Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 19 June 2002
Accepted 9 September 2002

The method of joint probability distribution functions applied to SIR–MIR and to SIRAS–MIRAS cases

Carmelo Giacovazzo[a]* and Dritan Siliqi[a,b]

[a]Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, and [b]Laboratory of X-ray Diffraction, Department of Inorganic Chemistry, Faculty of Natural Sciences, Tirana, Albania. Correspondence e-mail: c.giacovazzo@area.ba.cnr.it

SIR–MIR and SIRAS–MIRAS cases are studied by application of the joint probability distribution method. The final results are conditional probability distributions of the protein phases given the structure-factor moduli of the protein and of the derivatives, and the structure factors of the heavy-atom substructures. The approach is able to treat errors arising from measurements, from the heavy-atom structure model and from the lack of isomorphism. The relations between the present approach and previous methods are described. The formulas have been implemented in a procedure that is able to automatically phase protein reflections up to protein resolution [Giacovazzo, Ladisa & Siliqi (2002). *Acta Cryst.* A**58**, 598–604].

1. Notation

f_j : scattering factor of the jth atom.
Σ_p, Σ_d, $\Sigma_H = \sum f_j^2$, where the summation is extended to the protein atoms, to the derivative, and to the heavy-atom structure.
$F_p = \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) = |F_p| \exp(i\phi_p)$: structure factor of the native protein.
$E_p = A_p + iB_p = R_p \exp(i\phi_p) = F_p/\Sigma_p^{1/2}$: normalized structure factor of the native protein.
$F_H = \sum f_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) = |F_H| \exp(i\phi_H)$: structure factor of the heavy-atom structure.
$E_H = A_H + iB_H = R_H \exp(i\phi_H) = F_H/\Sigma_p^{1/2}$: E_H is the structure factor of the heavy-atom structure, pseudo-normalized with respect to the protein scattering power.
$F_d = \sum g_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) = |F_d| \exp(i\phi_d)$: structure factor of the derivative.
$E_d = A_d + iB_d = R_d \exp(i\phi_d) = F_d/\Sigma_p^{1/2}$: E_d is the structure factor of the derivative, pseudo-normalized with respect to the protein scattering power.
$\Delta_{\rm iso} = (|F_d| - |F_p|)$.

2. Introduction

The isomorphous replacement method is essentially based on two steps: first, the determination of the positions of the heavy atoms, and then the estimation of the native protein phases *via* the combined use of F_H, $|F_p|$, $|F_d|$. The fundamental paper by Blow & Crick (1959) has been the milestone for the second step: the paper has been revisited by several authors, among whom are Terwilliger & Eisenberg (1987), who give a more detailed analysis of the errors due to lack of isomorphism, inadequacy of the heavy-atom model and observational errors. In particular, functions were obtained that estimate the conditional probability for the native protein phases given the prior information on F_H, $|F_p|$, $|F_d|$.

The irruption of direct methods in the macromolecular area (Hauptman, 1982) suggested that the classical two-step technique used by SIR and MIR methods could be replaced by a one-step approach: the protein phases could be directly obtained by application of the joint probability distributions

$$P(E_p, E_d) \tag{1}$$

$$P(E_{p\mathbf{h}}, E_{p\mathbf{k}}, E_{p\mathbf{h+k}}, E_{d\mathbf{h}}, E_{d\mathbf{k}}, E_{d\mathbf{h+k}}). \tag{2}$$

Accordingly, the recovery of the heavy-atom positions was no longer a necessary preliminary step for the phase assignment. The distributions (1) and (2) were obtained under the following assumptions: the scattering power of the heavy-atom structure may be (at least roughly) estimated, and no lack of isomorphism occurs. Under the same assumptions, a recent series of papers (see Giacovazzo & Siliqi, 1997, and references therein) made the direct-methods treatment of the SIR case practicable in real cases.

In this paper, we will apply the joint probability distribution method to derive the phase distribution functions useful for SIR, MIR, SIRAS and MIRAS cases. The method has already been applied: (*a*) by Giacovazzo & Siliqi (2001*a*,*b*,*c*) to treat the SAD (single anomalous dispersion) and the MAD (multiple anomalous dispersion) cases. New efficient probabilistic formulas were proposed to estimate protein phases given the anomalous scatterer positions; (*b*) by Burla *et al.* (2002), for finding the anomalous scatterer positions given the experimental diffraction moduli.

In this paper, we will adopt, for the SIR and MIR cases (§3 of this paper), the probabilistic scenario described by Giacovazzo & Siliqi (2001*a*,*b*,*c*), according to which

$$|F_d|\exp(i\phi_d) = |F_p|\exp(i\phi_p) + |F_H|\exp(i\phi_H) + |\mu|\exp(i\theta), \qquad (3)$$

where $|\mu|\exp(i\theta)$ represents the cumulative error, the components of which are errors due to lack of isomorphism, errors in measurements and errors in the heavy-atom substructure (Terwilliger & Eisenberg, 1983).

For the SIRAS and MIRAS cases (§4 of this paper), we will assume that

$$|F_d^+|\exp(i\phi_d^+) = |F_p|\exp(i\phi_p) + |F_H^+|\exp(i\phi_H) + |\mu^+|\exp(i\theta^+) \qquad (4)$$

and

$$|F_d^-|\exp(i\phi_d^-) = |F_p|\exp(-i\phi_p) + |F_H^-|\exp(i\phi_H^-) + |\mu^-|\exp(i\theta^-). \qquad (5)$$

3. SIR and MIR cases

3.1. The SIR case in $P\bar{1}$. The joint probability distribution $P(E_p, E_d|E_H)$ and related distributions

Let us assume that

(*a*) the atomic positions of the native protein are the primitive random variables of our probabilistic approach;

(*b*) some (or all) heavy atoms have been located, and F_H is the structure factor corresponding to them;

(*c*) the assumptions (4) hold, with $\langle\mu\rangle = 0$.

The characteristic function of the distribution $P(E_p, E_d|E_H)$ may be written as (see Appendix *A*)

$$C(u_p, u_d) = \langle\exp i(u_pE_p + u_dE_d)\rangle \approx \exp(iu_dE_H)\exp\{-[u_p^2 + u_d^2(1+\sigma^2) + 2u_pu_d]/2\}, \qquad (6)$$

where u_p and u_d are carrying variables associated with E_p and E_d, respectively,

$$\sigma^2 = |\mu|^2/\Sigma_p.$$

The Fourier transform of (6) leads to

$$P(E_p, E_d|E_H) \approx (2\pi)^{-1}\sigma^{-1}\exp\{-(1/2\sigma^2)[(E_d - E_H)^2 + (1+\sigma^2)E_p^2 - 2E_p(E_d - E_H)]\}. \qquad (7)$$

Equation (7) is basic for all the conditional distributions. From (7), we first derive the marginal distribution $P(E_p, R_d|E_H)$ and then the conditional $P(E_p|R_d, E_H)$:

$$P(E_p|R_d, E_H) \approx L\exp\{-[R_d^2 + (E_p + E_H)^2 + \sigma^2R_p^2]/(2\sigma^2)\} \times \cosh[R_d(E_H + E_p)/\sigma^2], \qquad (8)$$

where L is a suitable scaling factor. Then the probability that the sign s_p of E_p is ±1 is given by

$$P(s_p = +1|\ldots) \approx \exp(-R_pE_H/\sigma^2)\cosh[R_d(E_H + R_p)/\sigma^2],$$
$$P(s_p = -1|\ldots) \approx \exp(R_pE_H/\sigma^2)\cosh[R_d(E_H - R_p)/\sigma^2],$$

from which

$$P(s_p = 1|\ldots)/P(s_p = -1|\ldots) \approx \exp(-2R_pE_H/\sigma^2)\frac{\cosh[R_d(E_H + R_p)/\sigma^2]}{\cosh[R_d(E_H - R_p)/\sigma^2]}. \qquad (9)$$

Expression (9) does not coincide with formula (16) in Blow & Crick (1959), where sinh replaces our cosh function: the two expressions converge only when the arguments of cosh are sufficiently large.

The normalization of the sign probabilities [*i.e.* by imposing $P(s_p = +1|\ldots) + P(s_p = -1|\ldots) = 1$] leads to

$$P(s_p = +1|\ldots) \approx \left\{1 + \exp(2R_pE_H/\sigma^2)\frac{\cosh[R_d(E_H - R_p)/\sigma^2]}{\cosh[R_d(E_H + R_p)/\sigma^2]}\right\}^{-1}. \qquad (10)$$

In terms of structure factors, the probability function (10) may be written as

$$P(s_p = +1|\ldots) \approx \left\{1 + \exp(2|F_p|F_H/\langle|\mu|^2\rangle)\frac{\cosh[|F_d|(F_H - |F_p|)/\langle|\mu|^2\rangle]}{\cosh[|F_d|(F_H + |F_p|)/\langle|\mu|^2\rangle]}\right\}^{-1}. \qquad (11)$$

If the arguments of the cosh functions are large enough, then cosh x may be approximated by 0.5 exp$|x|$, and the probability (10) may be replaced by two very simple expressions:

$$P(\phi_p = \phi_H|\ldots) \approx 0.5 + 0.5\tanh\{R_H(R_d - R_p)/\sigma^2\} \quad \text{if} \quad R_p > R_H, \qquad (12a)$$
$$P(\phi_p = \phi_H|\ldots) \approx 0.5 + 0.5\tanh\{R_p(R_d - R_H)/\sigma^2\} \quad \text{if} \quad R_p < R_H. \qquad (12b)$$

The probability (12*b*) takes care of the cases in which 'overcross' occurs.

Let us now show that (8) encompasses the Blow & Crick (1959) distribution. On assuming that the approximation $\cosh x \approx 0.5\exp|x|$ is valid, (12*b*) transforms into

$$P(E_p|R_d, E_H) \approx L\exp(-R_p^2/2)\exp\{-(R_d - R_{d\text{calc}})^2/(2\sigma^2)\},$$

where $E_{d\text{calc}} = E_p + E_H$. In terms of structure factors,

$$P(\phi_p\big||F_d|, F_H) \approx L\exp\{-(|F_d| - |F_{d\text{calc}}|)^2/(2\langle|\mu|^2\rangle)\},$$

which agrees with the Blow & Crick (1959) formula.

3.2. The SIR case in $P1$. The joint probability distribution $P(E_p, E_d|E_H)$ and related distributions

Under the same assumptions specified for the centric case, the characteristic function of the distribution $P(E_p, E_d|E_H)$ in $P1$ may be written as

$$C(u_p, v_p, u_d, v_d) = \langle \exp i(u_pA_p + v_pB_p + u_dA_d + v_dB_d)\rangle = \exp\{i(u_dA_H + v_dB_H)\}\exp\{-\tfrac{1}{4}[u_p^2 + v_p^2 + (1+|\sigma^2|)(u_d^2+v_d^2) + 2u_pu_d + 2v_pv_d]\},$$

where u_p, v_p, u_d, v_d are carrying variables associated with A_p, B_p, A_d, B_d, respectively. The change of variables

$$\begin{cases} u_p = \rho_p \cos\psi_p \\ v_p = \rho_p \sin\psi_p \end{cases} \begin{cases} u_d = \rho_d \cos\psi_d \\ v_d = \rho_d \sin\psi_d \end{cases}, \begin{cases} A_p = R_p \cos\phi_p \\ B_p = R_p \sin\phi_p \end{cases}, \begin{cases} A_d = R_d \cos\phi_d \\ B_d = R_d \sin\phi_d \end{cases}, \begin{cases} A_H = R_H \cos\phi_H \\ B_H = R_H \sin\phi_H \end{cases}$$

leads to

$$C(\rho_p, \rho_d, \psi_p, \psi_d)\,d\rho_p\,d\rho_d\,d\psi_p\,d\psi_d = \rho_p\rho_d \exp[iR_H\rho_d\cos(\psi_d - \phi_H)] \times \exp\{-\tfrac{1}{4}[\rho_p^2 + (1+\sigma^2)\rho_d^2 + 2\rho_p\rho_d\cos(\psi_d - \psi_p)]\}.$$

The Fourier inversion of the characteristic function gives

$$P(R_p, R_d, \phi_p, \phi_d|E_H) \approx \pi^{-2}(R_pR_d/|\sigma|^2)\exp\big((-1/\sigma^2)\{[R_d^2 + R_H^2 - 2R_dR_H\cos(\phi_d - \phi_H)] + (1+\sigma^2)R_p^2 - 2R_pR_d\cos(\phi_d - \phi_p) + 2R_pR_H\cos(\phi_p - \phi_H)\}\big), \quad (13)$$

which is the required joint probability distribution function. Then the marginal distribution

$$P(R_p, R_d, \phi_p|E_H) \approx (\pi\sigma^2)^{-1}2R_pR_d\exp\{(-1/\sigma^2)(R_d^2 + R_{d\rm calc}^2 + \sigma^2R_p^2)\}I_0(z) \quad (14)$$

is obtained, where I_0 is the modified Bessel function of order zero,

$$z = 2R_dR_{d\rm calc}/\sigma^2, \quad (15)$$
$$R_{d\rm calc} = [R_p^2 + R_H^2 + 2R_pR_H\cos(\phi_p - \phi_H)]^{1/2}.$$

Then a very simple phase probability is obtained,

$$P(\phi_p|R_p, R_d, E_H) \approx L\exp\left\{\frac{-2R_pR_H}{\sigma^2}\cos(\phi_p - \phi_H)\right\}I_0(z), \quad (16)$$

where L is a suitable normalizing factor that may be calculated *via* numerical methods.

In terms of structure factors, the distribution (16) becomes

$$P(\phi_p\big||F_p|, |F_d|, F_H) \approx L\exp\left\{\frac{-2|F_pF_H|}{|\mu|^2}\cos(\phi_p - \phi_H)\right\}I_0(z), \quad (17)$$

where

$$z = 2|F_dF_{d\rm calc}|/|\mu|^2 \quad (18)$$

and

$$|F_{d\rm calc}| = [|F_p|^2 + |F_H|^2 + 2|F_pF_H|\cos(\phi_p - \phi_H)]^{1/2}.$$

Equation (17) is the required conditional phase distribution and constitutes one of the main results of this paper.

3.3. About the conditional probability distribution $P(\phi_p||F_p|, |F_d|, F_H)$

The distribution (17) deserves to be discussed with regard to:

(*a*) its applicative aspects;

(*b*) its relation with classical Blow & Crick (1959) and Terwilliger & Eisenberg (1987) distributions.

For the point (*a*), we note that, since $\sigma^2 \ll 1$, z is usually a quite large number for any value of ϕ_p. For $|z|$ sufficiently large, the following approximation (Abramowitz & Stegun, 1972) may be used:

$$I_0(x) = \exp|x|/(2\pi|x|)^{1/2}. \quad (19)$$

Then numerical techniques (*i.e.* by calculating P in stepped ϕ_p values between 0 and 2π) can be applied to derive the best phase estimate and the relative variance. The simpler approximation

$$I_0(z) \approx \exp(z^2/4)$$

is discouraged because z is usually quite a large number for any value of ϕ_p.

The favorable results obtained by us (Giacovazzo & Siliqi, 2001*a*, *b*,*c*) in the MAD case suggest a simplification in the use of (17) by introducing in (13) the approximation

$$\phi_d \approx \phi_p.$$

Then the simple conditional distribution

$$P(\phi_p|R_p, R_d, E_H) \approx [2\pi I_o(G)]^{-1}\exp[G\cos(\phi_p - \phi_H)] \quad (20)$$

is derived, where

$$G = 2(R_d - R_p)R_H/\sigma^2 = 2\Delta_{\rm iso}|F_H|/|\mu|^2. \quad (21)$$

According to (20), ϕ_H is the expected value of ϕ_p if $\Delta_{\rm iso} > 0$, otherwise ϕ_p is expected to be about $\phi_H + \pi$. The larger the product $|\Delta_{\rm iso}R_H|/\sigma^2$, the more accurate the expectation will be. Equation (21) will be the formula we will use for the SIR case in our practical applications.

To answer the point (*b*), we show that expression (14) encompasses a Blow & Crick (1959) distribution. This may be obtained from (14) *via* two approximations: first (14) is factorized as

$$P(R_p, R_d, \phi_p|E_H) \approx [R_p/(\pi\sigma)]\exp\{-R_p^2\}2R_d\sigma^{-1} \times \exp\{-(R_d^2 + R_{d\rm calc}^2)/\sigma^2\}I_0(2R_dR_{d\rm calc}/\sigma^2)$$

and then (19) is applied. We obtain

$$P(R_p, R_d, \phi_p|E_H) \approx [R_p]\exp\{-R_p^2\}(\pi)^{-3/2}(R_d/R_{d\rm calc})^{1/2} \times \exp\{-(R_d - R_{d\rm calc})^2/\sigma^2\}.$$

Then

$$P(\phi_p|R_p, R_d, E_H) \approx L(R_d/R_{d\rm calc})^{1/2}\exp\{-(R_d - R_{d\rm calc})^2/\sigma^2\}, \quad (22)$$

which, in terms of structure factors, transforms into

$$P(\phi_p \big| |F_p|, |F_d|, F_H) \approx L(|F_d|/|F_{d\text{calc}}|)^{1/2} \exp\{-(|F_d| - |F_{d\text{calc}}|)^2/|\mu|^2\}. \quad (23)$$

Distribution (23) differs from the distribution obtained by Blow & Crick (1959) owing to the presence of the factor $(R_d/R_{d\text{calc}})^{1/2}$. The effect of this factor on the distribution (23) is negligible in practice since it is combined with the more rapidly varying exponential function, unless $|F_p|$ is very small. This last condition reduces the practical impact of (23); however, the above calculations show that the Blow & Crick (1959) probability function constitutes an approximation of the distribution (17) provided by the more rigorous method of the joint probability distribution functions.

3.4. The MIR case. The probability distribution $P(E_p, \mathbf{E}_d|\mathbf{E}_H)$ in $P\bar{1}$

Suppose that:

(*a*) the diffraction data of *n* derivatives have been collected; accordingly,

$$\mathbf{E}_d = \{E_{d1}, E_{d2}, \ldots, E_{dn}\};$$

(*b*) the following relation may be established for each *i*th derivative:

$$|F_{di}| \exp(i\phi_{di}) = |F_{pi}| \exp(i\phi_{pi}) + |F_{Hi}| \exp(i\phi_{Hi}) + |\mu_i| \exp(i\vartheta_i);$$

(*c*) μ_i is uncorrelated with μ_j for $i \neq j$, and $\langle\mu_i\rangle = 0$ for any *i*.

Then the characteristic function of the distribution

$$P(E_p, \mathbf{E}_d|\mathbf{E}_H) \equiv P(E_p, E_{d1}, \ldots, E_{dn}|E_{H1}, \ldots, E_{Hn}) \quad (24)$$

is

$$C(u_p, u_{d1}, \ldots, u_{dn}) = \exp\left\{i\left(\sum_{j=1}^{n} u_{dj}E_{Hj}\right) - \tfrac{1}{2}u_p^2 - \tfrac{1}{2}\sum_{j=1}^{n} e_j u_{dj}^2 - u_p\sum_{j=1}^{n} u_{dj} - \sum_{j,q=1}^{n} u_{dj}u_{dq}\right\},$$

where u_p and u_{dj} are carrying variables associated with E_p and E_{dj}, respectively,

$$e_j = 1 + \sigma_j^2, \quad \sigma_j^2 = |\mu_j|^2/\Sigma_p.$$

The joint probability distribution (24) is then

$$P(E_p, \mathbf{E}_{dj}|\mathbf{E}_{Hj}) = (2\pi)^{-(n+1)} \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \exp\left\{-i\left[u_pE_p + \sum_{j=1}^{n} u_{dj}(E_{dj} - E_{Hj})\right] - \tfrac{1}{2}\left(u_p^2 + \sum_{j=1}^{n} e_j u_{dj}^2 + 2u_p\sum_{j=1}^{n} u_{dj} + 2\sum_{\substack{j,q=1\\ j<q}}^{n} u_{dj}u_{dq}\right)\right\}. \quad (25)$$

In a shorter form, (25) may be rewritten as

$$P(E_p, \mathbf{E}_d|\mathbf{E}_H) \approx (2\pi)^{-(n+1)} \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \exp\{-i\overline{\mathbf{T}}\mathbf{U} - \tfrac{1}{2}\overline{\mathbf{U}}\mathbf{K}\mathbf{U}\}\, \mathrm{d}\mathbf{U},$$

where

$$\overline{\mathbf{T}} = [E_p, (E_{d1} - E_{H1}), \ldots, (E_{dn} - E_{Hn})],$$

$$\overline{\mathbf{U}} = [u_p, u_{d1}, \ldots, u_{dn}],$$

$$\mathbf{K} = \begin{vmatrix} 1 & 1 & \ldots & 1 \\ 1 & e_1 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & \ldots & e_n \end{vmatrix},$$

$$\det \mathbf{K} = \prod_{j=1}^{n} \sigma_j^2.$$

Then

$$P(E_p, \mathbf{E}_d|\mathbf{E}_H) \approx (2\pi)^{-(n+1)/2}(\det \mathbf{K})^{-1/2} \exp\left\{-\tfrac{1}{2}\lambda_{11}E_p^2 - \tfrac{1}{2}\sum_{j=1}^{n}\lambda_{j+1,j+1}(E_{dj} - E_{Hj})^2 - E_p\sum_{j=1}^{n}\lambda_{1,j+1}(E_{dj} - E_{Hj}) - \sum_{\substack{j,q=1\\ j<q}}^{n}\lambda_{j+1,q+1}(E_{dj} - E_{Hj})(E_{dq} - E_{Hq})\right\}, \quad (26)$$

where λ_{jq} are the elements of the matrix $\mathbf{K}^{-1}$.

We note:

$$\lambda_{11} = 1 + \sum_{j=1}^{n}(1/\sigma_j^2),$$

$$\lambda_{1,j+1} = -(1/\sigma_j^2) \quad \text{if} \quad n > 1,$$

$$\lambda_{j+1,j+1} = 1/\sigma_j^2.$$

Accordingly, the following relations hold:

$$\lambda_{11} = 1 - \sum_{j=1}^{n}\lambda_{1,j+1}, \quad \lambda_{j+1,j+1} = -\lambda_{1,j+1}. \quad (27)$$

Putting (27) into (26) gives

$$P(E_p, \mathbf{E}_d|\mathbf{E}_H) \approx (2\pi)^{-(n+1)/2}(\det \mathbf{K})^{-1/2} \exp\left\{-\tfrac{1}{2}E_p^2 + \tfrac{1}{2}\sum_{j=1}^{n}\lambda_{1,j+1}[E_p - (E_{dj} - E_{Hj})]^2 - \sum_{\substack{j,q=1\\ j<q}}^{n}\lambda_{j+1,q+1}(E_{dj} - E_{Hj})(E_{dq} - E_{Hq})\right\}. \quad (28)$$

The conditional sign probability $P(s_p \big| |\mathbf{E}_d|, \mathbf{E}_H)$ may be obtained (s_p is the sign of $\mathbf{E}_p$) by first calculating the joint probability

$$\sum_{s_{dj}=\pm 1} P(E_p, s_{d1}|E_{d1}|, \ldots, s_{dn}|E_{dn}|\big|\mathbf{E}_H)$$

and then by deriving the marginal distribution. The conclusive formula is rather intricate. We prefer to introduce in (26) the approximation

$$\phi_{dj} = \phi_p \quad \text{for} \quad j = 1, \ldots n.$$

Then the marginal distribution

$$P(E_p|\mathbf{E}_H) \approx L \exp\left\{-\tfrac{1}{2}E_p^2 + \tfrac{1}{2}\sum_{j=1}^{n} \lambda_{1,j+1}[s_p \Delta^n_{\mathrm{iso}j} - E_{Hj}]^2\right\} \quad (29)$$

is obtained, where $\Delta^n_{\mathrm{iso}j} = (R_{dj} - R_p)$. The distribution (29) integrates the Wilson component [say $\exp(-E_p^2/2)$] with the contribution provided by the prior knowledge of the $\Delta^n_{\mathrm{iso}j}$. Now

$$\begin{aligned} P(s_p = s_H) &= 0.5 + 0.5 \tanh\left\{-\sum_j \lambda_{1,j+1} R_{Hj} \Delta^n_{\mathrm{iso}j}\right\} \\ &= 0.5 + 0.5 \tanh\left\{\frac{\sum_j R_{Hj}\Delta^n_{\mathrm{iso}j}}{\sigma_j^2}\right\} \\ &= 0.5 + 0.5 \tanh\left\{\frac{\sum_j |F_{Hj}|\Delta_{\mathrm{iso}j}}{\mu_j^2}\right\}. \quad (30) \end{aligned}$$

Equation (30) is our conclusive formula for the centric case.

3.5. The MIR case. The probability distribution $P(E_p, \mathbf{E}_d\,|\,\mathbf{E}_H)$ in $P1$

The characteristic function of the distribution

$$P(A_p, A_{d1}, \ldots, A_{dn}, B_{d1}, \ldots, B_{dn}|E_{H1}, \ldots, E_{Hn}) \quad (31)$$

is

$$\begin{aligned} &C(u_p, u_{d1}, \ldots, u_{dn}, \mathrm{v}_p, \mathrm{v}_{d1}, \ldots, \mathrm{v}_{dn}) \\ &\approx \exp\left\{i\left(\sum_{j=1}^{n} u_{dj}A_{Hj} + \mathrm{v}_{dj}B_{Hj}\right) - \tfrac{1}{4}u_p^2 \right. \\ &\quad - \tfrac{1}{4}\sum_{j=1}^{n} e_j(u_{dj}^2 + \mathrm{v}_{dj}^2) - \tfrac{1}{2}u_p\sum_{j=1}^{n} u_{dj} - \tfrac{1}{2}\mathrm{v}_p\sum_{j=1}^{n}\mathrm{v}_{dj} \\ &\quad \left. - \tfrac{1}{2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}(u_{di}u_{dj} + \mathrm{v}_{di}\mathrm{v}_{dj})\right\}. \quad (32) \end{aligned}$$

The joint probability distribution function (31) may be obtained by Fourier inversion of (32): we have

$$\begin{aligned} &P(E_p, \mathbf{E}_d|\mathbf{E}_H) \\ &\approx (2\pi)^{-2(n+1)}2^{(n+1)}\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty}\exp\{-i\overline{\mathbf{T}}\mathbf{U} - \tfrac{1}{2}\overline{\mathbf{U}}\mathbf{K}\mathbf{U}\}\,\mathrm{d}\overline{\mathbf{U}} \\ &\approx \pi^{-(n+1)}(\det\mathbf{K})^{-1/2}\exp\{-\tfrac{1}{2}\overline{\mathbf{T}}\mathbf{K}^{-1}\mathbf{T}\}, \quad (33) \end{aligned}$$

where

$$\begin{aligned} \overline{\mathbf{T}} &= [2^{1/2}A_p, 2^{1/2}(A_{d1} - A_{H1}), \ldots, 2^{1/2}(A_{dn} - A_{Hn}), \\ &\quad 2^{1/2}B_p, 2^{1/2}(B_{d1} - A_{H1}), \ldots, 2^{1/2}(B_{dn} - B_{Hn})], \\ \overline{\mathbf{U}} &= [u_p, u_{d1}, \ldots, u_{dn}, v_p, v_{d1}, \ldots, \mathrm{v}_{dn}] \\ \mathbf{K} &= \begin{vmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{vmatrix}, \\ \mathbf{Q} &= \begin{vmatrix} 1 & 1 & \ldots & 1 \\ 1 & e_1 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & \ldots & e_n \end{vmatrix}, \\ \det\mathbf{K} &= \left[\prod_{j=1}^{n}\sigma_j^2\right]^2. \end{aligned}$$

We have

$$\begin{aligned} P(E_p, \mathbf{E}_d|\mathbf{E}_H) &\approx \pi^{-(n+1)}(\det\mathbf{K})^{-1/2}\exp\left\{-\lambda_{11}(A_p^2 + B_p^2)\right. \\ &\quad - \sum_{j=1}^{n}\lambda_{j+1,j+1}[(A_{dj} - A_{Hj})^2 + (B_{dj} - B_{Hj})^2] \\ &\quad - \sum_{j=1}^{n}\lambda_{1,j+1}[A_p(A_{dj} - A_{Hj}) + B_p(B_{dj} - B_{Hj})] \\ &\quad - \sum_{\substack{j,q=1\\ j<q}}^{n}\lambda_{j+1,q+1}[(A_{dj} - A_{Hj})(A_{dq} - A_{Hq}) \\ &\quad \left. + (B_{dj} - B_{Hj})(B_{dq} - B_{Hq})]\right\}, \quad (34) \end{aligned}$$

where λ_{jq} are the elements of the matrix $\mathbf{K}^{-1}$. Since

$$\mathbf{K}^{-1} = \begin{vmatrix} \mathbf{Q}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{-1} \end{vmatrix},$$

the expressions for the λ_{ij} stated in $P\bar{1}$ hold also for $P1$. We can then rewrite (34) in the form

$$\begin{aligned} P(E_p, \mathbf{E}_d|\mathbf{E}_H) &\approx \pi^{-(n+1)}(\det\mathbf{K})^{-1/2}\exp\left\{-(A_p^2 + B_p^2)\right. \\ &\quad + \sum_{j=1}^{n}\lambda_{1,j+1}|E_p - (E_{dj} - E_{Hj})|^2 \\ &\quad - \sum_{\substack{j,q=1\\ j<q}}^{n}\lambda_{j+1,q+1}[(A_{dj} - A_{Hj})(A_{dq} - A_{Hq}) \\ &\quad \left. + (B_{dj} - B_{Hj})(B_{dq} - B_{Hq})]\right\}. \quad (35) \end{aligned}$$

A mathematical procedure similar to that used in $P\bar{1}$ (*i.e.* $\phi_d \approx \phi_p$) leads to

$$P(\phi_p|R_p, \mathbf{R}_d, \mathbf{E}_H) \approx [2\pi I_o(G)]^{-1}\exp[\alpha_p\cos(\phi_p - \vartheta_p)], \quad (36)$$

where

$$\tan\vartheta_p = \frac{\sum_{j=1}^{n} G_j \sin\phi_{Hj}}{\sum_{j=1}^{n} G_j\cos\phi_{Hj}} = \frac{T}{B}, \quad (37)$$

$$G_j = 2R_{Hj}\Delta^n_{\mathrm{iso}j}/\sigma_j^2 = 2|F_{Hj}|\Delta_{\mathrm{iso}j}/|\mu_j|^2, \quad (38)$$

$$\alpha_p = (T^2 + B^2)^{1/2}. \quad (39)$$

ϑ_p is the most probable value of ϕ_p and α_p its reliability parameter. The relations (36)–(39) will be our tools for the experimental applications in the next paper (Giacovazzo *et al.*, 2002).

4. SIRAS and MIRAS cases

4.1. The SIRAS case in $P1$

The following additional notation will be used to treat the SIRAS–MIRAS cases:

$$
\begin{aligned}
f_j &= f_j^0 + \Delta f_j + if_j'' = f_j' + if_j'',\\
F^+ &= |F^+| \exp(i\phi^+) = F_{\mathbf{h}},\\
F^- &= |F^-| \exp(i\phi^-) = F_{-\mathbf{h}},\\
E_d^+ &= A_d^+ + iB_d^+,\\
E_d^- &= A_d^- + iB_d^-.
\end{aligned}
$$

E_d^+ and E_d^- are pseudo-normalized structure factors (*i.e.* normalized with respect to the native protein). Let us study the conditional probability distribution

$$P(A_p, A_d^+, A_d^-, B_p, B_d^+, B_d^- | A_H^+, A_H^-, B_H^+, B_H^-),$$

in short P, under the following mathematical model:

$$
\begin{aligned}
A_d^+ &= A_p + A_H^+ + |\sigma^+| \cos\vartheta^+\\
&= \left[\sum_{j=1}^{N} f_j \cos 2\pi\mathbf{h}\cdot\mathbf{r}_j + \sum_{j=1}^{H}(f_j' \cos 2\pi\mathbf{h}\cdot\mathbf{r}_j - f_j'' \sin 2\pi\mathbf{h}\cdot\mathbf{r}_j)\right.\\
&\quad \left. + |\mu^+| \cos\vartheta^+\right] \Big/ \Sigma_p^{1/2},\\
B_d^+ &= B_p + B_H^+ + |\sigma^+| \sin\vartheta^+\\
&= \left[\sum_{j=1}^{N} f_j \sin 2\pi\mathbf{h}\cdot\mathbf{r}_j + \sum_{j=1}^{H}(f_j' \sin 2\pi\mathbf{h}\cdot\mathbf{r}_j + f_j'' \cos 2\pi\mathbf{h}\cdot\mathbf{r}_j)\right.\\
&\quad \left. + |\mu^+| \sin\vartheta^+\right] \Big/ \Sigma_p^{1/2},\\
A_d^- &= A_p + A_H^- + |\sigma^-| \cos\vartheta^-\\
&= \left[\sum_{j=1}^{N} f_j \cos 2\pi\mathbf{h}\cdot\mathbf{r}_j + \sum_{j=1}^{H}(f_j' \cos 2\pi\mathbf{h}\cdot\mathbf{r}_j + f_j'' \sin 2\pi\mathbf{h}\cdot\mathbf{r}_j)\right.\\
&\quad \left. + |\mu^-| \cos\vartheta^-\right] \Big/ \Sigma_p^{1/2},\\
B_d^- &= -B_p + B_H^- + |\sigma^-| \sin\vartheta^-\\
&= \left[-\sum_{j=1}^{N} f_j \sin 2\pi\mathbf{h}\cdot\mathbf{r}_j + \sum_{j=1}^{H}(-f_j' \sin 2\pi\mathbf{h}\cdot\mathbf{r}_j + f_j'' \cos 2\pi\mathbf{h}\cdot\mathbf{r}_j)\right.\\
&\quad \left. + |\mu^-| \sin\vartheta^-\right] \Big/ \Sigma_p^{1/2},
\end{aligned}
$$

where $\sigma^{\pm} = \mu^{\pm}/\Sigma_p^{1/2}$. The characteristic function of the distribution P is given by

$$
\begin{aligned}
&C(u_p, u_d^+, u_d^-, \mathrm{v}_p, \mathrm{v}_d^+, \mathrm{v}_d^-)\\
&\quad \approx \exp i(u_d^+ A_H^+ + u_d^- A_H^- + \mathrm{v}_d^+ B_H^+ + \mathrm{v}_d^- B_H^-)\\
&\qquad \times \exp\{-\tfrac{1}{4}[(u_p^2 + \mathrm{v}_p^2) + e_d^+(u_d^{+2} + \mathrm{v}_d^{+2}) + e_d^-(u_d^{-2} + \mathrm{v}_d^{-2})\\
&\qquad + 2u_p(u_d^+ + u_d^-) + 2\mathrm{v}_p(\mathrm{v}_d^+ - \mathrm{v}_d^-) + 2(u_d^+ u_d^- - \mathrm{v}_d^+ \mathrm{v}_d^-)]\},
\end{aligned}
$$

where $e^{\pm} = 1 + |\sigma^{\pm}|^2$. The Fourier transform of C gives

$$P = \pi^{-3}(e^+e^-)^{-1}(\det\mathbf{K})^{-1/2} \exp\{-\tfrac{1}{2}(\overline{\mathbf{E}}'\mathbf{K}^{-1}\mathbf{E}')\}, \qquad (40)$$

where

$$
\mathbf{K} = \begin{vmatrix}
1 & (e^+)^{-1/2} & (e^-)^{-1/2} & 0 & 0 & 0\\
(e^+)^{-1/2} & 1 & (e^+e^-)^{-1/2} & 0 & 0 & 0\\
(e^-)^{-1/2} & (e^+e^-)^{-1/2} & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 1 & (e^+)^{-1/2} & (-e^-)^{-1/2}\\
0 & 0 & 0 & (e^+)^{-1/2} & 1 & -(e^+e^-)^{-1/2}\\
0 & 0 & 0 & -(e^-)^{-1/2} & -(e^+e^-)^{-1/2} & 1
\end{vmatrix},
$$

$$
\begin{aligned}
\overline{E}' = [&A_p 2^{1/2}, (A_d^+ - A_H^+)(2/e^+)^{1/2}, (A_d^- - A_H^-)(2/e^-)^{1/2},\\
&B_p 2^{1/2}, (B_d^+ - B_H^+)(2/e^+)^{1/2}, (B_d^- - B_H^-)(2/e^-)^{1/2}].
\end{aligned}
$$

After some calculations, (40) reduces to

$$
\begin{aligned}
P \approx\ & \pi^{-3}(e^+e^-)^{-1}(\det\mathbf{K})^{-1/2} \exp\{-[1 + (|\sigma^+|)^{-2} + (|\sigma^-|)^{-2}]\\
&\times (A_p^2 + B_p^2) - (|\sigma^+|)^{-2}[(A_d^+ - A_H^+)^2 + (B_d^+ - B_H^+)^2]\\
&- (|\sigma^-|)^{-2}[(A_d^- - A_H^-)^2 + (B_d^- - B_H^-)^2]\\
&+ 2(|\sigma^+|)^{-2}[A_p(A_d^+ - A_H^+) + B_p(B_d^+ - B_H^+)]\\
&+ 2(|\sigma^-|)^{-2}[A_p(A_d^- - A_H^-) - B_p(B_d^- - B_H^-)]\}. \qquad (41)
\end{aligned}
$$

Equation (41) may be rewritten in a simpler form:

$$
\begin{aligned}
P \approx\ & \pi^{-3}(e^+e^-)^{-1}(\det\mathbf{K})^{-1/2} \exp\{-|E_p|^2 - (|\sigma^+|)^{-2}\\
&\times |E_d^+ - (E_p + E_H^+)|^2 + (|\sigma^-|)^{-2}|E_d^- - (\overline{E}_p + E_H^-)|^2\}. \qquad (42)
\end{aligned}
$$

The change of variables

$$
\begin{aligned}
A_p &= R_p \cos\phi_p, \quad A_d^+ = R_d^+ \cos\phi_d^+ \quad A_d^- = R_d^- \cos\phi_d^-\\
B_p &= R_p \sin\phi_p, \quad B_d^+ = R_d^+ \sin\phi_d^+, \quad B_d^- = R_d^- \sin\phi_d^-
\end{aligned}
$$

reduces equation (41) to

$$
\begin{aligned}
P \approx\ & \pi^{-3}(|\sigma^+\sigma^-|)^{-2} R_p R_d^+ R_d^- \exp\{-R_p^2 - (|\sigma^+|)^{-2}\\
&\times [R_p^2 + R_d^{+2} + R_H^{+2} - 2R_d^+ R_H^+ \cos(\phi_d^+ - \phi_H^+)\\
&+ 2R_p R_d^+ \cos(\phi_d^+ - \phi_p) - 2R_p R_H^+ \cos(\phi_p^+ - \phi_H^+)]\\
&- (|\sigma^-|)^{-2}[R_p^2 + R_d^{-2} + R_H^{-2} - 2R_d^- R_H^- \cos(\phi_d^- - \phi_H^-)\\
&+ 2R_p R_d^- \cos(\phi_d^- + \phi_p) - 2R_p R_H^- \cos(\phi_p^+ + \phi_H^-)]. \qquad (43)
\end{aligned}
$$

The use (see §3.5) of the phase relations

$$\phi_d^+ \approx \phi_p, \quad \phi_d^- \approx -\phi_p$$

gives

$$
\begin{aligned}
P(\phi_p | \ldots) \approx\ & \exp\{2(|\sigma^+|)^{-2} R_H^+(R_d^+ - R_p)\cos(\phi_p - \phi_H^+)\\
&+ 2(|\sigma^-|)^{-2} R_H^-(R_d^- - R_p)\cos(\phi_p + \phi_H^-)\}. \qquad (44)
\end{aligned}
$$

In terms of $|F|$'s, (44) is rewritten as

$$
\begin{aligned}
P(\phi_p | \ldots) &\approx \exp\{G^+ \cos(\phi_p - \phi_H^+) + G^- \cos(\phi_p + \phi_H^-)\}\\
&\approx \exp\{X \cos(\phi_p - \vartheta_p)\}, \qquad (45)
\end{aligned}
$$

where

$$
\begin{aligned}
G^+ = 2|F_H^+|\Delta_{\rm iso}^+/|\mu^+|^2, &\quad G^- = 2|F_H^-|\Delta_{\rm iso}^-/|\mu^-|^2,\\
\Delta_{\rm iso}^+ = |F_d^+| - |F_p|, &\quad \Delta_{\rm iso}^- = |F_d^-| - |F_p|
\end{aligned}
$$

$$\tan\vartheta_p = \frac{(G^+ \sin\phi_H^+ - G^- \sin\phi_H^-)}{(G^+ \cos\phi_H^+ + G^- \cos\phi_H^-)} = \frac{T}{B}, \qquad (46)$$

$$X = (T^2 + B^2)^{1/2}. \qquad (47)$$

ϑ_p is the best estimate of ϕ_p, X is the reliability factor.

Equation (46) is of very simple use [as well as (30) and (37) for the SIR and the MIR cases]; however, some details are necessary to disclose its internal mechanism. Let us:

(i) rewrite (46) in the most useful form

$$\tan\vartheta_p = \frac{\Delta_{\rm iso}^+ \sin\phi_H^+ + \Delta_{\rm iso}^- \sin\phi_H^{-*}}{\Delta_{\rm iso}^+ \cos\phi_H^+ + \Delta_{\rm iso}^- \cos\phi_H^{-*}}, \qquad (48)$$

where $\phi_H^{-*} = -\phi_H^-$;

(ii) denote by F'_H the heavy-atom structure factor arising from the real scattering (*i.e.* from $f'_H = f^0_H + \Delta f'_H$) and by F''_H that from the imaginary scattering (*i.e.* from f''_H). We will assume only one species of heavy atom and we will consider a few didactical cases:

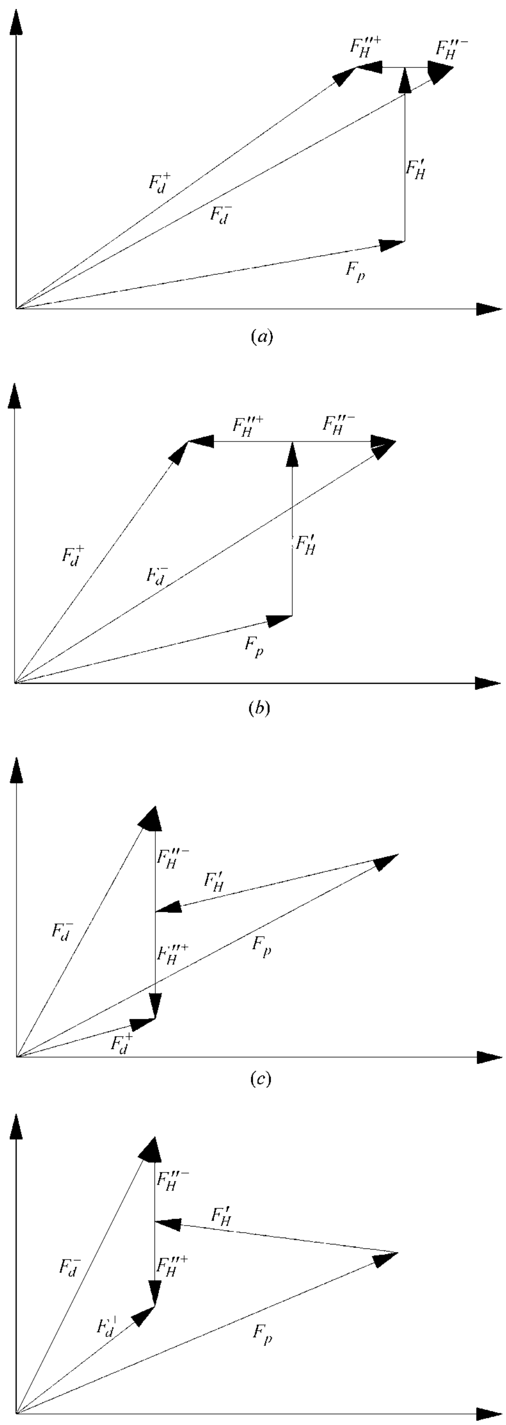

Figure 1
Four situations for the SIRAS case are depicted: (*a*) $|F'_H| \gg |F''_H|$; (*b*) F''_H compared with $|F'_H|$, $\Delta^-_{\rm iso} > 0$; (*c*) $|F''_H|$ compared with $|F'_H|$, $\Delta^-_{\rm iso} < 0$, $\Delta^-_{\rm iso} < 0$; (*d*) $|F''_H|$ compared with $|F'_H|$, $\Delta^+_{\rm iso} < 0$, $\Delta^-_{\rm iso} > 0$.

(*a*) $f'_H \gg f''_H$ (then $|F'_H| \gg |F''_H|$). In this case, $\Delta^+_{\rm iso} \approx \Delta^-_{\rm iso}$, $\phi^+_H \approx \phi^{-*}_H$ and (48) reduces to

$$\tan \vartheta_p \approx (\Delta^+_{\rm iso} \sin \phi^+_H)/(\Delta^+_{\rm iso} \cos \phi^+_H).$$

Accordingly, $\vartheta_p \approx \varphi^+_H$ if $\Delta^+_{\rm iso} \approx \Delta^-_{\rm iso} > 0$, $\vartheta_p \approx \phi^+_H + \pi$ if $\Delta^+_{\rm iso} \approx \Delta^-_{\rm iso} < 0$. This is nothing but the classical SIR case: the anomalous scattering does not add any valuable information to the phase indication provided by (30) (see Fig. 1*a*).

(*b*) f''_H is comparable with f'_H (then $|F''_H|$ is also comparable with $|F'_H|$), $\Delta^+_{\rm iso} > 0$, $\Delta^-_{\rm iso} > 0$ with $|\Delta^-_{\rm iso}| > |\Delta^+_{\rm iso}|$. Then (48) may be written as

$$\tan \vartheta_p \approx (|\Delta^+_{\rm iso}| \sin \phi^+_H + |\Delta^-_{\rm iso}| \sin \phi^{-*}_H) \times (|\Delta^+_{\rm iso}| \cos \phi^+_H + |\Delta^-_{\rm iso}| \cos \phi^-_H)^{-1}, \qquad (49)$$

according to which φ_p is estimated between ϕ^+_H and ϕ^{-*}_H, closer to φ^{-*}_p. This situation is illustrated in Fig. 1*b*);

(*c*) f''_H is comparable with f'_H, $\Delta^+_{\rm iso} < 0$, $\Delta^-_{\rm iso} < 0$ with $|\Delta^-_{\rm iso}| > |\Delta^+_{\rm iso}|$. Then (48) reduces to

$$\tan \vartheta_p \approx [|\Delta^+_{\rm iso}| \sin(\phi^+_H + \pi) + |\Delta^-_{\rm iso}| \sin(\phi^{-*}_H + \pi)] \times [|\Delta^+_{\rm iso}| \cos(\phi^+_H + \pi) + |\Delta^-_{\rm iso}| \cos(\phi^{-*}_H + \pi)]^{-1},$$

which estimates ϕ_p between $(\varphi^+_H + \pi)$ and $(\phi^{-*}_H + \pi)$, closer to $(\phi^{-*}_H + \pi)$. This situation is illustrated in Fig. 1*c*).

(*d*) f''_H is comparable with f'_H, $\Delta^+_{\rm iso} < 0$, $\Delta^-_{\rm iso} > 0$ with $|\Delta^-_{\rm iso}| > |\Delta^+_{\rm iso}|$. In this case, (48) reduces to

$$\tan \vartheta_p \approx [|\Delta^+_{\rm iso}| \sin(\phi^+_H + \pi) + |\Delta^-_{\rm iso}| \sin \phi^{-*}_H] \times [|\Delta^+_{\rm iso}| \cos(\phi^+_H + \pi) + |\Delta^-_{\rm iso}| \cos \phi^{-*}_H]^{-1},$$

which estimates φ_p between $(\varphi^+_H + \pi)$ and ϕ^{-*}_H, closer to ϕ^{-*}_H (see Fig. 1*d*).

All four figures indicate that (46) is a sensible way of assigning the phases given $\Delta^+_{\rm iso}$, $\Delta^-_{\rm iso}$, F_H. We check how accurate (46) may be in an ideal case [we postpone to the paper by Giacovazzo *et al.* (2002) the applications to real data]. To this purpose, we use as protein data the calculated data of 1srv (Walsh *et al.*, 1999), space group $C222_1$, unit-cell parameters $a = 63.47$, $b = 65.96$, $c = 75.03$ Å, 1186 non-H atoms and three S atoms in 145 amino acids. The SIRAS case was simulated by substituting Se for S and by selecting values of $\Delta f' = -8$, $\Delta f'' = 3$. The results are shown in Table 1, where we give the average phase error $\langle|\Delta\phi|\rangle_{\rm SIRAS}$ for the NREF1 reflections with a weight $I_1(X)/I_o(X)$ larger than the W value. The value $\langle|\Delta\phi|\rangle_{\rm SIRAS}$ is the average difference (in °) between the phase estimates provided by (46) and the true phases (*i.e.* the values calculated from the published crystal structure). $I_i(X)$ is the modified Bessel function of order i calculated at X, as given by (47). To evaluate the efficiency of (46), we have also simulated the SIR case, by eliminating the anomalous dispersion from the Se atoms. The corresponding figures are tabulated in columns 4 and 5. In particular, $\langle|\Delta\phi|\rangle_{\rm SIR}$ is the average phase error for the NREF2 reflections with a weight $I_1(G)/I_o(G)$ larger than W [the argument G is defined by (21)]. The comparison shows quite evidently the larger efficiency of (46): about 41° is the error for the 7303 reflections if (21) is applied, only 22° if (46) is used.

Table 1
1srv calculated data; the reliability of the estimates (46) is compared with that of the SIR estimates [equation (38)].

W	NREF2	$\langle\Delta\lvert\phi\rvert\rangle_{\rm SIRAS}$ (°)	NREF1	$\langle\lvert\Delta\phi\rvert\rangle_{\rm SIR}$ (°)
0.0	7303	22	7303	41
0.1	6648	21	6648	40
0.2	6144	21	6144	39
0.3	5698	21	5698	38
0.4	5239	21	5239	37
0.5	4763	21	4763	36
0.6	4312	20	4312	35
0.7	3806	20	3806	33
0.8	3043	19	3043	31
0.9	1861	19	1861	27

4.2. The MIRAS case in *P*1

The extension of the theory described in §4.1 to the MIRAS case is straightforward. We only provide the conclusive formulas when more derivatives with heavy atoms as anomalous scatterers are available. In this case, (45) is replaced by

$$P(\phi_p|\ldots) \approx \exp\{X\cos(\phi_p - \vartheta_p)\}, \qquad (50)$$

where

$$\tan\vartheta_p = \frac{\sum_j (G_j^+ \sin\phi_{Hj}^+ - G_j^- \sin\phi_{Hj}^-)}{\sum_j (G_j^+ \cos\phi_{Hj}^+ + G_j^- \cos\phi_{Hj}^-)} = \frac{T}{B} \qquad (51)$$

$$X = (T^2 + B^2)^{1/2} \qquad (52)$$

The summation over j varies over the different derivatives.

5. Conclusions

The rigorous use of the joint probability distribution functions enabled us to provide new probabilistic formulas estimating the protein phases for the SIR–MIR and SIRAS–MIRAS cases. The formulas are of very simple cases, include the treatment of the errors and, as we prove in the following paper (Giacovazzo *et al.*, 2002), are very efficient.

The main results of the paper have been presented at the Euroconference on *Phasing Biological Macromolecules* (*PHABIO*) held in Martina Franca (TA), 23–27 June 2001.

APPENDIX *A*

In the case of perfect isomorphism, $F_d = F_p + F_H$. If $C(u_p, u_d)$ is the characteristic function in $P\bar{1}$ of $P(E_p, E_d|E_H)$, then (see §3.1)

$$C(u_p, u_d) = \exp(iu_d E_H)\exp\{-(u_p^2 + u_d^2 + 2u_p u_d)/2\}. \qquad (53)$$

The Fourier transform of (53) gives

$$\begin{aligned} P(E_p, E_d|E_H) &= (2\pi)^{-2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{-0.5[u_p^2 + u_d^2 + 2u_p u_d] \\ &\quad - i[E_p u_p + (E_d - E_H)u_d]\}\,{\rm d}u_p\,{\rm d}u_d \\ &= (2\pi)^{-1/2}\exp(-E_p^2/2)(2\pi)^{-1} \\ &\quad \times \int_{-\infty}^{\infty} \exp\{iu_p[E_d - (E_p + E_H)]\}\,{\rm d}u_p \\ &= (2\pi)^{-1/2}\exp(-E_p^2/2)\delta[E_d - (E_p + E_H)], \qquad (54) \end{aligned}$$

where δ is the Dirac delta function.

The distribution (54) may be so interpreted: E_p satisfies the usual Wilson distribution because it is not constrained by the prior knowledge of E_H. *Vice versa*, such a prior information constrains E_d to be distributed according to the Dirac delta function centered on $E_p + E_H$. This last property is trivial and does not add any additional information to the definition $F_d = F_p + F_H$.

References

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions*. New York: Dover.
Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Siliqi, D. (2002). *Acta Cryst.* D**58**, 928–935
Giacovazzo, C., Ladisa, M. & Siliqi, D. (2002). *Acta Cryst.* A**58**, 598–604.
Giacovazzo, C. & Siliqi, D. (1997). *Acta Cryst.* A**53**, 789–798.
Giacovazzo, C. & Siliqi, D. (2001*a*). *Acta Cryst.* A**57**, 40–46.
Giacovazzo, C. & Siliqi, D. (2001*b*). *Acta Cryst.* A**57**, 414–419.
Giacovazzo, C. & Siliqi, D. (2001*c*). *Acta Cryst.* A**57**, 700–707.
Hauptman, H. (1982). *Acta Cryst.* A**38**, 289–294.
Terwilliger, T. C. & Eisenberg, D. (1983). *Acta Cryst.* A**39**, 813–817.
Terwilliger, T. C. & Eisenberg, D. (1987). *Acta Cryst.* A**43**, 6–13.
Walsh, M. A., Dementieva, I., Evans, O., Sanishili, R. & Joachimiak, A. (1999). *Acta Cryst.* D**55**, 1168–1173.