where $f_j(\rho)$ is defined in (6), to an arbitrary precision. Real data, however, are subject to a statistical uncertainty, and it is pointless to demand agreement that is more precise than this uncertainty. Previously described procedures (*e.g.* Livesey & Skilling, 1985) for finding a maximum entropy distribution have defined the set of feasible points by a constraint of the form $\sum\{[|F_{obs}(\mathbf{h}_j)| - |f_j(\rho)|]/\sigma_j\}^2 = m$, where σ_j is the standard deviation of $|F_{obs}(\mathbf{h}_j)|$ and the sum is over m reflections. This constraint contains nothing that requires the quantities within the braces to have a distribution that in any way resembles a normal distribution, and, further, has resulted in problems in computation. By contrast, the lack of precision in the data is relevant to the procedure described here only in the definition of the stopping rule. Refinement can continue until $[|F_{obs}(\mathbf{h}_j)| - |f_j(\rho)|]^2/\sigma_j^2 < 1$ for all j.

It can be shown (Luenberger, 1984) that any set of starting phases will lead to a unique maximum entropy map. As Bricogne (1984) has shown, however, different sets of starting phases may lead to different maps, each of which is equally consistent with the data. In these circumstances the total entropy of the map provides a criterion for choosing among different possible sets of starting phases.

The entropy maximization procedure described here is a means of finding an everywhere-positive electron distribution for which the amplitude of the structure factor, $|\langle\exp(2\pi i\mathbf{h}.\mathbf{r})\rangle|$, is equal to $|F_{obs}(\mathbf{h})|$ for an arbitrarily large set of reflections. It does not, however, make any use of the fact that a crystal is composed of atoms that have definite well known physical and chemical properties. It is thus a tool for obtaining a map into which an atomic model may be fitted, with subsequent refinement by least-squares methods. If non-negativity of electron density is a sufficient condition to determine a unique atomic structure by direct methods (Woolfson, 1987, and references therein), it follows logically that entropy maximization, which is a stronger condition, will find

the same atomic structure. A characteristic of the maximum entropy distribution is that it maximizes the minimum density in the unit cell, thereby minimizing the probability that some reflection in the unmeasurable region of reciprocal space will have an amplitude for which there is no phase that will not cause the density at some point to be negative. Maximum entropy is not a necessary condition for an acceptable structure, so structures whose entropies are far from the global maximum cannot be ruled out. Nevertheless, it is at least a plausible conjecture that, in a centrosymmetric structure, the sign combination for the strongest reflections that has the highest entropy is likely to be the correct one, and that, in a non-centrosymmetric structure, the correct set of phases will give a distribution that has an entropy close to the maximum. Maximum entropy *is* an efficient way to express the mutual phase implications of a large set of amplitudes simultaneously.

References

BRICOGNE, G. (1984). *Acta Cryst.* A40, 410–445.
COLLINS, D. M. (1982). *Nature (London)*, 298, 49–51.
GILL, P. E., MURRAY, W. & WRIGHT, M. M. (1981). *Practical Optimization.* New York: Academic Press.
JAYNES, E. T. (1979). *The Maximum Entropy Formalism*, edited by R. D. LEVINE & M. TRIBUS, pp. 15–118. Cambridge, MA: Massachusetts Institute of Technology.
LIVESEY, A. K. & SKILLING, J. (1985). *Acta Cryst.* A41, 113–122.
LUENBERGER, D. G. (1984). *Linear and Nonlinear Programming.* Reading, MA: Addison Wesley.
PRINCE, E. (1982). *Mathematical Techniques in Crystallography and Materials Science.* New York: Springer-Verlag.
PRINCE, E., SJÖLIN, L. & ALENLJUNG, R. (1988). *Acta Cryst.* A44, 216–222.
SHANNON, C. E. (1948). *Bell Syst. Tech. J.* pp. 379–423, 623–656.
SHORE, J. E. & JOHNSON, R. W. (1980). *IEEE Trans. Inf. Theory*, IT-26, 26–37; IT-29, 942–943.
STEWART, G. W. (1973). *Introduction to Matrix Computations.* New York: Academic Press.
WILKINS, S. W., VARGHESE, J. N. & LEHMANN, M. S. (1983). *Acta Cryst.* A39, 47–60.
WOOLFSON, M. M. (1987). *Acta Cryst.* A43, 593–612.

The Estimation of Triplet Invariants from Multi-Wavelength Data

By E. A. KLOP, H. KRABBENDAM AND J. KROON

Laboratorium voor Kristal- en Structuurchemie, Rijksuniversiteit, Padualaan 8, 3584 CH Utrecht, The Netherlands

Abstract

A two-step procedure is presented for the estimation of triplet invariants from multi-wavelength data. In the first step wavelength-independent structure-factor magnitudes of both the total structure and the light-atom substructure, together with associated structure-factor phase differences, are calculated explicitly *via* a modified Singh & Ramaseshan [*Acta Cryst.* (1968), B24, 35–40] procedure. In the second step these

quantities are employed as conditional information in the probability distribution of a triplet invariant which can be derived from Hauptman's [*Acta Cryst.* (1982), A38, 289-294] joint probability distribution for an isomorphous pair of structures. Test calculations on the structure of the iron-containing protein ferredoxin show the feasibility of the procedure.

Introduction

Karle (1984) proposed to estimate triplet phase invariants *via* an algebraic analysis of multi-wavelength data. By use of a set of simultaneous equations he showed that it is possible to calculate wavelength-independent structure-factor magnitudes of the substructure formed by the anomalous scatterers and of the substructure formed by the non-anomalously scattering atoms, as well as phase differences associated with the phases of these structure factors. Triplet invariants are subsequently estimated under the assumption that the triplet invariants of the anomalous scatterers are close to zero.

In the present paper we will estimate triplet invariants of the total structure *via* a probability distribution in which wavelength-independent structure-factor magnitudes and phase differences calculated from multi-wavelength data are used as conditional information. We will use a modified version of the Singh & Ramaseshan (1968) method to obtain wavelength-independent quantities explicitly. The procedure has been tested using artificial two-wavelength diffraction data of the protein ferredoxin.

Definitions

F^+_{ij} Structure factor for reflection h_i at wavelength λ_j, not necessarily on absolute scale. [$F^+_{ij} \equiv F(h_i, \lambda_j)$.]

F^-_{ij} Complex conjugate of the structure factor for reflection $-h_i$ at wavelength λ_j, on the same scale as F^+_{ij}. [$F^-_{ij} \equiv F^*(-h_i, \lambda_j)$.]

F^N_i Contribution to F^+_{ij} based on the normal parts of the scattering factors of all atoms (including anomalously scattering atoms) on absolute scale.

F^L_i Contribution to F^+_{ij} of the light atoms (*i.e.* the non-anomalously scattering atoms) on absolute scale.

A subscript i denotes the reciprocal vector h_i. Other quantities are defined in the text.

The procedure

In a multi-wavelength experiment magnitudes $|F^+_{ij}|$ and $|F^-_{ij}|$ ($j = 1, \ldots, U$) can be obtained from data collected on a crystal containing anomalous scatterers. Here U is the number of wavelengths. These magnitudes may be used to calculate wavelength-independent quantities $|F^N_i|$, $|F^L_i|$ and ν_i explicitly as will be shown below. The doublet invariant ν_i is defined as $\varphi^N_i - \varphi^L_i$ where φ^N_i and φ^L_i are the phases of F^N_i and F^L_i respectively.

Subsequent normalization of $|F^N_i|$ and $|F^L_i|$ yields $|E^N_i|$ and $|E^L_i|$, written as S_i and T_i respectively. Using S_i, T_i and ν_i ($i = 1, 2, 3$), subject to $h_1 + h_2 + h_3 = 0$, as conditional information, the conditional probability distribution $P[\Phi^N | S_i, T_i, \nu_i (i = 1, 2, 3)]$ of the triplet invariant $\Phi^N = \varphi^N_1 + \varphi^N_2 + \varphi^N_3$ may be employed for the estimation of triplet invariants. This distribution can be expressed as

$$P[\Phi^N | S_i, T_i, \nu_i (i = 1, 2, 3)]$$
$$= [2\pi I_0(\kappa)]^{-1} \exp[\kappa \cos(\Phi^N - \xi)] \qquad (1)$$

where κ and ξ follow from

$$\kappa \cos \xi = X \quad \kappa \sin \xi = Y \quad \kappa > 0$$

and

$$X = 2\beta_0 S_1 S_2 S_3 + 2\beta_1 (T_1 S_2 S_3 \cos \nu_1$$
$$+ S_1 T_2 S_3 \cos \nu_2 + S_1 S_2 T_3 \cos \nu_3)$$
$$+ 2\beta_2 [T_1 T_2 S_3 \cos(\nu_1 + \nu_2)$$
$$+ T_1 S_2 T_3 \cos(\nu_1 + \nu_3)$$
$$+ S_1 T_2 T_3 \cos(\nu_2 + \nu_3)]$$
$$+ 2\beta_3 T_1 T_2 T_3 \cos(\nu_1 + \nu_2 + \nu_3)$$

$$Y = 2\beta_1 (T_1 S_2 S_3 \sin \nu_1 + S_1 T_2 S_3 \sin \nu_2$$
$$+ S_1 S_2 T_3 \sin \nu_3)$$
$$+ 2\beta_2 [T_1 T_2 S_3 \sin(\nu_1 + \nu_2)$$
$$+ T_1 S_2 T_3 \sin(\nu_1 + \nu_3)$$
$$+ S_1 T_2 T_3 \sin(\nu_2 + \nu_3)]$$
$$+ 2\beta_3 T_1 T_2 T_3 \sin(\nu_1 + \nu_2 + \nu_3).$$

I_0 is the modified Bessel function of the first kind and of order zero. This distribution can be obtained from Hauptman's (1982a) joint probability distribution [equation (3.4)] in the same way as equation (15) of Klop, Krabbendam & Kroon (1987) or alternatively directly from that equation (15) by analogy. Although the present notation differs from that used in the latter paper, both notations coincide if the definitions of the f structure and g structure are interchanged in the latter paper, so that the f structure includes the heavy atoms and the g structure does not.

Since the light-atom structure is part of the total structure the coefficients β_0, β_1, β_2 and β_3 defined

by Hauptman (1982a) are simplified to

$$\beta_0 = \frac{\sigma_3 - \tau_3}{(1-\alpha^2)^3 \sigma_2^{3/2}}$$

$$\beta_1 = -\alpha\beta_0$$

$$\beta_2 = \alpha^2\beta_0$$

$$\beta_3 = [\tau_3\sigma_2^3 - 3\tau_3\sigma_2^2\tau_2 + 3\tau_3\sigma_2\tau_2^2 - \sigma_3\tau_2^3]$$

$$\times [(1-\alpha^2)^3 \tau_2^{3/2}\sigma_2^3]^{-1}$$

where

$$\sigma_x = \sum_{j=1}^{N} (f_j)^x \quad \tau_x = \sum_{j=1}^{N\prime} (f_j)^x$$

and α as defined by Hauptman (1982a) is simplified to

$$\alpha = (\tau_2/\sigma_2)^{1/2}.$$

N is the total number of atoms in the unit cell. The superscript (') specifies that the summation is restricted to the non-anomalous scatterers.

The calculation of wavelength-independent quantities

The calculation of wavelength-independent quantities from multi-wavelength data was advocated by Karle (1980, 1984). In this section we will present the approach of Singh & Ramaseshan (1968) modified to include scale factors and to obtain explicit expressions for $|F_i^N|$, $|F_i^L|$ and the phase difference ν_i. It is assumed that the anomalous scatterers are identical.

Let $c_{ij}^{1/2}$ be the local scale factor to put $|F_{ij}^+|$ and $|F_{ij}^-|$ on an absolute scale, and let G_i be defined as

$$G_i = \sum_{k=1}^{N\prime\prime} n_k T_k \exp(2\pi i \mathbf{h}_i . \mathbf{r}_k),$$

where ('') specifies that the summation is restricted to the anomalous scatterers, n_k is the occupation factor of atom k and T_k is its isotropic temperature factor. If f_j' and f_j'' are the real and imaginary parts of the scattering factor of the heavy atoms at wavelength λ_j we have

$$c_{ij}^{1/2} F_{ij}^\pm = F_i^N + (f_j' \pm i f_j'') G_i \quad (j = 1, \ldots, U)$$

$$= [(a_i + f_j') + i(b_i \pm f_j'')] G_i \quad G_i \neq 0$$

where $a_i + i b_i \equiv F_i^N / G_i$.
So for $G_i \neq 0$,

$$c_{ij}|F_{ij}^\pm|^2 = [(a_i + f_j')^2 + (b_i \pm f_j'')^2] g_i^2 \quad (2a,b)$$

where g_i is the magnitude of G_i. Now, let

$$\Delta I_{ij} \equiv |F_{ij}^+|^2 - |F_{ij}^-|^2$$

and

$$M_{ij} \equiv (|F_{ij}^+|^2 + |F_{ij}^-|^2)/2.$$

Adding and subtracting the two equations (2a) and (2b) leads to

$$c_{ij}M_{ij} = [(a_i + f_j')^2 + b_i^2 + f_j''^2] g_i^2 \quad (3)$$

$$c_{ij}\Delta I_{ij} = 4 f_j'' b_i g_i^2. \quad (4)$$

From (4) it follows that

$$c_{ij} = (f_j'' \Delta I_{i1} / f_1'' \Delta I_{ij}) c_{i1} \quad \text{for } \Delta I_{ij} \neq 0. \quad (5)$$

For the two-wavelength case ($U = 2$), the quantities a_i and b_i are subsequently eliminated and a quadratic equation in g_i^2 is obtained with solutions

$$g_{i,\pm}^2 = (Q_i/P) \pm (Q_i^2 - R_i P)^{1/2}/P. \quad (6)$$

The parameters P, Q_i and R_i are

$$P = [(\Delta f')^2 - (f_1''^2 - f_2''^2)]^2 + 4(\Delta f')^2 f_1''^2$$

$$Q_i = (\Delta f')^2 (c_{i1}M_{i1} + c_{i2}M_{i2})$$

$$+ (f_1''^2 - f_2''^2)(c_{i1}M_{i1} - c_{i2}M_{i2})$$

$$R_i = (c_{i1}M_{i1} - c_{i2}M_{i2})^2 + \rho_i^2(\Delta f')^2 \quad (7)$$

with ρ_i and $\Delta f'$ defined as

$$\rho_i \equiv c_{i2}\Delta I_{i2}/2 f_2'' \quad \text{and} \quad \Delta f' \equiv f_1' - f_2'. \quad (8)$$

The scale factor c_{ij} of reflections for which $\Delta I_{ij} \neq 0$ can be expressed in terms of c_{i1} via (5). In order to reduce the statistical error in the scale factor for reflection $\mathbf{h}_i$ one may calculate a scale factor by averaging local scale factors of reflections in a region in reciprocal space that contains the reciprocal-lattice point corresponding to $\mathbf{h}_i$. This procedure can be used to calculate the scale factor of centrosymmetrical and pseudo-centrosymmetrical reflections. Such reflections have $\Delta I_{ij} = 0$, so (5) cannot be used. If the data are free of systematic errors, global scaling (Matthews & Czerwinski, 1975) can be applied: $c_{ij} = c_j$.

For the parameters a_i and b_i we obtain

$$a_{i,\pm} = \{c_{i1}M_{i1} - c_{i2}M_{i2} - [(f_1'^2 - f_2'^2) + (f_1''^2 - f_2''^2)]g_{i,\pm}^2\}$$

$$\times (2\Delta f' g_{i,\pm}^2)^{-1} \quad (9a)$$

$$b_{i,\pm} = \rho_i / 2 g_{i,\pm}^2 \quad (9b)$$

by using (3), (4) and (8). For each reflection two solutions are obtained, viz

$$a_{i,-}, b_{i,-}, g_{i,-} \quad \text{and} \quad a_{i,+}, b_{i,+}, g_{i,+}$$

analogous to the heavy-atom lower estimate (HLE) and heavy-atom upper estimate (HUE) in protein crystallography for single isomorphous replacement combined with anomalous scattering (SIRAS). In the two-wavelength case, as in the SIRAS case, the HLE solution (i.e. $a_{i,-}, b_{i,-}, g_{i,-}$) is chosen as the correct one. (There is no ambiguity if $U \geq 3$.)

Equation (6) was derived for $g_i \neq 0$. If $g_i = 0$, we have $c_{ij}^{1/2} F_{ij}^\pm = F_i^N$ so $M_{i1} = M_{i2}$ (if there are no scaling errors) and $\Delta I_{ij} = 0$ so $R_i = 0$ and, by (6), $g_{i,-} = 0$, $g_{i,+} = 2Q_i/P$. Hence (6) is also valid for $g_i = 0$.

Omitting the subscripts ($\pm$), we have by definition

$$a_i + ib_i = F_i^N / G_i,$$

so

$$|F_i^N| = (a_i^2 + b_i^2)^{1/2} g_i. \tag{10}$$

Let f^H be the normal part of the heavy-atom scattering factor and $F_i^H \equiv f^H G_i$; then

$$F_i^L = F_i^N - F_i^H = (a_i - f^H + ib_i) G_i$$

so

$$|F_i^L| = [(a_i - f^H)^2 + b_i^2]^{1/2} g_i. \tag{11}$$

The phase difference ν_i is easily obtained as the argument of F_i^N / F_i^L:

$$F_i^N / F_i^L = [a_i^2 + b_i^2 - a_i f^H - ib_i f^H]$$
$$\times [(a_i - f^H)^2 + b_i^2]^{-1} \tag{12a}$$

$$\nu_i = \varphi_i^N - \varphi_i^L = \text{Arg} (F_i^N / F_i^L). \tag{12b}$$

If c_{i1} is unknown then $g_i / c_{i1}^{1/2}$, $|F_i^N| / c_{i1}^{1/2}$ and $|F_i^L| / c_{i1}^{1/2}$ are obtained by application of (6), (10) and (11), i.e. amplitudes are obtained on the relative scale of the diffraction data collected at wavelength λ_1 instead of on an absolute scale. This has no disturbing effect either on the application of direct methods or on the application of Patterson methods. In the former case normalization of data in the usual way brings the calculated values of $|F_i^N|$ and $|F_i^L|$ on to an absolute scale. Note that a_i and b_i and hence ν are independent of c_{i1}, i.e. insensitive to scaling errors if (5) is used.

Practical application

The procedure was tested on artificial two-wavelength diffraction data of the iron-containing protein ferredoxin (Sieker, Adman & Jensen, 1972) which crystallizes in $P2_12_12_1$ with $a = 30 \cdot 52$, $b = 37 \cdot 75$, $c = 39 \cdot 7$ Å and molecular weight $M_r \sim 6000$. The eight iron atoms in the molecule, located in two Fe-S clusters, are assumed to be the only anomalous scatterers. When synchroton radiation is used an f' value of about -10 and an f'' value of about $+6$ can be obtained for Fe by tuning the wavelength around its K edge (Phillips et al., 1977). The artificial data consist of structure factors calculated using the parameters listed in Table 1. Practical tests of the procedure were made without and with the introduction of errors. In the latter case random errors normally distributed with standard deviation $\sigma = 0 \cdot 01, 0 \cdot 02$ and $0 \cdot 04$ are independently applied to each structure factor. The R factors in Table 1 are defined as

$$R_j \equiv \frac{\sum \{|F_{ij}^+(\text{error})| + |F_{ij}^-(\text{error})|\}}{\sum (|F_{ij}^+| + |F_{ij}^-|)}$$

where $F_{ij}^\pm(\text{error})$ is the random error applied to $F_{ij}^\pm$.

In the calculation of wavelength-independent quantities from the artificial diffraction data a reflection was rejected if: (a) $g_{i,-}$ is imaginary; (b) $g_{i,-} < 2 \cdot 0$; or (c) $g_{i,-} > 32 \cdot 0$ (32 Fe atoms in the unit cell).

Table 1. *Parameters used to calculate two-wavelength diffraction data*

Anomalous scatterers: Fe
B (overall) $= 10 \cdot 00$ Å^2
Resolution: 2 Å
3328 reflections

$\lambda_1 = 1 \cdot 541$ Å	$f_1' = -1 \cdot 18$	$f_1'' = 3 \cdot 20$	$c_1 = 1 \cdot 0$
$\lambda_2 = 1 \cdot 743$ Å	$f_2' = -10 \cdot 00$	$f_2'' = 6 \cdot 00$	$c_2 = 1 \cdot 0$

σ (%)	R_1 (%)	R_2 (%)
1	0·74	0·72
2	1·48	1·44
4	2·95	2·87

Table 2 lists the number of reflections used in the calculations for each chosen standard deviation together with the following R parameters:

$$R_g \equiv \frac{\sum |g_{i,-} - g_{i,\text{true}}|}{\sum g_{i,\text{true}}} \qquad R_{FN} \equiv \frac{\sum |F_{i,-}^N - F_{i,\text{true}}^N|}{\sum |F_{i,\text{true}}^N|}$$

$$R_{FL} \equiv \frac{\sum |F_{i,-}^L - F_{i,\text{true}}^L|}{\sum |F_{i,\text{true}}^L|}$$

and the average error $\langle \nu(\text{error}) \rangle$ in the phase differences. It turns out that the calculated heavy-atom contributions g are more sensitive to errors than are the calculated magnitudes $|F^N|$ and $|F^L|$. The calculation of wavelength-independent HLE/HUE quantities is an exact algebraic procedure, so the R indices should be zero for $\sigma = 0 \cdot 0$ if the HLE/HUE ambiguity did not exist. The low values of the R indices for $\sigma = 0 \cdot 0$ suggest that the resolution of this ambiguity by choosing the HLE solution is a good procedure. The number of reflections for which the HUE solution instead of the HLE solution is correct is rather small (98 out of 2331 reflections for ferredoxin, $\sigma = 0 \cdot 0$). Furthermore these reflections appear to have small $|F^N|$ values ($|F^N| < 140 \cdot 6$ compared with a maximum $|F^N|$ of $2181 \cdot 6$). For 319 reflections the HUE solution is impossible since $g_+ > 32 \cdot 0$; for these reflections there is no ambiguity.

After normalization of the $|F^N|$ and $|F^L|$ values obtained by application of (11) and (12), 500 reflections with largest $|E^N|$ values were used to generate $\sum_2$ relationships (see Table 3). Note that triple products generated from reflections with large $|E^N|$ values may not always have large κ values since κ also depends on $|E^L|$ and ν. Therefore retaining only those reflections with large $|E^N|$ is a computationally efficient but not necessarily optimal procedure.

Triplet invariants were estimated by application of (1) and subsequently sorted on κ. Table 4 summarizes the results of 8000 invariants with largest κ and Table 5 shows a sample of ten invariants with their κ values.

Discussion

The test results in the previous section show that triplet invariants can be estimated accurately via the probability distribution $P[\Phi^N | S_i, T_i, \nu_i (i = 1, 2, 3)]$

Table 2. *The calculation of wavelength-independent quantities: standard deviation σ of the error (%), number of reflections, indices R_g, R_{FN} and R_{FL} (%) and average error $\langle \nu(\text{error}) \rangle$ in the phase difference ν (°) for each family of reflections*

Family of reflections	σ	Number of reflections	R_g	R_{FN}	R_{FL}	$\langle \nu(\text{error}) \rangle$
Non-centrosymmetric reflections	0	1908	1·46	0·02	0·54	1·5
	1	1897	5·20	0·67	2·29	3·5
	2	1880	8·86	1·31	4·04	5·3
	4	1897	16·76	2·63	7·50	8·9
Centrosymmetric reflections	0	423	3·80	0·11	1·83	8·6
	1	422	7·33	0·68	3·88	10·8
	2	422	10·89	1·25	5·85	12·9
	4	409	17·66	2·40	9·75	16·2
All reflections	0	2331	1·96	0·04	0·81	2·8
	1	2319	5·65	0·67	2·63	4·8
	2	2302	9·29	1·30	4·43	6·7
	4	2306	16·95	2·58	7·97	10·2

Table 3. *Minimum value of 500 largest $|E^N|$ magnitudes and number of relationships*

σ	Minimum $|E^N|$	Number of relationships
0	1·24	19 458
1	1·24	19 654
2	1·24	19 481
4	1·27	20 047

Table 4. *Average κ values, Cochran κ values (kapco) (calculated from $|E^N|$) and magnitudes of error in the estimated values of triplet invariants*

Top	σ	$\langle \kappa \rangle$	$\langle$kapco$\rangle$	$\langle \Phi(\text{error}) \rangle$
1000	0	6·77	0·55	16·0
	1	6·76	0·56	17·0
	2	7·13	0·55	17·7
	4	8·00	0·56	21·1
2000	0	5·10	0·51	20·7
	1	5·06	0·51	21·2
	2	5·36	0·50	21·8
	4	5·97	0·51	26·9
8000	0	2·47	0·41	39·6
	1	2·45	0·42	40·8
	2	2·62	0·41	41·8
	4	2·88	0·42	48·1

Table 5. *A sample of ten triplet invariants ($\sigma = 0$)*

Serial number of invariant	κ	kapco	Φ^{est}	Φ^{true}	Error
1	27·14	0·32	−164	−166	2
101	10·20	0·43	−11	−17	6
201	8·37	0·87	75	105	30
301	7·24	0·74	72	70	2
401	6·51	0·28	−24	−20	4
501	5·98	0·41	−6	−26	20
601	5·43	0·29	−83	−106	23
701	5·07	0·42	−50	−66	16
801	4·78	0·34	24	0	24
901	4·47	0·53	−58	−88	30

after calculation of the conditional information from two-wavelength data $|F_{ij}^+|$, $|F_{ij}^-|$ ($j = 1, 2$).

Hauptman (1983) suggested an alternative approach in which $|F_{ij}^+|$, $|F_{ij}^-|$ ($j = 1, 2$) is used as conditional information after normalization. This approach consists of the estimation of 64 (not-

independent) triplet phases $\Omega_1 = \varphi_{1,1} + \varphi_{2,1} + \varphi_{3,1}$, $\Omega_2 = \varphi_{-1,1} + \varphi_{2,1} + \varphi_{3,1}$, $\Omega_3 = \varphi_{1,2} + \varphi_{2,1} + \varphi_{3,1}, \ldots$, $\Omega_{64} = \varphi_{-1,2} + \varphi_{-2,2} + \varphi_{-3,2}$ ($\varphi_{\pm i,j}$ is the phase of $F_{ij}^{\pm}$) for each $\sum_2$ relationship *via* 64 twelve-magnitude probability distributions $P_k[\Omega_k||E_{ij}^+|, |E_{ij}^-|$ ($i = 1, 2, 3; j = 1, 2$)], $k = 1, \ldots, 64$, analogous to the six-magnitude probabilistic formulae for the one-wavelength case derived by Hauptman (1982*b*) and Giacovazzo (1983). The inevitable approximations involved in these extremely laborious derivations pose a serious problem since they may restrict the usefulness of the phase-determining formulae. On the other hand, different types of anomalous scatterers are more easily incorporated in the approach suggested by Hauptman. In the two-step procedure described in the present paper the complexity of the problem is greatly reduced by the algebraic calculation of wavelength-independent data, prior to using direct methods. Furthermore, since the algebraic pre-processing of the data can easily cope with data collected at more than two wavelengths (the HLE/HUE ambiguity is removed) the use of 216 (in the three-wavelength case) eighteen-magnitude distributions can be avoided.

In the multi-wavelength procedure proposed by Karle (1984) solution of a set of linear equations gives the heavy-atom contributions together with phase differences $\varepsilon_i \equiv \varphi_i^L - \varphi_i^H$ where φ_i^L and φ_i^H are the phases of F_i^L and F_i^H respectively. Triplet invariants are estimated using the relation $\Phi^L = \Phi^H + \varepsilon_1 + \varepsilon_2 + \varepsilon_3$ with $\Phi^H \equiv \varphi_1^H + \varphi_2^H + \varphi_3^H$. If the Cochran κ value of the heavy-atom contribution is large then $\Phi^H \simeq 0$ so $\Phi^L \simeq \varepsilon_1 + \varepsilon_2 + \varepsilon_3$. The procedure described in the present paper differs from Karle's procedure in several respects. (i) In the latter procedure the set of linear equations is singular in the two-wavelength case so that at least three wavelengths are required. The former procedure uses non-linear equations which can be solved in the two-wavelength case as well. (ii) The phase-determining information is $|F_i^N|$, $|F_i^L|$ and $\varphi_i^N - \varphi_i^L$ in the former procedure *versus* g_i and ε_i in the latter. With one type of

anomalous scatterer, $|F_{ij}^{\pm}|$ is uniquely determined by $|F_i^N|$, $|F_i^L|$, $\varphi_i^N - \varphi_i^L$, f_j' and f_j'', but not so, however, by g_i, ε_i, f_j' and f_j''. This suggests that the phase-determining information in the former procedure is more powerful than in the latter. (iii) Triplet invariants of which the contributing reflections have small g values cannot be estimated in Karle's procedure. However, in our procedure these invariants can also be estimated since for these reflections $F^N \simeq F^L$ and $\nu \simeq 0$, i.e. (1) can still be used.

Instead of using $|F^N|$, $|F^L|$ and $\varphi^N - \varphi^L$ it is possible to use $|F^N|$, $|F^H|$ and $\varphi^N - \varphi^H$ or $|F^L|$, $|F^H|$ and $\varphi^L - \varphi^H$ as conditional information (after normalization of the magnitudes). With each of these choices, $|F_{ij}^{\pm}|$ is determined if there is one type of anomalous scatterer. Hence with one type of anomalous scatterer the information contained in $|F_{ij}^{\pm}|$ is also present in each of the three choices mentioned. The choice for $|F^N|$, $|F^L|$ and $\varphi^N - \varphi^L$ was motivated by the fact that it resembles most closely the conditional information used in Hauptman's (1982a) distribution for single isomorphous replacement from which our equation (1) was derived.

In protein crystallography it is customary to try to solve the heavy-atom structure first from isomorphous replacement data, often supplemented by anomalous dispersion. With multi-wavelength data it is possible to solve the heavy-atom structure by standard direct or Patterson methods using the calculated heavy-atom magnitudes g. Subsequently, with known heavy-atom positions, protein structure-factor phases can be determined except for reflections with small heavy-atom contributions. The procedure proposed in the present paper allows accurate estimation of triplet invariants and does not require the positions of the heavy atoms. The derivation of structure-factor phases from these triplet phase invariants will be the subject of a forthcoming paper.

References

GIACOVAZZO, C. (1983). *Acta Cryst.* A39, 585-592.
HAUPTMAN, H. (1982a). *Acta Cryst.* A38, 289-294.
HAUPTMAN, H. (1982b). *Acta Cryst.* A38, 632-641.
HAUPTMAN, H. (1983). Private communication.
KARLE, J. (1980). *Int. J. Quantum Chem.* 7, 357-367.
KARLE, J. (1984). *Acta Cryst.* A40, 526-531.
KLOP, E. A., KRABBENDAM, H. & KROON, J. (1987). *Acta Cryst.* A43, 810-820.
MATTHEWS, B. W. & CZERWINSKI, E. W. (1975). *Acta Cryst.* A31, 480-487.
PHILLIPS, J. C., WLODAWER, A., GOODFELLOW, J. M., WATENPAUGH, K. D., SIEKER, L. C., JENSEN, L. H. & HODGSON, K. O. (1977). *Acta Cryst.* A33, 445-455.
SIEKER, L. C., ADMAN, E. T. & JENSEN, L. H. (1972). *Nature (London)*, 235, 40-42.
SINGH, A. K. & RAMASESHAN, S. (1968). *Acta Cryst.* B24, 35-40.

SHORT COMMUNICATIONS

Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 1000 words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible.

On the orthogonal transformation used for structural comparisons.

By SIMON K. KEARSLEY, *Department of Chemistry, Yale University, 225 Prospect Street, New Haven, CT 06511, USA*

Abstract

Rotation matrices that minimize or maximize the sum of the squared distances between corresponding atoms for two structures are found using a constrained least-squares procedure solved analytically as an eigenvalue problem in quaternion parameters.

Orthogonal transformations have been used widely to compare molecular structures (Cox, 1967; McLachlan, 1972, 1979; Rao & Rossmann, 1973; Nyburg, 1974; Hendrickson, 1979; Kenknight, 1984; Honzatko, 1986; Lesk, 1986). Such a transformation superimposes structures to facilitate visual comparisons and to give a quantitative measure of shape similarity as the root mean square (r.m.s.) deviation of distances between corresponding atoms. A prerequisite for the comparison is that the atom–atom correspondence has been determined between molecules. Then a criterion for the best superposition of two molecular structures is that the sum of the squared distances between these atoms be a minimum. Finding the optimal orthogonal transformation requires determination of a rotation matrix R and a translation vector that will superimpose two sets of coordinates to meet this criterion. To obtain optimal overlap with respect to the translation vector the two coordinate sets should first have their centroids moved to the origin. The problem of finding the rotation matrix, the elements of which are not linearly independent, invariably involves some iterative optimization procedure, although Kabsch (1976, 1978) solves for R by a direct method using Lagrange multipliers to impose orthogonality constraints on the transformation