process

The phase problem of X-ray crystallography may be defined as the problem of determining the phases  $\varphi$  of the normalized structure factors  $E$  when only the magnitudes  $|E|$  are given. Owing to the atomicity of crystal structures and the redundancy of the known magnitudes, the phase problem is overdetermined

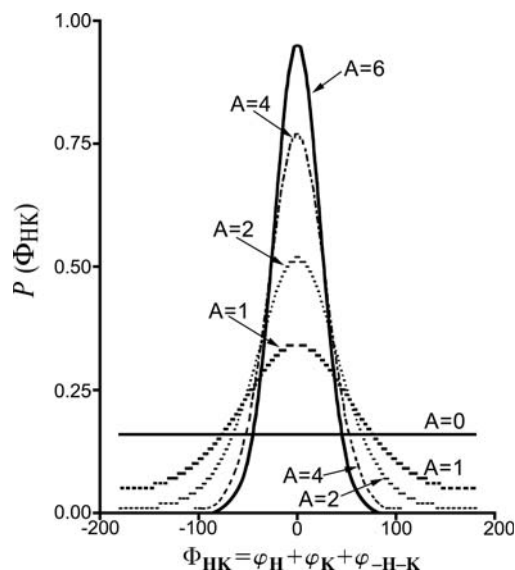


Figure 16.1.3.1

The conditional probability distribution,  $P(\Phi_{\mathbf{HK}})$ , of the three-phase structure invariants,  $\Phi_{\mathbf{HK}}$ , having associated parameters  $A_{\mathbf{HK}}$  with values of 0, 1, 2, 4 and 6. When  $A \approx 0$ , all values of  $\Phi_{\mathbf{HK}}$  are equally likely, and no information useful for phase determination is available. However, the sum of the three phases for most invariants with  $A \approx 6$  is close to  $0^\circ$ , and an estimate of one phase can be made if the other two are known.

and is, therefore, solvable in principle. This overdetermination implies the existence of relationships among the  $E$ 's and, since the magnitudes  $|E|$  are presumed to be known, there exist identities among the phases that are dependent on the known magnitudes alone. The techniques of probability theory lead to the joint probability distributions of arbitrary collections of  $E$  from which the conditional probability distributions of selected sets of phases, given the values of suitably chosen magnitudes  $|E|$ , may be inferred.

## 16.1.3.1. Structure invariants

The magnitude-dependent entities that constitute the foundation of direct methods are linear combinations of phases called *structure invariants*. The term 'structure invariant' stems from the fact that the values of these quantities are independent of the choice of origin. The most useful of the structure invariants are the three-phase or *triplet invariants*,

$$\Phi_{\mathbf{HK}} = \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}}, \quad (16.1.3.1)$$

the conditional probability distribution (Cochran, 1955), given  $A_{\mathbf{HK}}$ , of which is

$$P(\Phi_{\mathbf{HK}}) = [2\pi I_0(A_{\mathbf{HK}})]^{-1} \exp(A_{\mathbf{HK}} \cos \Phi_{\mathbf{HK}}), \quad (16.1.3.2)$$

where

$$A_{\mathbf{HK}} = (2/N^{1/2}) |E_{\mathbf{H}} E_{\mathbf{K}} E_{\mathbf{H}+\mathbf{K}}| \quad (16.1.3.3)$$

and  $N$  is the number of atoms, here presumed to be identical, in the primitive unit cell. This distribution is illustrated in Fig. 16.1.3.1. The expected value of the cosine of a particular triplet,  $\Phi_{\mathbf{HK}}$ , is given by the ratio of modified Bessel functions,  $I_1(A_{\mathbf{HK}})/I_0(A_{\mathbf{HK}})$ .

Estimates of the invariant values are most reliable when the normalized structure-factor magnitudes ( $|E_{\mathbf{H}}|$ ,  $|E_{\mathbf{K}}|$  and  $|E_{-\mathbf{H}-\mathbf{K}}|$ ) are large and the number of atoms in the corresponding primitive unit cell,  $N$ , is small. This is one important reason why direct phasing is more difficult for macromolecules than it is for small molecules. Four-phase or quartet invariants have proven helpful in small-molecule structure determination, particularly when