Acta Cryst. (1996). A**52**, 77–87

The Estimation of Four-Phase Structure Invariants Using the Single Difference of Isomorphous Structure Factors

Christos E. Kyriakidis,* René Peschar and Henk Schenk

Laboratory for Crystallography, Amsterdam Institute for Molecular Studies, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands. E-mail: chrisk@crys.chem.uva.nl

(Received 8 June 1993; accepted 10 August 1995)

Abstract

The recently introduced approach of using the difference between isomorphous structure factors as a random variable in the derivation of joint probability distributions of isomorphous structure factors has been extended to secure the conditional joint probability distribution of the quartet phase sums present among isomorphous structure factors. It is shown for calculated data sets (native and heavy-atom derivative) of the proteins avian pancreatic polypeptide and cytochrome c that, with single-wavelength anomalous-scattering data, reliable estimates can be obtained for the quartet phase sums, even if the estimates are based on the structure-factor differences of the four quartet main-term reflections only.

1. Introduction

The crystal structure of relatively small molecules is readily determined from the diffraction intensities by means of direct methods (Schenk, 1991). Direct methods have been less successful in solving macromolecular structures but in the last decade it was acknowledged than an efficient direct-methods solution of macromolecular structures should allow the simultaneous utilization of various sources of phasing information. This led to the development of expressions based on the joint probability distribution of isomorphously related structure factors (Hauptman, 1982*a*,*b*; Giacovazzo, 1983; Giacovazzo, Cascarano & Zheng, 1988; Fortier & Nigam, 1989; Peschar & Schenk, 1991). The test results were encouraging, though not unexpected, since more data were involved (Hauptman, Potter & Weeks, 1982; Hauptman, 1982*b*; Giacovazzo, 1983; Furey, Chandrasekhar, Dyda & Sax, 1990). The probabilistic approach leading to these expressions relies on using individual random variables for heavily correlated (normalized) structure factors, *e.g.* F_H and F_{-H}, in the presence of a few anomalous scatterers. Since both individual structure factors are a function of all N atoms and because of their heavy correlation, the final probabilistic quantities turn out to be complicated functions of the scattering factors of all N atoms [see, for example, the definition of functions (3.20)–(3.37) in Hauptman (1982*b*) and the functions in §3 of Giacovazzo (1983)]. Recently, a different approach was taken by Kyriakidis, Peschar & Schenk (1993*c*) by exploiting the difference structure factor F^d_ν of two isomorphous structure factors F^l_ν and F^m_ν as a random variable:

$$\begin{aligned} F^d_\nu &= F^l_\nu - F^m_\nu \\ &= \sum_{j=1}^{N} f^l_{j\nu} \exp[2\pi i \mathbf{H}_\nu \cdot \mathbf{r}_j] \\ &\quad - \sum_{j=1}^{N} f^m_{j\nu} \exp[2\pi i \mathbf{H}_\nu \cdot \mathbf{r}_j] \\ &= \sum_{j=1}^{N} (f^l_{j\nu} - f^m_{j\nu}) \exp[2\pi i \mathbf{H}_\nu \cdot \mathbf{r}_j] \\ &= |F^d_\nu| \exp(i\varphi^d_\nu). \end{aligned} \qquad (1)$$

The subscript ν refers to a particular reflection, the superscripts l and m refer to two individual isomorphous structure factors while d denotes a dependence on the difference between the isomorphous structure factors only. The atomic scattering factors include anomalous-dispersion corrections,

$$\begin{aligned} f_{j\nu} \equiv f_j(H_\nu) &= f^0_{j\nu} + f'_j + if''_j \\ &= f^r_j + if''_j \\ &= |f_{j\nu}| \exp(i\delta_{j\nu}). \end{aligned} \qquad (2)$$

Both the magnitude $|F^d_\nu|$ and the phase φ^d_ν of F^d_ν are functions of the magnitudes and phases of F^l_ν and F^m_ν. From (1), it follows that

$$|F^d_\nu|^2 = |F^l_\nu|^2 + |F^m_\nu|^2 - 2|F^l_\nu||F^m_\nu| \cos(\psi^d_\nu) \qquad (3)$$

with the doublet phase sum

$$\begin{aligned} \psi^d_\nu &= \varphi^l_\nu + s^d \varphi^m_\nu \\ s^d &= \begin{cases} -1 & \text{if } H^l = H^m \\ +1 & \text{if } H^l = -H^m. \end{cases} \end{aligned} \qquad (4)$$

Various combinations of two isomorphously related structure factors can be defined, *e.g.* F_H and F^*_{-H} (* means complex conjugation) if a few anomalous scatterers are present. F^d_ν is defined in such a way that only those (n) atoms that have an appreciably non-zero atomic scattering-factor difference in (1) will contribute to the final probabilistic expression. With F_v^d as a random variable, in effect a reduction of N to n is achieved, which is expected to improve the quality of the estimates since, in the case of normal diffraction data (non-isomorphous and no anomalous scattering), the reliability of the triplet and quartet phase-sum estimates is a function of $N^{-(1/2)}$ and N^{-1}, respectively. An additional advantage of the difference-structure-factor approach is that the mathematical calculations are simplified. It has been shown for calculated structure-factor data that, by taking the F_v^d as a random variable, reliable estimates of the triplet phase sums present among isomorphous data sets can be obtained, even if the diffraction ratio is small (Kyriakidis, Peschar & Schenk, 1993*a*). An additional improvement is achieved by supplementing the estimation of the doublet phase sums with vectors from a difference Patterson synthesis (Kyriakidis, Peschar & Schenk, 1993*b*,*c*).

In the current paper, the difference-structure-factor approach is extended in order to obtain estimates of quartet phase sums present amongst isomorphous structure factors. These quartets are expected to be important for the application of direct methods in macromolecular crystallography (Sheldrick, 1993).

2. The quartet phase sum in direct methods

The three-dimensional quartet phase-sum relation

$$\psi_{1234} = \varphi_1 + \varphi_2 + \varphi_3 + \varphi_4 \qquad (5)$$

with the subscripts 1 to 4 referring to four reflections H_1, H_2, H_3 and $H_4 = -H_1 - H_2 - H_3$, whose indices add to 0, was introduced by Hauptman & Karle (1953) as being potentially more appropriate in solving three-dimensional structures than the two-dimensional triplet phase sum

$$\psi_{123} = \varphi_1 + \varphi_2 + \varphi_3, \qquad (6)$$

in which the subscripts 1 to 3 refer to three reflections H_1, H_2 and $H_3 = -H_1 - H_2$, whose indices add up to 0. Some years later, Simerska (1956) derived the quartet relationship from a generalization of the Sayre–Hughes equation for products of three reflections instead of two. Both Hauptman & Karle and Simerska showed that (5) lies more probably near zero for larger values of

$$E_q = |E_1E_2E_3E_4|N^{-1}. \qquad (7)$$

The triplet relationship (6) is expected to be estimated more reliably because the E_t values,

$$E_t = |E_1E_2E_3|N^{-1/2}, \qquad (8)$$

which determine the reliability of the triplet estimation, are in general larger than the E_q values, which depend on N^{-1} only. Therefore, quartets were not used for practical purposes until Schenk (1973*a*) showed that quartets can also be formed by summing two triplets with one phase in common. In this way, quartet (5) depends not only on $|E_1|$, $|E_2|$, $|E_3|$ and $|E_4|$ but also on the so-called cross terms $|E_5|$ $(H_5 = H_1 + H_2)$, $|E_6|$ $(H_6 = H_1 + H_3)$ and $|E_7|$ $(H_7 = H_2 + H_3)$. It was shown that quartets with large cross-term magnitudes most probably lie near 0, while quartets with small cross-term magnitudes are expected to lie near π (Schenk, 1973*a*,*b*; Schenk & De Jong, 1973; Schenk, 1974; Hauptman, 1974). This new point of view led to the development of improved joint probability distributions for estimating the quartet phase sum (Hauptman, 1975*a*,*b*, 1976; Giacovazzo, 1975; Giacovazzo, 1976*a*,*b*) and later on to the formulation of the neighbourhood principle (Hauptman, 1975*b*) and the representation theory (Giacovazzo, 1977*b*). The latter theories identify those structure factors upon which the phase sum of a structure (sem)invariant most sensitively depends. In practice, the approaches of Hauptman and Giacovazzo are closely related and often lead to the same results in spite of different starting points (Heinerman, 1977; Giacovazzo, 1977*a*). As an alternative to the closed exponential expressions of order N^{-1}, Peschar (1987) investigated the incorporation of higher-order terms in the series-expansion form of the joint probability distribution. A comparison with the results of Hauptman and Giacovazzo showed that the estimation based on the series expansion lies systematically in between those of the distributions of Hauptman and Giacovazzo which underestimate and overestimate, respectively, the quartets to be 0 or π.

In direct-method routines, quartets have been applied in particular to starting-set procedures and figures of merit (Schenk, 1973*a*; Schenk & De Jong, 1973; Schenk, 1974; De Titta, Edmonds, Langs & Hauptman, 1975; Gilmore, 1977; van der Putten & Schenk, 1979; Freer & Gilmore, 1980; Cascarano, Giacovazzo & Viterbo, 1987). More recently, Sheldrick (1993) used quartets together with triplets to solve some small macromolecules.

2.1. *The joint probability distribution and the conditional probability distribution of the quartet phase sum in the presence of anomalous scattering*

2.1.1. *Four structure factors.* As indicated above, the probability distribution of the quartet phase sum involving the structure factors of the four main-term reflections only (H_1, H_2, H_3 and $H_4 = -H_1 - H_2 - H_3$) has not been used extensively because it was clearly inferior to the seven-structure-factor expressions. However, as will be discussed later in this paper, the four-structure-factor expression is an important starting point to obtain a conditional joint probability expression of the quartet phase sum present among two isomorphous data sets. Let us denote by R_i and Φ_i the random variables for the structure-factor magnitude $|F_i|$ and phase φ_i, respectively. If, in the structure-factor expressions, complex-valued atomic scattering factors are allowed for, the joint probability expression of the magnitudes R_1, R_2, R_3 and R_4 and the phases Φ_1, Φ_2, Φ_3 and Φ_4 of the four quartet main-term structure factors can be expressed as

$$P(\Phi_1,\ldots,\Phi_4,R_1,\ldots,R_4) \propto \exp[2W_{1234}\cos(\Phi_1+\Phi_2+\Phi_3+\Phi_4+\Delta_{1234})], \quad (9)$$

$$W_{1234}=R_1R_2R_3R_4|Z_{1234}|,$$

which involves the following definitions:

$$z_v=\sum_{j=1}^{N}|f_{jv}|^2 \quad (10)$$

$$Z_{1234}=|Z_{1234}|\exp[i\Delta_{1234}] = (z_1z_2z_3z_4)^{-1}\sum_{j=1}^{N}(f_{j1}f_{j2}f_{j3}f_{j4})^*, \quad (11)$$

with * being the complex conjugation and the atomic scattering factors are as defined in (2). With the random variable for the quartet phase sum ψ_{1234} defined to be Ψ_{1234},

$$\Psi_{1234}=\Phi_1+\Phi_2+\Phi_3+\Phi_4, \quad (12)$$

the conditional probability distribution of Ψ_{1234} given R_1, R_2, R_3 and R_4 can be expressed as

$$P(\Psi_{1234}|R_1,R_2,R_3,R_4) = L_4^{-1}\exp[2W_{1234}\cos(\Psi_{1234}+\Delta_{1234})] \quad (13)$$

with L_4 a normalization constant. From (13), an expectation value for Ψ_{1234} is readily obtained as

$$\langle\exp[i\Psi_{1234}]\rangle=\mathrm{Br}(2W_{1234})\exp[-i\Delta_{1234}] \quad (14)$$

with Br(x) the ratio $I_1(x)/I_0(x)$ of the modified Bessel functions I_1 and I_0. The distribution of Ψ_{1234} is centred around $-\Delta_{1234}$ and Br($2W_{1234}$) acts as a statistical weight.

2.1.2. *Seven-structure-factor expression.* The joint probability distribution of the four main-term and three cross-term structure factors (*e.g.* Hauptman, 1975*a*, 1976) is readily generalized if the atomic scattering factors are complex valued,

$$\begin{aligned}P(R_1,\ldots,R_7,\Phi_1,\ldots,\Phi_7) \propto \exp[&2W_{125}\cos(\Phi_1+\Phi_2-\Phi_5+\Delta_{125})\\ &+2W_{345}\cos(\Phi_3+\Phi_4+\Phi_5+\Delta_{345})\\ &+2W_{136}\cos(\Phi_1+\Phi_3-\Phi_6+\Delta_{136})\\ &+2W_{246}\cos(\Phi_2+\Phi_4+\Phi_6+\Delta_{246})\\ &+2W_{237}\cos(\Phi_2+\Phi_3-\Phi_7+\Delta_{237})\\ &+2W_{147}\cos(\Phi_1+\Phi_4+\Phi_7+\Delta_{147})\\ &+2W_{1\text{-}7}\cos(\Phi_1+\Phi_2+\Phi_3+\Phi_4+\Delta_{1\text{-}7})] \quad (15)\end{aligned}$$

with

$$W_{abc}=R_aR_bR_c|Z_{abc}|, \quad (16)$$

$$Z_{abc}=|Z_{abc}|\exp[i\Delta_{abc}]=(z_az_bz_c)^{-1}\sum_{j=1}^{N}(f_{ja}f_{jb}f_{jc})^*, \quad (17)$$

$$W_{1\text{-}7}\exp(i\Delta_{1\text{-}7})=R_1R_2R_3R_4Z_{1\text{-}7} \quad (18)$$

and

$$Z_{1\text{-}7}=|Z_{1\text{-}7}|\exp(i\Delta_{1\text{-}7}) = Z_{1234}-Z_{125}Z_{345}-Z_{136}Z_{246}-Z_{147}Z_{237}. \quad (19)$$

In the case of absent cross terms R_5, R_6 and/or R_7, the W and Z terms containing these quantities should be omitted from (15) and (19). It should be noted that only terms have been included in (15) that are of importance for the subsequent calculations of the conditional joint probability distribution of the quartet phase sum (12).

3. A conditional joint probability of the quartet phase sum

The conditional joint probability distribution of the quartet phase sum given the seven structure-factor magnitudes $R_1,\ldots,R_7$ is often arrived at by integrating over the random variables for the cross-term phases in the joint probability distribution of the structure factors (Hauptman, 1975*b*). When applied to (15), this approach leads to

$$P(\Psi_{1234}|R_1,\ldots,R_7) \propto \exp[2W_{1\text{-}7}\cos(\Psi_{1234}+\Delta_{1\text{-}7})]I_0(2Z_5)I_0(2Z_6)I_0(2Z_7) \quad (20)$$

with Z_5, Z_6 and Z_7 as follows:

$$\begin{aligned}Z_5^2&=W_{125}^2+W_{345}^2+2W_{125}W_{345}\cos(\Psi_{1234}+\Delta_{125}+\Delta_{345})\\ Z_6^2&=W_{136}^2+W_{246}^2+2W_{136}W_{246}\cos(\Psi_{1234}+\Delta_{136}+\Delta_{246})\\ Z_7^2&=W_{147}^2+W_{237}^2+2W_{147}W_{237}\cos(\Psi_{1234}+\Delta_{147}+\Delta_{237}).\end{aligned} \quad (21)$$

If one (or more) of the cross terms R_5, R_6 and R_7 is absent, the corresponding Bessel function I_0 containing this magnitude should be omitted. The application of (21) is not straightforward because a numerical integration is required to get an expectation value for the quartet phase sum. Therefore, we have chosen a different approach.

An efficient technique to get a conditional joint probability distribution of an invariant phase sum starting from the joint probability distribution of the structure factors involved was introduced by Peschar (1987). In the case of the quartet phase sum, the procedure starts from (15) but instead of integrating with respect to the cross-term random variables Φ_5, Φ_6 and Φ_7, the expectation values $\langle\exp(i\Phi_5)\rangle$, $\langle\exp(i\Phi_6)\rangle$ and $\langle\exp(i\Phi_7)\rangle$ are calculated from the triplet terms present and subsequently introduced in (15) as known cross-term phase information. Following this approach, the conditional joint probability of the quartet phase sum becomes

$$P(\Psi_{1234}|R_1,\ldots,R_7)\propto\exp[2G_{1\text{-}7}\cos(\Psi_{1234}+\Lambda_{1\text{-}7})], \quad (22)$$

$$\begin{aligned} G_{1\text{-}7}\exp[i\Lambda_{1\text{-}7}] = {} & W_{1\text{-}7}\exp(i\Delta_{1\text{-}7}) \\ & + \{W_{125}\mathrm{Br}(2W_{345}) \\ & + W_{345}\mathrm{Br}(2W_{125})\}\exp[i(\Delta_{125}+\Delta_{345})] \\ & + \{W_{136}\mathrm{Br}(2W_{246}) \\ & + W_{246}\mathrm{Br}(2W_{136})\}\exp[i(\Delta_{136}+\Delta_{246})] \\ & + \{W_{147}\mathrm{Br}(2W_{237}) \\ & + W_{237}\mathrm{Br}(2W_{147})\}\exp[i(\Delta_{147}+\Delta_{237})]. \end{aligned}$$

3.1. *The conditional probability distribution of the quartet phase sums present among four difference structure factors*

The joint probability theory of structure factors while allowing for complex-valued atomic scattering factors is directly applicable to the difference structure factors of isomorphous structure factors. Let us define R_1^d, R_2^d, R_3^d and R_4^d to be random variables for the four magnitudes $|F_H^d|$, $|F_K^d|$, $|F_L^d|$ and $|F_{-H-K-L}^d|$, respectively, and Ψ_{1234}^d to be the random variable of the difference-structure-factor quartet ψ_{1234}^d,

$$\Psi_{1234}^d = \Phi_1^d + \Phi_2^d + \Phi_3^d + \Phi_4^d. \qquad (23)$$

Analogous to (9)–(14), the conditional joint probability distribution of Ψ_{1234}^d can be expressed as

$$\begin{aligned} & P(\Psi_{1234}^d|R_1^d, R_2^d, R_3^d, R_4^d) \\ & \quad \propto \exp[2W_{1234}^d\cos(\Psi_{1234}^d + \Delta_{1234}^d)], \qquad (24) \end{aligned}$$

$$W_{1234}^d = R_1^dR_2^dR_3^dR_4^d|Z_{1234}^d|,$$

which involves the following definitions:

$$z_\nu^d = \sum_{j=1}^N |f_{j\nu}^l - f_{j\nu}^m|^2 \quad (\text{for } \nu = 1, 2, 3, 4) \qquad (25)$$

$$\begin{aligned} z_{1234}^d &= |z_{1234}^d|\exp[i\Delta_{1234}^d] \\ &= (z_1^dz_2^dz_3^dz_4^d)^{-1}\sum_{j=1}^N\prod_{\nu=1}^4(f_{j\nu}^l - f_{j\nu}^m)^* \qquad (26) \end{aligned}$$

with * denoting complex conjugation.

Expression (24) is a function of both data sets. In order to arrive at an expression that is a function of variables of one data set only, we follow now a procedure introduced by Kyriakidis *et al.* (1993*c*). The product of the random variables R_1^d, R_2^d, R_3^d, R_4^d and $\exp[i\Psi_{1234}^d]$ in terms of the structure factors is

$$R_1^dR_2^dR_3^dR_4^d\exp[i\Psi_{1234}^d] = \prod_{\nu=1}^4(F_\nu^l - F_\nu^m). \qquad (27)$$

The right-hand side of (27) can be written as a sum of the 16 contributing isomorphous quartets,

$$\psi_{1234}^{lmnq} = \varphi_1^l + \varphi_2^m + \varphi_3^n + \varphi_4^q \quad (l, m, n, q = 1, 2). \qquad (28)$$

Each of these 16 terms is now expressed exclusively in one of the 16 isomorphous quartets. After replacing in (27) the doublet phase sums $\exp(i\psi_\nu^d)$ by the estimates $\langle\exp(i\Psi_\nu^d)\rangle$, with Ψ_ν^d the random variable for the doublet phase sum,

$$\langle\exp(i\Psi_\nu^d)\rangle = \langle\exp[i(\Phi_\nu^l + s^d\Phi_\nu^m)]\rangle = \exp(i\lambda_\nu^d) \qquad (29)$$

with

$$\lambda_\nu^d = \cos^{-1}\left(\frac{|F_\nu^l|^2 + |F_\nu^m|^2 - z_\nu^d}{2|F_\nu^l||F_\nu^m|}\right) \quad l, m = 1, 2,$$

the random variable Φ_ν^m is readily expressed in Φ_ν^l and λ_ν^d.

For example, if all terms are expressed in the quartet Ψ_{1234}^{1111}, this leads to

$$\begin{aligned} & (F_1^1 - F_1^2)(F_2^1 - F_2^2)(F_3^1 - F_3^2)(F_4^1 - F_4^2) \\ & \quad = \exp[i\Psi_{1234}^{1111}]\{|F_1^1F_2^1F_3^1F_4^1| - |F_1^1F_2^1F_3^1F_4^2|\exp[-i\lambda_4^d] \\ & \qquad - |F_1^1F_2^1F_3^2F_4^1|\exp[-i\lambda_3^d] \\ & \qquad + |F_1^1F_2^1F_3^2F_4^2|\exp[-i(\lambda_3^d + \lambda_4^d)] \\ & \qquad - |F_1^1F_2^2F_3^1F_4^1|\exp[-i\lambda_2^d] \\ & \qquad + |F_1^1F_2^2F_3^1F_4^2|\exp[-i(\lambda_2^d + \lambda_4^d)] \\ & \qquad + |F_1^1F_2^2F_3^2F_4^1|\exp[-i(\lambda_2^d + \lambda_3^d)] \\ & \qquad - |F_1^1F_2^2F_3^2F_4^2|\exp[-i(\lambda_2^d + \lambda_3^d + \lambda_4^d)] \\ & \qquad - |F_1^2F_2^1F_3^1F_4^1|\exp[-i\lambda_1^d] \\ & \qquad + |F_1^2F_2^1F_3^1F_4^2|\exp[-i(\lambda_1^d + \lambda_4^d)] \\ & \qquad + |F_1^2F_2^1F_3^2F_4^1|\exp[-i(\lambda_1^d + \lambda_3^d)] \\ & \qquad - |F_1^2F_2^1F_3^2F_4^2|\exp[-i(\lambda_1^d + \lambda_3^d + \lambda_4^d)] \\ & \qquad + |F_1^2F_2^2F_3^1F_4^1|\exp[-i(\lambda_1^d + \lambda_2^d)] \\ & \qquad - |F_1^2F_2^2F_3^1F_4^2|\exp[-i(\lambda_1^d + \lambda_2^d + \lambda_4^d)] \\ & \qquad - |F_1^2F_2^2F_3^2F_4^1|\exp[-i(\lambda_1^d + \lambda_2^d + \lambda_3^d)] \\ & \qquad - |F_1^2F_2^2F_3^2F_4^2|\exp[-i(\lambda_1^d + \lambda_2^d + \lambda_3^d + \lambda_4^d)]\}. \qquad (30) \end{aligned}$$

The term beween { } in (30) does not depend on the quartet Ψ_{1234}^{1111} itself and can be expressed as $A_{1234}^{1111}\times\exp[i\Lambda_{1234}^{1111}]$. In this way, combining (27) with (30) gives

$$|R_1^dR_2^dR_3^dR_4^d|\exp[i\Psi_{1234}^d] = A_{1234}^{1111}\exp[i(\Psi_{1234}^{1111} + \Lambda_{1234}^{1111})]. \qquad (31)$$

Finally, after (31) is combined with (24), the distribution of Ψ_{1234}^{1111} becomes

$$\begin{aligned} & P(\Psi_{1234}^{1111}|R_1^d, R_2^d, R_3^d, R_4^d) \\ & \quad \propto \exp[2G_{1234}^{1111}\cos(\Psi_{1234}^{1111} + \Lambda_{1234}^{1111} + \Delta_{1234}^d)], \qquad (32) \end{aligned}$$

$$G_{1234}^{1111} = A_{1234}^{1111}|Z_{1234}^d|.$$

Expressions for the other quartets in (28) can be set up in a similar way.

3.2. *The conditional probability distribution of the quartet phase sum among isomorphous data sets in the case of seven difference structure factors*

Starting from the difference-structure-factor analogue of (15),

$$\begin{aligned}P(\Phi_1^d,\ldots,\Phi_7^d,R_1^d,\ldots,R_7^d)&\\ \propto \exp[2W_{125}^d\cos(\Phi_1^d+\Phi_2^d-\Phi_5^d+\Delta_{125}^d)&\\ +2W_{345}^d\cos(\Phi_3^d+\Phi_4^d+\Phi_5^d+\Delta_{345}^d)&\\ +2W_{136}^d\cos(\Phi_1^d+\Phi_3^d-\Phi_6^d+\Delta_{136}^d)&\\ +2W_{246}^d\cos(\Phi_2^d+\Phi_4^d+\Phi_6^d+\Delta_{246}^d)&\\ +2W_{237}^d\cos(\Phi_2^d+\Phi_3^d-\Phi_7^d+\Delta_{237}^d)&\\ +2W_{147}^d\cos(\Phi_1^d+\Phi_4^d+\Phi_7^d+\Delta_{147}^d)&\\ +2W_{1\text{-}7}^d\cos(\Phi_1^d+\Phi_2^d+\Phi_3^d+\Phi_4^d+\Delta_{1\text{-}7}^d)],& \qquad (33)\end{aligned}$$

various conditional joint probability expressions can be obtained for the quartet Ψ_{1234}^{1111}, dependent on whether the random variables R_5^d, R_6^d and R_7^d, denoted now simply as $R_{\rm cross}^d$, of the cross-term difference structure factors are (assumed to be) known completely or their magnitudes $|R_{\rm cross}^d|$ only.

3.2.1. R_{cross}^d *known.* As explained in §3.1, it is convenient to express all terms in the probabilistic expression in those of a single data set only. For example, if the triplet term 125 in (33) is expressed completely in quantities of data set 1 by means of (29), the result is

$$\begin{aligned}W_{125}^d&\cos(\Phi_1^d+\Phi_2^d-\Phi_5^d+\Delta_{125}^d)\\ &=G_{125}^{111}\cos(\Phi_1^1+\Phi_2^1-\Phi_5^1+\Delta_{125}^d+\Lambda_{125}^{111}), \qquad (34)\end{aligned}$$

$$G_{125}^{111}=|Z_{125}^d|A_{125}^{111}$$

with

$$\begin{aligned}R_1^dR_2^dR_5^d&\exp[i(\Phi_1^d+\Phi_2^d-\Phi_5^d)]\\ &=A_{125}^{111}\exp[i(\Phi_1^1+\Phi_2^1-\Phi_5^1+\Lambda_{125}^{111})]. \qquad (35)\end{aligned}$$

The same technique applied to the quartet term in (33) results in

$$\begin{aligned}W_{1\text{-}7}^d&\cos(\Phi_1^d+\Phi_2^d+\Phi_3^d+\Phi_4^d+\Delta_{1\text{-}7}^d)\\ &=G_{1\text{-}7}^{1111}\cos(\Phi_1^1+\Phi_2^1+\Phi_3^1+\Phi_4^1+\Delta_{1\text{-}7}^d+\Lambda_{1234}^{1111}),\\ &\qquad (36)\end{aligned}$$

$$G_{1\text{-}7}^{1111}=|Z_{1\text{-}7}^d|A_{1234}^{1111}.$$

As a result, the conditional joint probability expression of the quartet Ψ_{1234}^{1111} becomes

$$P(\Psi_{1234}^{1111}|R_1^d,\ldots,R_7^d)\propto\exp[2T_{1\text{-}7}^{1111}\cos(\Psi_{1234}^{1111}+\Omega_{1\text{-}7})], \qquad (37)$$

$$\begin{aligned}T_{1\text{-}7}^{1111}\exp[i\Omega_{1\text{-}7}^{1111}]=G_{1\text{-}7}^{1111}&\exp[i(\Delta_{1\text{-}7}^d+\Lambda_{1234}^{1111})]\\ &+\{G_{125}^{111}{\rm Br}(2G_{345}^{111})+G_{345}^{111}{\rm Br}(2G_{125}^{111})\}\\ &\times\exp[i(\Delta_{345}^d+\Lambda_{345}^{111}+\Delta_{125}^d+\Lambda_{125}^{111})]\\ &+\{G_{136}^{111}{\rm Br}(2G_{246}^{111})+G_{246}^{111}{\rm Br}(2G_{136}^{111})\}\\ &\times\exp[i(\Delta_{246}^d+\Lambda_{246}^{111}+\Delta_{136}^d+\Lambda_{136}^{111})]\\ &+\{G_{147}^{111}{\rm Br}(2G_{237}^{111})+G_{237}^{111}{\rm Br}(2G_{147}^{111})\}\\ &\times\exp[i(\Delta_{237}^d+\Lambda_{237}^{111}+\Delta_{147}^d+\Lambda_{147}^{111})].\end{aligned}$$

In the cases of single-wavelength anomalous scattering (SAS) and two-wavelength anomalous scattering (2DW), the majority of the doublet phase sums tend to be positive. The assumption that the doublet sign is positive together with the estimate for the doublet phase-sum magnitude (29) leads to a completely available $R_{\rm cross}^d$. For simplicity, it has been assumed in (33)–(37) that all three cross-term difference structure factors are known. For absent cross terms, the relevant terms in (37) should be omitted.

3.2.2. *Only* $|R_{cross}^d|$ *known.* For SAS or 2DW data, it may in some instances also be useful to consider the case that only $|R_{\rm cross}^d|$ is available, *e.g.* when $|R_{\rm cross}^d|$ is small. Taking again, as an example, the triplet 125 term in (33), the left-hand side of (34) can be expressed as

$$\begin{aligned}2R_1^dR_2^dR_5^d&\exp[i(\Phi_1^d+\Phi_2^d-\Phi_5^d)]\\ &=(F_1^1-F_1^2)(F_2^1-F_2^2)(R_5^d)^*\\ &=2|R_5^d|\exp[i(\Phi_1^1+\Phi_2^1-\Phi_5^d)]A_{12}^{11}\exp[i\Lambda_{12}^{11}] \qquad (38)\end{aligned}$$

with

$$\begin{aligned}A_{12}^{11}\exp[i\Lambda_{12}^{11}]=|F_1^1F_2^1|&-|F_1^2F_2^1|\exp[-i\lambda_1^d]\\ &-|F_1^1F_2^2|\exp[-i\lambda_2^d]\\ &+|F_1^2F_2^2|\exp[-i(\lambda_1^d+\lambda_2^d)]. \qquad (39)\end{aligned}$$

So the triplet 125 term becomes

$$\begin{aligned}2W_{125}^d&\cos(\Phi_1^d+\Phi_2^d-\Phi_5^d+\Delta_{125}^d)\\ &=2G_{125}^{11d}\cos(\Phi_1^1+\Phi_2^1-\Phi_5^d+\Delta_{125}^d+\Lambda_{125}^{11}), \qquad (40)\end{aligned}$$

$$G_{125}^{11d}=Z_{125}^dA_{12}^{11}|R_5^d|.$$

Similarly, $2R_3^dR_4^dR_5^d\exp[i(\Phi_3^d+\Phi_4^d+\Phi_5^d+\Delta_{345}^d)]$ becomes

$$2G_{345}^{11d}\exp[i(\Phi_3^1+\Phi_4^1+\Phi_5^d+\Lambda_{34}^{11}+\Delta_{345}^d)]. \qquad (41)$$

From (40) and (41), estimates for the cross-term phase random variables can be obtained:

$$\langle\exp[i\Phi_5^d]\rangle={\rm Br}(2G_{125}^{11d})\exp[i(\Phi_1^1+\Phi_2^1+\Lambda_{12}^{11}+\Delta_{125}^d)] \qquad (42)$$

and

$$\langle\exp[-i\Phi_5^d]\rangle={\rm Br}(2G_{345}^{11d})\exp[i(\Phi_3^1+\Phi_4^1+\Lambda_{34}^{11}+\Delta_{345}^d)],$$

respectively. Insertion of (42) in the triplet terms 345 and

125 in (33) gives

$$\exp[i(\Phi^{1111}_{1234} + \Lambda^{11}_{12} + \Delta^{d}_{125} + \Lambda^{11}_{34} + \Delta^{d}_{345})] \times \{G^{11d}_{125}\mathrm{Br}(2G^{11d}_{345}) + G^{11d}_{345}\mathrm{Br}(2G^{11d}_{125})\}. \quad (43)$$

The triplet terms 136 and 246 and the triplet terms 147 and 237 can be handled in the same way so the final conditional joint probability expression can be expressed in a form similar to (38):

$$P(\Psi^{1111}_{1234}|R^{d}_{1}, \ldots, R^{d}_{4}, |R^{d}_{5}|, |R^{d}_{6}|, |R^{d}_{7}|) \propto \exp[2T^{1111}_{1\text{-}7} \cos(\Psi^{1111}_{1234} + \Omega_{1\text{-}7})] \quad (44)$$

$$\begin{aligned} T^{1111}_{1234} \exp[i\Omega^{1111}_{1234}] = {} & G^{1111}_{1\text{-}7} \exp[i(\Delta^{d}_{1\text{-}7} + \Lambda^{1111}_{1234})] \\ & + \{G^{11d}_{125}\mathrm{Br}(2G^{11d}_{345}) + G^{11d}_{345}\mathrm{Br}(2G^{11d}_{125})\} \\ & \times \exp[i(\Delta^{d}_{345} + \Delta^{d}_{125} + \Lambda^{11}_{12} + \Lambda^{11}_{34})] \\ & + \{G^{11d}_{136}\mathrm{Br}(2G^{11d}_{246}) + G^{11d}_{246}\mathrm{Br}(2G^{11d}_{136})\} \\ & \times \exp[i(\Delta^{d}_{246} + \Delta^{d}_{136} + \Lambda^{11}_{24} + \Lambda^{11}_{13})] \\ & + \{G^{11d}_{147}\mathrm{Br}(2G^{11d}_{237}) + G^{11d}_{237}\mathrm{Br}(2G^{11d}_{147})\} \\ & \times \exp[i(\Delta^{d}_{237} + \Delta^{d}_{147} + \Lambda^{11}_{14} + \Lambda^{11}_{23})]. \end{aligned}$$

The above two cases are not readily applicable to single isomorphous replacement data without anomalous-scattering (SIRNAS) data or including anomalous-scattering (SIRAS) data because then neither cross-term doublet signs nor main-term doublet signs are available in a straightforward way.

4. Results and discussion

In order to assess the predictive quality of (32), (37) and (44), extensive tests have been performed with calculated structural data from two small proteins from the Protein Data Bank (Bernstein *et al.*, 1977; Abola, Bernstein, Bryant, Koetzle & Weng, 1987). The data sets used in the tests involve both native and heavy-atom-derivative data of APP [avian pancreatic polypeptide (Blundell, Pitts, Tickle, Wood & Wu, 1981); in the PDB release of 1991 known as 1PPT] and C550 [cytochrome c from *Paracoccous denitrificans* (Timkovich & Dickerson, 1976); in the PDB release of 1991 known as 155C].

For each structure, four different types of two isomorphous data sets have been constructed:

(i) *SAS case.* The isomorphous data sets are the Friedel-related-index sets $\{H(S_1)\}$ and $\{-H(S_1)\}$ both of the heavy-atom derivative (denoted as S_1) and using Cu $K\alpha$ radiation.

(ii) *2DW case.* The isomorphous data sets used are: $\{H(\lambda_1)\}$ and $\{H(\lambda_2)\}$ with $\lambda_1 = \mathrm{Cr}\ K\alpha$ and $\lambda_2 = \mathrm{Cu}\ K\alpha$ radiation, both selected for the heavy-atom derivative S_1.

(iii) *SIRAS case.* The isomorphous data set cases are defined as $\{H(S_1)\}$ and $\{H(S_2)\}$ with S_1 the heavy-atom (Hg) derivative and S_2 the native protein. Anomalous-dispersion corrections have been applied for all atoms.

Table 1. *Abbreviations and procedures employed in Tables 2–6*

SD4	Quartet estimation *via* (32)
SD7	Quartet estimation *via* (37)
SD7*	Quartet estimation *via* (44)
ALG	Doublets estimated *via* the algebraic technique [Kyriakidis *et al.*, 1993*b*, equation (15*b*)]
PAT	Doublets estimated *via* the Patterson-improved algebraic technique [Kyriakidis *et al.*, 1993*b*, §3.3]
TRUE	Calculated doublets used
DR	Diffraction ratio (Kyriakidis *et al.*, 1993*a*)
W	Reliability factor of the estimates
PQ	Positive quartets only
NQ	Negative quartets only
NQR	Cunulative number of quartets involved in the statistics
AER	Mean absolute error in invariant phase sum estimates [in mc, see equation (45)]
ERR	Mean error in invariant phase sum estimates [in mc, see equation (45)]
NS	Number of wrong doublet signs

(iv) *SIRNAS case.* The isomorphous data sets are defined as $\{H(S_1)\}$ and $\{H(S_2)\}$ with S_1 the heavy-atom derivative and S_2 the native protein. No anomalous-dispersion corrections have been applied.

In Kyriakidis *et al.* (1993*b*), it was shown that algebraically based estimates of doublet phase sums can be useful to get correct estimates of triplet phase sums present among isomorphous data sets; in particular, if the estimation of the doublet phase sums are based on vector information from a special difference Patterson synthesis. For the benefit of the current paper, the two algebraic approaches in Kyriakidis *et al.* (1993*b*) have been tested: (*a*) the algebraic doublet estimation (ALG) and (*b*) the algebraic estimation improved by difference-Patterson vectors (PAT). For SAS and 2DW data, the ALG doublet estimation depends only on the imaginary dispersion correction of the anomalous scatterers. For reference, and to establish the theoretical limits of (32), (37) and (44), the actual calculated doublet values (TRUE) have also been used.

In the statistics, only quartets of type Ψ^{1111}_{1234} have been included. In each cumulative statistics, four quantities are listed: the reliability underlimit W, the cumulative number of quartets (NQR) with a reliability factor above this underlimit, the absolute mean difference AER in mc ($1000\,\mathrm{mc} = 2\pi\,\mathrm{rad}$),

$$\mathrm{AER} = \langle ||\Psi_4|_{\mathrm{true}} - |\Psi_4|_{\mathrm{est}}| \rangle,$$

and the mean difference in mc, (45)

$$\mathrm{ERR} = \langle |\Psi_{4\mathrm{true}} - \Psi_{4\mathrm{est}}| \rangle.$$

4.1. *Results for APP*

Table 1 lists abbreviations used in Tables 2–6.

The native protein APP crystallizes in space group $C2$. It has 302 atoms in the asymmetric unit of which only Zn^{2+} contributes to the anomalous scattering. For this small protein and its heavy-atom (Hg) derivative, data up to 2.27 Å resolution (1454 reflections for each set) have been calculated using atomic coordinates from the PDB.

Table 2. *APP*: $|E^1|$ *and doublet phase-sum estimate statistics*

		PAT			ALG		
Nos.	$\|E_\nu^1\|$ range	AER	ERR	NS	AER	ERR	NS
1–250	$1.33 \le \|E\| \le 3.01$	1.5	1.8	9	5.5	5.7	7
251–500	$1.04 \le \|E\| \le 1.33$	1.9	2.5	15	7.5	8.0	14
501–750	$0.82 \le \|E\| \le 1.04$	2.6	4.5	28	10.3	12.0	28
751–1000	$0.60 \le \|E\| \le 0.82$	3.8	7.1	34	15.2	18.8	36
1001–1250	$0.37 \le \|E\| \le 0.60$	5.0	10.8	50	22.3	28.0	50
1251–1454	$0.02 \le \|E\| \le 0.37$	18.7	80.9	71	60.4	116.7	72

4.1.1. *SAS case*. In the SAS case, quartet phase-sum relations have been generated for the heavy-atom (Hg) derivative. When generating the quartet phase-sum relations, six different sets of main-term reflections were considered (see Table 2) in order to investigate any dependence of the estimates on $|E^1|$ and $|E^2|$. For each set, 25 000 positive and 25 000 negative quartets were generated. Only quartets with at least two observed cross-term structure factors present in the data set were accepted. The classification of quartets as being positive or negative should be done in principle according to $|E^d|$ but because of its dependence on $\langle\cos(\Psi^d)\rangle$ this is not convenient. Moreover, in the SAS and 2DW cases, the majority of the doublets are quite small so the decrease in $|E^d|$ is determined mainly by the decrease of both $|E^1|$ (and $|E^2|$). Therefore, the classification of quartets as being positive or negative was carried out with the $|E^1|$'s, using criteria holding for normal diffraction data (Peschar, 1987).

An excerpt of the quartet phase-sum statistics listed in Table 3 for the main term classes 1–250 and 751–1000 show that, in spite of the small diffraction ratio (0.11), reliable estimates of quartet phase sums can be obtained when W is larger than 0.2 (0.3 in the case of negative quartets). When going from the set 1–250 to 751–1000, the general reliability of the quartet estimates decreases somewhat, even if correct doublets are used. This deterioration becomes worse if the PAT estimation is used, while the ALG estimation is clearly insufficient for the (501–750 and) 751–1000 set. This general deterioration is the result of increasingly incorrect doublet estimates, as demonstrated by the PAT and ALG doublet errors in Table 2. Indeed, for the 1251–1454 main-term set, it is virtually impossible to get correct quartet estimates by means of PAT and/or ALG doublet estimates while the use of correct doublets still leads to acceptable results.

A remarkable result is that the overall estimation error for the quartets turns out to be the same for (32) and (37) while those of (42) are approximately the same, and only occasionally slightly worse, than those of (37). This may seem to be an unexpected result that contradicts the usual considerations on the estimation of quartets using cross terms.

However, the current case differs at several points from normal diffraction data. Most important is the role of the doublets, because of which (32) leads to unique estimates on the interval $(-\pi, \pi)$. For normal diffraction data, doublets do not occur and the only possible estimates for the quartet phase sum are 0 or π, dependent on the cross-term magnitudes. For isomorphous data sets, however, the doublets play the role of the cross terms. It is well known that in order to estimate phase-sum invariants of a certain order, *e.g.* a quartet of order N^{-1}, cross-term-magnitude information is required. In a generalization and practical application of this principle, it has been shown that the estimation of a higher-order phase-sum invariant (quartet or quintet) can be improved by considering only the estimates of cross-term phases on the basis of the lowest-order invariant available. For normal diffraction data, these are the triplet phase sums (Peschar, 1987). As it turns out, this same principle seems to apply to the difference structure factor. The lowest-order invariants available now are the doublet phase sums and the reliability weights of their estimates are much larger than those of the triplets. As a result, expressing the main-term phases of the second data set $\Phi_i(2)$ in those of the first $\Phi_i(1)$ *via* the doublets [see (29)] will have a much larger influence than including cross-term information, which enters only *via* triplets for which the reliability factor is much smaller.

A second essential difference with the normal diffraction case should also be pointed out. In (37), it is explicitly assumed that the complete difference structure factor is available. The availability of the phase part of the difference structure factor depends on whether the doublet sign is (assumed to be) known or not. In the SAS and 2DW cases, the majority of the doublets have a positive sign, provided the data sets are chosen in an appropriate way: in the SAS case, $\{H\}$ should be first and $\{-H\}$ second while in the 2DW case the wavelength resulting in the largest anomalous effects should be selected first. For the benefit of this paper, we restricted the tests to two extreme cases: either complete estimates of the cross-term difference structure factors are assumed to be available, as in (37), or only an estimate of their magnitudes, which leads to (44).

From Table 3, it is evident that the PAT-estimated doublets lead to better results than those with ALG-estimated doublets. In view of Table 2, this difference seems to be mostly due to a better estimation of the doublet magnitude in the PAT case because the number of incorrect doublet signs is almost the same in the ALG and PAT cases. The sudden breakdown of the reliability of the SD4(PAT) and SD7(PAT) estimates in the case 1251–1454 (compared with 1001–1250) can be attributed to the same effect.

4.1.2. *2DW case*. In Table 4, cumulative statistics of quartet phase-sum estimates are listed for a 2DW case ($\lambda_1 = \mathrm{Cr}K\alpha$, $\lambda_2 = \mathrm{Cu}K\alpha$) for the Hg derivative of APP. The diffraction ratio is only 0.047 but still acceptable results can be obtained. This ratio corresponds with doublet values of about 4–5 mc. Additional tests done in

Table 3. *Cumulative statistics of the quartet phase-sum estimates of the heavy-atom (Hg) derivative of the protein APP in the SAS case*

		1–250 (PQ)			1–250 (NQ)			751–1000 (PQ)			751–1000 (NQ)		
	W	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR
SD4													
TRUE	0.6	7586	13	13	4514	17	17	462	10	11	491	13	13
	0.3	12084	15	16	10389	25	30	2392	12	12	1820	15	15
	0.2	13783	21	27	13503	36	71	4147	20	28	3765	36	56
	0.1	18813	38	59	18618	75	126	7806	58	100	9449	104	188
	0.0	25000	62	98	25000	99	162	25000	122	189	25000	136	226
SD7													
TRUE	0.6	10849	14	14	8174	17	18	465	11	11	818	14	14
	0.3	13287	16	17	11968	27	33	3670	13	14	3052	16	17
	0.2	16125	22	30	15332	51	80	5395	25	35	5162	45	73
	0.1	20333	43	66	20197	81	135	8889	62	106	10882	105	188
	0.0	25000	61	98	25000	98	162	25000	122	189	25000	136	226
SD4													
PAT	0.6	5536	26	28	3135	35	40	674	30	40	324	29	35
	0.3	11428	32	35	9047	42	49	2852	67	92	2131	72	99
	0.2	12752	36	43	11818	58	77	4837	78	111	4448	94	138
	0.1	17494	51	69	16500	82	124	8850	106	162	10942	133	208
	0.0	25000	78	114	25000	112	173	25000	143	214	25000	152	234
SD7													
PAT	0.9	6507	27	29	3657	38	44	666	34	44	335	37	45
	0.6	9848	30	33	6994	40	47	1627	55	74	1056	57	76
	0.3	12511	33	38	10891	44	51	4234	72	100	3443	78	109
	0.2	14528	38	47	13239	61	84	6135	84	120	6196	104	154
	0.1	19194	56	78	18443	89	137	10153	111	167	12670	135	210
	0.0	25000	78	114	25000	111	173	25000	143	214	25000	152	232
SD4													
ALG	0.4	27	88	95	31	85	92	–	–	–	–	–	–
	0.3	1499	82	116	1802	94	134	530	126	189	590	165	230
	0.2	24682	94	129	24483	123	183	24799	154	223	24591	161	237
	0.0	25000	94	129	25000	123	182	25000	154	223	25000	160	237
SD7													
ALG	0.5	331	72	88	466	84	100	39	127	164	42	171	229
	0.4	3847	85	118	4480	99	147	1731	145	212	2270	157	225
	0.3	24581	94	128	24430	123	183	24726	154	223	24493	161	237
	0.0	25000	94	128	25000	122	182	25000	154	223	25000	160	237
SD7*													
PAT	0.9	5115	28	31	4039	41	47	667	49	67	445	51	69
	0.6	9156	30	33	7202	42	49	1678	55	74	1257	61	85
	0.3	12499	34	39	10855	44	52	4042	71	98	3561	80	112
	0.2	14351	40	49	13581	63	88	6197	87	127	6674	111	167
	0.1	18982	56	80	18703	91	140	10336	113	172	13169	137	214
	0.0	25000	78	114	25000	111	173	25000	143	214	25000	152	232
SD7*													
ALG	0.9	196	60	70	228	69	80	49	170	221	38	177	236
	0.6	2189	71	88	2948	81	105	1569	160	217	1656	153	215
	0.3	15559	86	116	16379	108	155	13686	152	219	13579	157	231
	0.2	24891	94	128	24890	123	182	24920	154	223	24869	160	237
	0.0	25000	94	128	25000	123	182	25000	154	223	25000	160	237

the SAS and 2DW cases for both the native APP, in which Zn is the only anomalous scatterer, and the heavy-atom derivative show that if the diffraction ratio becomes smaller than 0.04 the quality of the estimates breaks down progressively. At a diffraction ratio of 0.01, even the use of correct doublets does not yield any useful results any more.

4.1.3. *SIRAS and SIRNAS case.* In both the SIRAS and SIRNAS cases, the diffraction ratio is large (0.60 and 0.64, respectively). Although the doublet magnitudes can be estimated in a reliable way, a lack of knowledge of the doublet signs prevents a correct estimation of the quartet phase sums. The ERR results in the ALG case (see Table 5) suggest the opposite but a close inspection of the individual estimates shows that almost all quartets with large W are estimated to be either 0 or π. In spite of this sign ambiguity, the low AER data for the PAT and ALG cases shows that (32), (37) and (42) are quite effective in predicting the quartet phase-sum magnitudes. The results in the SIRNAS case (DR = 0.64) are quite similar to those in the SIRAS case and are therefore not listed.

Table 4. *Cumulative statistics of the estimates for the positive quartet phase sums for the heavy-atom (Hg) derivative of the protein APP in the 2DW case ($\lambda_1 = Cr\,K\alpha$, $\lambda_2 = Cu\,K\alpha$)*

DR = 0.047; negative doublets: 5; the first 25 000 positive quartets generated among the strongest 250 $|E_v^1|$'s have been included in the statistics. *: $|R_{\rm cross}^d|$ assumed to be known.

	1–250 SD4 (TRUE)			1–250 SD7 (TRUE)			1–250 SD4 (PAT)			1–250 SD7 (PAT)		
W	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR
0.9	2632	28	30	7932	31	34	–	–	–	5333	45	50
0.6	7569	31	34	11445	35	38	4142	42	45	9152	49	54
0.3	14793	50	59	17255	56	69	11005	52	58	12377	54	61
0.2	18304	63	83	20048	68	91	12750	58	67	14806	65	78
0.1	21859	76	103	22778	78	108	18586	85	109	20449	91	119
0.0	25000	87	120	25000	87	120	25000	106	142	25000	104	141

	SD4 (ALG)			SD7 (ALG)			SD7* (PAT)			SD7* (ALG)		
W	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR
0.9	–	–	–	–	–	–	4474	47	52	124	74	86
0.6	–	–	–	–	–	–	8398	49	54	1866	80	99
0.5	–	–	–	416	97	118	9781	50	57	4026	86	110
0.4	135	109	130	4024	96	128	11100	52	59	8138	92	119
0.3	2236	100	127	21437	104	141	12495	56	64	15375	99	131
0.2	22398	103	140	24912	104	141	20379	92	119	24503	104	141
0.0	25000	105	142	25000	104	141	25000	105	141	25000	105	141

Table 5. *Cumulative statistics of the quartet phase-sum estimates of the protein APP in the SIRAS case*

DR = 0.60. Negative doublets: 105. Positive quartets only generated among main-term reflections nos. 1–250.

	1–250 SD4 (TRUE)			1–250 SD7 (TRUE)			1–250 SD4 (PAT)			1–250 SD7 (PAT)		
W	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR
1.4	–	–	–	966	2	2	–	–	–	1285	59	124
1.2	–	–	–	2769	1	1	–	–	–	3244	66	127
1.0	–	–	–	5202	1	1	–	–	–	5632	70	132
0.8	897	2	2	7625	1	1	1224	59	125	8010	72	138
0.6	5457	1	1	9922	1	1	6026	70	136	10150	74	141
0.3	11451	1	1	12528	1	1	11635	76	142	12629	77	144
0.2	12837	6	10	14587	7	12	12948	80	147	14715	82	149
0.1	17686	21	38	19131	25	46	17734	92	161	19349	96	166
0.0	25000	47	88	25000	47	88	25000	108	182	25000	108	182

	1–250 SD7* (PAT)			1–250 SD4 (ALG)			1–250 SD7 (ALG)			1–250 SD7* (ALG)		
W	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR
2.3	1006	59	109	–	–	–	–	–	–	148	31	35
1.9	2205	62	118	–	–	–	–	–	–	491	42	57
1.4	5777	69	131	–	–	–	–	–	–	2028	53	89
1.2	6487	70	133	–	–	–	131	24	28	3484	60	105
1.0	7867	72	135	–	–	–	566	37	47	6030	79	134
0.8	9441	74	139	40	16	18	2277	46	83	10099	87	145
0.6	11059	75	142	832	36	53	8512	79	135	15377	99	162
0.5	11717	76	142	2657	45	90	15336	100	163	18268	105	170
0.4	12472	77	144	8150	78	136	22552	114	182	21481	111	178
0.0	25000	108	182	25000	118	188	25000	118	188	25000	118	188

4.2. *Results for C550*

The protein C550 has molecular weight $M_r = 14\,500$ (1017 atoms in the asymmetric unit) and crystallizes in space group $P2_12_12_1$, structure factors up to 2.5 Å were calculated. The native protein contains 4 Fe and 24 S atoms which scatter anomalously at the wavelengths used. In the original heavy-atom derivative, the anomalously scattering group $(PtCl_4)^{2-}$ was present but, for the sake of simplicity, in the current tests this group has been replaced by a Po atom that has the same effective Z value (84) at $\sin(\theta)/\lambda = 0$.

The isomorphous data sets of C550 in the SAS and SIRAS cases are characterized by a large number of small doublet values. This is partly due to the large amount of phase-restricted reflections in the space group

Table 6. *Cumulative statistics of the doublet and quartet phase-sum estimates of the protein cytochrome c in the SAS (heavy-atom derivative) and SIRAS (heavy-atom derivative and native) cases*

Cu $K\alpha$ radiation. Resolution: 2.5 Å. Positive quartets only generated among main-term reflections nos. 1–500 and 501–1000.

Doublets

	Nos.	$\|E^1\|$ range	AER	ERR	NS
(I) SAS	1–500	$5.66 \leq \|E\| \leq 1.4$	3.2	4.0	45
(II) SAS	501–1000	$1.4 \leq \|E\| \leq 1.13$	4.0	5.5	75
(III) SIRAS	1–500	$5.66 \leq \|E\| \leq 1.4$	3.4	21.5	207
(IV) SIRAS	501–1000	$1.4 \leq \|E\| \leq 1.13$	3.6	32.0	235

Quartets

	(I) SD4 (PAT)			(II) SD4 (PAT)			(III) SD4 (PAT)			(IV) SD7 (PAT)		
W	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR	NQR	AER	ERR
6.0	–	–	–	–	–	–	266	103	151	23	85	172
5.0	93	80	110	9	47	47	502	101	154	86	94	168
4.0	302	77	104	25	58	70	1000	109	169	206	96	167
3.0	845	78	104	162	71	80	2003	113	176	606	111	193
2.0	2550	97	130	847	94	120	4347	123	192	1944	133	218
1.5	4679	112	154	2030	115	157	6793	132	206	3626	142	228
1.0	9256	133	188	5362	137	194	10878	141	218	7151	149	232
0.5	18220	147	213	14298	154	225	17345	148	228	13857	156	238
0.0	25000	152	222	25000	158	233	25000	153	235	25000	160	242

$P2_12_12_1$. In Table 6, some examples are listed of results obtained in the SAS and SIRAS cases with Cu$K\alpha$. Although the diffraction ratio is quite small in the SAS case (0.09), the quartet phase-sum estimates are still acceptable provided the Patterson-improved (PAT) doublet estimates are used. Attempts with the ALG-estimated doublets were not successful. In the SIRAS case with a diffraction ratio of 0.36, the results are somewhat less successful. This can be attributed mainly to the fact that the doublet signs cannot be predicted in the current probabilistic approach. An additional problem arises from the large amount of phase-restricted reflections. Although the difference-structure-factor approach can be set up in a similar way as for general reflections (see Appendix I), an analysis of $\langle|F^d|^2\rangle$ leads to the conclusion that, in addition to the usual interatomic vector terms upon which the PAT-doublet estimation is based, other terms are also present that in general are not available in a straightforward way. In spite of the sign ambiguity in the SIRAS case, it is hopeful that the magnitude of the estimated quartets correspond quite well with the actual quartet phase-sum magnitudes. In the 501–1000 main-term set, less quartets are found at a higher reliability level but those that do occur are comparable in reliability with those from the 1–500 set.

In conclusion, by using the technique of difference structure factors, it is possible to obtain reliable estimates of quartet phase sums present among isomorphous data sets. The estimates are unique on the interval $-\pi$ to π in the case of SAS or 2DW data provided the diffraction ratio is large enough (at least 0.04). In contrast to the normal diffraction case, involving only a single data set and no anomalous scattering, for the estimation of quartet phase sums among isomorphous structure factors knowledge of cross-term magnitudes is not essential because their role is fulfilled by the doublet phase-sum estimates. In the SIRAS and SIRNAS cases, doublet magnitudes are estimated correctly but the doublet signs cannot be estimated from the current probabilistic approach. Nevertheless, the SIRAS data suggest that quartet phase sums can be sorted out that lie close to 0 or π, provided the doublet estimates are supplemented by Patterson vectors.

The authors gladly acknowledge that the comments of one of the referees helped to improve the quality of this paper.

APPENDIX *A*
The difference structure factor for phase-restricted isomorphous structure factors

Although the derivations of the expressions in this paper do hold formally only if the normal structure factors are not phase restricted, a formulation of the difference-structure-factor approach for these types of reflection can be set up in a similar way. [A discussion on the estimation of phase-restricted doublets *via* the usual probabilistic technique can be found in Giacovazzo (1987).]

The expression for a structure factor with a phase restriction $0/\pi$ is

$$F_H = 2\sum_{j=1}^{N/2} f_{jH}\cos[2\pi\mathbf{H}\cdot\mathbf{r}_j] \qquad (46)$$

so $F_{-H} = F_H$. Since $F_H^* \neq F_H$, a difference structure factor can be defined as

$$F_H^d = F_H - F_H^* = 2\sum_{j=1}^{N/2} f_{jH}^d\cos[2\pi\mathbf{H}\cdot\mathbf{r}_j] \qquad (47)$$

with $f_{jH}^d = f_{jH} - f_{jH}^*$. Because of anomalous scattering, the phase φ_H will deviate slightly from its phase restriction φ_r: $\varphi_H = \varphi_r + \delta_H$.

From (47), $|F_H^d|^2 = 2|F_H|^2[1 - \cos(2\delta_H)]$ so the same functional form is obtained as for general reflections [see (13) in Kyriakidis *et al.*, 1993*b*]:

$$\langle\cos(2\delta_H)\rangle = (2|F_H|^2 - \langle|F_H^d|^2\rangle)/2|F_H|^2. \qquad (48)$$

Expressing $|F_H^d|^2$ in the atomic contributions leads to

$$\begin{aligned}|F_H^d|^2 = 2\sum_{j1=1}^{N/2}\sum_{j2=1}^{N/2} & f_{jH}^d(f_{jH}^d)^* \\ & \times \{\cos[2\pi\mathbf{H}\cdot(\mathbf{r}_{j1} - \mathbf{r}_{j2})] \\ & + \cos[2\pi\mathbf{H}\cdot(\mathbf{r}_{j1} + \mathbf{r}_{j2})]\}. \qquad (49)\end{aligned}$$

The first cosine term in (49) contributes to the doublet estimation but the second, involving $\mathbf{r}_{j1} + \mathbf{r}_{j2}$, is not available and must therefore be neglected. Structure factors with a different phase restriction can be dealt with in a similar way. For example, for reflections restricted on $\pm\pi/2$, the difference structure factor can be defined as

$$F_H^d = F_H - F_{-H}^* \qquad (50)$$

with

$$F_H = 2i\sum_{j=1}^{N/2} f_{jH}\sin[2\pi\mathbf{H}\cdot\mathbf{r}_j].$$

References

Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987). *Crystallographic Databases – Information Content, Software Systems, Scientific Applications*, edited by F. H. Allen, G. Bergerhoff & S. Sievers. Bonn/Cambridge/Chester: IUCr.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. & Wu, C. W. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4175–4179.

Cascarano, G., Giacovazzo, C. & Viterbo, D. (1987). *Acta Cryst.* A**43**, 22–29.

De Titta, G. T., Edmonds, J. W., Langs, D. A. & Hauptman, H. (1975). *Acta Cryst.* A**31**, 472–479.

Fortier, S. & Nigam, G. D. (1989). *Acta Cryst.* A**45**, 247–254.

Freer, A. A. & Gilmore, C. J. (1980). *Acta Cryst.* A**36**, 470–475.

Furey, W. Jr, Chandrasekhar, K., Dyda, F. & Sax, M. (1990). *Acta Cryst.* A**46**, 560–567.

Giacovazzo, C. (1975). *Acta Cryst.* A**31**, 252–259.

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

Giacovazzo, C. (1976*b*). *Acta Cryst.* A**32**, 958–966.

Giacovazzo, C. (1977*a*). *Acta Cryst.* A**33**, 50–54.

Giacovazzo, C. (1977*b*). *Acta Cryst.* A**33**, 933–944.

Giacovazzo, C. (1983). *Acta Cryst.* A**39**, 585–592.

Giacovazzo, C. (1987). *Acta Cryst.* A**43**, 73–75.

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

Gilmore, C. J. (1977). *Acta Cryst.* A**33**, 712–716.

Hauptman, H. (1974). *Acta Cryst.* A**30**, 472–476.

Hauptman, H. (1975*a*). *Acta Cryst.* A**31**, 671–679.

Hauptman, H. (1975*b*). *Acta Cryst.* A**31**, 680–687.

Hauptman, H. (1976). *Acta Cryst.* A**32**, 877–882.

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

Hauptman, H. (1982*b*). *Acta Cryst.* A**38**, 632–641.

Hauptman, H. & Karle, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal. ACA Monograph* No. 3. New York: Polycrystal Book Service.

Hauptman, H., Potter, S. & Weeks, C. M. (1982). *Acta Cryst.* A**38**, 294–300.

Heinerman, J. J. L. (1977). PhD thesis, University of Utrecht, The Netherlands.

Kyriakidis, C. E., Peschar, R. & Schenk, H. (1993*a*). *Acta Cryst.* A**49**, 350–358.

Kyriakidis, C. E., Peschar, R. & Schenk, H. (1993*b*). *Acta Cryst.* A**49**, 359–369.

Kyriakidis, C. E., Peschar, R. & Schenk, H. (1993*c*). *Acta Cryst.* A**49**, 557–569.

Peschar, R. (1987). PhD thesis, University of Amsterdam, The Netherlands.

Peschar, R. & Schenk, H. (1991). *Acta Cryst.* A**47**, 428–440.

Putten, N. van der & Schenk, H. (1979). *Acta Cryst.* A**35**, 381–387.

Schenk, H. (1973*a*). *Acta Cryst.* A**29**, 77–82.

Schenk, H. (1973*b*). *Acta Cryst.* A**29**, 480–482.

Schenk, H. (1974). *Acta Cryst.* A**30**, 477–482.

Schenk, H. (1991). Editor. *Direct Methods for Solving Crystal Structures.* New York: Plenum Press.

Schenk, H. & De Jong, J. G. H. (1973). *Acta Cryst.* A**29**, 31–34.

Sheldrick, G. M. (1993). *Crystallographic Computing 6. A Window on Modern Crystallography*, edited by H. D. Flack, L. Parkanyi & K. Simon, pp. 100–110. IUCr/Oxford University Press.

Simerska, M. (1956). *Czech. J. Phys.* **6**, 1–7.

Timkovich, R. & Dickerson, R. E. (1976). *J. Biol. Chem.* **251**, 4033–4046.