the real-space part is able to lead to the correct solution even when the tangent formula does not provide favourable phase values.

2.2.10.3. *Integration of direct methods with isomorphous replacement techniques*

The modulus of the isomorphous difference

$$\Delta F = |F_{PH}| - |F_P|$$

may be assumed at a first approximation as an estimate of the heavy-atom s.f. F_H. Normalization of $|\Delta F|$'s and application of the tangent formula may reveal the heavy-atom structure (Wilson, 1978).

The theoretical basis for integrating the techniques of direct methods and isomorphous replacement was introduced by Hauptman (1982*a*). According to his notation let us denote by f_j and g_j atomic scattering factors for the atom labelled j in a pair of isomorphous structures, and let $E_{\mathbf{h}}$ and $G_{\mathbf{h}}$ denote corresponding normalized structure factors. Then

$$E_{\mathbf{h}} = |E_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}) = \alpha_{20}^{-1/2} \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j),$$

$$G_{\mathbf{h}} = |G_{\mathbf{h}}| \exp(i\psi_{\mathbf{h}}) = \alpha_{02}^{-1/2} \sum_{j=1}^{N} g_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j),$$

where

$$\alpha_{mn} = \sum_{j=1}^{N} f_j^m g_j^n.$$

The conditional probability of the two-phase structure invariant $\Phi = \varphi_{\mathbf{h}} - \psi_{\mathbf{h}}$ given $|E_{\mathbf{h}}|$ and $|G_{\mathbf{h}}|$ is (Hauptman, 1982*a*)

$$P(\Phi|\ |E|, |G|) \simeq [2\pi I_0(Q)]^{-1} \exp(Q \cos \Phi),$$

where

$$Q = |EG|[2\alpha/(1-\alpha^2)],$$
$$\alpha = \alpha_{11}/(\alpha_{20}^{1/2}\alpha_{02}^{1/2}).$$

Three-phase structure invariants were evaluated by considering that eight invariants exist for a given triple of indices $\mathbf{h}$, $\mathbf{k}$, $\mathbf{l}$ ($\mathbf{h}+\mathbf{k}+\mathbf{l}=0$):

$$\Phi_1 = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} \qquad \Phi_2 = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \psi_{\mathbf{l}}$$
$$\Phi_3 = \varphi_{\mathbf{h}} + \psi_{\mathbf{k}} + \varphi_{\mathbf{l}} \qquad \Phi_4 = \psi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}}$$
$$\Phi_5 = \varphi_{\mathbf{h}} + \psi_{\mathbf{k}} + \psi_{\mathbf{l}} \qquad \Phi_6 = \psi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \psi_{\mathbf{l}}$$
$$\Phi_7 = \psi_{\mathbf{h}} + \psi_{\mathbf{k}} + \varphi_{\mathbf{l}} \qquad \Phi_8 = \psi_{\mathbf{h}} + \psi_{\mathbf{k}} + \psi_{\mathbf{l}}.$$

So, for the estimation of any Φ_j, the joint probability distribution

$$P(E_{\mathbf{h}}, E_{\mathbf{k}}, E_{\mathbf{l}}, G_{\mathbf{h}}, G_{\mathbf{k}}, G_{\mathbf{l}})$$

has to be studied, from which eight conditional probability densities can be obtained:

$$P(\Phi_i||E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{l}}|, |G_{\mathbf{h}}|, |G_{\mathbf{k}}|, |G_{\mathbf{l}}|)$$
$$\simeq [2\pi I_0(Q_j)]^{-1} \exp[Q_j \cos \Phi_j]$$

for $j = 1, \ldots, 8$.

The analytical expressions of Q_j are too intricate and are not given here (the reader is referred to the original paper). We only say that Q_j may be positive or negative, so that reliable triplet phase estimates near 0 or near π are possible: the larger $|Q_j|$, the more reliable the phase estimate.

A useful interpretation of the formulae in terms of experimental parameters was suggested by Fortier *et al.* (1984): according to them, distributions do not depend, as in the case of the traditional three-phase invariants, on the total number of atoms per unit cell but rather on the scattering difference between the native protein and the derivative (that is, on the scattering of the heavy atoms in the derivative).

Hauptman's formulae were generalized by Giacovazzo *et al.* (1988): the new expressions were able to take into account the resolution effects on distribution parameters. The formulae are completely general and include as special cases native protein and heavy-atom isomorphous derivatives as well as X-ray and neutron diffraction data. Their complicated algebraic forms are easily reduced to a simple expression in the case of a native protein heavy-atom derivative: in particular, the reliability parameter for Φ_1 is

$$Q_1 = 2[\sigma_3/\sigma_2^{3/2}]_P |E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{l}}| + 2[\sigma_3/\sigma_2^{3/2}]_H \Delta_{\mathbf{h}} \Delta_{\mathbf{k}} \Delta_{\mathbf{l}}, \quad (2.2.10.2)$$

where indices P and H warn that parameters have to be calculated over protein atoms and over heavy atoms, respectively, and

$$\Delta = (F_{PH} - F_P)/(\textstyle\sum f_j^2)_H^{1/2}.$$

Δ is a pseudo-normalized difference (with respect to the heavy-atom structure) between moduli of structure factors.

Equation (2.2.10.2) may be compared with Karle's (1983) qualitative rule: if the sign of

$$[(F_{\mathbf{h}})_{PH} - (F_{\mathbf{h}})_P][(F_{\mathbf{k}})_{PH} - (F_{\mathbf{k}})_P][(F_{\mathbf{l}})_{PH} - (F_{\mathbf{l}})_P]$$

is plus then the value of Φ_1 is estimated to be zero; if its sign is minus then the expected value of Φ_1 is close to π. In practice Karle's rule agrees with (2.2.10.2) only if the Cochran-type term in (2.2.10.2) may be neglected. Furthermore, (2.2.10.2) shows that large reliability values do not depend on the triple product of structure-factor differences, but on the triple product of pseudo-normalized differences. A series of papers (Giacovazzo, Siliqi & Ralph, 1994; Giacovazzo, Siliqi & Spagna, 1994; Giacovazzo, Siliqi & Platas, 1995; Giacovazzo, Siliqi & Zanotti, 1995; Giacovazzo *et al.*, 1996) shows how equation (2.2.10.2) may be implemented in a direct procedure which proved to be able to estimate the protein phases correctly without any preliminary information on the heavy-atom substructure.

Combination of direct methods with the two-derivative case is also possible (Fortier *et al.*, 1984) and leads to more accurate estimates of triplet invariants provided experimental data are of sufficient accuracy.

2.2.10.4. *Integration of anomalous-dispersion techniques with direct methods*

If the frequency of the radiation is close to an absorption edge of an atom, then that atom will scatter the X-rays anomalously (see Chapter 2.4) according to $f = f' + if''$. This results in the breakdown of Friedel's law. It was soon realized that the Bijvoet difference could also be used in the determination of phases (Peerdeman & Bijvoet, 1956; Ramachandran & Raman, 1956; Okaya & Pepinsky, 1956). Since then, a great deal of work has been done both from algebraic (see Chapter 2.4) and from probabilistic points of view. In this section we are only interested in the second.

We will mention the following different cases:

(1) The OAS (one-wavelength anomalous scattering) case, also called SAS (single-wavelength anomalous scattering).

(2) The SIRAS (single isomorphous replacement combined with anomalous scattering) case. Typically, native protein and heavy-atom-derivative data are simultaneously available, with heavy atoms as anomalous scatterers.

(3) The MIRAS case, which generalizes the SIRAS case.

(4) The MAD case, a multiple-wavelength technique.

references