Acta Cryst. (1990). A46, 521-526

A Multi-Solution Approach with Invariants Estimated from Two-Wavelength Data

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

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

(*Received* 23 *June* 1989; *accepted* 30 *January* 1990)

Abstract

The two-step procedure for estimating triplet invariants from two-wavelength data [Klop, Krabbendam & Kroon (1989). *Acta Cryst.* A45, 203-208] is applied to artificial data from the protein ferredoxin. Two different probability distributions for the estimation of triplet invariants are compared. The possibility of estimating triplet phase invariants for which one or more reflections have a small heavy-atom contribution is demonstrated. The estimated triplet invariants are used in a convergence procedure and a generalized tangent formula. It is shown that accurate phase determination is possible even at low resolution.

Introduction

The solution of macromolecular crystal structures still relies heavily on additional experimental methods such as isomorphous-replacement and anomalous-scattering techniques. Several authors report that direct methods can be successfully applied for extension and refinement of protein phases using the tangent formula in combination with structure-factor phase information derived from anomalous scattering and isomorphous replacement. [*e.g.* Hendrickson (1973); Woolfson & Yao Jia-xing (1988), using the Sayre-equation tangent formula]. These methods require prior structural information as the heavy-atom substructure must be solved to obtain an initial set of structure-factor phases.

In recent years, anomalous scattering and isomorphous replacement have been integrated with direct methods to produce triplet invariant phase estimates (Kroon, Spek & Krabbendam, 1977; Hauptman, 1982*a*, *b*; Giacovazzo, 1983). The integration of multi-wavelength anomalous scattering with direct methods was a further step in this direction (Karle, 1984; Klop, Krabbendam & Kroon, 1989*a*). In these methods, solution of the heavy-atom substructure is no longer required.

In a previous paper (Klop *et al.*, 1989*a*), it was shown that triplet invariants can be estimated from two-wavelength data using a conditional probability distribution based on Hauptman's joint probability distribution for single isomorphous replacement. In the first step of the procedure, wavelength-independent structure-factor magnitudes and phase differences are calculated algebraically from wavelength-dependent diffraction data (see also Klop, Krabbendam & Kroon, 1989*b*). The wavelength-independent quantities thus obtained are used as conditional information in the second step of the procedure, the estimation of triplet invariants. In the present paper two different distributions for the estimation of triplet invariants are compared. The multi-wavelength phasing procedure is completed by using the estimated triplet invariants in a multi-solution procedure to generate structure-factor phases.

Definitions

In the following definitions a subscript i denotes the reciprocal vector $\mathbf{h}_i$.

F_{ij}^{+} Structure factor at wavelength λ_j.

F_{ij}^{-} Complex conjugate of the structure factor for reflection $-\mathbf{h}_i$ at wavelength λ_j.

F_i^N Structure factor based on the normal (*i.e.* non-anomalous) parts of the scattering factors of all atoms (including anomalously scattering atoms).

F_i^L Structure factor based on the normal parts of the scattering factors of all atoms excluding anomalously scattering atoms.

F_i^H Structure factor based on the normal parts of the scattering factors of the anomalously scattering atoms.

Other quantities are defined in the text.

Method

Two-wavelength data $|F_{i1}^{+}|$, $|F_{i1}^{-}|$, $|F_{i2}^{+}|$ and $|F_{i2}^{-}|$ corresponding to reflections $\mathbf{h}_i$ were calculated from the known coordinates of the protein ferredoxin (Sieker, Adman & Jensen, 1972) using the parameters listed in Table 1. Ferredoxin is a small iron-containing protein with a molecular weight $M_r \sim 6000$ which crystallizes in space group $P2_12_12_1$. The eight Fe atoms in the molecule are located in two Fe-S clusters and are assumed to be the only anomalous scatterers.

Wavelength-independent structure-factor magnitudes $|F_i^N|$, $|F_i^L|$, $|F_i^H|$ and phase differences $\nu_i \equiv \varphi_i^N - \varphi_i^L$ and $\delta_i \equiv \varphi_i^N - \varphi_i^H$ were calculated from the

0108-7673/90/060521-06$03.00

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

Anomalous scatterers: Fe
$\lambda_1 = 1{\cdot}541$ Å $\quad f_1' = -1{\cdot}18 \quad f_1'' = 3{\cdot}20$
$\lambda_2 = 1{\cdot}743$ Å $\quad f_2' = -10{\cdot}0 \quad f_2'' = 6{\cdot}00$
B[overall] = 10·0
Resolution: 2 Å
3328 reflections

two-wavelength data as detailed by Klop *et al.* (1989*a*). The phases of F_i^N, F_i^L, F_i^H are denoted by φ_i^N, φ_i^L and φ_i^H, respectively. For reflections with small heavy-atom contributions F_i^H we have $F_i^N \simeq F_i^L$ and $\nu_i \simeq 0$. The ambiguity that exists with two-wavelength data is resolved by choosing the heavy-atom lower estimate (HLE). The resulting structure-factor magnitudes were normalized and are denoted by $|E_i^N|$, $|E_i^L|$ and $|E_i^H|$. Reflections with $|E_i^N| < 0{\cdot}2$ were omitted from subsequent calculations. The remaining reflections were used to generate Σ_2 relationships using the program *SULTAN*, an extended and adapted version of *MULTAN*80 (Main *et al.*, 1980).

Define the triplet invariant Φ^N by $\Phi^N = \varphi_1^N + \varphi_2^N + \varphi_3^N$ where $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0$. In separate runs of *SULTAN*, triplet invariants Φ^N and their reliabilities κ can be estimated by the program according to two different probability distributions, *viz* (i) the conditional probability distribution $P(\Phi^N || E_i^N|, |E_i^L|, \nu_i\,[i = 1, 3])$ of the triplet invariant Φ^N given the normalized structure-factor magnitudes $|E_i^N|$ and $|E_i^L|$ and the phase differences ν_i; (ii) the conditional probability distribution $P(\Phi^N || E_i^H|, \delta_i\,[i = 1, 3])$ of the triplet invariant Φ^N given the magnitudes $|E_i^H|$ and phase differences δ_i. Distribution (i) is given by Klop *et al.* (1989*a*) and reads

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

where κ and ξ follow from

$$\kappa \cos \xi = X, \quad \kappa \sin \xi = Y, \quad \kappa > 0 \quad (2a)$$

and

$$\begin{aligned} X = {} & 2\beta_0 S_1 S_2 S_3 + 2\beta_1 [T_1 S_2 S_3 \cos \nu_1 \\ & + S_1 T_2 S_3 \cos \nu_2 + S_1 S_2 T_3 \cos \nu_3] \\ & + 2\beta_2 [T_1 T_2 S_3 \cos(\nu_1 + \nu_2) + T_1 S_2 T_3 \cos(\nu_1 + \nu_3) \\ & + S_1 T_2 T_3 \cos(\nu_2 + \nu_3)] \\ & + 2\beta_3 T_1 T_2 T_3 \cos(\nu_1 + \nu_2 + \nu_3), \quad (2b) \end{aligned}$$

$$\begin{aligned} Y = {} & 2\beta_1 [T_1 S_2 S_3 \sin \nu_1 + S_1 T_2 S_3 \sin \nu_2 \\ & + S_1 S_2 T_3 \sin \nu_3] \\ & + 2\beta_2 [T_1 T_2 S_3 \sin(\nu_1 + \nu_2) \\ & + T_1 S_2 T_3 \sin(\nu_1 + \nu_3) + S_1 T_2 T_3 \sin(\nu_2 + \nu_3)] \\ & + 2\beta_3 T_1 T_2 T_3 \sin(\nu_1 + \nu_2 + \nu_3), \quad (2c) \end{aligned}$$

where $S_i \equiv |E_i^N|$, $T_i \equiv |E_i^L|$ and the coefficients β_k are given by Klop *et al.* (1989*a*). Distribution (ii) is defined as

$$\begin{aligned} & P(\Phi^N || E_i^H|, \delta_i\,[i = 1, 3]) \\ & \quad = [2\pi I_0(\kappa^H)]^{-1} \\ & \qquad \times \exp\{\kappa^H \cos[\Phi^N - (\delta_1 + \delta_2 + \delta_3)]\}, \quad (3) \end{aligned}$$

where κ^H is the Cochran kappa value of the heavy-atom structure,

$$\kappa^H = 2n^{-1/2} |E_1^H E_2^H E_3^H| \quad (4)$$

and n is the number of anomalous scatterers in the unit cell.

If the heavy-atom structure-factor contribution for a reflection with index $\mathbf{h}_i$ is small ($|F^H| < \frac{1}{16}|F^H|_{\max}$), the index is denoted by $\mathbf{h}_i'$. We can then discern four different types of Σ_2 relationships, characterized by

$$\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = \mathbf{0} \quad (5a)$$

$$\mathbf{h}_1' + \mathbf{h}_2 + \mathbf{h}_3 = \mathbf{0} \quad (5b)$$

$$\mathbf{h}_1' + \mathbf{h}_2' + \mathbf{h}_3 = \mathbf{0} \quad (5c)$$

$$\mathbf{h}_1' + \mathbf{h}_2' + \mathbf{h}_3' = \mathbf{0}. \quad (5d)$$

Because E_1^H is small for a relationship of types (5*b*)–(5*d*), κ^H is small and distribution (3) will not be useful. The latter distribution can therefore only be used to estimate triplet invariants of type (5*a*). However, distribution (1) may yield triplet-invariant estimates for the other types as well.

The estimated triplet invariants were sorted in decreasing order on κ. Convergence mapping was performed based on these κ values; origin-defining phases were chosen by the program. The enantiomorph is specified by the invariants so no enantiomorph-defining reflection is required. Magic integers were used to assign phases to several additional reflections.

The start sets of phases thus obtained were extended and refined using a generalized tangent formula

$$\tan \varphi_{\mathbf{h}} = \frac{\sum_{\mathbf{k}} w_{\mathbf{k}} w_{\mathbf{h}-\mathbf{k}} \kappa_{\mathbf{h},\mathbf{k}} \sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}} - \xi_{\mathbf{h},\mathbf{k}})}{\sum_{\mathbf{k}} w_{\mathbf{k}} w_{\mathbf{h}-\mathbf{k}} \kappa_{\mathbf{h},\mathbf{k}} \cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}} - \xi_{\mathbf{h},\mathbf{k}})}, \quad (6)$$

where $\xi_{\mathbf{h},\mathbf{k}}$ is the value of the triplet invariant Φ^N estimated *via* distribution (1) or (3) and

$$w_{\mathbf{h}} = \alpha_{\mathbf{h}}/5 \text{ for } \alpha_{\mathbf{h}} \leq 5, \qquad w_{\mathbf{h}} = 1 \text{ for } \alpha_{\mathbf{h}} \geq 5$$

and (7)

$$\alpha_{\mathbf{h}}^2 = (A_{\mathbf{h}}^2 + B_{\mathbf{h}}^2).$$

The numerator and denominator of the tangent formula are denoted by $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$ respectively. Note that $w_{\mathbf{h}}$ is the *MULTAN*80 default weight.

Table 2. *Minimum kappa values κ_{min}, average kappa values $\langle\kappa\rangle$, and average errors $\langle\Phi^{init}(error)\rangle$ in the values of NSR triplet invariants ($\kappa \geq \kappa_{min}$) estimated via distributions* (1) *and* (3) *for ferredoxin at* 2 Å *resolution*

Number of reflections (N_{ref})		2225
Number of Σ_2 relationships generated (NGR)		1 707 567
Number of strong Σ_2 relationships used (NSR)		80 000
Distribution used	(1)	(3)
β_0	1·826	
β_1	−1·622	
β_2	1·441	
β_3	−1·250	
κ_{min}	1·70	1·22
$\langle\kappa\rangle$	2·63	1·89
$\langle\Phi^{init}(error)\rangle$ (°)	35·1	34·4

Results

The results obtained by application of the tangent formula can be judged from the conventional figures of merit (ABSFOM, PSIZERO, RESID and CFOM) and, since the structure is known, from the average error $\langle\varphi^{tan}(error)\rangle$ in the phases and in the invariants $\langle\Phi^{tan}(error)\rangle$. The former is calculated for phases φ^{tan} having a final weight $w = 1{\cdot}0$ after tangent-formula phase generation and the latter is calculated for invariants Φ^{tan} that are generated from these phases (*i.e.* $\Phi^{tan} = \varphi_1^{tan} + \varphi_2^{tan} + \varphi_3^{tan}$). The average error in the initial estimates of the triplet invariants is denoted by $\langle\Phi^{init}(error)\rangle$.

Table 2 lists the average error $\langle\Phi^{init}(error)\rangle$ in the estimated triplet invariants for two different runs of the program using the distributions (1) and (3) respectively for ferredoxin. The data set consists of reflections having spacings greater than 2 Å and heavy-atom structure-factor contributions $|F^H| > \frac{1}{16}|F^H|_{max}$. The quality of the estimated triplet invariants is independent of the distribution chosen as can be verified by comparing the average error $\langle\Phi^{init}(error)\rangle$ for the two different runs of the program. This also applies for tests using data at other resolutions.

Distribution (1) allows the estimation of invariants of which a contributing reflection has a small heavy-atom contribution F^H [*i.e.* invariants of type (5*b*)–(5*c*)] as Table 3 shows. The present version of the program cannot use more than 80 000 invariants in the tangent expansion. This implies that only a small fraction of the invariants of type (5*b*)–(5*c*) can be used in a tangent expansion for a 2 Å data set. In Table 4 results are given obtained by the application of distribution (1) to a 3 Å data set which includes 356 reflections with $|F^H| < \frac{1}{16}|F^H|_{max}$. The approximations $F^N \simeq F^L$ and $\nu \simeq 0$ were used for the latter reflections. This corresponds with $F^H \simeq 0$ so distribution (3) cannot be applied. The second column of Table 4 pertains to the set of phases with the smallest average error $\langle\Phi^{tan}(error)\rangle$ [which coincides with the set with the smallest average error $\langle\varphi^{tan}(error)\rangle$; two other sets had a slightly better combined figure of merit (CFOM)]. For this phase set, 43 reflections with $|F^H| < \frac{1}{16}|F^H|_{max}$ are phased with average weight 0·29 after application of the tangent formula. The weighted average error for these reflections is 59·1°. In the following we will use distribution (1) to supply initial estimates for the values of the triplet invariants but we will exclude reflections with small heavy-atom contribution ($|F^H| < \frac{1}{16}|F^H|_{max}$) in view of the program limit referred to above.

Table 3. *Average error $\langle\Phi^{init}(error)\rangle$ and number of invariants estimated via distribution* (1) *in intervals of κ from* 0 *to* 5·0 *for ferredoxin at* 2 Å *resolution with* (i) *reflections with small heavy-atom contributions excluded,* (ii) *reflections with small heavy-atom contributions included*

κ	Number of invariants (i)	Number of invariants (ii)	$\langle\Phi^{init}(error)\rangle$ (°) (i)	$\langle\Phi^{init}(error)\rangle$ (°) (ii)
0·0–0·5	942 174	3 750 106	81·4	87·6
0·5–1·0	504 058	543 938	70·3	72·1
1·0–1·5	151 820	176 843	56·8	59·6
1·5–2·0	55 804	68 040	45·9	49·0
2·0–2·5	24 132	30 321	37·2	40·3
2·5–3·0	11 908	15 174	30·7	33·6
3·0–3·5	6485	8388	27·3	29·1
3·5–4·0	3753	4970	24·6	26·6
4·0–4·5	2302	3040	22·8	24·1
4·5–5·0	1548	1978	20·4	22·5

Table 4. *Results of tangent formula phasing of ferrodoxin at* 3 Å *resolution after estimating triplet invariants by distribution* (1)

Reflections with small heavy-atom contributions are included in the data set. N_{sets} is the number of start sets generated by the program. N_{ref} ($w = 1{\cdot}0$) is the number of reflections that have weight $w = 1{\cdot}0$ after tangent formula phasing.

N_{ref}	956	N_{ref} ($w = 1{\cdot}0$)	467
NGR	388 783	$\langle\varphi^{tan}(error)\rangle$ (°)	23·7
NSR	12 558	$\langle\Phi^{tan}(error)\rangle$ (°)	20·1
κ_{min}	0·80	ABSFOM	1·1860
$\langle\kappa\rangle$	1·49	PSIZERO	1·112
$\langle\Phi^{init}(error)\rangle$ (°)	43·7	RESID	15·90
N_{sets}	96	CFOM	2·8634

Table 5. *Average magnitude of the error $\langle\Phi^{init}(error)\rangle$ in estimated values of triplet invariants arranged in descending order of $\langle\kappa\rangle$ values for ferredoxin at* 2 Å *resolution*

Number of invariants	κ_{max}	κ_{min}	$\langle\kappa\rangle$	$\langle\Phi^{init}(error)\rangle$ (°)
1000	27·14	7·12	9·15	11·7
1000	7·12	5·88	6·41	15·5
1000	5·88	5·26	5·54	17·6
1000	5·26	4·85	5·04	19·2
1000	4·84	4·53	4·68	20·4
20 000	27·14	2·87	4·14	23·8
80 000	27·14	1·70	2·63	35·1

Table 5 shows the average errors in the 80 000 top ranking estimates for ferredoxin at 2 Å resolution. In

Table 6. *Tangent formula phasing for ferredoxin at 2 Å resolution for different magnitudes of error; s.d. is the standard deviation of the error in* $|F_{ij}^{\pm}|$; *'rank' indicates the position of the phase set when ranked according to decreasing COMFOM*

s.d. (%)	0	4	8
R_1	0·0	2·9	6·0
R_2	0·0	3·0	5·9
N_{ref}	2225	2210	2283
NGR	1 707 567	1 750 232	1 973 111
NSR	80 000	80 000	80 000
κ_{min}	1·70	1·78	2·23
$\langle\kappa\rangle$	2·63	2·76	3·30
$\langle\Phi^{init}(\text{error})\rangle$ (°)	35·1	41·4	57·2
N_{sets}	96	96	120
$N_{ref}(w=1·0)$	1800	1843	1934
$\langle\varphi^{tan}(\text{error})\rangle$ (°)	11·7	13·0	24·2
$\langle\Phi^{tan}(\text{error})\rangle$ (°)	12·6	15·5	32·0
ABSFOM	1·0110	0·9099	0·6475
PSIZERO	2·648	2·211	2·53
RESID	9·32	11·38	32·34
COMFOM	2·9796	2·9813	2·9269
Rank	1	2	1

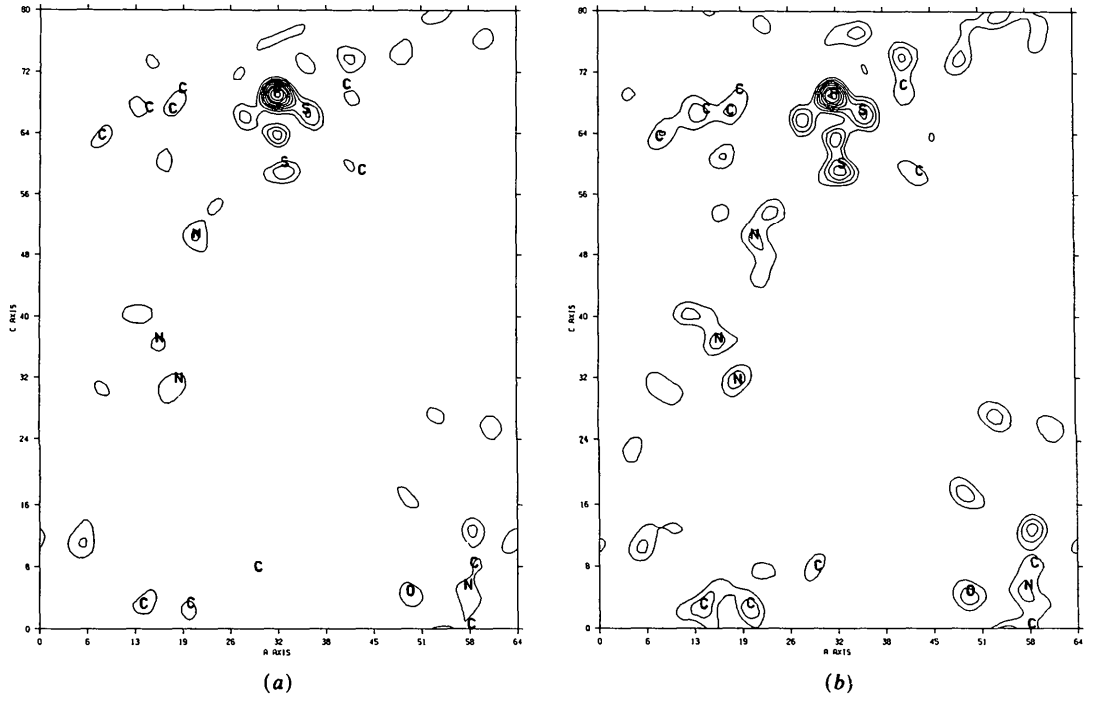
Table 7. *Tangent formula phasing for ferredoxin at 4 Å resolution for different magnitudes of error; s.d. is the standard deviation of the error in* $|F_{ij}^{\pm}|$; *'rank' indicates the position of the phase set when ranked according to decreasing COMFOM*

s.d. (%)	0	4	8
R_1	0·0	2·9	5·8
R_2	0·0	2·8	5·7
N_{ref}	304	320	331
NGR	29 656	33 908	36 906
NSR	9284	9284	9284
κ_{min}	0·60	0·71	1·13
$\langle\kappa\rangle$	1·42	1·63	2·52
$\langle\Phi^{init}(\text{error})\rangle$ (°)	40·9	48·5	63·3
N_{sets}	96	108	72
$N_{ref}(w=1·0)$	281	291	294
$\langle\varphi^{tan}(\text{error})\rangle$ (°)	7·1	15·3	34·9
$\langle\Phi^{tan}(\text{error})\rangle$ (°)	10·0	23·5	55·9
ABSFOM	1·2253	1·0596	0·7774
PSIZERO	1·625	1·75	1·93
RESID	16·60	16·02	25·68
COMFOM	2·8727	2·7411	2·8183
Rank	1	3	3

Tables 6 and 7 results are given for data sets at resolutions of 2 and 4 Å. To study the effect of errors in the data, tests were made without and with the introduction of errors. In the latter case, random errors normally distributed with standard deviation $\sigma = 0·04$ and 0·08 are independently applied to each structure factor. The R factors listed in Tables 6 and 7 are defined as

$$R_j \equiv \sum_i [|F_{ij}^{+}(\text{error})| + |F_{ij}^{-}(\text{error})|] / \sum_i (|F_{ij}^{+}| + |F_{ij}^{-}|),$$

where $F_{ij}^{\pm}(\text{error})$ is the random error applied to $F_{ij}^{\pm}$. The tangent formula results in Tables 6 and 7 pertain to the phase set with the smallest average error $\langle\varphi^{tan}(\text{error})\rangle$. Of 96 resultant phase sets at 2 Å resolution (error-free data), six gave average errors $\langle\varphi^{tan}(\text{error})\rangle$ less than 15·0° with an average error of 11·7° for the phase set with the smallest average error. The latter set appeared to have the best combined figure of merit. Fig. 1(*a*) shows a section of the E map calculated from this phase set. For comparison,



(*a*) (*b*)

Fig. 1. E map calculated for 2 Å ferredoxin, (*a*) using the phase set obtained by application of the tangent formula, (*b*) using the true phase set (3328 reflections). The true positions of the atoms are indicated.

Fig. 1(*b*) displays a section of the E map calculated from the full 2 Å sphere with its 'true' phases.

Discussion

In a previous paper (Klop *et al.*, 1989*a*), wavelength-independent structure-factor magnitudes and phase differences were calculated from artificial data and test results were presented in which triplet invariants were estimated (from $|E^N|$, $|E^L|$ and ν) using the 500 reflections with the largest $|E^N|$ values (excluding reflections with small heavy-atom structure-factor contribution). In the present paper, this procedure is replaced by employing wavelength-independent quantities of the (larger) set of reflections with $|E^N| > 0{\cdot}2$ to estimate triplet invariants. Comparison of Table 4 of Klop *et al.* (1989*a*) with Table 5 of the present paper shows that a far greater number of invariants with large κ values is obtained with the larger set of reflections.

Karle (1984) proposed the estimation of triplet invariants Φ^L from multi-wavelength data by employing the relation

$$\Phi^L - \Phi^H = \varepsilon_1 + \varepsilon_2 + \varepsilon_3, \qquad (8)$$

where $\Phi^L = \varphi_1^L + \varphi_2^L + \varphi_3^L$, $\Phi^H = \varphi_1^H + \varphi_2^H + \varphi_3^H$ and $\varepsilon_i = \varphi_i^L - \varphi_i^H$. The phase differences ε_i and the magnitudes $|F_i^L|$ and $|F_i^H|$ are calculated from multi-wavelength data *via* an algebraic procedure using a set of linear equations which requires at least three-wavelength data $|F_{i1}^{\pm}|$, $|F_{i2}^{\pm}|$ and $|F_{i3}^{\pm}|$ or the use of statistical estimates for some of the unknowns (Karle, 1989). In the procedure described by Karle (1984), triplet invariants Φ^H are estimated according to a suitable probability distribution. Karle concluded that the estimate $\Phi^H \cong 0$ with reliability κ^H can be employed. From relationship (8), the triplet invariants Φ^L are estimated as

$$\Phi^L \cong \varepsilon_1 + \varepsilon_2 + \varepsilon_3 \qquad (9)$$

with reliability κ^H. Estimation of invariants Φ^N *via* (3) in the present paper is closely related to Karle's procedure since it amounts to using

$$\Phi^N - \Phi^H = \delta_1 + \delta_2 + \delta_3 \qquad (10)$$

together with $\Phi^H \cong 0$ with reliability κ^H. Our conditional information $|E_i^H|$ and δ_i, however, is calculated from *two*-wavelength data *via* a set of non-linear equations. The small average errors listed in Table 2 obtained after employing distribution (3) to estimate triplet invariants Φ^N show that Karle's procedure of calculating the difference of two triplet invariants algebraically and using the estimate $\Phi^H \cong 0$ with reliability κ^H can also be applied in the two-wavelength case by using a set of non-linear equations.

Reference to Table 2 shows that distributions (1) and (3) are comparable with respect to their ability to estimate triplet invariants (reflections with small heavy-atom contribution being excluded from the data set). Fortier, Weeks & Hauptman (1984) showed that if Hauptman's (1982*a*, *b*) joint probability distribution is applied to the case where the derivative consists of the native protein plus the heavy-atom content, then the conditional probability distributions for single isomorphous replacement depend on the scattering difference between the native protein and the derivative, that is on the scattering of the heavy atoms in the derivative. Since distribution $P(\Phi^N \,||E_i^N|, |E_i^L|, \nu_i\, [i = 1, 3])$ [(1)] was derived from Hauptman's (1982*a*, *b*) joint probability distribution it is expected that (1) will depend on the difference in scattering power associated with the structure factors F^N and F^L. Distribution $P(\Phi^N \,||E_i^H|, \delta_i\, [i = 1, 3])$ [(3)] exploits the difference in scattering power (*i.e.* the scattering of the heavy atoms) associated with the structure factors F^N and F^L. The observation that (1) yields similar results to (3) is consistent with the view that (1) depends on the difference in scattering power associated with the structure factors F^N and F^L.

The possibility of estimating triplet invariants of the types (5*b*), (5*c*) or (5*d*) *via* distribution (1) is clearly demonstrated by the results for 2 Å data in Table 3. Application of the tangent formula to obtain structure-factor phases φ^N for reflections with small heavy-atom contribution is somewhat disappointing as can be concluded from the large weighted average error (59·1°) and the small average weight (0·29) of these phases for a 3 Å data set. It should be remarked, however, that the default weighting scheme of *MULTAN* was employed in the tangent formula. The influence of an adequate weighting scheme was demonstrated by Pitts, Tickle, Wood & Blundell (1982) who attempted to employ the tangent formula to extend phases of avian pancreatic polypeptide at 2·04 Å resolution. The default *MULTAN* weighting scheme led to the introduction of spurious noise peaks in the electron density map. However, the use of the Hull & Irwin (1978) weighting scheme together with a device that damped large changes in the phase or weight of the reflections allowed phase extension from 2·04 to 1·37 Å. The determination of an optimal weighting scheme for our purposes is non-trivial and requires further studies.

From Tables 6 and 7 it can be judged that accurate phasing is possible not only at 2 Å resolution but, provided that the errors in the data are not too large, even at the much lower resolution of 4 Å. The close similarity of the two Fourier syntheses (Fig. 1) calculated with true and estimated phases (obtained with error-free data) supports this conclusion. The set with the smallest phase error was always among the three sets with the highest combined figure of merit, which indicates that the conventional figures of merit are useful for structures with the size of small proteins.

The error in the phases increases slightly if the error in the data increases from $\sigma = 0{\cdot}00$ to $\sigma = 0{\cdot}04$ whereas from $\sigma = 0{\cdot}04$ to $\sigma = 0{\cdot}08$ the increase in phase error is much stronger. In reality, however, the standard deviation of the (random) error will be smaller than 0·08, but the data will be affected by systematic errors as well.

The results reported in the present paper show that it is possible to estimate triplet invariants from artificial two-wavelength data (which may be corrupted by random errors) and use the invariants in a multi-solution procedure to obtain structure-factor phases without the need to solve the heavy-atom structure.

References

FORTIER, S., WEEKS, C. M. & HAUPTMAN, H. (1984). *Acta Cryst.* A**40**, 544-548.

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

HAUPTMAN, H. (1982*a*). *Acta Cryst.* A**38**, 289-294.

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

HENDRICKSON, W. A. (1973). *Biophysical Applications of Crystallography Techniques*, pp. 61-83. *Trans. Am. Crystallogr. Assoc.*

HULL, E. & IRWIN, M. J. (1978). *Acta Cryst.* A**34**, 863-870.

KARLE, J. (1984). *Acta Cryst.* A**40**, 526-531.

KARLE, J. (1989). *Acta Cryst.* A**45**, 303-307.

KLOP, E. A., KRABBENDAM, H. & KROON, J. (1989*a*). *Acta Cryst.* A**45**, 203-208.

KLOP, E. A., KRABBENDAM, H. & KROON, J. (1989*b*). *Acta Cryst.* A**45**, 609-613.

KROON, J., SPEK, A. L. & KRABBENDAM, H. (1977). *Acta Cryst.* A**33**, 382-385.

MAIN, P., FISKE, S. J., HULL, S. E., LESSINGER, L., GERMAIN, G., DECLERQ, J.-P. & WOOLFSON, M. M. (1980). *MULTAN*80. *A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Univs. of York, England, and Louvain, Belgium.

PITTS, J. E., TICKLE, I. J., WOOD, S. P. & BLUNDELL, T. L. (1982). *Sov. Phys. Crystallogr.* **27**(1), 56-62.

SIEKER, L. C., ADMAN, E. T. & JENSEN, L. H. (1972). *Nature (London)*, **235**, 40-42.

WOOLFSON, M. M. & YAO JIA-XING (1988). *Acta Cryst.* A**44**, 410-413.

Acta Cryst. (1990). A**46**, 526-534

X-ray Transmissivity Measurements using Dynamical Diffraction under Grazing-Incidence Conditions

BY F. RIEUTORD

Institut Laue-Langevin, BP 156*X*, 38042 *Grenoble CEDEX, France*

(*Received* 12 *October* 1989; *accepted* 5 *February* 1990)

Abstract

The dependence of the X-ray transmission coefficient of a thin-film coating as a function of incidence has been measured in the grazing-angle range. The method is based on the use of a substrate Bragg reflection to redirect the incident or transmitted beams. It allows grazing incidence from both outside and inside the substrate to be performed. The geometry of the experiment is described. The results are interpreted by means of dynamical theory combined with an optical formalism for stratified systems. Experimental results and applications are compared with reflectivity data.

1. Introduction

Surface and thin-film studies using X-ray or neutrons in glancing-incidence geometries have undergone considerable development in recent years. In these studies, one makes use of the different signals from a sample when struck by a grazing beam such as, for instance, specular reflection in the reflectivity technique (Parratt, 1954), diffracted beams in grazing-incidence scattering (GIS) (Marra, Eisenberger & Cho, 1979) or even fluorescence (Brunel, 1986). Glancing angles are used in these techniques to limit the penetration depth and thus to enhance the sensitivity to the near-surface volume. Therefore, bulk substrate scattering is often considered as a nuisance in these experiments. The aim of this paper is to show that it is possible to take advantage of the substrate to perform measurements of the transmission coefficient which can provide useful information, complementing data from other techniques.

The interpretation of the data of grazing-incidence techniques is often made by means of kinematical theory (Born approximation). Such an approximation is justified by the weak coupling between X-rays (or neutrons) and matter. When the incidence angle is close to the critical angle for total reflection, this approximation has to be removed since multiple-scattering effects become important. Improvement of simple Born approximations can be made [*e.g.* distorted-wave Born approximation (Vineyard, 1982)] but the most general way of performing the calculations

0108-7673/90/060526-09$03.00