Acta Cryst. (1997). A**53**, 799–806

The Translation Problem in Molecular Replacement Techniques. II. The Estimate of Quartet Invariants

C. GIACOVAZZO,* L. MANNA AND D. SILIQI†

Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy. E-mail: crisg01@area.ba.cnr.it

(*Received 28 January 1997; accepted 30 June 1997*)

Abstract

The method of joint probability distribution functions of structure factors has been used to estimate quartet invariants when prior information on the orientation of molecular fragments is available. The mathematical approach makes use of the Gram–Charlier expansion of the characteristic function, as described by Giacovazzo [*Acta Cryst.* (1976), A**32**, 91–99] for deriving quartet estimates in the absence of prior information. The conclusive formula is a von Mises distribution: the expected value of the quartet phase may lie anywhere between 0 and 2π. The reliability parameter may be large even for proteins, provided the fractionary scattering power of the molecular fragments with known orientation is sufficiently large. The first practical applications prove the correctness of the probabilistic approach and suggest the usefulness of the quartet information even in molecular replacement methods when a model molecule has been oriented by some rotation function and needs to be translated into a proper position.

1. Introduction

In macromolecular crystallography, it frequently occurs (after the application of a rotation function or related methods) that the orientation of a molecule is known while its absolute position is unknown. The problem has been solved by observing that, when a molecule is translated in the unit cell, symmetry-related molecules move accordingly. As a consequence, all intermolecular vectors change (while intramolecular vectors remain unmodified) and the absolute position of a molecule would correspond to a maximum of the overlapping between the calculated cross vectors of the model and those of the observed Patterson map. The mathematical tool that checks the overlapping is usually called the translation function: several functions have been proposed, which have been reviewed by various authors, among which we quote Argos & Rossmann (1980) and Beurskens, Gould, Bruins Slot & Bosman (1987). Of particular interest for practical applications are the techniques used in *X-PLOR* (Brünger, 1992) based on previous work by Brünger (1990). To the knowledge of the authors, the joint probability distribution function of structure factors was never used to locate well oriented macromolecules up to a recent paper by Giacovazzo, Manna, Siliqi, Bolognesi & Rizzi (1997). A pioneering paper by Main (1976) (see also Heinerman, 1977) showed how the prior information on the orientation of a molecule can be exploited to estimate triplet invariants. Subsequent applications by other authors were mainly addressed to shifting small molecules correctly oriented but wrongly located by the tangent formula. The problem of locating a well oriented molecule is not simple because the translation function can show many maxima and the correct solution may not correspond to the largest one. Thus, some alternative approaches deserve to be checked, among which direct methods are a respectable candidate. Giacovazzo, Manna, Siliqi, Bolognesi & Rizzi (1997) have made some preliminary tests on the use of triplet invariants in the translation problem. Some real cases met in macromolecular crystallography have been revisited; for them, direct methods succeeded in locating the model macromolecules described in the original publications after being oriented by application of rotation functions. Besides triplet invariants, a useful tool for direct methods is expected to be the quartet invariants. While the probabilistic theory of quartets in the absence of prior information is already available (Hauptman, 1975*a*,*b*; Giacovazzo, 1975, 1976*a*,*b*; Hauptman & Green, 1976), no attempt has been made so far to estimate quartet phases when molecular fragments have been previously oriented. Such estimates cannot be obtained by simple application of the central limit theorem but require the use of the more complex method of the joint probability distribution functions of structure factors. This paper is devoted to the application of such a probabilistic approach to the quartet estimation: the first practical tests will also be described. For the sake of simplicity, we will not consider in our calculations the effect of space-group symmetry. To take into account such additional information, we should integrate our probabilistic approach with the representation method (Giacovazzo, 1977, 1980*a*). As a consequence, our theoretical results will be strictly valid in $P1$ but they may be applied in any space group.

† Permanent address: Laboratory of X-ray Diffraction, Department of Inorganic Chemistry, Faculty of Natural Sciences, Tirana University, Tirana, Albania.

2. Symbols and abbreviations

$F_{\mathbf{h}}$: structure factor with vectorial index $\mathbf{h}$
$\phi_{\mathbf{h}}$: phase of $F_{\mathbf{h}}$
$f_j(\mathbf{h})$: scattering factor of the jth atom
$\mathbf{C}_s \equiv (\mathbf{R}_s, \mathbf{T}_s)$: sth symmetry operator. $\mathbf{R}_s$ is the rotational part, $\mathbf{T}_s$ the translational part
m: order of the point group of the space group
N: number of atoms in the unit cell
N_f: number of molecular fragments (symmetry independent) with unknown position and fixed orientation
n_i: number of atoms in the ith molecular fragment
t_q: number of atoms (symmetry independent) whose positions are completely unknown
q: number of atoms (symmetry-equivalent included) whose positions are completely unknown
$\sum_q \mathbf{h} = \sum_{j=1}^{q} f_j^2(\mathbf{h})$: scattering power of the q atoms with completely unknown positions

$$\sum_N \mathbf{h} = \sum_{j=1}^{N} f_j^2(\mathbf{h})$$

$$\sum_{3q}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = \sum_{j=1}^{q} f_j(\mathbf{h}_1) f_j(\mathbf{h}_2) f_j(\mathbf{h}_3)$$

$$\sum_{3N}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = \sum_{j=1}^{N} f_j(\mathbf{h}_1) f_j(\mathbf{h}_2) f_j(\mathbf{h}_3)$$

$$\sum_{4q}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = \sum_{j=1}^{q} f_j(\mathbf{h}_1) f_j(\mathbf{h}_2) f_j(\mathbf{h}_3) f_j(\mathbf{h}_4)$$

$$\sum_{4N}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = \sum_{j=1}^{N} f_j(\mathbf{h}_1) f_j(\mathbf{h}_2) f_j(\mathbf{h}_3) f_j(\mathbf{h}_4)$$

$\sigma_i = \sum_{j=1}^{N} Z_j^i$, where Z_j is the atomic number of the jth atom
$\varepsilon_{\mathbf{h}}$: Wilson's factor responsible for the enhancement or depression of the intensity of certain subsets of reflections due to particular symmetry elements

$$\Phi = \phi_{\mathbf{h}_1} + \phi_{\mathbf{h}_2} + \phi_{\mathbf{h}_3} + \phi_{\mathbf{h}_4}$$

with $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4 = 0$
$I_0(x)$: modified Bessel function of order zero

3. The primitive random variables

The crystal structure can be divided into two parts: the first includes N_f molecular fragments with known orientation and their symmetry equivalents. The generic jth atom belonging to the ith fragment has trial positional vector $\mathbf{u}_j$ and $\boldsymbol{\tau}_i$ is the shift to be applied to the ith fragment to translate atoms to the correct positions

$$\mathbf{u}_j + \boldsymbol{\tau}_i, \quad \text{for} \quad j = 1, \ldots, n_i, \; i = 1, \ldots, N_f. \tag{1}$$

The second part of the structure includes q atoms whose positions $\mathbf{r}_j$ are completely unknown. Then the structure factor may be written as

$$F_{\mathbf{h}} = \sum_{i=1}^{N_f} g_i(\mathbf{h}) + F_{q\mathbf{h}} = F_{p\mathbf{h}} + F_{q\mathbf{h}}, \tag{2}$$

where

$$g_i(\mathbf{h}) = \sum_{s=1}^{m} \sum_{j=1}^{n_i} f_j(\mathbf{h}) \exp[2\pi i \mathbf{h} \mathbf{C}_s(\mathbf{u}_j + \boldsymbol{\tau}_i)]$$

$$= \sum_{s=1}^{m} g_{is}(\mathbf{h}) \exp(2\pi i \mathbf{h} \mathbf{R}_s \boldsymbol{\tau}_i),$$

$$g_{is}(\mathbf{h}) = \sum_{j=1}^{n_i} f_j(\mathbf{h}) \exp(2\pi i \mathbf{h} \mathbf{C}_s \mathbf{u}_j) \tag{3}$$

$$F_{q\mathbf{h}}(\mathbf{h}) = \sum_{s=1}^{m} \sum_{j=1}^{t_q} f_j(\mathbf{h}) \exp(2\pi i \mathbf{h} \mathbf{C}_s \mathbf{r}_j). \tag{4}$$

We note: (*a*) The primitive random variables in our probabilistic approach are the N_f shifts $\boldsymbol{\tau}_i$ and the t_q atomic positions $\mathbf{r}_j$. Accordingly, the use of the prior information reduces the number of primitive random variables from $(n_i N_f + t_q)$ random positions in the absence of prior information to $t = N_f + t_q$. (*b*) For the sake of simplicity, the mathematical model does not allow that atoms occupy special positions. This is not critical in most cases. (*c*) The factors $g_{is}(\mathbf{h})$ do not depend on $\boldsymbol{\tau}_i$ and may be calculated on the basis of the prior information for any i and s. (*d*) Our primitive random variables are considered statistically independent of each other and are randomly distributed in the unit cell. Forbidden domains (eventually generated by the symmetry elements) are not taken into consideration. Structure factors may be normalized by observing (Main, 1976) that

$$\langle |F_{\mathbf{h}}|^2 \rangle = \varepsilon_{\mathbf{h}} \left[\sum_{s=1}^{m} \sum_{i=1}^{N_f} |g_{is}(\mathbf{h})|^2 + \sum_q (\mathbf{h}) \right]. \tag{5}$$

Then,

$$E_{\mathbf{h}} = f_{\mathbf{h}} / \langle |F_{\mathbf{h}}|^2 \rangle^{1/2}.$$

Let us now make more explicit the notation (1)–(4) since it will be heavily used in the probabilistic calculations described in the following section. Let a_{is} and b_{is} be the real and imaginary parts of g_{is}, respectively, then,

$$a_{is}(\mathbf{h}) = \sum_{j=1}^{n_i} f_j(\mathbf{h}) \cos[2\pi \mathbf{h}(\mathbf{R}_s \mathbf{u}_j + \mathbf{T}_s)] = |g_{is}| \cos \phi_{is},$$

$$b_{is}(\mathbf{h}) = \sum_{j=1}^{n_i} f_j(\mathbf{h}) \sin[2\pi \mathbf{h}(\mathbf{R}_s \mathbf{u}_j + \mathbf{T}_s)] = |g_{is}| \sin \phi_{is},$$

$$\tan \phi_{is} = b_{is}/a_{is}.$$

Accordingly, the real and imaginary parts of $E_{\mathbf{h}}$, say

$$E_{\mathbf{h}} = A_{\mathbf{h}} + iB_{\mathbf{h}},$$

may be written as

$$A_{\mathbf{h}} = \left\{ \sum_{s=1}^{m} \sum_{i=1}^{N_f} [a_{is} \cos(2\pi \mathbf{h} \mathbf{R}_s \tau_i) - b_{is} \sin(2\pi \mathbf{h} \mathbf{R}_s \tau_i)] + \sum_{s=1}^{m} \sum_{i=1}^{t_q} \cos[2\pi \mathbf{h}(\mathbf{R}_s \mathbf{r}_j + \mathbf{T}_s)] \right\} \Big/ \langle |F_{\mathbf{h}}|^2 \rangle^{1/2} \quad (6)$$

$$B_{\mathbf{h}} = \left\{ \sum_{s=1}^{m} \sum_{i=1}^{N_f} [b_{is} \cos(2\pi \mathbf{h} \mathbf{R}_s \tau_i) - a_{is} \sin(2\pi \mathbf{h} \mathbf{R}_s \tau_i)] + \sum_{s=1}^{m} \sum_{i=1}^{t_q} \sin[2\pi \mathbf{h}(\mathbf{R}_s \mathbf{r}_j + \mathbf{T}_s)] \right\} \Big/ \langle |F_{\mathbf{h}}|^2 \rangle^{1/2}. \quad (7)$$

Expressions like (6) and (7) play a central role in the process for deriving the quartet phase distribution.

4. The joint probability distribution $P(E_{\mathbf{h}_1}, E_{\mathbf{h}_2}, E_{\mathbf{h}_3}, E_{\mathbf{h}_4}, E_{\mathbf{h}_1+\mathbf{h}_2}, E_{\mathbf{h}_1+\mathbf{h}_3}, E_{\mathbf{h}_1+\mathbf{h}_4})$ when molecular fragments have been oriented

We will use the following notation:

$$E_1 = A_1 + iB_1 = R_1 \exp(i\phi_1) = R_{\mathbf{h}_1} \exp(i\phi_{\mathbf{h}_1})$$
$$E_2 = A_2 + iB_2 = R_2 \exp(i\phi_2) = R_{\mathbf{h}_2} \exp(i\phi_{\mathbf{h}_2})$$
$$E_3 = A_3 + iB_3 = R_3 \exp(i\phi_3) = R_{\mathbf{h}_3} \exp(i\phi_{\mathbf{h}_3})$$
$$E_4 = A_4 + iB_4 = R_4 \exp(i\phi_4) = R_{\mathbf{h}_4} \exp(i\phi_{\mathbf{h}_4})$$
$$E_5 = A_5 + iB_5 = R_5 \exp(i\phi_5) = R_{\mathbf{h}_1+\mathbf{h}_2} \exp(i\phi_{\mathbf{h}_1+\mathbf{h}_2})$$
$$E_6 = A_6 + iB_6 = R_6 \exp(i\phi_6) = R_{\mathbf{h}_1+\mathbf{h}_3} \exp(i\phi_{\mathbf{h}_1+\mathbf{h}_3})$$
$$E_7 = A_7 + iB_7 = R_7 \exp(i\phi_7) = R_{\mathbf{h}_1+\mathbf{h}_4} \exp(i\phi_{\mathbf{h}_1+\mathbf{h}_4}),$$

where

$$\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4 = 0.$$

The joint probability distribution

$$P(\phi_1, \phi_2, \ldots, \phi_7, R_1, R_2, \ldots, R_7) \quad (8)$$

will be derived in the following way.

(*a*) The characteristic function

$$C(u_1, u_2, \ldots, u_7, v_1, v_2, \ldots, v_7) \quad (9)$$

of the distribution

$$P(A_1, A_2, \ldots, A_7, B_1, B_2, \ldots, B_7)$$

will be calculated *via* a Gram–Charlier expansion, by including terms up to t^{-1} order. u_i, v_i, $i = 1, 2, \ldots, 7$, are the carrying variables associated with A_i and B_i, $i = 1, 2, \ldots, 7$, respectively.

(*b*) The characteristic function

$$C(\theta_1, \theta_2, \ldots, \theta_7, \rho_1, \rho_2, \ldots, \rho_7) \quad (10)$$

of the distribution (8) will be obtained *via* the following change of variables:

$$u_i = 2^{1/2} \rho_i \cos\theta_i, \quad v_i = 2^{1/2} \rho_i \sin\theta_i, \quad i = 1, 2, \ldots, 7,$$
$$A_i = R_i \cos\phi_i, \quad B_i = R_i \sin\phi_i, \quad i = 1, 2, \ldots, 7,$$

θ_i, ρ_i, $i = 1, 2, \ldots, 7$, are carrying variables associated with ϕ_i and R_i, $i = 1, 2, \ldots, 7$, respectively.

(*c*) The distribution (8) will be finally derived by calculating the Fourier transform of (10). The final result is (see Giacovazzo, 1980*b*)

$$P(\phi_1, \phi_2, \ldots, \phi_7, R_1, R_2, \ldots, R_7) \cong (2\pi)^{-14} 2^7 R_1 \ldots R_7 \int_0^\infty \ldots \int_0^\infty \int_0^{2\pi} \ldots \int_0^{2\pi} \rho_1 \rho_2 \ldots \rho_7 \times \exp\{-i[2^{1/2} \rho_1 R_1 \cos(\phi_1 - \theta_1) + \ldots + 2^{1/2} \rho_7 R_7 \cos(\phi_7 - \theta_7)] - (\rho_1^2 + \ldots + \rho_7^2)/2\} \times [1 + S_3 + (S_4 + S_3^2/2)], \quad (11)$$

where

$$S_\mu = \sum_{r+s+\ldots+w=\mu} 2^{\mu/2} \sum_j ({}^j k_{rs\ldots w}/r!s!\ldots w!)(i^\mu) \times (\rho_1 \cos\theta_1)^r (\rho_2 \cos\theta_2)^s \ldots (\rho_7 \cos\theta_7)^w.$$

${}^j k_{rs\ldots w}$ are the cumulants of the distribution and the summation over j goes over the symmetry-independent atoms belonging to some of the molecular fragments, with known orientation or without. Lengthy calculations, here not reported for brevity, lead to the following expression:

$$\begin{aligned} P(\phi_1, \phi_2, \ldots, \phi_7, R_1, R_2, \ldots, R_7) &\cong (2\pi)^{-14} 2^7 R_1 \ldots R_7 \int_0^\infty \ldots \int_0^\infty \int_0^{2\pi} \ldots \int_0^{2\pi} \rho_1 \rho_2 \ldots \rho_7 \\ &\times \exp\{-i[2^{1/2} \rho_1 R_1 \cos(\phi_1 - \theta_1) + \ldots \\ &+ 2^{1/2} \rho_7 R_7 \cos(\phi_7 - \theta_7)] - (\rho_1^2 + \ldots + \rho_7^2)/2\} \\ &\times \{1 + i2^{-1/2}[c_{125} \rho_1 \rho_2 \rho_5 \cos(\theta_1 + \theta_2 - \theta_5 - \zeta_{125}) \\ &+ c_{345} \rho_3 \rho_4 \rho_5 \cos(\theta_3 + \theta_4 + \theta_5 - \zeta_{345}) \\ &+ c_{136} \rho_1 \rho_3 \rho_6 \cos(\theta_1 + \theta_3 - \theta_6 - \zeta_{136}) \\ &+ c_{246} \rho_2 \rho_4 \rho_6 \cos(\theta_2 + \theta_4 + \theta_6 - \zeta_{246}) \\ &+ c_{237} \rho_2 \rho_3 \rho_7 \cos(\theta_2 + \theta_3 - \theta_7 - \zeta_{237}) \\ &+ c_{147} \rho_1 \rho_4 \rho_7 \cos(\theta_1 + \theta_4 + \theta_7 - \zeta_{147})] \\ &+ 2^{-1} c_{1234} \rho_1 \rho_2 \rho_3 \rho_4 \cos(\theta_1 + \theta_2 + \theta_3 + \theta_4 - \zeta_{1234}) \\ &- 2^{-2}[c_{125} c_{345} \rho_1 \rho_2 \rho_3 \rho_4 \rho_5^2 \cos(\theta_1 + \theta_2 + \theta_3 + \theta_4 \\ &- \zeta_{125} - \zeta_{345}) + c_{136} c_{246} \rho_1 \rho_2 \rho_3 \rho_4 \rho_6^2 \\ &\times \cos(\theta_1 + \theta_2 + \theta_3 + \theta_4 - \zeta_{136} - \zeta_{246}) + c_{237} c_{147} \\ &\times \rho_1 \rho_2 \rho_3 \rho_4 \rho_7^2 \cos(\theta_1 + \theta_2 + \theta_3 + \theta_4 - \zeta_{237} - \zeta_{147})] \\ &- 2^{-3}[c_{125}^2 \rho_1^2 \rho_2^2 \rho_5^2 + c_{345}^2 \rho_3^2 \rho_4^2 \rho_5^2 + c_{136}^2 \rho_1^2 \rho_3^2 \rho_6^2 \\ &+ c_{246}^2 \rho_2^2 \rho_4^2 \rho_6^2 + c_{147}^2 \rho_1^2 \rho_4^2 \rho_7^2 + c_{237}^2 \rho_2^2 \rho_3^2 \rho_7^2] \\ &+ \ldots\} \, \mathrm{d}\phi_1 \ldots \mathrm{d}\phi_7 \mathrm{d}\rho_1 \ldots \mathrm{d}\rho_7. \end{aligned} \quad (12)$$

Terms not essential for the estimate of the quartet phase invariant have been omitted from (12). We used the following notation:

$$
\begin{aligned}
c_{klp} &= (c'^2_{klp} + c''^2_{klp})^{1/2},\\
c'_{klp} &= [\langle|F_{\mathbf{h}_k}|^2\rangle\langle|F_{\mathbf{h}_l}|^2\rangle\langle|F_{\mathbf{h}_p}|^2\rangle]^{-1/2}\\
&\quad\times\Bigg\{\sum_{s=1}^{m}\sum_{i=1}^{N_f}|g_{is}(\mathbf{h}_k)g_{is}(\mathbf{h}_l)g_{is}(\mathbf{h}_p)|\\
&\quad\times\cos[\phi_{is}(\mathbf{h}_k)+\phi_{is}(\mathbf{h}_l)+\phi_{is}(\mathbf{h}_p)]\\
&\quad+\textstyle\sum_{3q}(\mathbf{h}_k,\mathbf{h}_l,\mathbf{h}_p)\Bigg\}\\
c''_{klp} &= [\langle|F_{\mathbf{h}_k}|^2\rangle\langle|F_{\mathbf{h}_l}|^2\rangle\langle|F_{\mathbf{h}_p}|^2\rangle]^{-1/2}\\
&\quad\times\Bigg\{\sum_{s=1}^{m}\sum_{i=1}^{N_f}|g_{is}(\mathbf{h}_k)g_{is}(\mathbf{h}_l)g_{is}(\mathbf{h}_p)|\\
&\quad\times\sin[\phi_{is}(\mathbf{h}_k)+\phi_{is}(\mathbf{h}_l)+\phi_{is}(\mathbf{h}_p)]\Bigg\}\\
\zeta_{klp} &= \tan^{-1}(c''_{klp}/c'_{klp}).
\end{aligned}
$$

The indices k, l, p vary over the combinations present in (12) for which

$$\mathbf{h}_h+\mathbf{h}_l+\mathbf{h}_p=0.$$

Furthermore,

$$
\begin{aligned}
c_{1234} &= (c'^2_{1234}+c''^2_{1234})^{1/2}\\
c'_{1234} &= [\langle|F_{\mathbf{h}_1}|^2\rangle\langle|F_{\mathbf{h}_2}|^2\rangle\langle|F_{\mathbf{h}_3}|^2\rangle\langle|F_{\mathbf{h}_4}|^2\rangle]^{-1/2}\\
&\quad\times\Bigg\{\sum_{s=1}^{m}\sum_{i=1}^{N_f}|g_{is}(\mathbf{h}_1)g_{is}(\mathbf{h}_2)g_{is}(\mathbf{h}_3)g_{is}(\mathbf{h}_4)|\\
&\quad\times\cos[\phi_{is}(\mathbf{h}_1)+\phi_{is}(\mathbf{h}_2)+\phi_{is}(\mathbf{h}_3)+\phi_{is}(\mathbf{h}_4)]\\
&\quad+\textstyle\sum_{4q}(\mathbf{h}_1,\mathbf{h}_2,\mathbf{h}_3,\mathbf{h}_4)\Bigg\}\\
c''_{1234} &= [\langle|F_{\mathbf{h}_1}|^2\rangle\langle|F_{\mathbf{h}_2}|^2\rangle\langle|F_{\mathbf{h}_3}|^2\rangle\langle|F_{\mathbf{h}_4}|^2\rangle]^{-1/2}\\
&\quad\times\Bigg\{\sum_{s=1}^{m}\sum_{i=1}^{N_f}|g_{is}(\mathbf{h}_1)g_{is}(\mathbf{h}_2)g_{is}(\mathbf{h}_3)g_{is}(\mathbf{h}_4)|\\
&\quad\times\sin[\phi_{is}(\mathbf{h}_1)+\phi_{is}(\mathbf{h}_2)+\phi_{is}(\mathbf{h}_3)+\phi_{is}(\mathbf{h}_4)]\Bigg\}\\
\zeta_{1234} &= \tan^{-1}(c''_{1234}/c'_{1234}).
\end{aligned}
$$

The integration of the right-hand side of (12) may be performed by applying well known mathematical techniques here not reported for brevity. After some calculations, we obtain

$$
\begin{aligned}
&P(\phi_1,\phi_2,\ldots,\phi_7,R_1,R_2,\ldots,R_7)\\
&\quad\cong\pi^{-7}R_1R_2\ldots R_7\exp(-R_1^2-R_2^2-\ldots-R_7^2)\\
&\qquad\times\{1+2[c_{125}R_1R_2R_5\cos(\phi_1+\phi_2-\phi_5-\zeta_{125})\\
&\qquad+c_{345}R_3R_4R_5\cos(\phi_3+\phi_4+\phi_5-\zeta_{345})\\
&\qquad+c_{136}R_1R_3R_6\cos(\phi_1+\phi_3-\phi_6-\zeta_{136})\\
&\qquad+c_{246}R_2R_4R_6\cos(\phi_2+\phi_4+\phi_6-\zeta_{246})\\
&\qquad+c_{237}R_2R_3R_7\cos(\phi_2+\phi_3-\phi_7-\zeta_{237})\\
&\qquad+c_{147}R_1R_4R_7\cos(\phi_1+\phi_4+\phi_7-\zeta_{147})]\\
&\qquad-2R_1R_2R_3R_4[c_{1234}\cos(\phi_1+\phi_2+\phi_3+\phi_4-\zeta_{1234})\\
&\qquad+c_{125}c_{345}(R_5^2-1)\cos(\phi_1+\phi_2+\phi_3+\phi_4\\
&\qquad-\zeta_{125}-\zeta_{345})+c_{136}c_{246}(R_6^2-1)\\
&\qquad\times\cos(\phi_1+\phi_2+\phi_3+\phi_4-\zeta_{136}-\zeta_{246})+c_{237}c_{147}\\
&\qquad\times(R_7^2-1)\cos(\phi_1+\phi_2+\phi_3+\phi_4-\zeta_{237}-\zeta_{147})]\\
&\qquad+Q+\ldots\},
\end{aligned}
\tag{13}
$$

where

$$
\begin{aligned}
Q &= c_{125}^2(R_1^2-1)(R_2^2-1)(R_5^2-1)\\
&\quad+c_{345}^2(R_3^2-1)(R_4^2-1)(R_5^2-1)\\
&\quad+c_{136}^2(R_1^2-1)(R_3^2-1)(R_6^2-1)\\
&\quad+c_{246}^2(R_2^2-1)(R_4^2-1)(R_6^2-1)\\
&\quad+c_{237}^2(R_2^2-1)(R_3^2-1)(R_7^2-1)\\
&\quad+c_{147}^2(R_1^2-1)(R_4^2-1)(R_7^2-1).
\end{aligned}
$$

Equation (13) is the main result of this paper. If no well oriented molecular fragment exists then

$$\textstyle\sum_{3q}(\mathbf{h}_k,\mathbf{h}_l,\mathbf{h}_p)=\sum_{3N}(\mathbf{h}_k,\mathbf{h}_l,\mathbf{h}_p),$$

$$\textstyle\sum_q(\mathbf{h})=\sum_N(\mathbf{h}),\quad\langle|F_{\mathbf{h}}|^2\rangle=\sum_N(\mathbf{h}),$$

$$c'_{klp}=\textstyle\sum_{3N}(\mathbf{h}_k,\mathbf{h}_l,\mathbf{h}_p)\big[\sum_N(\mathbf{h}_k)\sum_N(\mathbf{h}_l)\sum_N(\mathbf{h}_p)\big]^{-1/2},$$

$$c''_{klp}=0,$$

$$
\begin{aligned}
c'_{1234} &= \textstyle\sum_{4N}(\mathbf{h}_1,\mathbf{h}_2,\mathbf{h}_3,\mathbf{h}_4)\\
&\quad\times\big[\sum_N(\mathbf{h}_1)\sum_N(\mathbf{h}_2)\sum_N(\mathbf{h}_3)\sum_N(\mathbf{h}_4)\big]^{-1/2},
\end{aligned}
$$

$$c''_{1234}=0,$$

$$c_{klp}=\sigma_3/\sigma_2^{3/2}\cong N^{-1/2},\quad c_{1234}\cong\sigma_4/\sigma_2^2\cong N^{-1}.$$

In these conditions, distribution (13) coincides with equation (8) of Giacovazzo (1976*a*).

5. The conditional probability distribution $P(\Phi|R_1,R_2,\ldots,R_7)$

We are often interested in the conditional distribution

$$P(\Phi|R_1,R_2,\ldots,R_7).\tag{14}$$

Equation (14) may be derived in four steps:

(*a*) By integrating (13) over ϕ_5, ϕ_6 and ϕ_7 so obtaining the marginal distribution

$$P(\phi_1, \phi_2, \phi_3, \phi_4, R_1, R_2, \ldots, R_7). \tag{15}$$

Equation (15) does not contain triplet terms.

(*b*) By stating in (15) the identity

$$\begin{aligned}
&2R_1R_2R_3R_4[c_{1234}\cos(\phi_1+\phi_2+\phi_3+\phi_4-\zeta_{1234})\\
&\quad+c_{125}c_{345}(R_5^2-1)\cos(\phi_1+\phi_2+\phi_3+\phi_4-\zeta_{125}\\
&\quad-\zeta_{345})+c_{136}c_{246}(R_6^2-1)\cos(\phi_1+\phi_2+\phi_3+\phi_4\\
&\quad-\zeta_{136}-\zeta_{246})+c_{237}c_{147}(R_7^2-1)\\
&\quad\times\cos(\phi_1+\phi_2+\phi_3+\phi_4-\zeta_{237}-\zeta_{147})]\\
&\quad=\alpha\cos(\phi_1+\phi_2+\phi_3+\phi_4-\xi),
\end{aligned}$$

where

$$\begin{aligned}
\alpha\cos\xi &= 2R_1R_2R_3R_4[c_{1234}\cos\zeta_{1234}\\
&\quad+c_{125}c_{345}(R_5^2-1)\cos(\zeta_{125}+\zeta_{345})\\
&\quad+c_{136}c_{246}(R_6^2-1)\cos(\zeta_{136}+\zeta_{246})\\
&\quad+c_{237}c_{147}(R_7^2-1)\cos(\zeta_{237}+\zeta_{147})],\\
\alpha\sin\xi &= 2R_1R_2R_3R_4[c_{1234}\sin\zeta_{1234}\\
&\quad+c_{125}c_{345}(R_5^2-1)\sin(\zeta_{125}+\zeta_{345})\\
&\quad+c_{136}c_{246}(R_6^2-1)\sin(\zeta_{136}+\zeta_{246})\\
&\quad+c_{237}c_{147}(R_7^2-1)\sin(\zeta_{237}+\zeta_{147})],
\end{aligned}$$

$$\begin{aligned}
\alpha &= [(\alpha\cos\xi)^2+(\alpha\sin\xi)^2]^{1/2},\\
\xi &= \tan^{-1}[(\alpha\sin\xi)/(\alpha\cos\xi)].
\end{aligned}$$

(*c*) By integrating such a modified distribution over ϕ_1, ϕ_2, ϕ_3, ϕ_4 under the condition that $\phi_1+\phi_2+\phi_3+\phi_4=\Phi$. Then the distribution

$$\begin{aligned}
&P(\Phi, R_1, R_2, \ldots, R_7)\\
&\quad\cong 64\pi^{-1}R_1\ldots R_7\exp(-R_1^2-\ldots-R_7^2)\\
&\qquad\times[1+Q+\alpha\cos(\Phi-\xi)]
\end{aligned}$$

is obtained.

(*d*) By calculating

$$\begin{aligned}
&P(\Phi|R_1, R_2, \ldots, R_7)\\
&\quad=P(\Phi, R_1, R_2, \ldots, R_7)\Big/\int_0^{2\pi}P(\Phi, R_1, R_2, \ldots, R_7)\,\mathrm{d}\Phi\\
&\quad=2\pi^{-1}[1+\alpha(1+Q)^{-1}\cos(\Phi-\xi)],
\end{aligned}$$

which may be approximated by the normalized von Mises distribution

$$P(\Phi|R_1, R_2, \ldots, R_7)\cong[2\pi I_0(G)]^{-1}\exp[G\cos(\Phi-\xi)], \tag{16}$$

where

$$G=\alpha/(1+Q). \tag{17}$$

We note: (i) the expected value of Φ is ξ, which may assume any value between 0 and 2π. (ii) G is the reliability parameter: it is always positive. A large value of G indicates that the phase relationship $\Phi\approx\xi$ is reliable. (iii) Q is a scale factor that modulates the reliability of the phase indication. It may be settled to zero when it is found to be negative (Giacovazzo, 1980*b*). (iv) If no molecular fragment is oriented, (16) reduces to the well known Giacovazzo (1976*a*, 1980*b*) quartet formula.

6. A geometric interpretation of the quartet phase distribution function

The estimate of Φ *via* (16) depends on an intricate interrelationship between the seven magnitudes R_i, the 12 parameters c_{klp} and ζ_{klp}, and the two parameters c_{1234} and ζ_{1234}. In order to understand the potentiality and the main characteristics of the formula, we describe here a geometric interpretation that will help the reader to better understand our results and to design its possible use.

Let us calculate $\langle E_{\mathbf{h}_k}E_{\mathbf{h}_l}E_{\mathbf{h}_p}\rangle$ when $\mathbf{h}_k+\mathbf{h}_l+\mathbf{h}_p=0$. According to (1) and (2),

$$\begin{aligned}
\langle E_{\mathbf{h}_k}E_{\mathbf{h}_l}E_{\mathbf{h}_p}\rangle &= \left\langle\left[\sum_{i=1}^{N_f}g_i(\mathbf{h}_k)+F_{q\mathbf{h}_k}\right]\left[\sum_{i=1}^{N_f}g_i(\mathbf{h}_l)+F_{q\mathbf{h}_l}\right]\right.\\
&\quad\left.\times\left[\sum_{i=1}^{N_f}g_i(\mathbf{h}_p)+F_{q\mathbf{h}_p}\right]\right\rangle\\
&\quad\times(\langle|F_{\mathbf{h}_k}|^2\rangle\langle|F_{\mathbf{h}_l}|^2\rangle\langle|F_{\mathbf{h}_p}|^2\rangle)^{-1/2}\\
&=\left[\sum_i^{N_f}\sum_{s=1}^{m}g_{is}(\mathbf{h}_k)g_{is}(\mathbf{h}_l)g_{is}(\mathbf{h}_p)\right.\\
&\quad\left.+\sum\nolimits_{3q}(\mathbf{h}_k, \mathbf{h}_l, \mathbf{h}_p)\right]\\
&\quad\times(\langle|F_{\mathbf{h}_k}|^2\rangle\langle|F_{\mathbf{h}_l}|^2\rangle\langle|F_{\mathbf{h}_p}|^2\rangle)^{-1/2}\\
&=c'_{klp}+ic''_{klp}\\
&=c_{klp}\exp(i\zeta_{klp}).
\end{aligned} \tag{18}$$

Similarly,

$$\langle E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}E_{\mathbf{h}_4}\rangle=c'_{1234}+ic''_{1234}=c_{1234}\exp(i\zeta_{1234}). \tag{19}$$

According to (18), c_{klp} is the expected value of $|E_{\mathbf{h}_k}E_{\mathbf{h}_l}E_{\mathbf{h}_p}|$ and ζ_{klp} is the expected value of $(\phi_{\mathbf{h}_k}+\phi_{\mathbf{h}_l}+\phi_{\mathbf{h}_p})$. Similarly, c_{1234} is the expected value of $|E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}E_{\mathbf{h}_4}|$ and ζ_{1234} is the expected value of $(\phi_{\mathbf{h}_1}+\phi_{\mathbf{h}_2}+\phi_{\mathbf{h}_3}+\phi_{\mathbf{h}_4})$ when $|E_{\mathbf{h}_1+\mathbf{h}_2}|$, $|E_{\mathbf{h}_1+\mathbf{h}_3}|$ and $|E_{\mathbf{h}_2+\mathbf{h}_3}|$ are all unknown. If such cross magnitudes are known, the quartet $E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}E_{\mathbf{h}_4}$ may be considered

equal (but for the real positive factor $R^2_{\mathbf{h}_1+\mathbf{h}_2}$) to the product $(E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2})(E_{\mathbf{h}_3}E_{\mathbf{h}_4}E_{\bar{\mathbf{h}}_1+\bar{\mathbf{h}}_2})$. Therefore,

$$\langle E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}E_{\mathbf{h}_4}\rangle \cong L\langle E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}\cdot E_{\mathbf{h}_3}E_{\mathbf{h}_4}E_{\bar{\mathbf{h}}_1+\bar{\mathbf{h}}_2}\rangle,$$

where L is a scale factor. If the two triplets are considered independent of each other then

$$\langle E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}E_{\mathbf{h}_4}\rangle \cong L\langle E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}\rangle\langle E_{\mathbf{h}_3}E_{\mathbf{h}_4}E_{\bar{\mathbf{h}}_1+\bar{\mathbf{h}}_2}\rangle$$
$$\cong Lc_{125}c_{345}\exp i(\zeta_{125}+\zeta_{345}). \qquad (20)$$

In fact, the two triplets are not independent and our distribution suggests replacing (20) by

$$\cong c_{125}c_{345}(R_5^2-1)\exp i(\zeta_{125}+\zeta_{345}).$$

If the same considerations are applied to the cross reflections E_6 and E_7, we could conclude that the complex parameter $G\exp(i\xi)$, as defined in (17), may be recovered [unless the scaling factor is $(1+Q)^{-1}$] by adding four vectors (see Fig. 1):

(I) the vector $\mathbf{c}_1$ with modulus $2R_1R_2R_3R_4c_{1234}$ and phase ζ_{1234};

(II) the vector $\mathbf{c}_2$ with modulus $2R_1R_2R_3R_4\times c_{125}c_{345}(R_5^2-1)$ and phase $\zeta_{125}+\zeta_{345}$;

(III) the vector $\mathbf{c}_3$ with modulus $2R_1R_2R_3R_4\times c_{136}c_{246}(R_6^2-1)$ and phase $\zeta_{136}+\zeta_{246}$;

(IV) the vector $\mathbf{c}_4$ with modulus $2R_1R_2R_3R_4\times c_{147}c_{237}(R_7^2-1)$ and phase $\zeta_{147}+\zeta_{237}$.

It may be observed that: (*a*) $2R_1R_2R_3R_4$ is contained in all the four moduli c_i. Thus, a large value of $R_1R_2R_3R_4$ is a necessary condition for the reliability of the phase relationship $\Phi\approx\xi$. (*b*) The values of the factors c_{1234}, $c_{125}c_{345}$, $c_{136}c_{246}$ and $c_{147}c_{237}$ depend on the structure complexity and on the scattering power of the molecular fragment in fixed orientation. When such a power is small, the four factors are close to N^{-1} and (16) is of no use in macromolecular crystallography. If the power is large enough, the resultant modulus G may be sufficiently large to be useful for proteins too. (*c*) Large G values are obtained if the following conditions are satisfied:

(i) $\zeta_{1234}\approx\zeta_{125}+\zeta_{345}\approx\zeta_{136}+\zeta_{246}\approx\zeta_{147}+\zeta_{237}$;

(ii) (R_5^2-1), (R_6^2-1) and (R_7^2-1) are all positive or all negative.

For small molecules, the conditions (i) and (ii) are satisfied if the moduli R_i, $i=1,2,\ldots,7$, are sufficiently large. That is not probable for macromolecules. (*d*) Unreliable phase indications are obtained when the four vectors $\mathbf{c}_i$ are randomly distributed over the trigonometric circle.

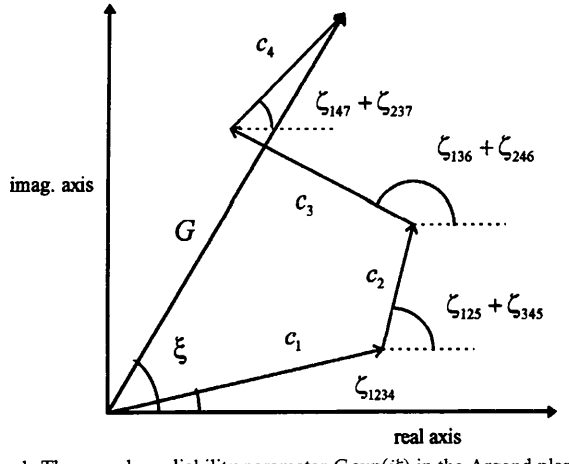

Fig. 1. The complex reliability parameter $G\exp(i\xi)$ in the Argand plane in terms of component vectors.

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

NREF is the number of symmetry-independent reflections.

Code name	Space group	NREF	Resolution (Å)
M-FABP[(a)]	$P2_12_12_1$	7595	2.14
LPH[(b)]	$P2_1$	17352	1.50
STM[(c)]	$P2_12_12_1$	9758	1.97
XSD[(d)]	$P2_12_12_1$	19056	2.01

References: (*a*) Zanotti, Scapin, Spadon, Veerkamp & Sacchettini (1992); (*b*) Rizzi, Wittemberg, Coda, Fasano, Ascenzi & Bolognesi (1994); (*c*) Nardini, Tarricone, Rizzi, Lania, Desideri, De Sanctis, Coletta, Petruzzelli, Ascenzi, Coda & Bolognesi (1995); (*d*) Djinovic Carugo, Collyer, Coda, Carri, Battistoni, Bottaro, Polticelli, Desideri & Bolognesi (1993).

7. The first applications

We used four protein structures to test the efficiency of (16): code name, space group and other useful data are in Table 1. We checked the correctness of the formula (16) in ideal conditions: the full structure is the search model, the structure factors are calculated from the published crystal structure up to experimental resolution.

It is well known (see Giacovazzo, Burla & Cascarano, 1992; Burla, Cascarano & Giacovazzo, 1992; Cascarano, Giacovazzo, Moliterni & Polidori, 1994; and literature quoted therein) that quartets with large cross magnitudes are strongly correlated with triplets: accordingly, the simultaneous use of triplets and of large-cross-magnitude quartet invariants is not advised. Even if this result was established when no prior information on the orientation of a molecule is available, it is very likely that it holds even in the case treated in this paper. We therefore focused our attention on the quartet invariants with small cross magnitudes only.

For each of the four test structures, NLAR reflections (those with the largest $|E|$ values) are selected, among which the basis vectors of the quartet invariants are found. The computing technique (widely used in literature) requires the computation of the psi-zero triplets, each triplet having two strong-magnitude reflections in the set NLAR and one small-magnitude reflection. A quartet is considered in our statistical

Table 2. *Statistical calculations for triplet invariants [estimated via (21)] and quartet invariants [estimated via (16)] for all the test structures using the full published structure as search model*

NR is the number of triplets (quartets) having $T >$ ARG ($Q >$ ARG), $\langle|\Delta\Phi|\rangle$ is average error (°), % is the percentage of triplets with $|\Delta\Phi| > \pi/2$.

M-FABP

	Triplet invariants			Quartet invariants		
ARG	NR	%	$\langle\|\Delta\Phi\|\rangle$	NR	%	$\langle\|\Delta\Phi\|\rangle$
0.0	21897	99.0	20	200000	99.0	18
0.4	21754	99.1	19	19728	99.4	18
2.0	15173	100.0	15	7920	100.0	13
3.2	5552	100.0	12	910	100.0	10
4.4	809	100.0	8	23	100.0	7
6.5	3	100.0	4	–	–	–

LPH

	Triplet invariants			Quartet invariants		
ARG	NR	%	$\langle\|\Delta\Phi\|\rangle$	NR	%	$\langle\|\Delta\Phi\|\rangle$
0.0	10455	99.5	12	8497	98.8	15
0.4	10438	99.5	12	8474	98.9	15
2.0	10205	99.8	11	7940	99.5	14
3.2	8958	99.9	11	5881	99.7	13
4.4	3471	99.9	10	2537	99.8	12
6.5	204	99.5	14	365	99.5	14
15.0	4	100.0	8	9	100.0	8

STM

	Triplet invariants			Quartet invariants		
ARG	NR	%	$\langle\|\Delta\Phi\|\rangle$	NR	%	$\langle\|\Delta\Phi\|\rangle$
0.0	17455	99.8	15	20000	99.8	18
0.4	17406	99.8	15	19802	99.4	17
2.0	13198	100.0	13	7727	100.0	13
3.2	4535	100.0	9	756	100.0	9
4.4	348	100.0	4	22	100.0	6

XSD

	Triplet invariants			Quartet invariants		
ARG	NR	%	$\langle\|\Delta\Phi\|\rangle$	NR	%	$\langle\|\Delta\Phi\|\rangle$
0.0	11357	100.0	11	20000	99.9	13
0.4	11357	100.0	11	19974	99.9	13
2.0	10877	100.0	10	15972	100.0	12
3.2	7710	100.0	9	5658	100.0	10
4.4	2462	100.0	7	585	100.0	7
6.5	10	100.0	2	2	100.0	3

calculations only if

$$R_5^2 + R_6^2 + R_7^2 < 2.$$

A statistical analysis of the results is shown in Table 2. N_q is the number of the quartets having $G >$ ARG,

$$\langle|\Delta\Phi|\rangle = \langle|\Phi_{\text{true}} - \Phi_{\text{est}}|\rangle$$

is the corresponding average of the absolute difference between the 'true' (corresponding to the published test structure) quartet phase and the quartet phase estimated *via* (16), % is the percentage of quartets for which $\langle|\Delta\Phi|\rangle$ is smaller than $\pi/2$. For reader usefulness, quartet results are always compared with the corresponding triplet invariants (calculated for the same NLAR reflections). Triplets were estimated *via* the Main (1976) formula

$$P(\Phi_3) \approx [2\pi I_0(T)]^{-1} \exp[T\cos(\Phi_3 - \xi_3)], \qquad (21)$$

where

$$\Phi_3 = \phi_{\mathbf{h}_1} + \phi_{\mathbf{h}_2} + \phi_{\mathbf{h}_3} \quad \text{with} \quad \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0$$

$$T = 2|E_{M\mathbf{h}_1} E_{M\mathbf{h}_2} E_{M\mathbf{h}_3}|[(T'^2 + T''^2) \times (\langle|F_{\mathbf{h}_1}|^2\rangle_M \langle|F_{\mathbf{h}_2}|^2\rangle_M \langle|F_{\mathbf{h}_3}|^2\rangle_M)^{-1}]^{1/2}$$

$$T' = \Re\left[\sum_{s=1}^{m}\sum_{j=1}^{N_f} g_{is}(\mathbf{h})_1 g_{is}(\mathbf{h}_2) g_{is}(\mathbf{h}_3) + \sum_{3q}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)\right]$$

$$T'' = \Im\left[\sum_{s=1}^{m}\sum_{j=1}^{N_f} g_{is}(\mathbf{h})_1 g_{is}(\mathbf{h}_2) g_{is}(\mathbf{h}_3)\right]$$

$$\tan\xi_3 = T''/T'.$$

Two features in Table 2 should be noted: the first is that the reliability parameter of the quartets spans over large ARG intervals, even if with a frequency slightly smaller than that of the triplet parameter. The second feature to note is that, for equal values of the reliability parameter, triplets and quartets show a similar reliability. For example, for M-FABP, the percentage of quartets with $G > 0.4$ for which $|\Delta\Phi| < \pi/2$ is 99.1; the corresponding percentage for triplets is 99.4. It may be concluded that (16) works correctly. However, the quartet behaviour is quite different from that corresponding to the so-called negative quartets calculated in the absence of prior information. Indeed, in this case, negative quartet reliability is much smaller than triplet reliability and rapidly decreases with increasing unit-cell size. We guess that the prior knowledge of the molecule orientation provides negative quartets with a big supplement of information that can make them competitive with triplet relationships. In order to check the quartet reliability when the prior information on the molecular orientation concerns only a fragment (and not all the structure), we suppose that only 60% of the asymmetric unit is well oriented and the rest is unknown. The statistical analysis of the triplet and quartet reliability is shown in Table 3. We observe: (*a*) as expected, the average triplet and quartet reliabilities decrease with respect to Table 2; (*b*) even if lower than triplet reliability, quartet reliability is potentially useful. The above results open the way to a possible application of small-cross-magnitude quartets in molecular replacement techniques.

8. Conclusions

The probabilistic theory of the quartet invariants when the orientation of a molecular fragment is known while

Table 3. *Statistical calculations for triplet invariants [estimated via (21)] and quartet invariants [estimated via (16)] for all the test structures using 60% of the structure as search model*

NR is the number of triplets (quartets) having T > ARG (Q > ARG), $\langle|\Delta\Phi|\rangle$ is average error (°), % is the percentage of triplets with $|\Delta\Phi| > \pi/2$.

M-FABP

	Triplet invariants			Quartet invariants		
ARG	NR	%	$\langle\|\Delta\Phi\|\rangle$	NR	%	$\langle\|\Delta\Phi\|\rangle$
0.0	25889	71.8	64	20000	63.5	73
0.4	22926	73.7	62	12081	68.0	69
2.0	1575	91.0	40	196	80.6	54
3.2	123	96.7	32	18	88.9	40
4.4	4	100.0	30	-	-	-

LPH

	Triplet invariants			Quartet invariants		
ARG	NR	%	$\langle\|\Delta\Phi\|\rangle$	NR	%	$\langle\|\Delta\Phi\|\rangle$
0.0	13063	78.8	36	20000	68.0	69
0.4	12538	79.7	55	15769	71.5	65
2.0	4512	89.8	42	2209	83.7	50
3.2	1221	92.6	37	414	82.1	52
4.4	220	90.5	37	75	73.3	63
6.5	14	57.1	75	3	33.3	133

STM

	Triplet invariants			Quartet invariants		
ARG	NR	%	$\langle\|\Delta\Phi\|\rangle$	NR	%	$\langle\|\Delta\Phi\|\rangle$
0.0	19010	73.8	62	20000	63.4	74
0.4	16680	75.7	59	10355	69.2	67
2.0	891	92.8	41	78	84.6	47
3.2	55	94.5	31	2	100.0	17
4.4	2	100.0	1	-	-	-

XSD

	Triplet invariants			Quartet invariants		
ARG	NR	%	$\langle\|\Delta\Phi\|\rangle$	NR	%	$\langle\|\Delta\Phi\|\rangle$
0.0	11474	77.0	59	20000	66.7	70
0.4	10794	78.0	57	15157	70.2	66
2.0	2184	90.8	40	717	86.8	49
3.2	287	95.8	31	64	90.6	46
4.4	36	94.4	32	5	100.0	38

its absolute position is unknown has been described. The conclusive formula estimating the quartet phase is of von Mises type: the expected phase value may lie anywhere between 0 and 2π, the reliability parameter may be large enough even for proteins, provided a sufficiently large fragment has known orientation. Accordingly, quartets can find useful applications in molecular replacement methods as an alternative to the widely used translation functions. The role of the quartet invariants in molecular replacement methods has still to be established. There are two main problems to solve: (i) Recently (Giacovazzo, Manna, Siliqi, Bolognesi & Rizzi, 1997), it has been shown that only if the Main (1976) formula for triplets is suitably modified can triplet invariants be usefully applied to solve the translation problem. Similar modifications to (16) would probably be necessary to make it sufficiently robust to succeed in practical use (where errors in measured data are combined with errors in the model fragment and in its orientation). (ii) The quality of the information yielded by negative quartet invariants in practical cases has still to be compared with that provided by triplet invariants. In other words, we have to establish if the use of additional computing resources necessary for the estimation and the use of quartet invariants in the phasing procedure are justified by better results. Both the above problems will be investigated in a future paper.

References

Argos, P. & Rossmann, M. G. (1980). *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 381–389. New York: Plenum Press.

Beurskens, P. T., Gould, R. O., Bruins Slot, H. J. & Bosman, W. P. (1987). *Z. Kristallogr.* **179**, 127–159.

Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.

Brünger, A. T. (1992). *X-PLOR* version 3.1. *A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT, USA.

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

Cascarano, G., Giacovazzo, C., Moliterni, A. G. & Polidori, G. (1994). *Acta Cryst.* A**50**, 22–27.

Djinovic Carugo, K., Collyer, C., Coda, A., Carri, M. T., Battistoni, A., Bottaro, G., Polticelli, F., Desideri, A. & Bolognesi, M. (1993). *Biochem. Biophys. Res. Commun.* **194**, 1008–1011.

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). *Acta Cryst.* A**33**, 933–944.

Giacovazzo, C. (1980*a*). *Acta Cryst.* A**36**, 362–372.

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

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

Giacovazzo, C., Manna, L., Siliqi, D., Bolognesi, M. & Rizzi, M. (1997). *Acta Cryst.* A**53**. Submitted

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

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

Hauptman, H. & Green, E. A. (1976). *Acta Cryst.* A**32**, 45–49.

Heinerman, J. J. L. (1977). *Acta Cryst.* A**33**, 100–106.

Main, P. (1976). *Crystallographic Computing Techniques*, pp. 97–105. Munksgaard: Copenhagen.

Nardini, M., Tarricone, C., Rizzi, M., Lania, A., Desideri, G., De Sanctis, G., Coletta, M., Petruzzelli, R., Ascenzi, P., Coda, A. & Bolognesi, M. (1995). *J. Mol. Biol.* **247**, 459–465.

Rizzi, M., Wittemberg, J. B., Coda, A., Fasano, M., Ascenzi, P. & Bolognesi, M. (1994). *J. Mol. Biol.* **244**, 86–99.

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