Acta Cryst. (1996). A**52**, 143–151

On Integrating Direct Methods and Isomorphous Replacement Techniques. II. The Quartet Invariant Estimate

CARMELO GIACOVAZZO[a,b]* AND DRITAN SILIQI[c]

[a]*Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy,* [b]*Istituto di Ricerca per lo Sviluppo di Metodologie Cristallografiche, CNR, c/o Dipartimento Geomineralogico, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, and* [c]*Laboratory of X-ray Diffraction, Department of Inorganic Chemistry, Faculty of Natural Sciences, Tirana University, Tirana, Albania. E-mail: giacovazzo@area.ba.cnr.it*

(Received 2 May 1995; accepted 30 August 1995)

Abstract

In a preceding paper [Giacovazzo & Siliqi (1996). *Acta Cryst.* A**52**, 133–142], the joint probability distribution of seven pairs of isomorphous structure factors was derived. Its complicated mathematical expressions are here simplified by introducing the assumption that isomorphism is due to heavy-atom addition to the native structure. The reliability of the conclusive formula for calculated error-free data perfectly agrees with expectations. The formula, however, is not robust against lack of isomorphism and errors in measurements: in the paper, theoretical reasons are given justifying this behaviour. The use of the prior information on the heavy-atom structure markedly improves the formula, which then proves suitable for practical applications.

1. Symbols and notation

The notation is that used in the paper by Giacovazzo & Siliqi (1996) (from now on denoted as paper I).

2. Introduction

In paper I, the joint probability distribution function

$$P(\phi_1,\ldots,\phi_7,\psi_1,\ldots,\psi_7,R_1,\ldots,R_7,S_1,\ldots,S_7) \quad (1)$$

has been derived [see equation (I.18)]. From (1), the conditional probability

$$P(\Phi|R_1,\ldots,R_7,S_1,\ldots,S_7), \quad (2)$$

where

$$\Phi=\phi_1+\phi_2+\phi_3+\phi_4$$

may be obtained as follows.

(*a*) The marginal distribution

$$P(\phi_1,\ldots,\phi_4,R_1,\ldots,R_7,S_1,\ldots,S_7)$$

is obtained by integrating (1) over the ten variables $\phi_5,\phi_6,\phi_7,\psi_1,\ldots,\psi_7$.

Table 1. *Code name, space group and crystallochemical data for test structures*

Structure code	Reference	Space group	Molecular formula	Z
APP	(1)	$C2$	$C_{190}N_{53}O_{58}Zn$	4
CARP	(2)	$C2$	$C_{513}N_{131}O_{121}Ca_2S$	4
E2	(3)	$F432$	$C_{1170}N_{310}O_{366}S_7$	96
M-FABP	(4)	$P2_12_12_1$	$C_{667}N_{170}O_{261}S_3$	4

References: (1) Glover, Haneef, Pitts, Wood, Moss, Tickle & Blundell (1983); (2) Kretsinger & Nockolds (1973); (3) Mattevi, Obmolova, Schulze, Kalk, Wesphal, De Kok & Hol (1992); (4) Zanotti, Scapin, Spadon, Veerkamp & Sacchettini (1992).

(*b*) The conditional distribution

$$\begin{aligned}P(\phi_1,\ldots,\phi_4|R_1,\ldots,R_7,S_1,\ldots,S_7)\\ =P(\phi_1,\ldots,\phi_4,R_1,\ldots,R_7,S_1,\ldots,S_7)\\ \times\left[\int_0^{2\pi}\ldots\int_0^{2\pi}P(\phi_1,\ldots,\phi_4,R_1,\ldots,R_7,S_1,\ldots,S_7)\right.\\ \left.\mathrm{d}\phi_1\ldots\mathrm{d}\phi_4\right]^{-1}\end{aligned} \quad (3)$$

is calculated.

(*c*) The distribution (3) is integrated over $\phi_1,\phi_2,\phi_3,\phi_4$ under the condition that

$$\phi_1+\phi_2+\phi_3+\phi_4=\Phi.$$

The calculations may be performed *via* the formulas quoted in Appendix *A* of paper I. However, the conclusive distribution turns out to be too complicated for most routine applications. We prefer to derive a simpler result by introducing the following basic assumption: the derivative structure is obtained by addition of heavy atoms to the basic structure. This is the classical case of isomorphism between the native protein and its heavy-atom derivative. We will show that in this situation the joint probability distribution (I.18) will become useful and will reveal the characteristic features to be exploited in practical applications. The experimental diffraction data of some proteins and their derivatives will be used to check the mathematical approach and the final probabilistic formulas. The code names, the space group and the crystallochemical data of the test structures are given in Table 1, the relevant parameters of the diffraction data are shown in Table 2.

3. A simplified P_7 distribution

For a native-protein–heavy-atom-derivative case, the following relations can be used for making the distribution (I.18) simpler [see Giacovazzo, Cascarano & Zheng (1988), hereafter GCZ, for the triplet parameters].

$$\alpha_i = [\sigma_2]_p^{1/2}/[\sigma_2]_d^{1/2}$$

$$(1-\alpha_i^2) = [\sigma_2]_H/[\sigma_2]_d$$

$$\alpha_i^2/(1-\alpha_i^2) = [\sigma_2]_p/[\sigma_2]_H$$

$$\alpha_i/(1-\alpha_i^2) = [\sigma_2]_p^{1/2}[\sigma_2]_d^{1/2}/[\sigma_2]_H$$

$$\begin{aligned}\gamma_{\bar{i}jl}/\alpha_i &= \gamma_{i\bar{j}l}/\alpha_j = \gamma_{ij\bar{l}}/\alpha_l = \gamma_{\bar{i}\bar{j}l}/\alpha_i\alpha_j = \gamma_{\bar{i}j\bar{l}}/\alpha_i\alpha_l \\ &= \gamma_{i\bar{j}\bar{l}}/\alpha_j\alpha_l = \gamma_{ijl} = \gamma_p = [\sigma_3/\sigma_2^{3/2}]_p\end{aligned}$$

$$\gamma_{\bar{i}\bar{j}\bar{l}} = \{[\sigma_3/\sigma_2^{3/2}]_p + [\sigma_3/\sigma_2^{3/2}]_H[\sigma_2^{3/2}]_H/[\sigma_2^{3/2}]_p\}\alpha_i\alpha_j\alpha_l$$

$$\beta_{ijl} = [\sigma_3/\sigma_2^{3/2}]_p - [\sigma_3/\sigma_2^{3/2}]_H[\sigma_2^{3/2}]_p/[\sigma_2^{3/2}]_H$$

$$\begin{aligned}\beta_{\bar{i}jl}/(\alpha_j\alpha_l) &= \beta_{i\bar{j}l}/(\alpha_i\alpha_l) = \beta_{ij\bar{l}}/(\alpha_i\alpha_j) = -\beta_{i\bar{j}\bar{l}}/\alpha_i \\ &= \beta_{\bar{i}j\bar{l}}/\alpha_j = -\beta_{\bar{i}\bar{j}l}/\alpha_l = \beta_{\bar{i}\bar{j}\bar{l}}\end{aligned}$$

$$\begin{aligned}\beta_{\bar{i}\bar{j}\bar{l}} &= \alpha_i\alpha_j\alpha_l[\sigma_3/\sigma_2^{3/2}]_H[\sigma_2^{3/2}]_H[\sigma_2^{3/2}]_p^{-1}[(1-\alpha_i^2)(1-\alpha_j^2) \\ &\quad\times(1-\alpha_l^2)]^{-1} = [\sigma_3/\sigma_2^{3/2}]_H[\sigma_2^{3/2}]_d/[\sigma_2^{3/2}]_H\end{aligned}$$

$$\gamma_{1234} = [\sigma_4/\sigma_2^2]_p$$

$$\begin{aligned}\gamma_{123\bar{4}}/\alpha_4 &= \gamma_{12\bar{3}4}/\alpha_3 = \ldots = \gamma_{12\bar{3}\bar{4}}/\alpha_3\alpha_4 \\ &= \gamma_{1\bar{2}3\bar{4}}/\alpha_2\alpha_4 = \ldots = \gamma_{\bar{1}\bar{2}\bar{3}4}/\alpha_1\alpha_2\alpha_3 = \gamma_{1234}\end{aligned}$$

$$\gamma_{\bar{1}\bar{2}\bar{3}\bar{4}} = \{[\sigma_4/\sigma_2^2]_p + [\sigma_4/\sigma_2^2]_H[\sigma_2]_H^2[\sigma_2]_p^{-2}\}\alpha_1\alpha_2\alpha_3\alpha_4$$

$$\beta_{1234} = [\sigma_4/\sigma_2^2]_p + [\sigma_4/\sigma_2^2]_H[\sigma_2]_p^2[\sigma_2]_H^{-2}$$

$$\begin{aligned}-\beta_{123\bar{4}}/(\alpha_1\alpha_2\alpha_3) &= -\beta_{12\bar{3}4}/(\alpha_1\alpha_2\alpha_4) = -\beta_{1\bar{2}34}/(\alpha_1\alpha_2\alpha_3) \\ &= \ldots = \beta_{\bar{1}\bar{2}34}/(\alpha_3\alpha_4) = \beta_{\bar{1}2\bar{3}4}/(\alpha_2\alpha_4) \\ &= \ldots = -\beta_{\bar{1}\bar{2}\bar{3}4}/\alpha_4 = \beta_{\bar{1}\bar{2}\bar{3}\bar{4}}\end{aligned}$$

$$\beta_{\bar{1}\bar{2}\bar{3}\bar{4}} = [\sigma_4/\sigma_2^2]_H[\sigma_2]_p^2[\sigma_2]_H^{-2}(\alpha_1\alpha_2\alpha_3\alpha_4)^{-1}$$

$$\begin{aligned}B_{1234i} &= \prod_{i=1}^{4}\gamma_{12i}\gamma_{34i}(1-\alpha_4^2-\alpha_3^2-\ldots+\alpha_1^2\alpha_2^2\alpha_3^2\alpha_4^2) \\ &= [\sigma_3/\sigma_2^{3/2}]_p^2\end{aligned}$$

$$\begin{aligned}B_{123\bar{4}i} &= \prod_{i=1}^{4}\gamma_{12i}\gamma_{34i}(\alpha_4-\alpha_4-\alpha_3^2\alpha_4+\ldots+\alpha_3^2\alpha_4+\ldots \\ &\quad+\alpha_1^2\alpha_2^2\alpha_3^2\alpha_4) = 0\end{aligned}$$

$$\vdots \qquad \vdots$$

$$B_{\bar{1}\bar{2}\bar{3}\bar{4}i} = 0.$$

In the following, we will always neglect $[\sigma_3/\sigma_2^{3/2}]_p$ with respect to $[\sigma_3/\sigma_2^{3/2}]_H$ and $[\sigma_2^2/\sigma_4]_p$ with respect to $[\sigma_2^2/\sigma_4]_p$. Furthermore, from Appendix *A*, the following simplifications can be derived:

(*a*) B_{1234i} is negligible with respect to $B_{1234\bar{i}}$, which may be approximated to

$$B_{1234\bar{i}} \simeq \alpha_i^2[\sigma_3/\sigma_2^{3/2}]_H^2[\sigma_2]_p/[\sigma_2]_H.$$

(*b*) $B_{1234i}^{\mathrm{mod}(1)} \simeq B_{1234i}^{\mathrm{mod}(2)}$ are negligible when compared with $B_{1234\bar{i}}$.

(*c*) $B_{123\bar{4}i}$, $B_{123\bar{4}i}^{\mathrm{mod}(1)}$, $B_{123\bar{4}i}^{\mathrm{mod}(2)}$ are negligible with respect to $B_{123\bar{4}\bar{i}}$ *etc.*

In conclusion, we can approximate (I.18) by the expression

$$\begin{aligned}P_7 &\simeq \prod_{i=1}^{7}\left((1/\pi^2)[R_iS_i/(1-\alpha_i^2)]\exp\{-[1/(1-\alpha_i^2)\right. \\ &\quad\left.\times[R_i^2+S_i^2-2\alpha_iR_iS_i\cos(\psi_i-\phi_i)]\}\right) \\ &\quad\times\left\{1+\sum_{i,j,l}[2\beta_{ijl}R_iR_jR_l\cos(\phi_i+\phi_j+\phi_l)\right. \\ &\quad+2\beta_{\bar{i}jl}S_iR_jR_l\cos(\psi_i+\phi_j+\phi_l) \\ &\quad+2\beta_{i\bar{j}l}R_iS_jR_l\cos(\phi_i+\psi_j+\phi_l) \\ &\quad+2\beta_{ij\bar{l}}R_iR_jS_l\cos(\phi_i+\phi_j+\psi_l) \\ &\quad+2\beta_{i\bar{j}\bar{l}}R_iS_jS_i\cos(\phi_i+\psi_j+\psi_l) \\ &\quad+2\beta_{\bar{i}j\bar{l}}S_iR_jS_l\cos(\psi_i+\phi_j+\psi_l) \\ &\quad+2\beta_{\bar{i}\bar{j}l}S_iS_jR_l\cos(\psi_i+\psi_j+\phi_l) \\ &\quad+2\beta_{\bar{i}\bar{j}l}S_iS_jS_l\cos(\psi_i+\psi_j+\psi_l)]+\sum_{i,j,l}(B_{ijl}+B_{\bar{i}jl} \\ &\quad+B_i+B_{ij\bar{l}}+B_{i\bar{j}\bar{l}}+B_{\bar{i}j\bar{l}}+B_{\bar{i}\bar{j}l}+B_{\bar{i}\bar{j}\bar{l}}) \\ &\quad+2R_1R_2R_3R_4\left[\beta_{1234}+\sum_{i=5}^{7}(1-\alpha_i^2)^{-1}B_{1234\bar{i}}L_{\bar{i}}\right] \\ &\quad\times\cos(\phi_i+\phi_2+\phi_3+\phi_4) \\ &\quad+2R_1R_2R_3S_4\left[\beta_{123\bar{4}}+\sum_{i=5}^{7}(1-\alpha_i^2)^{-1}B_{123\bar{4}\bar{i}}L_{\bar{i}}\right] \\ &\quad\times\cos(\phi_1+\phi_2+\phi_3+\psi_4)+\ldots+2S_1S_2S_3S_4 \\ &\quad\times\left[\beta_{\bar{1}\bar{2}\bar{3}\bar{4}}+\sum_{i=5}^{7}(1-\alpha_i^2)^{-1}B_{\bar{1}\bar{2}\bar{3}\bar{4}\bar{i}}L_{\bar{i}}\right] \\ &\quad\left.\times\cos(\psi_1+\psi_2+\psi_3+\psi_4)\right\}.\end{aligned} \qquad (4)$$

Equation (4) is our basic distribution: it represents a good approximation of the more complicated distribution

Table 2. *Relevant parameters for diffraction data of test structures*

Structure code	Native RES (Å)	Native NREFL	Derivative Heavy atom	Derivative $[\sigma_2]_H/[\sigma_2]_p$	Derivative RES (Å)	Derivative NREFL
APP	0.99	17058	Hg	0.46	2.0	2086
CARP	1.70	5056	Hg	0.09	2.0	4687
E2	3.00	10388	Hg	0.08	3.0	9179
M-FABP	2.14	7595	Hg	0.06	3.0	7125

(I.18) when we deal with a native protein and a heavy-atom derivative.

4. The quartet distribution

We integrate over ten variables $\phi_5, \phi_6, \phi_7, \psi_1, \ldots, \psi_7$ in order to obtain, according to (3), the conditional distribution

$$
\begin{aligned}
&P(\phi_1, \phi_2, \phi_3, \phi_4 | R_1, \ldots, R_7, S_1, \ldots, S_7) \\
&\quad \simeq Q^{-1}\Bigg(1 + \sum_{i,j,l}(B_{ijl} + B_{\bar{i}jl} + B_{i\bar{j}l} \\
&\qquad + B_{ij\bar{l}} + B_{i\bar{j}\bar{l}} + B_{\bar{i}j\bar{l}} + B_{\bar{i}\bar{j}l} + B_{\bar{i}\bar{j}\bar{l}}) \\
&\qquad + \Bigg\{2R_1R_2R_3R_4\left[\beta_{1234} + \sum_{i=5}^{7}(1-\alpha_i^2)^{-1}B_{1234\bar{i}}\langle L_{\bar{i}}\rangle\right] \\
&\qquad + 2R_1R_2R_3S_4\left[\beta_{123\bar{4}} + \sum_{i=5}^{7}(1-\alpha_i^2)^{-1}B_{123\bar{4}\bar{i}}\langle L_{\bar{i}}\rangle\right]D_{14} \\
&\qquad + \ldots + 2S_1S_2S_3S_4\left[\beta_{\bar{1}\bar{2}\bar{3}\bar{4}} + \sum_{i=5}^{7}(1-\alpha_i^2)^{-1}B_{\bar{1}\bar{2}\bar{3}\bar{4}\bar{i}}\langle L_{\bar{i}}\rangle\right] \\
&\qquad \times D_{11}D_{12}D_{13}D_{14}\Bigg\}\cos(\phi_1+\phi_2+\phi_3+\phi_4)\Bigg), \qquad (5)
\end{aligned}
$$

where Q^{-1} is a suitable scale factor,

$$D_{1i} = I_1[2\alpha_i R_i S_i/(1-\alpha_i^2)]/I_0[2\alpha_i R_i S_i/(1-\alpha_i^2)]$$

and

$$\langle L_{\bar{i}}\rangle = (1-\alpha_i^2)^{-1}[S_i^2 + \alpha_i R_i^2 - 2\alpha_i R_i S_i D_{1i}] - 1.$$

Then,

$$
\begin{aligned}
&P(\Phi | R_1, \ldots, R_7, S_1, \ldots, S_7) \\
&\quad \simeq [2\pi I_0(A)]^{-1}\exp(A\cos\Phi), \qquad (6)
\end{aligned}
$$

where

$$A = T/(1+B), \qquad (7)$$

$$
\begin{aligned}
T &= 2R_1R_2R_3R_4\left[\beta_{1234} + \sum_{i=5}^{7}(1-\alpha_i^2)^{-1}B_{1234\bar{i}}\langle L_{\bar{i}}\rangle\right] \\
&\quad + 2R_1R_2R_3S_4\left[\beta_{123\bar{4}} + \sum_{i=5}^{7}(1-\alpha_i^2)^{-1}B_{123\bar{4}\bar{i}}\langle L_{\bar{i}}\rangle\right]D_{14} \\
&\quad + \ldots + 2S_1S_2S_3S_4\left[\beta_{\bar{1}\bar{2}\bar{3}\bar{4}} + \sum_{i=5}^{7}(1-\alpha_i^2)^{-1}B_{\bar{1}\bar{2}\bar{3}\bar{4}\bar{i}}\langle L_{\bar{i}}\rangle\right] \\
&\quad \times D_{11}D_{12}D_{13}D_{14}, \qquad (8)
\end{aligned}
$$

$$
\begin{aligned}
B &= 1 + \sum_{i,j,l}(B_{ijl} + B_{\bar{i}jl} + B_{i\bar{j}l} + B_{ij\bar{l}} \\
&\quad + B_{i\bar{j}\bar{l}} + B_{\bar{i}j\bar{l}} + B_{\bar{i}\bar{j}l} + B_{\bar{i}\bar{j}\bar{l}}). \qquad (9)
\end{aligned}
$$

The mathematical implications concerning the passage from the linear expression (5) to the exponential expression (6) are exactly those described by Giacovazzo (1976) for the quartet invariant estimate in the absence of derivative data.

The distribution (6) assumes a particularly attractive form if a variable change is made. We replace in (6)

$$R_i, S_i, \quad \text{for } i = 1, \ldots, 7$$

by

$$R_i' = F_{p_i}/\Sigma_H^{1/2} \quad \text{and} \quad S_i' = F_{d_i}/\Sigma_H^{1/2},$$

which are pseudo-normalized (with respect to the heavy-atom structure) structure factors. Accordingly,

$$R_i = R_i'[\sigma_2]_H^{1/2}/[\sigma_2]_p^{1/2}, \quad S_i = S_i'[\sigma_2]_H^{1/2}/[\sigma_2]_d^{1/2}.$$

Then, (6) still holds, and its parameters may be rewritten in a simple form:

$$
\begin{aligned}
A &= [2\Delta_1'\Delta_2'\Delta_3'\Delta_4'/(1+B')]\{[\sigma_4/\sigma_2^2]_H \\
&\quad + [\sigma_3/\sigma_2^{3/2}]_H^2[\langle L_{\bar{5}}'\rangle + \langle L_{\bar{6}}'\rangle + \langle L_{\bar{7}}'\rangle]\}, \qquad (10)
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta_i' &= S_i'D_{1i} - R_i' \\
\langle L_{\bar{i}}'\rangle &= (S_i'^2 + R_i'^2 - 2R_i'S_i'D_{1i}') - 1 \\
D_{1i}' &= I_1(2R_i'S_i')/I_0(2R_i'S_i') \\
B' &\simeq \tfrac{1}{2}[\sigma_3/\sigma_2^{3/2}]_H^2[\langle L_{\bar{1}}'\rangle\langle L_{\bar{2}}'\rangle\langle L_{\bar{5}}'\rangle + \langle L_{\bar{3}}'\rangle\langle L_{\bar{4}}'\rangle\langle L_{\bar{5}}'\rangle \\
&\quad + \langle L_{\bar{1}}'\rangle\langle L_{\bar{3}}'\rangle\langle L_{\bar{6}}'\rangle + \langle L_{\bar{2}}'\rangle\langle L_{\bar{4}}'\rangle\langle L_{\bar{6}}'\rangle + \langle L_{\bar{1}}'\rangle\langle L_{\bar{4}}'\rangle\langle L_{\bar{7}}'\rangle \\
&\quad + \langle L_{\bar{2}}'\rangle\langle L_{\bar{3}}'\rangle\langle L_{\bar{7}}'\rangle].
\end{aligned}
$$

The reliability parameter A of the quartet phase Φ now has a simpler expression and may be conveniently studied.

5. About the main features of the quartet formula

Let us consider the quartet formula derived by Giacovazzo (1976, 1980) for the case in which derivative data are not available. Then,

$$P(\Phi) = [2\pi I_0(A)^{-1}]\exp\{A\cos\Phi\},$$

where

$$
\begin{aligned}
A &= [2R_1R_2R_3R_4/(1+B)]\{\sigma_4/\sigma_2^2 + (\sigma_3/\sigma_2^{3/2})^2 \\
&\quad \times (\varepsilon_5 + \varepsilon_6 + \varepsilon_7)\}, \\
\varepsilon_i &= |R_i|^2 - 1, \qquad (10') \\
B &= \tfrac{1}{2}[\sigma_3/\sigma_2^{3/2}]^2(\varepsilon_1\varepsilon_2\varepsilon_5 + \varepsilon_3\varepsilon_4\varepsilon_5 + \varepsilon_1\varepsilon_3\varepsilon_6 + \varepsilon_2\varepsilon_4\varepsilon_6 \\
&\quad + \varepsilon_1\varepsilon_4\varepsilon_7 + \varepsilon_2\varepsilon_3\varepsilon_7).
\end{aligned}
$$

B is a positive scaling factor; it is assumed $B = 0$ if $B \le 0$.

Equation (10) has a mathematical form similar to (10′): thus, the well known features of (10′) may be used to describe the role of (10). Examples are given below.

(*a*) In (10), we will set $B' = 0$ if $B' \le 0$.

(*b*) If the heavy atoms are of the same type, (10) may be replaced by

$$A = (2/N_H)[\Delta'_1\Delta'_2\Delta'_3\Delta'_4/(1+B')] \times \{1 + [\langle L'_5\rangle + \langle L'_6\rangle + \langle L'_7\rangle]\},$$

where

$$B' = (1/2N_H)[\langle L'_1\rangle\langle L'_2\rangle\langle L'_5\rangle + \ldots + \langle L'_2\rangle\langle L'_3\rangle\langle L'_7\rangle]$$

and N_H is the number of heavy atoms in the primitive unit cell.

(*c*) If some of the cross terms are unknown, then suitable marginal distributions should be calculated. The final result is: if the pairs (R'_5, S'_5) and/or (R'_6, S'_6) and/or (R'_7, S'_7) are not among the measured data, then A may be updated by omitting in (10) the terms $\langle L'_5\rangle$ and/or $\langle L'_6\rangle$ and/or $\langle L'_7\rangle$.

(*d*) if $\langle L'_5\rangle$, $\langle L'_6\rangle$, $\langle L'_7\rangle$ are large enough, the sign of the quartet will probably coincide with the sign of the product $\Delta'_1\Delta'_2\Delta'_3\Delta'_4$. An extreme situation is that for which $|\Delta'_i|$ and $R'_iS'_i$, for $i = 5, 6, 7$, are large. Then,

$$\langle L'_i\rangle \simeq \Delta'^2_i - 1$$

and (10) assumes the more familiar expression

$$A = [2\Delta'_1\Delta'_2\Delta'_3\Delta'_4/(1+B')]\{[\sigma_4/\sigma_2^2]_H + [\sigma_3/\sigma_2^{3/2}]^2_H \times [(\Delta'^2_5 - 1) + (\Delta'^2_6 - 1) + (\Delta'^2_7 - 1)]\}, \quad (11)$$

where

$$\begin{aligned} B' = \tfrac{1}{2}[\sigma_3/\sigma_2^{3/2}]^2_H[&(\Delta'^2_1 - 1)(\Delta'^2_2 - 1)(\Delta'^2_5 - 1) \\ &+ (\Delta'^2_3 - 1)(\Delta'^2_4 - 1)(\Delta'^2_5 - 1) \\ &+ (\Delta'^2_1 - 1)(\Delta'^2_3 - 1)(\Delta'^2_6 - 1) \\ &+ (\Delta'^2_2 - 1)(\Delta'^2_4 - 1)(\Delta'^2_6 - 1) \\ &+ (\Delta'^2_1 - 1)(\Delta'^2_4 - 1)(\Delta'^2_7 - 1) \\ &+ (\Delta'^2_2 - 1)(\Delta'^2_3 - 1)(\Delta'^2_7 - 1)]. \end{aligned}$$

It should be noted that positive values of $(\Delta'^2_5 - 1) + (\Delta'^2_6 - 1) + (\Delta'^2_7 - 1)$ do not select positive quartets because the sign of the quartet cosine is expected to coincide with the sign of $\Delta'_1\Delta'_2\Delta'_3\Delta'_4$.

(*e*) If $\langle L'_5\rangle$, $\langle L'_6\rangle$, $\langle L'_7\rangle$ assume highly negative values, then the quartet is expected to have cosine sign opposite to that of the product $\Delta'_1\Delta'_2\Delta'_3\Delta'_4$. A typical case is that for which R_iS_i, for $i = 5, 6, 7$, is large and $|\Delta'_i|$ is small. In this case, again,

$$\langle L'_i\rangle \simeq \Delta'^2_i - 1 \quad \text{for } i = 5, 6, 7.$$

(*f*) We will refer to the quartets described in (*d*) as *large-cross quartets*; the quartets described in (*e*) will be called *small-cross quartets*. The first provide information strictly correlated with that supplied by the triplets (as estimated by GCZ). For example, let us suppose that

$$\Phi_{T_1} = \phi_1 + \phi_2 + \phi_5 \quad \text{and} \quad \Phi_{T_2} = \phi_3 + \phi_4 - \phi_5$$

are two triplets characterized by large values of $|\Delta'_1\Delta'_2\Delta'_3|$ and $|\Delta'_3\Delta'_4\Delta'_5|$, respectively. Let SIGN(1) and SIGN(2) be the signs of $\Delta'_1\Delta'_2\Delta'_5$ and $\Delta'_3\Delta'_4\Delta'_5$, respectively. The cosine sign of the quartet

$$\Phi = \Phi_{T_1} + \Phi_{T_2} = \phi_1 + \phi_2 + \phi_3 + \phi_4$$

is expected to have the same sign of SIGN(1) × SIGN(2), which coincides with the sign of $\Delta'_1\Delta'_2\Delta'_3\Delta'_4$. Thus, the combination of the triplet estimates supply a quartet estimate coincident with that provided by (11). On the contrary, small-cross quartets provide information statistically independent of that provided by the triplets. Indeed: (i) triplets with small $|\Delta|$ values are usually not used in the phasing process (Giacovazzo, Siliqi & Platas, 1995) because they are highly unreliable; (ii) the phase indications provided by the small-cross quartets are expected to be opposite to those provided by the triplets if these were used in the phasing process.

6. Check of the probabilistic formulas by calculated diffraction data

Let us first consider the role of (10′) in direct procedures applied to small molecules. It has been found (Burla, Cascarano & Giacovazzo, 1992; Giacovazzo, Burla & Cascarano, 1992) that: (*a*) the reliability of the quartets estimated positive by (10′) is nearly equal to the triplet reliability. However, such quartets carry into the phasing process information strictly correlated with the triplet information. The combined use of the positive quartets and of the triplets was therefore not advised; (*b*) the reliability of the quartets estimated negative by (10′) is remarkably smaller than the triplet reliability; furthermore, the number of reliable negative quartets is in general smaller than the corresponding number of reliable triplets. However, they carry into the phasing procedure an important amount of information uncorrelated with the information supplied by the triplets. The combined use of the triplets and negative quartets was strongly advised, and often makes the difference between success and failure.

The above observations suggest the following conclusions about the role of (10): (*a*) the large-cross quartets are expected to be as reliable as triplets, but their use is not advised in the phasing process. For the sake of brevity, we do not check the formula (10) on this type of quartet; (*b*) the small-cross quartets are expected to be less reliable than the triplets but their combined use with triplet relationships in the phasing process may be advisable. Thus, we will focus our tests on the small-cross quartets alone. They are generated as the sum of two psi-zero triplets by means of the program recently described by Giacovazzo, Siliqi & Platas (1995). Triplets are found by combining two of the about 800 reflections with the largest values of $|\Delta'|$, with one reflection chosen among those having $|\Delta'| \leq 0.3$.

We apply (10) to calculated data in order to check its efficiency in case of perfect isomorphism and in the

Table 3. *APP: statistical calculations for small-cross quartet and triplet invariants*

Calculated error-free data for native and derivative structures are used. For the quartets, the reliability parameters A and A_c as given by (10) and (16) are employed; for the triplet estimation, the GCZ formula is used. NR is the number of phase relationships having $|A|$ or $|A_c|$ larger than ARG, % is the percentage of phase relationships whose cosine sign is correctly estimated and $\langle|\Phi|^\circ\rangle$ is the average of the absolute values of the triplet or quartet phase Φ.

	Positive quartet [equation (10)]			Negative quartet [equation (10)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.4	9615	74.0	62	10385	72.0	115
0.8	7994	75.1	62	8292	73.2	116
1.6	2280	83.0	53	2426	79.9	124
3.2	1215	90.7	45	11238	89.5	135
5.5	5	100.0	30	3	100.0	134
9.0	–	–	–	–	–	–

	Positive triplets (GCZ)			Negative triplets (GCZ)		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
1.6	28923	76.3	60	21077	75.5	120
3.2	15655	79.7	55	8583	83.2	129
5.5	5026	87.9	46	2167	92.4	140
9.0	684	98.2	30	218	100.0	155

	Positive quartet [equation (16)]			Negative quartet [equation (16)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
1.6	9483	79.1	57	10517	77.1	121
3.2	3403	83.2	52	3626	80.4	125
5.5	850	90.7	46	946	89.7	134
9.0	34	100.0	24	39	100.0	152

Table 4. *CARP: statistical calculations for small-cross quartet and triplet invariants*

See Table 3 for the description of the protocol used for the calculations.

	Positive quartet [equation (10)]			Negative quartet [equation (10)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.2	2569	36.4	105	2512	35.5	74
0.4	297	11.1	135	294	9.9	44
0.8	1	0.0	110	2	0.0	46

	Positive triplets (GCZ)			Negative triplets (GCZ)		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.4	16963	91.3	42	12511	90.4	137
0.8	10537	95.2	37	4152	97.8	145
1.6	181	100.0	17	3	100.0	177

	Positive quartet [equation (16)]			Negative quartet [equation (16)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.2	10147	98.2	32	9853	97.8	147
0.4	4885	98.1	34	4813	97.6	147
0.8	2	100.0	8	4	100.0	169

absence of experimental errors in the measurements. Since triplet and quartet relationships could co-work in the phasing process, it is also useful to compare triplet (as

Table 5. *E2: statistical calculations for small-cross quartet and triplet invariants*

See Table 3 for the description of the protocol used for the calculations.

	Positive quartet [equation (10)]			Negative quartet [equation (10)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.2	10009	64.8	73	9991	65.8	108
0.4	2743	68.0	69	2639	69.2	113
0.8	92	78.3	51	72	76.4	124

	Positive triplets (GCZ)			Negative triplets (GCZ)		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
1.2	25208	89.2	44	22078	89.4	136
2.0	2313	95.3	35	1933	95.3	145
2.6	303	99.0	30	286	98.6	152

	Positive quartet [equation (16)]			Negative quartet [equation (16)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.4	10048	79.1	57	9952	80.1	125
0.8	1901	89.1	45	1845	90.5	136
1.2	258	97.7	31	242	92.6	145
1.6	33	100.0	25	20	100.0	173

Table 6. *M-FABP: statistical calculations for small-cross quartet and triplet invariants*

See Table 3 for the description of the protocol used for the calculations.

	Positive quartet [equation (10)]			Negative quartet [equation (10)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.2	2339	63.7	72	2414	65.1	109
0.4	375	64.3	69	369	66.9	112
0.8	9	44.4	84	7	42.9	69

	Positive triplets (GCZ)			Negative triplets (GCZ)		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.4	21132	96.2	32	16691	96.5	146
1.2	8086	98.9	28	4818	98.7	152
2.0	968	100.0	20	551	100.0	160

	Positive quartet [equation (16)]			Negative quartet [equation (16)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.2	9981	97.4	30	10019	97.4	150
0.4	7861	98.2	29	7825	98.0	151
0.8	1956	99.8	23	2033	99.9	156
1.2	95	100.0	17	112	100.0	163

estimated by GCZ) and quartet reliability. The results are shown in Tables 3–6. We observe the following.

(*a*) In accordance with our definition of small-cross quartets, we omit from the tables all the quartets for which

$$[\sigma_4/\sigma_2^2]_H + X[\sigma_3/\sigma_2^{3/2}]_H^2 X[\langle L_5'\rangle + \langle L_6'\rangle + \langle L_7'\rangle] \geq 0.$$

The number of small-cross quartets turns out to be markedly smaller than the number of large-cross quartets.

Table 7. *APP: statistical calculations for small-cross quartet and triplet invariants*

Experimental data for native and derivative structures are used. For the description of the protocol used for the calculations, see Table 3.

	Positive quartet [equation (10)]			Negative quartet [equation (10)]		
ARG	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$
0.8	9783	39.9	102	10217	40.9	80
1.6	5407	37.1	105	5856	39.0	77
3.2	812	30.7	113	1040	36.5	74
5.5	44	40.9	106	74	40.5	78
9.0	2	50.0	74	5	40.0	82

	Positive triplets (GCZ)			Negative triplets (GCZ)		
ARG	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$
1.6	7725	75.9	61	5420	73.3	118
3.2	3720	79.9	56	1880	76.8	121
5.5	612	85.9	48	226	76.5	125
9.0	39	94.9	36	2	100.0	180

	Positive quartet [equation (16)]			Negative quartet [equation (16)]		
ARG	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$
1.6	3075	68.5	69	3391	68.9	112
3.2	1400	70.9	66	1497	72.1	116
5.5	458	74.5	61	498	76.5	120
9.0	91	90.1	48	92	82.6	127

Table 8. *CARP: statistical calculations for small-cross quartet and triplet invariants*

Experimental data for native and derivative structures are used. For the description of the protocol used for the calculations, see Table 3.

	Positive quartet [equation (10)]			Negative quartet [equation (10)]		
ARG	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$
0.2	4292	41.3	100	4192	40.3	79
0.8	1013	37.2	104	999	39.7	78
1.6	135	27.4	116	116	40.5	75
2.6	23	30.4	122	18	38.9	82

	Positive triplets (GCZ)			Negative triplets (GCZ)		
ARG	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$
1.6	14429	74.6	61	12458	73.6	117
3.2	4360	74.3	61	3361	72.3	116
5.5	322	74.2	61	251	63.7	108

	Positive quartet [equation (16)]			Negative quartet [equation (16)]		
ARG	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$
1.2	8034	67.0	70	8032	66.0	108
2.0	3739	68.1	69	3624	66.5	109
3.2	1235	70.2	69	1161	68.1	112
5.5	183	71.0	69	163	60.1	106

(*b*) The reliability of the small-cross quartets is inferior to the triplet reliability. Their number is often markedly smaller than the number of triplets, in agreement with our expectations.

(*c*) For CARP, the small-cross quartets completely fail: the number of cosine signs incorrectly estimated is larger than the number of the correct ones. This cannot be understood if we do not pay attention to the assumptions that implicitly allow one to obtain (5) from (4).

Table 9. *E2: statistical calculations for small-cross quartet and triplet invariants*

Experimental data for native and derivative structures are used. For the description of the protocol used for the calculations, see Table 3.

	Positive quartet [equation (10)]			Negative quartet [equation (10)]		
ARG	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$
0.2	606	55.3	84	624	51.1	92
0.4	46	56.5	80	64	59.4	103
0.8	1	0.0	175	–	–	–

	Positive triplets (GCZ)			Negative triplets (GCZ)		
ARG	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$
0.8	25380	73.8	62	22057	73.6	117
1.2	5032	80.0	55	3819	79.9	125
1.6	805	85.5	49	624	86	131

	Positive quartet [equation (16)]			Negative quartet [equation (16)]		
ARG	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$	NR	%	$\langle\lvert\Phi^\circ\rvert\rangle$
0.2	10012	61.4	77	9988	61.0	103
0.4	2331	66.2	71	2265	67.8	110
0.8	83	90.4	43	91	80.2	123

7. About the use of the prior information on the heavy-atom positions

The marginal distribution

$$P(\phi_1, \phi_2, \phi_3, \phi_4 | R_1, \ldots, S_1, \ldots, S_7)$$

[see (5)] was obtained in the absence of any prior information by integrating the distribution

$$P(\phi_1, \ldots, \phi_7, \psi_1, \ldots, \psi_7 | R_1, \ldots, R_7, S_1 \ldots, S_7) \quad (12)$$

over the ten variables ϕ_5, ϕ_6, ϕ_7, ψ_1, …, ψ_7. Such a mathematical operation can be described in a qualitative way as follows. Since (Hauptman, 1982)

$$\langle \cos(\phi_i - \psi_i) \rangle = D_{1i} \equiv D'_{1i}, \quad (13)$$

the integration over ψ_5, ψ_6, ψ_7 is equivalent to replacing

$$\cos(\phi_1 + \phi_2 + \phi_3 + \psi_4) \text{ by its expected value}$$
$$D'_{14} \cos(\phi_1 + \phi_2 + \phi_3 + \phi_4)$$
$$\vdots$$
$$\cos(\psi_1 + \psi_2 + \psi_3 + \psi_4) \text{ by its expected value}$$
$$D'_{11} D'_{12} D'_{13} D'_{14} \cos(\phi_1 + \phi_2 + \phi_3 + \phi_4).$$

The reliability of (13) varies according to the value of $2R'_iS'_i$; indeed,

$$\mathrm{var}[\cos(\phi_i - \psi_i)] = \tfrac{1}{2} + \tfrac{1}{2} D'_{2i} - D'^2_{1i},$$

where $D'_{2i} = I_2(2R'_iS'_i)/I_0(2R'_iS'_i)$. For large values of $(2R'_iS'_i)$, the experimental value of $\cos(\phi_i - \psi_i)$ is

Table 10. *M-FABP: statistical calculations for small-cross quartet and triplet invariants*

Experimental data for native and derivative structures are used. For the description of the protocol used for the calculations, see Table 3.

	Positive quartet [equation (10)]			Negative quartet [equation (10)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.4	10096	47.3	93	9904	48.5	88
0.8	3387	48.1	92	3361	48.6	88
1.6	519	50.5	91	538	49.4	89
3.2	30	50.0	86	23	69.6	100

	Positive triplets (GCZ)			Negative triplets (GCZ)		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
1.2	13364	64.7	73	11656	61.8	104
2.0	8116	66.4	71	6536	63.2	105
3.2	1718	69.8	67	1381	63.1	107
4.4	314	71.3	64	239	63.2	105

	Positive quartet [equation (16)]			Negative quartet [equation (16)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.4	9909	52.7	87	10091	52.7	93
1.2	4300	53.2	86	4556	54.2	95
2.0	979	54.3	85	988	52.9	93
3.2	165	49.7	89	175	49.7	89
4.4	62	51.6	82	54	51.9	85

expected to be close to D'_{1i} so that the reliability parameter (10) will result in a useful discriminator. This may not occur if $(2R'_iS'_i)$ is small: in this case, the passage from (4) to (5) will make the reliability of the distribution deteriorate. A typical case in which (10) fails is depicted in Fig. 1: $|\Delta'| = |S' - R'|$ is small but the normalized structure factor of the heavy-atom structure has large modulus so that (13) is not obeyed.

The passage from (4) to (5) can be performed without loss of information if the heavy-atom structure is *a priori* known. Since (see Fig. 1 again)

$$|E_H|^2 = S'^2 + R'^2 - 2R'S'\cos(\phi_i - \psi_i), \qquad (14)$$

we can integrate (12) over ψ_5, ψ_6, ψ_7 by constraining them to satisfy (14). The final formula is

$$P(\Phi|R_1, \ldots, R_7, S_1, \ldots, S_7, |E_H|_5, \ldots, |E_H|_7) \simeq [2\pi I_0(A_c)]^{-1}\exp(A_c\cos\Phi), \qquad (15)$$

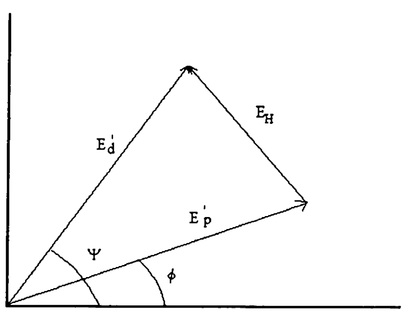

Fig. 1 The vectorial relationship between E'_p, E'_d and E_H.

Table 11. *Statistical calculations for small-cross quartet when in equation (16) the constraint (17) is used (experimental data)*

(*a*) APP

	Positive quartets [equation (10)]			Negative quartets [equation (10)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
1.6	3301	70.5	69	3621	71.6	114
3.2	1423	74.1	66	1577	75.7	119
5.5	156	84.6	61	181	86.7	131
9.0	–	–	–	5	80.0	142

(*b*) CARP

	Positive quartets (GCZ)			Negative quartets (GCZ)		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
1.2	6529	69.2	68	6453	68.4	111
2.0	2274	73.1	64	2247	70.9	114
3.2	619	76.9	60	655	73.7	116
5.5	29	79.3	61	32	75.0	113

(*c*) E2

	Positive quartets [equation (16)]			Negative quartets [equation (16)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.2	9921	65.8	72	10079	65.6	108
0.4	2103	73.2	63	2224	74.8	118
0.8	80	92.5	7	78	87.2	127

(*d*) M-FABP

	Positive quartets [equation (16)]			Negative quartets [equation (16)]		
ARG	NR	%	$\langle\|\Phi^\circ\|\rangle$	NR	%	$\langle\|\Phi^\circ\|\rangle$
0.4	9916	55.4	83	10084	54.9	96
1.2	2000	58.4	80	1993	57.7	95
2.0	287	55.1	84	268	54.5	93
3.2	42	57.1	76	47	63.8	89
4.4	6	66.7	67	13	69.2	85

$$A_c = [2\Delta'_1\Delta'_2\Delta'_3\Delta'_4/(1 + B'_c)]\{[\sigma_4/\sigma_2^2]_H + [\sigma_3/\sigma_2^{3/2}]_H^2[\varepsilon_{H5} + \varepsilon_{H6} + \varepsilon_{H7}]\}, \qquad (16)$$

where

$$B'_c = \tfrac{1}{2}[\sigma_3/\sigma_2^{3/2}]_H^2(\varepsilon_{H1}\varepsilon_{H2}\varepsilon_{H5} + \varepsilon_{H3}\varepsilon_{H4}\varepsilon_{H5} + \ldots + \varepsilon_{H2}\varepsilon_{H3}\varepsilon_{H7}),$$

$$\varepsilon_{Hi} = |E_{Hi}|^2 - 1.$$

The efficiency of (16) may be deduced from Tables 3–6. We note: (*a*) quartets estimated *via* (16) are as reliable as triplets estimated *via* the GCZ formula; (*b*) (16) is a much more efficient reliability parameter than (10) in all the cases. In particular, while (10) fails for CARP, (16) succeeds. We can conclude that the cosine sign of a quartet depends on ε_{H5}, ε_{H6}, ε_{H7} rather than on $\langle L'_5\rangle$, $\langle L'_6\rangle$, $\langle L'_7\rangle$. These last parameters are nothing but the expected values of $\varepsilon_{H5}, \varepsilon_{H6}, \varepsilon_{H7}$ in the absence of the prior information on the heavy-atom structure. When $|E_{H5}|, |E_{H6}|, |E_{H7}|$ are remarkably larger than $|\Delta'_5|, |\Delta'_6|, |\Delta'_7|$, respectively, then the reliability parameter (10) will suffer by a statistical bias and will fail. In order to

Table 12. *Values for parameters in equation (19)*

Structure code	$[\sigma_3/\sigma_2^{3/2}]_H^2$ $\{[\sigma_2]_p/[\sigma_2]_H\}$	$[\sigma_3/\sigma_2^{3/2}]_p[\sigma_3/\sigma_2^{3/2}]_H$ $\{[\sigma_2]_p^{1/2}/[\sigma_2]_H^{1/2}\}$	$[\sigma_3/\sigma_2^{3/2}]_H^2$
APP	4.39	0.10	0.24×10^{-2}
CARP	11.41	0.089×10^{-1}	0.70×10^{-3}
E2	0.52	0.36×10^{-2}	0.24×10^{-4}
M-FABP	7.78	0.44×10^{-1}	0.24×10^{-3}

confirm the above statement, we have calculated, for all the quartets in Tables 3–6, the average values of $|\Delta_i'|$ and $|E_{Hi}|$ for $i = 5, 6, 7$. We obtain

for APP: $\langle|\Delta'|\rangle = 0.08$ $\langle|E_H|\rangle = 0.23$
for CARP: $\langle|\Delta'|\rangle = 0.12$ $\langle|E_H|\rangle = 0.84$
for E2: $\langle|\Delta'|\rangle = 0.13$ $\langle|E_H|\rangle = 0.52$
for M-FABP: $\langle|\Delta'|\rangle = 0.11$ $\langle|E_H|\rangle = 0.59$.

The above data explain why, for APP, (10) works well and why CARP quartet estimates *via* (10) are not useful. When (16) is used, all the quartets for which

$$[\sigma_4/\sigma_2^2]_H + [\sigma_3/\sigma_2^{3/2}]_H^2[\varepsilon_{H5} + \varepsilon_{H6} + \varepsilon_{H7}] > 0$$

are expelled from the tables [indeed they are estimated positive *via* (16)], so contributing to the reliability of the remaining small-cross quartets.

8. Check of the probabilistic formulas by real diffraction data

The correctness of the quartet parameters (10) and (16) is supported by the statistical calculations quoted in Tables 3–6. However, their efficiency against lack of isomorphism between native and derivative structures, errors in experimental data and/or in their mathematical treatment is not proved so far. In order to do that we checked the formulas with real data. The outcome is shown in Tables 7–10. Note the following.

(*a*) Errors in measurements and lack of isomorphism strongly reduce the efficiency of the triplet estimates (compare Tables 3–6 with Tables 7–10). The deterioration is quite remarkable for M-FABP and CARP, less for APP and E2.

(*b*) Quartets are more sensitive than triplets to lack of isomorphism and errors in measurements. There are two main reasons for this. The first concerns the distorsion of the cross magnitudes: if we calculate, for the quartets in Tables 7–10, the average values of $|\Delta_i'|$ and $|E_{Hi}|$ for $i = 5, 6, 7$, we have:

for APP: $\langle|\Delta'|\rangle = 0.15$ $\langle|E_H|\rangle = 0.88$
for CARP: $\langle|\Delta'|\rangle = 0.15$ $\langle|E_H|\rangle = 0.88$
for E2: $\langle|\Delta'|\rangle = 0.13$ $\langle|E_H|\rangle = 0.70$
for M-FABP: $\langle|\Delta'|\rangle = 0.11$ $\langle|E_H|\rangle = 0.71$.

If the above values are compared with those obtained for calculated data, it is easily understood why the quartet estimates *via* (10) are so poor for the experimental data.

The second reason concerns the so-called 'inversion' of Δ': owing to lack of isomorphism, errors in measurements, treatment of the data *etc.*, the experimental sign of Δ' is opposite to the calculated one (see Giacovazzo, Siliqi & Spagna, 1994, for some statistics). If ν is the sign inversion frequency for the reflexions among which the basis quartet reflexions and the triplet reflexions are picked up, the inversion frequency for the triplet sign and for the quartet sign are approximately given by

$$\nu_T \simeq \nu^3 + 3\nu(1-\nu)^2$$
$$\nu_Q \simeq 4\nu^3(1-\nu) + 4\nu(1-\nu)^3,$$

respectively. For our test data, we found

for APP: $\nu \simeq 0.07$, $\nu_T = 0.16$, $\nu_Q = 0.22$
for CARP: $\nu \simeq 0.45$, $\nu_T = 0.38$, $\nu_Q = 0.42$
for E2: $\nu \simeq 0.03$, $\nu_T = 0.07$, $\nu_Q = 0.07$
for M-FABP: $\nu \simeq 0.13$, $\nu_T = 0.27$, $\nu_Q = 0.34$.

(*c*) The application of (10) to real data is never advised in practise.

(*d*) The reability parameter (16) guarantees more reliable estimates of the quartets even when (10) completely fails.

So far, we have used the prior information on the heavy-atom structure to modify (10) into (16). However, we can introduce a supplementary condition into (16): since $|\Delta_i'| \le |E_{Hi}|$ by definition, we can apply the following constraint to the basis magnitudes of the quartets:

$$\text{if} \quad |\Delta_i'| > |E_{Hi}|, \quad \text{then} \quad |\Delta_i'| = |E_{Hi}| \quad \text{for} \quad i = 1, \dots, 4. \qquad (17)$$

Table 11 shows the quartet statistics for the four test structures when (17) is applied. It is easily seen that the supplementary constraints improve the efficiency of (16) and make it suitable for practical applications.

9. Conclusions

Two probabilistic formulas have been obtained for estimating quartet phases. The first exploits the seven pairs of isomorphous reflexions (R_i, S_i), for $i = 1, \dots, 7$, the second benefits from the prior knowledge of the heavy-atom structure. While both formulas are reliable for error-free data, only the second type is sufficiently robust, against lack of isomorphism and experimental errors, to be safely applied to real experimental data. The theoretical reasons justifying such behaviour have been described. The paper also analyses the rule of the quartet invariants in a direct procedure aiming at phasing protein reflexions *via* native and derivative data. Emphasis has been given to the so-called small-cross quartets since they provide information statistically independent of that supplied by triplet invariants.

APPENDIX *A*

According to the definition in the text,

$$\begin{aligned}\beta_{1234\bar{i}} = {} & \left[\prod_{i=1}^{4}(1-\alpha_i^2)^{-1}\right]\{\gamma_{12\bar{i}}\gamma_{34\bar{i}} - \gamma_{12\bar{i}}\gamma_{3\bar{4}\bar{i}}\alpha_4 \\ & - \ldots + \gamma_{12\bar{i}}\gamma_{\bar{3}\bar{4}\bar{i}}\alpha_3\alpha_4 + \ldots + \gamma_{\bar{1}\bar{2}\bar{i}}\gamma_{34\bar{i}}\alpha_1\alpha_2 \\ & - \gamma_{1\bar{2}\bar{i}}\gamma_{\bar{3}\bar{4}\bar{i}}\alpha_2\alpha_3\alpha_4 - \gamma_{\bar{1}2\bar{i}}\gamma_{\bar{3}\bar{4}i}\alpha_1\alpha_3\alpha_4 \\ & - \gamma_{\bar{1}\bar{2}\bar{i}}\gamma_{\bar{3}4\bar{i}}\alpha_1\alpha_2\alpha_3 - \gamma_{\bar{1}\bar{2}\bar{i}}\gamma_{3\bar{4}\bar{i}}\alpha_1\alpha_2\alpha_4 \\ & + \gamma_{\bar{1}\bar{2}\bar{i}}\gamma_{\bar{3}\bar{4}\bar{i}}\alpha_1\alpha_2\alpha_3\alpha_4\} \\ = {} & \left[\prod_{i=1}^{4}(1-\alpha_i^2)^{-1}\right][\sigma_3/\sigma_2^{3/2}]_p^2\alpha_i^2 \\ & \times\{1-\alpha_4^2-\alpha_3^2-\alpha_2^2-\alpha_1^2\} \\ & + \left[\prod_{i=1}^{4}(1-\alpha_i^2)^{-1}\right][\sigma_3/\sigma_2^{3/2}]_p\alpha_i^2 \\ & \times\{\alpha_3^2\alpha_4^2-\alpha_2^2\alpha_3^2\alpha_4^2-\alpha_1^2\alpha_3^2\alpha_4^2\} \\ & \times\left\{[\sigma_3/\sigma_2^{3/2}]_p + [\sigma_3/\sigma_2^{3/2}]_H[\sigma_2]_H^{3/2}/[\sigma_2]_p^{3/2}\right\} \\ & + \left[\prod_{i=1}^{4}(1-\alpha_i^2)^{-1}\right][\sigma_3/\sigma_2^{3/2}]_p\alpha_i^2 \\ & \times\{\alpha_1^2\alpha_2^2-\alpha_1^2\alpha_2^2\alpha_4^2-\alpha_1^2\alpha_2^2\alpha_3^2\} \\ & \times\left\{[\sigma_3/\sigma_2^{3/2}]_p + [\sigma_3/\sigma_2^{3/2}]_H[\sigma_2]_H^{3/2}/[\sigma_2]_p^{3/2}\right\} \\ & + \left[\prod_{i=1}^{4}(1-\alpha_i^2)^{-1}\right]\alpha_1^2\alpha_2^2\alpha_3^2\alpha_4^2\left\{[\sigma_3/\sigma_2^{3/2}]_p\right. \\ & \left.+ [\sigma_3/\sigma_2^{3/2}]_H[\sigma_2]_H^{3/2}/[\sigma_2]_p^{3/2}\right\}^2 \\ = {} & \alpha_i^2\left([\sigma_3/\sigma_2^{3/2}]_p^2 + [\sigma_3/\sigma_2^{3/2}]_p[\sigma_3/\sigma_2^{3/2}]_H\right. \\ & \times\left\{[\sigma_2]_H^{3/2}/[\sigma_2]_p^{3/2}\right\}[\alpha_i^2/(1-\alpha_i^2)]^2 \\ & \left.+ \left[\prod_{i=1}^{4}\alpha_i^2/(1-\alpha_i^2)\right][\sigma_3/\sigma_2^{3/2}]_H^2\left\{[\sigma_2]_H^3/[\sigma_2]_p^3\right\}\right) \\ = {} & \alpha_i^2\left([\sigma_3/\sigma_2^{3/2}]_p^2 + [\sigma_3/\sigma_2^{3/2}]_p[\sigma_3/\sigma_2^{3/2}]_H\right. \\ & \left.\times\left\{[\sigma_2]_p^{1/2}/[\sigma_2]_H^{1/2}\right\} + [\sigma_3/\sigma_2^{3/2}]_H^2\{[\sigma_2]_p/[\sigma_2]_H\}\right).\end{aligned} \tag{18}$$

For the usual pairs, native-protein–heavy-atom derivative,

$$\begin{aligned}&[\sigma_3/\sigma_2^{3/2}]_H^2\{[\sigma_2]_p/[\sigma_2]_H\} \\ &\quad\gg [\sigma_3/\sigma_2^{3/2}]_p[\sigma_3/\sigma_2^{3/2}]_H\{[\sigma_2]_p^{1/2}/[\sigma_2]_H^{1/2}\} \\ &\quad\gg [\sigma_3/\sigma_2^{3/2}]_p^2.\end{aligned} \tag{19}$$

The reader can find in Table 12 the values of the various parameters in the inequality (19) calculated for the test structures. Therefore, we can approximate (18) to

$$B_{1234\bar{i}} = \alpha_i^2[\sigma_3/\sigma_2^{3/2}]_H^2\{[\sigma_2]_p/[\sigma_2]_H\}. \tag{20}$$

Since (see the main text)

$$B_{1234i} = [\sigma_3/\sigma_2^{3/2}]_p^2,$$

B_{1234i} may be neglected with respect to $B_{1234\bar{i}}$.

In an analogous way, it may be verified that

$$\begin{aligned}B_{1234i}^{\mathrm{mod}(1)} = {} & \alpha_i\big([\sigma_3/\sigma_2^{3/2}]_p^2 + [\sigma_3/\sigma_2^{3/2}]_p[\sigma_3/\sigma_2^{3/2}]_H^2 \\ & \times\left\{[\sigma_2]_p^{1/2}/[\sigma_2]_H^{1/2}\right\}\big).\end{aligned} \tag{21}$$

In accordance with (19), $B_{1234i}^{\mathrm{mod}(1)}$ may be neglected when compared with $B_{1234\bar{i}}$.

References

Burla, M. C., Cascarano, G. & Giacovazzo, C. (1992). *Acta Cryst.* A**48**, 906–912.

Giacovazzo, C. (1976). *Acta Cryst.* A**32**, 91–99.

Giacovazzo, C. (1980). *Direct Methods in Crystallography.* London: Academic Press.

Giacovazzo, C., Burla, M. C. & Cascarano, G. (1992). *Acta Cryst.* A**48**, 901–906.

Giacovazzo, C., Cascarano, G. & Zheng, C. (1988). *Acta Cryst.* A**44**, 45–51.

Giacovazzo, C. & Siliqi, D. (1996). *Acta Cryst.* A**52**, 133–142.

Giacovazzo, C., Siliqi, D. & Platas, J. G. (1995). *Acta Cryst.* A**51**, 811–820.

Giacovazzo, C., Siliqi, D. & Spagna, R. (1994). *Acta Cryst.* A**50**, 609–621.

Glover, I., Haneef, I., Pitts, J., Wood, S., Moss, D., Tickle, I. & Blundell, T. L. (1983). *Biopolymers*, **22**, 293–304.

Hauptman, H. (1982). *Acta Cryst.* A**38**, 289–294.

Kretsinger, R. H. & Nockolds, C. E. (1973). *J. Biol. Chem.* **248**, 3313–3326.

Mattevi, A., Obmolova, G., Schulze, E., Kalk, K. H., Westphal, A. H., De Kok, A. & Hol, W. G. J. (1992). *Science*, **255**, 1544–1550.

Zanotti, G., Scapin, G., Spadon, P., Veerkamp, J. H. & Sacchettini, J. C. (1992). *J. Biol. Chem.* **267**, 18541–18550.