Table 1. *Of* 68 *centrosymmetric reflexions, the* 34 *with the concentration parameter* α *(according to non-centrosymmetric formulae) larger than* 1·0 *are printed*

φ_t and φ_c are true and calculated phase values.
'*' labels wrong indications.

h	k	l	φ_t	φ_c	α
10	0	12	185·0	185·2	17·05
2	0	2	183·0	175·5	8·97
0	18	4	1·0	5·0	7·55
11	0	2	180·0	4·5*	5·82
8	0	12	184·0	184·2	4·53
0	12	12	6·0	4·4	3·90
0	6	17	183·0	184·9	3·79
11	0	4	181·0	185·2	3·54
0	2	4	4·0	3·5	3·00
0	4	6	4·0	183·4*	2·99
0	10	6	358·0	185·0*	2·37
0	18	2	181·0	185·4	2·23
10	14	0	8·0	5·3	2·13
0	14	12	189·0	184·0	1·52
2	7	0	1·0	3·8	1·24
8	1	0	180·0	182·7	1·18
0	14	3	181·0	2·2*	1·09
0	16	10	186·0	186·3	15·49
0	14	2	1·0	6·9	7·74
7	0	14	3·0	4·2	6·16
0	14	4	4·0	4·8	5·45
6	15	0	1·0	4·8	3·90
0	16	1	356·0	3·8	3·86
0	2	3	0·0	1·5	3·71
6	1	0	1·0	2·7	3·48
6	0	18	184·0	184·7	2·99
2	0	4	184·0	182·9	2·50
8	16	0	0·0	5·1	2·29
0	14	10	4·0	4·8	2·20
0	18	0	184·0	171·9	2·11
0	2	18	181·0	6·0*	1·44
0	8	16	181·0	20·8*	1·20
8	0	16	2·0	4·8	1·16
0	2	19	180·0	184·7	1·07

4. Concluding remarks

The limits of (3) and (4) may be described as: (*a*) they hold in $P\bar{1}$: in higher-symmetry space groups they hold only if restricted phases of type $(0, \pi)$ are involved; (*b*) for restricted phases of different type the values of x_1 and x_{01} have to be modified in accordance with space-group symmetry [see equation (10) in Giacovazzo (1987) for the generalized value of x_1].

Equations (3) and (4) are formally quite different from equivalent formulae working in non-centrosymmetrical space groups (Hauptman, 1982; Giacovazzo, 1983). The question arises whether: (*a*) information contained in centrosymmetric triplets is sufficiently large to be useful in practice for protein structure determination; (*b*) non-centrosymmetric formulae are sufficiently accurate to be used also for the estimation of centrosymmetric triplets.

In order to answer both questions 68 reflexions with restricted phase of type $(0, \pi)$ (from ferredoxin calculated data) have been estimated according to (4) by using 290 centrosymmetric triplets only. The same reflexions were also estimated according to non-centrosymmetric formulae. The outcome was practically identical for both types of formulae and is shown in Table 1. This table suggests that information contained in centrosymmetric triplets in dispersive structures is not negligible, and that non-centrosymmetric formulae can be used for estimating centrosymmetric as well as non-centrosymmetric triplets. Tests on real diffraction data have not been attempted; it is easy to foresee a reduced efficiency of the formulae according to the average error magnitude in the experimental data.

References

Adman, E. T., Sieker, L. C. & Jensen, L. M. (1973). *J. Biol. Chem.* **248**, 3987-3996.

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

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

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

Heinerman, J. J. L., Krabbendam, H., Kroon, J. & Spek, A. L. (1978). *Acta Cryst.* A**34**, 447-450.

Srinivasan, R. & Parthasarathy, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon Press.

Wilson, A. J. C. (1980). *Acta Cryst.* A**36**, 945-946.

Acta Cryst. (1988). A**44**, 294-300

New Probabilistic Formulas for Finding the Positions of Correctly Oriented Atomic Groups

By Carmelo Giacovazzo

Dipartimento Geomineralogico e Centro di Ricerca Interdipartimentale di Cristallografia, Università,
Campus Universitario, Via Salvemini, 70124 *Bari, Italy*

(*Received* 26 *June* 1987; *accepted* 8 *December* 1987)

Abstract

The probabilistic procedure described by Main [In *Crystallographic Computing Techniques* (1976), edited by F. R. Ahmed, pp. 97-105. Copenhagen: Munksgaard] has been reconsidered. In polar space groups some primitive random variables (atomic positions or shift vectors for molecular fragments) may be conveniently restricted to regions which are smaller than a unit cell. This introduces two new

0108-7673/88/030294-07$03.00

expressions for triplet relationships; furthermore, a new type of phase relationship is introduced which is not equivalent to invariant and seminvariant relationships.

Symbols and abbreviations

$\varphi_{\mathbf{h}}, \varphi_{\mathbf{k}}$: phase of the normalized structure factors $E_{\mathbf{h}}$, $E_{\mathbf{k}}$.

$R_{\mathbf{h}}, R_{\mathbf{k}}, \ldots$: moduli of $E_{\mathbf{h}}, E_{\mathbf{k}}, \ldots$.

N: number of atoms in the cell.

P: order of the space group.

m: number of molecular fragments (symmetry independent) with unknown position and fixed orientation.

n_i: number of atoms in the ith molecular fragment.

q: number of atoms (symmetry independent) whose positions are completely unknown.

$\mathbf{C}_s = (\mathbf{R}_s, \mathbf{T}_s)$: sth symmetry operator. $\mathbf{R}_s$ is its rotational part, $\mathbf{T}_s$ its translational part.

$\sum_q(\mathbf{h})$: $P \sum_{j=1}^{q} f_j^2(\mathbf{h})$ (scattering power of atoms with completely unknown position).

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

$\varepsilon_{\mathbf{h}}$: Wilson's factor responsible for the enhancement or depression of the intensity of certain subsets of reflexions due to particular symmetry elements.

1. Introduction

Orientation search methods in Patterson space (Braun, Hornstra & Leenhouts, 1969; Crowther, 1972; Nordman & Schilling, 1970) or trials by direct methods often provide the orientation of a molecular fragment with respect to the crystal axes. The position of the fragment with respect to the symmetry elements of the space group is usually achieved by: (*a*) translation functions in the vector space (Braun, Hornstra & Leenhouts, 1969; Huber, 1965; Nordman & Nakatsu, 1963) or in the reciprocal space (Tollin, 1966; Crowther & Blow, 1967; Karle, 1972; Langs, 1975; Harada, Lifchitz, Berthou & Jolles, 1981); or (*b*) special direct-methods procedures. Among these it is worth noting: (*b*1) a modified tangent formula (Karle, 1968) is used to recycle in $P1$ phases derived from the known fragment; (*b*2) reflection data are expanded in the space group $P1$ (Doesburg & Beurskens, 1983; Bruins Slot & Beurskens, 1984) and coefficients for Fourier synthesis are obtained by direct methods on difference structure factors. The position of the fragment relative to symmetry elements is deduced from the maximum of a suitable translation function; (*b*3) the correctly oriented but randomly positioned atomic groups are introduced as prior information in the probabilistic approach aimed at estimating triplet invariant phases (Main, 1976). The present paper is closely related to the last remarkable procedure: Main's approach will be reconsidered and new results will be derived.

The application of translation search techniques is rather simplified if restrictions (in accordance with space-group symmetry) are imposed on the admissible shifts. For example, shifts which leave invariant the algebraic form of the symmetry operators are superfluous: thus, thanks to the indeterminacy in choice of origin, the coordinates of the molecular centre of a fragment in $P2$ may be arbitrarily restricted to

$$0 \le x < \tfrac{1}{2}, \quad y = 0, \quad 0 \le z < \tfrac{1}{2}.$$

According to Hirshfeld (1968), for a convenient use of translation functions in vector space Cheshire groups and their unit cells have to be considered: in particular, Cheshire groups that derive from polar space groups have unit cells with one or more axes of vanishing length.

So far the above considerations have had no counterpart in the direct-methods techniques (*b*). Indeed, primitive random variables, according to their definitions, were always allowed to vary uniformly over the entire unit cell of the space group or over reciprocal space. We shall show here that imposing restrictions on the primitive random variables generates in polar space groups a new class of phase relationships which are not derivable from estimates of structure invariants or seminvariants. Such phase relationships will be studied by means of joint probability distribution methods.

2. Normalization of structure factors

(*a*) Let us divide the crystal structure into two parts: the first includes m molecular fragments with known orientation (and their symmetry equivalents), the second part comprises the atoms whose positions are completely unknown. Then the structure factor may be written down as

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

where

$$g_i(\mathbf{h}) = \sum_{j=1}^{n_i} f_j(\mathbf{h}) \sum_{s=1}^{P} \exp(2\pi i \mathbf{h} \mathbf{C}_s \mathbf{r}_j),$$

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

According to the stated postulates one may also write

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

where the $\mathbf{u}_j$'s are the trial atomic positions for the ith fragment and $\boldsymbol{\tau}_i$ is the shift to be applied to that fragment in order to translate atoms to the correct positions

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

A more useful form of $g_i(\mathbf{h})$ is

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

where

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

Factors $g_{is}(\mathbf{h})$ do not depend on $\boldsymbol{\tau}_i$ and may be calculated from prior information for any i and s. Thus the primitive random variables in our probabilistic approach are the m shifts $\boldsymbol{\tau}_i$ and q atomic positions $\mathbf{r}_j$. Without loss of generality, each primitive random variable will be considered statistically independent of the others.

From (1),

$$\begin{aligned}\langle|F_{\mathbf{h}}|^2\rangle &= \sum_{i,k=1}^{m}\langle g_i(\mathbf{h})g_k(-\mathbf{h})\rangle+\varepsilon_{\mathbf{h}}\textstyle\sum_q(\mathbf{h})\\ &=\sum_{i=1}^{m}\langle|g_i(\mathbf{h})|^2\rangle+\varepsilon_{\mathbf{h}}\textstyle\sum_q(\mathbf{h})\\ &=\sum_{i=1}^{m}\sum_{s_1,s_2=1}^{P}\langle g_{is_1}(\mathbf{h})g_{is_2}(-\mathbf{h})\\ &\quad\times\exp[2\pi i\mathbf{h}(\mathbf{R}_{s_1}-\mathbf{R}_{s_2})\boldsymbol{\tau}_i]\rangle+\varepsilon_{\mathbf{h}}\textstyle\sum_q(\mathbf{h}).\end{aligned} \tag{4}$$

The numerical value of $\varepsilon_{\mathbf{h}}$ differs from unity only when, for some $s_1\neq s_2$,

$$\mathbf{h}(\mathbf{R}_{s_1}-\mathbf{R}_{s_2})=0. \tag{5}$$

It is easy to show that

$$\langle|F_{\mathbf{h}}|^2\rangle=\varepsilon_{\mathbf{h}}\left\{\sum_{i=1}^{m}\sum_{s=1}^{P}|g_{is}(\mathbf{h})|^2+\textstyle\sum_q(\mathbf{h})\right\}. \tag{6}$$

The factor $\varepsilon_{\mathbf{h}}$ is not explicitly mentioned by Main (1976) in his equation (6).

(*b*) So far the m primitive random variables $\boldsymbol{\tau}_i$ as well as the q atomic positional variables $\mathbf{r}_i$ have been assumed to be uniformly distributed over the unit cell. Such an assumption may be considered unfavourable in space groups for which the allowed shifts of origin, consistent with the chosen algebraic form for the symmetry operators $\mathbf{C}_s$, are arbitrary displacements along any polar axes. Owing to the indeterminacy in the choice of origin, one of the shifts $\boldsymbol{\tau}_i$ may be restricted to a region that is smaller than the unit cell. As examples of different dimensionality we mention: (*a*) in *P*2 we are free to specify the origin along the diad axis by restricting $\boldsymbol{\tau}_1$ into the family of vectors of type $[x\,0\,z]$; (*b*) in *Pm* the origin may be fixed by restricting $\boldsymbol{\tau}_1$ to the class of vectors $[0\,y\,0]$; (*c*) in *P*1 $\boldsymbol{\tau}_1$ may be restricted to the vector $[0\,0\,0]$. From now on $\{\boldsymbol{\tau}_1\}$ will denote the family of restricted vectors $\boldsymbol{\tau}_1$ for a given space group. Once we have restricted $\boldsymbol{\tau}_1$, no other restriction can be assigned to the other $\boldsymbol{\tau}_i$'s, which may be assumed to be uniformly distributed over the unit cell.

In a polar space group let $\mathbf{h}$ satisfy

$$\mathbf{h}\boldsymbol{\tau}_1=0; \tag{7}$$

then, for every s, it will also be true that $\mathbf{h}\mathbf{R}_s\boldsymbol{\tau}_1=0$. Thus, according to (2),

$$g_1(\mathbf{h})=\sum_{s=1}^{P} g_{1s}(\mathbf{h})$$

is independent of $\boldsymbol{\tau}_1$, so that

$$\langle|F_{\mathbf{h}}|^2\rangle=|g_1(\mathbf{h})|^2+\varepsilon_{\mathbf{h}}\left\{\sum_{i=2}^{m}\sum_{s=1}^{P}|g_{is}(\mathbf{h})|^2+\textstyle\sum_q(\mathbf{h})\right\}. \tag{8}$$

In *P*2 and *Pm* the reflexions for which (7) is satisfied are $0k0$ and $h0l$ respectively; in *P*1 every reflexion satisfies (7).

The use of (8) in polar space groups rather than (6) is immediately understood in *P*1, where a molecular fragment with known orientation may be considered as a set of correctly positioned atoms. Thus $|g_1(\mathbf{h})|^2$ in (8) plays the same role as $|F_{p,\mathbf{h}}|^2$ in the normalization process when a molecular fragment is correctly located [see equation (2) of Giacovazzo (1983)]. Similar considerations can be applied in polar space groups with symmetry higher than *P*1, but only for subsets of reflexions satisfying (7).

(*c*) A normalized structure amplitude $E_{\mathbf{h}}$ is defined by

$$|E_{\mathbf{h}}|^2=AI_{\mathbf{h}}/\langle|F_{\mathbf{h}}|^2\rangle, \tag{9}$$

where $I_{\mathbf{h}}$ is the observed intensity reduced to an absolute scale by factor A, and $\langle|F_{\mathbf{h}}|^2\rangle$ is given by (6) or (8) according to the circumstances.

3. The conditional distribution $P(\varphi_{\mathbf{h}}|R_{\mathbf{h}})$ in polar space groups

In any polar space group a family of reflexions can always be found whose phases can be estimated from diffraction magnitudes only. For example: (*a*) In *P*1 one atom may be arbitrarily located on the origin: then all the phases may be estimated *via*, for instance, Sim's (1959) distribution. Their expected values will be 2π for any $\mathbf{h}$. (*b*) In *P*2 phases φ_{0k0} do not depend on (x,z) atomic coordinates: their values may be estimated by arbitrarily fixing one y atomic coordinate. (*c*) In *Pm* the phases φ_{h0l} do not depend on y coordinates, and may be estimated by arbitrarily fixing one pair of atomic coordinates (x,z).

In the absence of any prior structural information the above considerations are of limited use in practice, unless some few heavy atoms are in the structure. By contrast they are of great interest if the orientation of at least one molecular fragments is known. The probabilistic background is the following: for $\mathbf{h}$ satisfying (7) the conditional probability distribution of $\varphi_{\mathbf{h}}$ may be calculated on the assumption that $\boldsymbol{\tau}_1$ is a restricted random vector (according to §2) while $\boldsymbol{\tau}_i$, for $i=2,\ldots,m$, and the q positional vectors $\mathbf{r}_j$ are the primitive random variables, randomly distributed over the unit cell. Then

$$P(\varphi_{\mathbf{h}}|R_{\mathbf{h}})\simeq[2\pi I_0(G_{\mathbf{h}})]^{-1}\exp[G_{\mathbf{h}}\cos(\varphi_{\mathbf{h}}-\varphi_1)], \quad (10)$$

where $\varphi_{\mathbf{h}}$ and φ_1 are the phases of $F_{\mathbf{h}}$ and $g_1(\mathbf{h})$ respectively,

$$G_{\mathbf{h}}=2R_{\mathbf{h}}R_{1\mathbf{h}}/b_{\mathbf{h}},$$

$$R_{1\mathbf{h}}=|g_1(\mathbf{h})|/\langle|F_{\mathbf{h}}|^2\rangle^{1/2},$$

$$b_{\mathbf{h}}=\varepsilon_{\mathbf{h}}\left[\sum_{i=2}^{m}\sum_{s=1}^{P}|g_{is}(\mathbf{h})|^2+\textstyle\sum_q(\mathbf{h})\right]/\langle|F_{\mathbf{h}}|^2\rangle,$$

and $\langle|F_{\mathbf{h}}|^2\rangle$ is given by (8).

If $m=1$ and $q\to 0$ then $G_{\mathbf{h}}\to\infty$ as expected. If $m=1$ (only one correctly oriented molecular fragment) (10) reduces to the Sim (1959) distribution in the algebraic form proposed by Giacovazzo (1983). However, while Sim's distribution holds whatever $\mathbf{h}$ may be (provided some atoms are correctly located), our distribution (10) may only be applied in polar space groups and for $\mathbf{h}$ satisfying (7). Otherwise $P(\varphi_{\mathbf{h}})\simeq 1/(2\pi)$. Even if formally interesting, distribution (10) will be of limited use in practice unless phase information cannot be extended to other reflexions by means of phase relationships holding under the same hypotheses stated for (10). Such new phase relationships are described below.

4. The joint probability distribution of pairs of reflexions in polar space groups

In a polar space group let $\mathbf{h}_1$ and $\mathbf{h}_2$ be a pair of vectors both orthogonal to $\{\boldsymbol{\tau}_1\}$; then

$$\langle F_{\mathbf{h}_1}F_{\mathbf{h}_2}\rangle=g_1(\mathbf{h}_1)g_1(\mathbf{h}_2)$$

and also

$$\langle E_{\mathbf{h}_1}E_{\mathbf{h}_2}\rangle=g_1(\mathbf{h}_1)g_1(\mathbf{h}_2)/[\langle|F_{\mathbf{h}_1}|^2\rangle\langle|F_{\mathbf{h}_2}|^2\rangle]^{1/2}.$$

Thus an estimate for $\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2}$ is certainly derivable from the conditional distribution $P(\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2}|R_{\mathbf{h}_1}, R_{\mathbf{h}_2})$; however it does not significantly add to the phase information obtainable from the separate application of (10) to $E_{\mathbf{h}_1}$ and $E_{\mathbf{h}_2}$ respectively.

Assume now, for the same polar space group, that $\mathbf{h}_1$ and $\mathbf{h}_2$ are both vectors which are not orthogonal to $\{\boldsymbol{\tau}_1\}$ but which satisfy the condition

$$(\mathbf{h}_1+\mathbf{h}_2)\boldsymbol{\tau}_1=0. \quad (11)$$

Then the mean value of $F_{\mathbf{h}_1}F_{\mathbf{h}_2}$ is given by

$$\begin{aligned}\langle F_{\mathbf{h}_1}F_{\mathbf{h}_2}\rangle&=\langle g_1(\mathbf{h}_1)g_1(\mathbf{h}_2)\rangle\\&=\sum_{s_1,s_2=1}^{P}\langle g_{1s_1}(\mathbf{h}_1)g_{1s_2}(\mathbf{h}_2)\\&\quad\times\exp[2\pi i(\mathbf{h}_1\mathbf{R}_{s_1}+\mathbf{h}_2\mathbf{R}_{s_2})\boldsymbol{\tau}_1]\rangle. \quad (12)\end{aligned}$$

If condition (11) is verified, then (13) holds too for any s:

$$(\mathbf{h}_1+\mathbf{h}_2)\mathbf{R}_s\boldsymbol{\tau}_1=0, \quad (13)$$

and (12) becomes

$$\langle F_{\mathbf{h}_1}F_{\mathbf{h}_2}\rangle=\sum_{s=1}^{P}g_{1s}(\mathbf{h}_1)g_{1s}(\mathbf{h}_2). \quad (14)$$

Equation (14) shows us that $\theta_2=(\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2})$ can be estimated in polar space groups provided (11) is verified. Examples of such pairs are: (*a*) in $P1$ every pair $(\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2})$; (*b*) in $P2$ every pair $(\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2})$ for which $\mathbf{h}_1+\mathbf{h}_2=0k0$; (*c*) in Pm every pair $(\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2})$ for which $\mathbf{h}_1+\mathbf{h}_2=h0l$.

The joint probability distribution $P(\varphi_{\mathbf{h}_1}, \varphi_{\mathbf{h}_2}, R_{\mathbf{h}_1}, R_{\mathbf{h}_2})$ is derived in the Appendix [see (*A*5)], from which the conditional distribution of θ_2 may be obtained:

$$P(\theta_2|R_{\mathbf{h}_1}R_{\mathbf{h}_2})\simeq[2\pi I_0(\beta_2)]^{-1}\exp[\beta_2\cos(\theta_2-\delta_2)], \quad (15)$$

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

$$\beta_2=2R_{\mathbf{h}_1}R_{\mathbf{h}_2}d_2/(1-d_2^2),$$

$$d_2\exp(i\delta_2)=\frac{\sum_{s=1}^{P}[g_{1s}(\mathbf{h}_1)g_{1s}(\mathbf{h}_2)]}{[\langle|F_{\mathbf{h}_1}|^2\rangle\langle|F_{\mathbf{h}_2}|^2\rangle]^{1/2}},$$

and δ_2 is the phase of d_2.

Equation (15) allows phase extension to all measured reflexions. It may be observed that: (*a*) atoms with unknown positions do not give any information on the expected value of θ_2, but they help to define the variance of the phase relationship through the mean values $\langle|F_{\mathbf{h}}|^2\rangle$; (*b*) unreliable estimates are expected for too large ratios $\sum_q/\sum_N$; (*c*) while triplet-invariant phases are expected to lie near 2π (Cochran, 1955), δ_2 values uniformly span (in a statistical sense) the interval $(0, 2\pi)$. Their actual values, for a given pair $(\mathbf{h}_1, \mathbf{h}_2)$, depend on the orientation of the first molecular fragment.

5. The joint probability distribution of *n*-tuples of reflexions in polar space groups

In a polar space group let $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n$ be an n-tuple of vectors none of which is orthogonal to $\{\boldsymbol{\tau}_1\}$; they satisfy the condition

$$(\mathbf{h}_1+\mathbf{h}_2+\ldots+\mathbf{h}_n)\boldsymbol{\tau}_1=0.$$

The average value

$$\langle F_{\mathbf{h}_1}F_{\mathbf{h}_2}\dots F_{\mathbf{h}_n}\rangle = \sum_{s=1}^{m} g_{1s}(\mathbf{h}_1)g_{1s}(\mathbf{h}_2)\dots g_{1s}(\mathbf{h}_n) \qquad (16)$$

is then a non-vanishing quantity. Thus the conditional distribution of $\theta_n = \varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} + \dots + \varphi_{\mathbf{h}_n}$ given $R_{\mathbf{h}_1}, R_{\mathbf{h}_2}, \dots, R_{\mathbf{h}_n}$, may be calculated, in accordance with § 4, as

$$P(\theta_n \mid F_{\mathbf{h}_1}, R_{\mathbf{h}_2}, \dots, R_{\mathbf{h}_n}) \simeq [2\pi I_0(\beta_n)]^{-1}\exp[\beta_n\cos(\theta_n - \delta_n)] \qquad (17)$$

where

$$\beta_n = 2R_{\mathbf{h}_1}R_{\mathbf{h}_2}\dots R_{\mathbf{h}_n}d_n/(1-d_n^2),$$

$$d_n\exp(i\delta_n) = \frac{\sum\limits_{s=1}^{P} g_{1s}(\mathbf{h}_1)g_{1s}(\mathbf{h}_2)\dots g_{1s}(\mathbf{h}_n)}{[\langle|F_{\mathbf{h}_1}|^2\rangle\langle|F_{\mathbf{h}_2}|^2\rangle\dots\langle|F_{\mathbf{h}_n}|^2\rangle]^{1/2}}.$$

It may be observed that (a) (17) [as well as (15)] is a phase relationship of order $1/N^0$; its reliability does not depend on the structural complexity, but approximately on the ratio n_1/N; (b) the values δ_n uniformly span (in a statistical sense) the interval $(0, 2\pi)$; (c) β_n is a decreasing function of n (in an average sense). Thus relationships such as $\theta_n \simeq \delta_n$ are expected to be more reliable for small than for large n; (d) no phase information arises from molecular fragments with $i > 1$. Thus (17) is expected to be more effective when $m = 1$.

6. Conditional probability distribution of triplet invariants in non-polar space groups

Let $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$ satisfy the condition

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

In non-polar space groups the average value $\langle F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}\rangle$ may be calculated according to

$$\begin{aligned}\langle F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}\rangle &= \sum_{i=1}^{m}\langle g_i(\mathbf{h}_1)g_i(\mathbf{h}_2)g_i(\mathbf{h}_3)\rangle + \langle F_{q\mathbf{h}_1}F_{q\mathbf{h}_2}F_{q\mathbf{h}_3}\rangle\\ &= \sum_{i=1}^{m}\sum_{s_1,s_2,s_3=1}^{P}\langle g_{is_1}(\mathbf{h}_1)g_{is_2}(\mathbf{h}_2)g_{is_3}(\mathbf{h}_3)\\ &\quad\times\exp[2\pi i(\mathbf{h}_1\mathbf{R}_{s_1} + \mathbf{h}_2\mathbf{R}_{s_2} + \mathbf{h}_3\mathbf{R}_{s_3})\boldsymbol{\tau}_i]\rangle\\ &\quad+\sum_{j=1}^{q} f_j(\mathbf{h}_1)f_j(\mathbf{h}_2)f_j(\mathbf{h}_3)\\ &\quad\times\left\langle\sum_{s_1,s_2,s_3=1}^{P}\exp[2\pi i(\mathbf{h}_1\mathbf{C}_{s_1} + \mathbf{h}_2\mathbf{C}_{s_2}\right.\\ &\quad\left.+\mathbf{h}_3\mathbf{C}_{s_3})\mathbf{r}_j]\right\rangle.\end{aligned}$$

It is not difficult to show that

$$\langle F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}\rangle = w(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)\left\{\sum_{i=1}^{m}\sum_{s=1}^{P} g_{is}(\mathbf{h}_1)g_{is}(\mathbf{h}_2)g_{is}(\mathbf{h}_3) + \textstyle\sum_{3q}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)\right\} \qquad (18)$$

where $w(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$ is a weight factor which depends on $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$ and on the space group (Giacovazzo, 1974). The conditional probability distribution of $\Phi_3 = \varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} + \varphi_{\mathbf{h}_3}$ is given by

$$P(\Phi_3) \simeq [2\pi I_0(\alpha_3)]^{-1}\exp[\alpha_3\cos(\Phi_3 - \psi_3)], \qquad (19)$$

where

$$\begin{aligned}\alpha_3\exp(i\psi_3) &= 2R_{\mathbf{h}_1}R_{\mathbf{h}_2}R_{\mathbf{h}_3}w(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)\\ &\quad\times\left[\sum_{i=1}^{m}\sum_{s=1}^{P} g_{is}(\mathbf{h}_1)g_{is}(\mathbf{h}_2)g_{is}(\mathbf{h}_3) + \textstyle\sum_{3q}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)\right]\\ &\quad\times[\langle|F_{\mathbf{h}_1}|^2\rangle\langle|F_{\mathbf{h}_2}|^2\rangle\langle|F_{\mathbf{h}_3}|^2\rangle]^{-1/2};\end{aligned}$$

$\langle|F_{\mathbf{h}_i}|^2\rangle$ is defined according to (6). Equation (19) reduces to Main's result when $w(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = 1$. Equation (19) may be compared with (17) by observing that: (a) (19) is a phase relationship of order $N^{-1/2}$: thus its reliability depends on the structural complexity; (b) (19) is unreliable for large ratios $\sum_q/\sum_N$; (c) all p fragments with known orientation contribute to estimated Φ_3; (d) a non-vanishing real contribution $[\sum_{3q}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)]$ arises from atoms with unknown position. Thus q_3 values are expected to be more crowded around 2π than around π.

7. The joint probability distribution of triplet invariants in polar space groups

Let $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$ satisfy the condition $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0$. In polar space groups

$$\begin{aligned}\langle F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}\rangle &= g_1(\mathbf{h}_1)g_1(\mathbf{h}_2)g_1(\mathbf{h}_3)\\ &\quad+\sum_{i=2}^{m}\langle g_i(\mathbf{h}_1)g_i(\mathbf{h}_2)g_i(\mathbf{h}_3)\rangle + \langle F_{q\mathbf{h}_1}F_{q\mathbf{h}_2}F_{q\mathbf{h}_3}\rangle\\ &= \sum_{s_1,s_2,s_3=1}^{P} g_{1s_1}(\mathbf{h}_1)g_{1s_2}(\mathbf{h}_2)g_{1s_3}(\mathbf{h}_3)\\ &\quad\times\langle\exp[2\boldsymbol{\tau} i(\mathbf{h}_1\mathbf{R}_{s_1} + \mathbf{h}_2\mathbf{R}_{s_2} + \mathbf{h}_3\mathbf{R}_{s_3})\boldsymbol{\tau}_1]\rangle\\ &\quad+w(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)\left\{\sum_{i=2}^{m}\sum_{s=1}^{P} g_{is}(\mathbf{h}_1)g_{is}(\mathbf{h}_2)g_{is}(\mathbf{h}_3)\right.\\ &\quad\left.+\textstyle\sum_{3q}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)\right\}.\end{aligned} \qquad (20)$$

It is not difficult to show that (19) holds in polar space groups too without significant modifications.

If $\mathbf{h}_1$ and $\mathbf{h}_2$ and $\mathbf{h}_3$ are normal to $\{\boldsymbol{\tau}_1\}$ then

$$\begin{aligned}\langle F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}\rangle &= g_1(\mathbf{h}_1)g_1(\mathbf{h}_2)g_1(\mathbf{h}_3)\\ &\quad + w(\mathbf{h}_1,\mathbf{h}_2,\mathbf{h}_3)\left\{\sum_{i=2}^{m}\sum_{s=1}^{P} g_{is}(\mathbf{h}_1)g_{is}(\mathbf{h}_2)g_{is}(\mathbf{h}_3)\right.\\ &\quad \left. + \textstyle\sum_{3q}(\mathbf{h}_1,\mathbf{h}_2,\mathbf{h}_3)\right\}.\end{aligned}$$

The conditional distribution of Φ_3 given $R_{\mathbf{h}_1}$, $R_{\mathbf{h}_2}$, $R_{\mathbf{h}_3}$ may again be approximated by (19) provided that

$$\begin{aligned}\alpha_3\exp(i\psi_3) &= 2R_{\mathbf{h}_1}R_{\mathbf{h}_2}R_{\mathbf{h}_3}[\langle|F_{\mathbf{h}_1}|^2\rangle\langle|F_{\mathbf{h}_2}|^2\rangle\langle|F_{\mathbf{h}_3}|^2\rangle]^{-1/2}\\ &\quad\times\left\{g_1(\mathbf{h}_1)g_1(\mathbf{h}_2)g_1(\mathbf{h}_3) + w(\mathbf{h}_1,\mathbf{h}_2,\mathbf{h}_3)\right.\\ &\quad\times\left[\sum_{i=2}^{m}\sum_{s=1}^{P} g_{is}(\mathbf{h}_1)g_{is}(\mathbf{h}_2)g_{is}(\mathbf{h}_3)\right.\\ &\quad\left.\left. + \textstyle\sum_{3q}(\mathbf{h}_1,\mathbf{h}_2,\mathbf{h}_3)\right]\right\}.\end{aligned}$$

$\langle|F_{\mathbf{h}_i}|^2\rangle$ is defined according to (8).

It is easily seen that the estimation of Φ_3, for reflexions $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$ all orthogonal to $\{\boldsymbol{\tau}_1\}$, is not equivalent to that available after separate application of (10) to reflexions $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$. These highly informative triplets can play an important role in practical procedures if their number (depending on the point-group symmetry) is sufficiently large.

APPENDIX

Let $\mathbf{h}_1$ and $\mathbf{h}_2$ be two symmetry-independent vectors which are not orthogonal to the family of permitted $\boldsymbol{\tau}_1$ vectors and satisfy the condition $(\mathbf{h}_1+\mathbf{h}_2)\boldsymbol{\tau}_1=0$. Denote by $A_n=A_{\mathbf{h}_n}$, $B_n=B_{\mathbf{h}_n}$ the real and imaginary components of $E_{\mathbf{h}_n}$, and by $[\alpha_i(\mathbf{h}),\ \beta_{i\mathbf{h}}]$ and $[\alpha_{is}(\mathbf{h}),\ \beta_{is}(\mathbf{h})]$ the real and imaginary components of $g_i(\mathbf{h})$ and $g_{is}(\mathbf{h})$ respectively.

The characteristic function of the distribution $P(A_1,\ A_2,\ B_1,\ B_2)$ is given by

$$C(u_1,u_2,v_1,v_2)=\langle\exp i(u_1A_1+u_2A_2+v_1B_1+v_2B_2)\rangle, \tag{A1}$$

where u_i and v_i are carrying variables associated with A_i and B_i respectively. The first moments of $(A1)$ are

$$\langle A_i\rangle=\langle B_i\rangle=0 \quad \text{for } i=1,2,$$

$$\langle A_i^2\rangle=\langle B_i^2\rangle=\tfrac{1}{2} \quad \text{for } i=1,2,$$

$$\begin{aligned}\langle A_1A_2\rangle &= \sum_{i=1}^{m}\langle\alpha_i(\mathbf{h}_1)\alpha_i(\mathbf{h}_2)\rangle\\ &= \langle\alpha_1(\mathbf{h}_1)\alpha_1(\mathbf{h}_2)\rangle\\ &= [\langle|F_{\mathbf{h}_1}|^2\rangle\langle|F_{\mathbf{h}_2}|^2\rangle]^{-1/2}\sum_{j_1,j_2=1}^{n_1} f_{j_1}(\mathbf{h}_1)f_{j_2}(\mathbf{h}_2)\\ &\quad\times\sum_{s_1,s_2=1}^{P}\langle\cos[2\pi\mathbf{h}_1\mathbf{C}_{s_1}(\mathbf{u}_{j_1}+\boldsymbol{\tau}_1)]\\ &\quad\times\cos[2\pi\mathbf{h}_2\mathbf{C}_{s_2}(\mathbf{u}_{j_2}+\boldsymbol{\tau}_1)]\rangle,\end{aligned}$$

where $\langle|F_{\mathbf{h}_i}|^2\rangle$ is given by (6). Non-vanishing contributions to the average are obtained only when $\mathbf{C}_{s_1}=\mathbf{C}_{s_2}=\mathbf{C}_s$: thus

$$\begin{aligned}\langle A_1A_2\rangle &= \tfrac{1}{2}[\langle|F_{\mathbf{h}_1}|^2\rangle\langle|F_{\mathbf{h}_2}|^2\rangle]^{-1/2}\\ &\quad\times\sum_{j_1,j_2=1}^{n_1} f_{j_1}(\mathbf{h}_1)f_{j_2}(\mathbf{h}_2)\\ &\quad\times\sum_{s=1}^{P}\cos[2\pi(\mathbf{h}_1\mathbf{C}_s\mathbf{u}_{j_1}+\mathbf{h}_2\mathbf{C}_s\mathbf{u}_{j_2})]\\ &= \tfrac{1}{2}[\langle|F_{\mathbf{h}_1}|^2\rangle\langle|F_{\mathbf{h}_2}|^2\rangle]^{-1/2}\\ &\quad\times\mathscr{R}\left[\sum_{s=1}^{P} g_{1s}(\mathbf{h}_1)g_{1s}(\mathbf{h}_2)\right]=\tfrac{1}{2}d_1\end{aligned} \tag{A2}$$

where $\mathscr{R}$ stands for 'real part of'.

In a similar way the following relations are obtained:

$$\langle A_iB_i\rangle=0 \quad \text{for } i=1,2,$$

$$\langle B_1B_2\rangle=-\langle A_1A_2\rangle,$$

$$\begin{aligned}\langle A_1B_2\rangle &= \langle A_2B_1\rangle\\ &= \tfrac{1}{2}[\langle|F_{\mathbf{h}_1}|^2\rangle\langle|F_{\mathbf{h}_2}|^2\rangle]^{-1/2}\mathscr{I}\left[\sum_{s=1}^{m} g_{1s}(\mathbf{h}_1)g_{1s}(\mathbf{h}_2)\right]\\ &= \tfrac{1}{2}d_2,\end{aligned} \tag{A3}$$

where $\mathscr{I}$ stands for 'imaginary part of'.

The Fourier transform of $C(u_1,u_2,v_1,v_2)$ gives

$$\begin{aligned}P(A_1,A_2,B_1,B_2) &\simeq [1-(d_1^2+d_2^2)]^{-1/2}\pi^{-2}\\ &\quad\times\exp\{-[1-(d_1^2+d_2^2)]^{-1}[A_1^2+B_1^2+A_2^2+B_2^2\\ &\quad-2d_1(A_1A_2-B_1B_2)-2d_2(A_1B_2+A_2B_1)]\}.\end{aligned} \tag{A4}$$

Non-negativity of the term $1-(d_1^2+d_2^2)$ is always (also in the most critical case in which $p=1$ and $q=0$) ensured by the Schwarz inequality

$$\left|\sum_{s=1}^{P} g_{1s}(\mathbf{h}_1)g_{1s}(\mathbf{h}_2)\right|^2 \le \sum_{s=1}^{P}|g_{1s}(\mathbf{h}_1)|^2\sum_{s=1}^{P}|g_{1s}(\mathbf{h}_2)|^2.$$

A change of notation according to

$$A_i=R_i\cos\varphi_i, \qquad B_i=R_i\sin\varphi_i,$$

$$d_1=d\cos\delta, \qquad d_2=d\sin\delta$$

gives

$$P(R_1, R_2, \varphi_1, \varphi_2) \simeq \{R_1R_2/[\pi^2(1-d^2)^{1/2}]\} \times \exp\{-(1-d^2)^{-1}[R_1^2+R_2^2 - 2dR_1R_2\cos(\varphi_1+\varphi_2-\delta)]\}. \quad (A5)$$

References

BRAUN, P. B., HORNSTRA, J. & LEENHOUTS, J. I. (1969). *Philips Res. Rep.* **24**, 85-118.

BRUINS SLOT, H. J. & BEURSKENS, P. T. (1984). *Acta Cryst.* **A40**, 701-703.

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 174-177. New York: Gordon and Breach.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544-548.

DOESBURG, H. M. & BEURSKENS, P. T. (1983). *Acta Cryst.* **A39**, 368-376.

HARADA, Y., LIFCHITZ, A., BERTHOU, J. & JOLLES, P. (1981). *Acta Cryst.* **A37**, 398-406.

HIRSHFELD, F. L. (1968). *Acta Cryst.* **A24**, 301-311.

HUBER, R. (1965). *Acta Cryst.* **19**, 353-356.

GIACOVAZZO, C. (1974). *Acta Cryst.* **A30**, 631-634.

GIACOVAZZO, C. (1983). *Acta Cryst.* **A39**, 685-692.

KARLE, J. (1968). *Acta Cryst.* **B24**, 182-186.

KARLE, J. (1972). *Acta Cryst.* **B28**, 820-824.

LANGS, D. A. (1975). *Acta Cryst.* **A31**, 543-550.

MAIN, P. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 97-105. Copenhagen: Munksgaard.

NORDMAN, C. E. & NAKATSU, K. (1963). *J. Am. Chem. Soc.* **85**, 353-354.

NORDMAN, C. E. & SCHILLING, J. W. (1970). *Crystallographic Computing*, edited by F. R. AHMED, S. R. HALL & C. P. HUBER, pp. 110-114. Copenhagen: Munksgaard.

SIM, G. A. (1959). *Acta Cryst.* **12**, 813-815.

TOLLIN, P. (1966). *Acta Cryst.* **21**, 613-614.

Acta Cryst. (1988). **A44**, 300-308

Application of the Molecular Replacement Method to Multidomain Proteins. 2. Comparison of Various Methods for Positioning an Oriented Fragment in the Unit Cell

BY M. CYGLER* AND W. F. ANDERSON

MRC Group on Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

(*Received 7 May 1987; accepted 7 December 1987*)

Abstract

The capabilities of several different methods to determine the correct translation of a model for the application of the molecular replacement method of structure determination to multidomain proteins have been analyzed. The structure of the Fab fragment of the autoimmune anti-poly(dT)-specific antibody HED10 was determined using molecular replacement and provides an example for comparing different methods of determining the correct translation of the model and for evaluating the importance of the parameters used. Expansion to space group $P1$ and phasing with a correctly oriented randomly positioned model was found to be superior to either the Crowther-Blow translation function [Crowther & Blow (1967). *Acta Cryst.* **23**, 544-548] or a brute-force search when only a small part of the molecule was used as a model.

Introduction

In a previous paper (Cygler & Anderson, 1988) we have discussed the application of the rotation function (RF) to determine the orientation of fragments of a multidomain protein in a crystal. As an example of such a protein we have used the anti-poly(dT) immunoglobulin Fab fragment HED10 (Cygler, Boodhoo, Lee & Anderson, 1987). The Fab fragment is composed of two relatively rigid domains, the variable (V) and constant (C) domains, connected by two short polypeptide links that provide flexibility of the domain arrangement in the intact molecule (Amzel & Poljak, 1979; Table 1 of Cygler & Anderson, 1988). In the case of the HED10 Fab fragment the RF solutions have been obtained for V and C domains separately. Here we would like to present the various methods that we have applied to determine the position of these fragments in the unit cell, compare the effectiveness of these approaches and investigate what is the minimum size of the fragment that can be successfully used for such a purpose.

---

[*] Present address: Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec H4P 2R2, Canada.

0108-7673/88/030300-09$03.00