

that, in the case of flexible molecules, much clearer results are obtained by using a smaller but correct fragment, rather than a larger one that may differ significantly in a global sense from the unknown protein.

This work was supported by the Medical Research Council of Canada through the Group on Protein Structure and Function. The authors thank Dr D. R. Davies for providing them with coordinates of the J539 Fab fragment.

#### References

- AMZEL, L. M. & POLJAK, R. J. (1979). *Ann. Rev. Biochem.* **48**, 961-997.
- BI RU-CHANG, CUTFIELD, S. M., DODSON, E. J., DODSON, G. G., REYNOLDS, C. D. & TOLLEY, S. P. (1983). *Acta Cryst.* **B39**, 90-98.
- BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
- BOTT, R. & SARMA, R. (1976). *J. Mol. Biol.* **106**, 1037-1046.
- CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173-178. New York: Gordon & Breach.
- CYGLER, M., BOODHOO, A., LEE, J. S. & ANDERSON, W. F. (1987). *J. Biol. Chem.* **262**, 643-648.
- DEREWENDA, Z. S., DODSON, E. J., DODSON, G. G. & BRZozowski, A. M. (1981). *Acta Cryst.* **A37**, 407-413.
- DIJKSTRA, B. W., VAN NES, G. J. H., KALK, K. H., BRANDENBURG, N. P., HOL, W. G. J. & DRENTH, J. (1982). *Acta Cryst.* **B38**, 793-799.
- DODSON, E. (1979). Fast rotation function program. Univ. of York, England.
- DODSON, E., HARDING, M. M., HODGKIN, D. C. & ROSSMANN, M. G. (1966). *J. Mol. Biol.* **16**, 227-241.
- DRIESSEN, H. P. C. & WHITE, H. (1985). *Molecular Replacement*. Proceedings of the Daresbury Study Weekend, 15-16 February 1985, edited by P. A. MACHIN, pp. 27-32. Warrington: SERC Daresbury Laboratory.
- FUJINAGA, M., READ, R. J., SIELECKI, A., ARDELT, W., LASKOWSKI, M. JR & JAMES, M. N. G. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 4868-4872.
- HOPPE, W. (1957). *Acta Cryst.* **10**, 750-751.
- HUBER, R. (1965). *Acta Cryst.* **19**, 353-356.
- HUBER, R. & BENNETT, W. S. JR (1983). *Biopolymers*, **22**, 261-279.
- KABAT, E. A., WU, T. T., BILOFSKY, H., REID-MILER, M. & PERRY, H. (1983). *Sequences of Proteins of Immunological Interest*. US Department of Health and Human Services, Public Health Service, National Institute of Health.
- LEE, J. S., DOMBROSKI, D. F. & MOSMANN, T. R. (1982). *Biochemistry* **21**, 4940-4945.
- LEE, J. S., LEWIS, J. R., MORGAN, A. R., MOSMANN, T. R. & SINGH, B. (1981). *Nucleic Acids Res.* **9**, 1707-1721.
- MCPHALEN, C. A., SVENDSEN, I., JONASSEN, I. & JAMES, M. N. G. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 7242-7246.
- MARQUART, M., DEISENHOFER, J., HUBER, R. & PALM, W. (1980). *J. Mol. Biol.* **141**, 369-391.
- MUSICK, W. D. L. & ROSSMANN, M. G. (1979). *J. Biol. Chem.* **254**, 7611-7620.
- PHIZACKERLEY, R. P., WISHNER, B. C., BRYANT, S. H., AMZEL, L. M., LOPEZ DE CASTRO, J. A. & POLJAK, R. J. (1979). *Mol. Immunol.* **16**, 841-850.
- ROSSMANN, M. G. (1972). Editor. *The Molecular Replacement Method*. New York: Gordon and Breach.
- ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.
- SAUL, F., AMZEL, L. M. & POLJAK, R. J. (1978). *J. Mol. Biol.* **253**, 585-97.
- SCHIERBEEK, A. J., RENETSEDER, R., DIJKSTRA, B. W. & HOL, W. (1985). *Molecular Replacement*. Proceedings of the Daresbury Study Weekend, 15-16 February 1985, edited by P. A. MACHIN, pp. 16-21. Warrington: SERC Daresbury Laboratory.
- SCHIRMER, T., HUBER, R., SCHNEIDER, M., BODE, W., MILLER, M. & HACKERT, M. L. (1986). *J. Mol. Biol.* **188**, 651-676.
- SEGAL, D. M., PADLAN, E. A., COHEN, G. H., RUDIKOFF S., POTTER, M. & DAVIES, D. R. (1974). *Proc. Natl Acad. Sci. USA*, **71**, 4298-4302.
- STEIGEMANN, W. (1974). PhD Thesis, Technische Univ. Munich, Federal Republic of Germany.
- SUH, S. W., BHAT, T. N., NAIRA, M. A., COHEN, G. H., RAO, D. N., RUDIKOFF, S. & DAVIES, D. R. (1986). *Proteins Struct. Funct. Genet.* **1**, 74-80.
- WANG, D., BODE, W. & HUBER, R. Y. (1985). *J. Mol. Biol.* **185**, 595-624.

*Acta Cryst.* (1988). **A44**, 45-51

## On Integrating the Techniques of Direct Methods and Isomorphous Replacement. A New Probabilistic Formula for Triplet Invariants

BY C. GIACOVAZZO, G. CASCARANO AND ZHENG CHAO-DE\*

*Centro Interdipartimentale di Cristallografia, c/o Dipartimento Geomineralogico, Campus Universitario, Via Amendola, 70124 Bari, Italy*

(Received 22 April 1987; accepted 11 August 1987)

#### Abstract

The mathematical technique recently used [Hauptman (1982). *Acta Cryst.* **A38**, 289-294] for integrating

\* Permanent address: Institute of Physics, Academia Sinica, Beijing, China.

direct methods and isomorphous replacement techniques is reconsidered. The atomic positions are assumed to be the primitive random variables instead of the reciprocal vectors. A new probabilistic formula for estimating three-phase invariants given six magnitudes has been obtained which differs from the corresponding result of Hauptman. The first applications

suggest that the formula is quite efficient for picking up triplet cosines near 0 or  $\pi$ .

### 1. Introduction

The theoretical basis for integrating the techniques of direct methods and isomorphous replacement was introduced by Hauptman (1982) (henceforth referred to as H). In his notation, the atomic scattering factors for a corresponding pair of isomorphous structures are denoted by  $f_j$  and  $g_j$ , so that

$$E_h = |E_h| \exp(i\varphi_h) = (1/\alpha_{20}^{1/2}) \sum_{j=1}^N f_j \exp(2\pi i h r_j)$$

and

$$G_h = |G_h| \exp(i\psi_h) = (1/\alpha_{20}^{1/2}) \sum_{j=1}^N g_j \exp(2\pi i h r_j)$$

are respective normalized structure factors, where

$$\alpha_{mn} = \sum_{j=1}^N f_j^m g_j^n.$$

The main aim of Hauptman's paper was to initiate the probabilistic theory of two-phase and three-phase structure invariants for isomorphous pairs. In particular:

(a) the joint probability distribution

$$P(\varphi_h, \psi_h, |E_h|, |G_h|)$$

was found, from which the conditional distribution

$$P(\omega_h || E_h|, |G_h|)$$

may be obtained, where

$$\omega_h = \varphi_h - \psi_h$$

is a two-phase structure invariant;

(b) the joint probability distribution function

$$\begin{aligned} P(\varphi_1, \varphi_2, \varphi_3, \psi_1, \psi_2, \psi_3, R_1, R_2, R_3, S_1, S_2, S_3) \\ \approx [\pi^6 (1 - \alpha^2)^3]^{-1} R_1 R_2 R_3 S_1 S_2 S_3 \\ \times \exp\{-[1/(1 - \alpha^2)](R_1^2 + R_2^2 + R_3^2 \\ + S_1^2 + S_2^2 + S_3^2) + 2\beta[R_1 S_1 \cos(\varphi_1 - \psi_1) \\ + R_2 S_2 \cos(\varphi_2 - \psi_2) + R_3 S_3 \cos(\varphi_3 - \psi_3)] \\ + 2\beta_0 R_1 R_2 R_3 \cos(\varphi_1 + \varphi_2 + \varphi_3) \\ + 2\beta_1 [R_1 R_2 S_3 \cos(\varphi_1 + \varphi_2 + \psi_3) \\ + R_1 S_2 R_3 \cos(\varphi_1 + \psi_2 + \varphi_3) \\ + S_1 R_2 R_3 \cos(\psi_1 + \varphi_2 + \varphi_3)] \\ + 2\beta_2 [R_1 S_2 S_3 \cos(\varphi_1 + \psi_2 + \psi_3) \\ + S_1 R_2 S_3 \cos(\psi_1 + \varphi_2 + \psi_3) \\ + S_1 S_2 R_3 \cos(\psi_1 + \psi_2 + \varphi_3)] \\ + 2\beta_3 S_1 S_2 S_3 \cos(\psi_1 + \psi_2 + \psi_3)\} \end{aligned} \quad (1)$$

was secured, where  $\varphi_1, \varphi_2, \varphi_3, \psi_1, \psi_2, \psi_3$  stand for

$\varphi_h, \varphi_k, \varphi_l, \psi_h, \psi_k, \psi_l$  respectively,  $R_1, R_2, R_3, S_1, S_2, S_3$  represent  $|E_h|, |E_k|, |E_l|, |G_h|, |G_k|, |G_l|$  respectively, and

$$h + k + l = 0. \quad (2)$$

The parameters  $\alpha$  and  $\beta$  [equations (H2.5) and (H3.5)-(H3.9)] may be expressed in terms of the parameters  $\alpha_{mn}$ .

From (1) conditional probability distributions of the three-phase structure invariants

$$\omega_1 = \varphi_h + \varphi_k + \varphi_l$$

$$\omega_2 = \varphi_h + \varphi_k + \psi_l$$

$$\omega_3 = \varphi_h + \psi_k + \varphi_l$$

$$\omega_4 = \psi_h + \varphi_k + \varphi_l$$

$$\omega_5 = \varphi_h + \psi_k + \psi_l$$

$$\omega_6 = \psi_h + \varphi_k + \psi_l$$

$$\omega_7 = \psi_h + \psi_k + \varphi_l$$

$$\omega_8 = \psi_h + \psi_k + \psi_l,$$

given the six magnitudes  $R_1, R_2, R_3, S_1, S_2, S_3$ , may be found. The first applications of the method (Hauptman, Potter & Weeks, 1982) on error-free diffraction data from the protein cytochrome  $c_{550}$  from *Paracoccus denitrificans* (molecular weight  $M_r \approx 14\,500$ , space group  $P2_12_12_1$ ) and a single  $\text{PtCl}_4^{2-}$  derivative were successful. Reliable tens of thousands of triplet invariants having extreme values 0 or  $\pi$  were identified, so the method appeared to be of increasing importance for the determination of macromolecular structures.

A useful interpretation of the formulae in terms of experimental parameters was suggested by Fortier, Weeks & Hauptman (1984). It was shown that distributions do not depend, as in case of the traditional three-phase invariants, on the total number of atoms per unit cell but rather on the scattering difference between the native protein and the derivative (that is, on the scattering of the heavy atoms in the derivative).

Hauptman's approach is largely innovative and opened a new field for the application of direct methods. However, his mathematical procedure may be further modified in order to derive more accurate probabilistic formulas. As a first practical example, let us observe the parameter

$$\alpha = \alpha_{11}/(\alpha_{20}\alpha_{02})^{1/2},$$

which is deeply involved in distribution (1) (it is also employed to define the  $\beta_i$  parameters). While  $\alpha$  is a resolution-dependent parameter, its value in the procedure is calculated *via* zero-angle scattering factors of  $f$  and  $g$ . Indeed in (1) no difference is made among  $\alpha_h, \alpha_k, \alpha_l$  in spite of the fact they may be markedly different. The same considerations hold for the parameters  $\alpha_{mn}$  which define the more complex

parameters  $\beta$  and  $\beta_i$ ,  $i=0, 1, 2, 3$ , via equations (H3.5)–(H3.9).

The main aim of the present paper is therefore to obtain new probabilistic formulas by a mathematical approach which is able to take full account of the resolution effects on distribution parameters. It is worth mentioning that in Hauptman's approach the primitive random variable is the ordered triple  $(\mathbf{h}, \mathbf{k}, \mathbf{l})$  of reciprocal vectors which is assumed to be uniformly distributed over the subset (2) of the threefold Cartesian product  $S_r \times S_r \times S_r$  ( $S_r$  denotes reciprocal space). Then  $E_h, E_k, E_l, G_h, G_k, G_l$  are functions of the primitive random variables  $\mathbf{h}, \mathbf{k}, \mathbf{l}$ , so that they are themselves random variables.

In our approach atomic coordinates are chosen to be the primitive random variables while  $\mathbf{h}, \mathbf{k}, \mathbf{l}$  will be a fixed triple of reciprocal vectors. This enables us to calculate all the distribution parameters for that chosen triple, that is to say, at the resolution corresponding to the  $|\mathbf{h}|, |\mathbf{k}|, |\mathbf{l}|$  values.

The distribution based on atomic coordinates as the independent random variables is conceptually distinct from that in which it is assumed that the reciprocal vectors are random variables (see Giacobazzo, 1977). In most cases reciprocal vectors and atomic coordinates play a quite symmetric role [*i.e.* when no prior information is available, and it is only necessary that the fractional part of  $\mathbf{hr}$  be uniformly distributed in the interval  $(0, 1)$ ]. On the other hand prior information often breaks down the symmetry of the roles, so that the two distributions are deeply different [for example, when a structural fragment has a known position (Giacobazzo, 1983), or when superstructural effects are present (Cascarano, Giacobazzo & Luić, 1985, 1987)]. Symmetrization of the roles is still possible but some supplementary conditions on the reciprocal vectors are needed if they are used as primitive random variables. For example, in the case of substructural effects it is often necessary to restrict reciprocal vectors to subsets constituted by reflexions of fixed parity. In the present case, in which distribution parameters depend on the moduli  $|\mathbf{h}|, |\mathbf{k}|, |\mathbf{l}|$ , the ordered triple  $(\mathbf{h}, \mathbf{k}, \mathbf{l})$  has to be assumed to be uniformly distributed over the subset of  $S_r \times S_r \times S_r$ , defined by (2), but under the supplementary condition that  $|\mathbf{h}|, |\mathbf{k}|, |\mathbf{l}|$  have fixed chosen values.

In §§ 2 and 3 the main formulas are given, and in § 4 a first practical application is described.

## 2. The joint probability distribution of the six structure factors $E_h, E_k, E_l, G_h, G_k, G_l$ , where $\mathbf{h} + \mathbf{k} + \mathbf{l} = \mathbf{0}$

The atomic positions are chosen to be the primitive random variables. The mathematical formalism adopted here follows that described in a recent book (Giacobazzo, 1980) with some differences in detail, which for the sake of brevity are not described. One

obtains the result

$$\begin{aligned}
 P(\varphi_1, \varphi_2, \varphi_3, \psi_1, \psi_2, \psi_3, R_1, R_2, R_3, S_1, S_2, S_3) \\
 \approx (1/\pi^6) R_1 R_2 R_3 S_1 S_2 S_3 [(1 - \alpha_1^2)(1 - \alpha_2^2)(1 - \alpha_3^2)]^{-1} \\
 \times \exp \left\{ + \sum_{i=1}^3 [-(R_i^2 + S_i^2)/(1 - \alpha_i^2) \right. \\
 + 2\beta_{0i} R_i S_i \cos(\psi_i - \varphi_i)] \\
 + 2\beta_0 R_1 R_2 R_3 \cos(\varphi_1 + \varphi_2 + \varphi_3) \\
 + 2\beta_{11} S_1 R_2 R_3 \cos(\psi_1 + \varphi_2 + \varphi_3) \\
 + 2\beta_{12} R_1 S_2 R_3 \cos(\varphi_1 + \psi_2 + \varphi_3) \\
 + 2\beta_{13} R_1 R_2 S_3 \cos(\varphi_1 + \varphi_2 + \psi_3) \\
 + 2\beta_{21} R_1 S_2 S_3 \cos(\varphi_1 + \psi_2 + \psi_3) \\
 + 2\beta_{22} S_1 R_2 S_3 \cos(\psi_1 + \varphi_2 + \psi_3) \\
 + 2\beta_{23} S_1 S_2 R_3 \cos(\psi_1 + \psi_2 + \varphi_3) \\
 \left. + 2\beta_3 S_1 S_2 S_3 \cos(\psi_1 + \psi_2 + \psi_3) \right\} \quad (3)
 \end{aligned}$$

where

$$\begin{aligned}
 \beta_0 &= \left[ \prod_i (1 - \alpha_i^2) \right]^{-1} (-\alpha_1 \alpha_2 \alpha_3 \gamma_{456} \\
 &\quad + \alpha_1 \alpha_2 \gamma_{345} + \alpha_1 \alpha_3 \gamma_{246} + \alpha_2 \alpha_3 \gamma_{156} \\
 &\quad - \alpha_1 \gamma_{234} - \alpha_2 \gamma_{135} - \alpha_3 \gamma_{126} + \gamma_{123}), \\
 \beta_{11} &= \left[ \prod_i (1 - \alpha_i^2) \right]^{-1} (-\alpha_1 \alpha_2 \alpha_3 \gamma_{156} \\
 &\quad + \alpha_1 \alpha_2 \gamma_{135} + \alpha_1 \alpha_3 \gamma_{126} + \alpha_2 \alpha_3 \gamma_{456} \\
 &\quad - \alpha_1 \gamma_{123} - \alpha_2 \gamma_{345} - \alpha_3 \gamma_{246} + \gamma_{234}), \\
 \beta_{12} &= \left[ \prod_i (1 - \alpha_i^2) \right]^{-1} (-\alpha_1 \alpha_2 \alpha_3 \gamma_{246} \\
 &\quad + \alpha_1 \alpha_2 \gamma_{234} + \alpha_1 \alpha_3 \gamma_{456} + \alpha_2 \alpha_3 \gamma_{126} \\
 &\quad - \alpha_1 \gamma_{345} - \alpha_2 \gamma_{123} - \alpha_3 \gamma_{156} + \gamma_{135}), \\
 \beta_{13} &= \left[ \prod_i (1 - \alpha_i^2) \right]^{-1} (-\alpha_1 \alpha_2 \alpha_3 \gamma_{345} \\
 &\quad + \alpha_1 \alpha_2 \gamma_{456} + \alpha_1 \alpha_3 \gamma_{234} + \alpha_2 \alpha_3 \gamma_{135} \\
 &\quad - \alpha_1 \gamma_{246} - \alpha_2 \gamma_{156} - \alpha_3 \gamma_{123} + \gamma_{126}), \\
 \beta_{21} &= \left[ \prod_i (1 - \alpha_i^2) \right]^{-1} (-\alpha_1 \alpha_2 \alpha_3 \gamma_{234} \\
 &\quad + \alpha_1 \alpha_2 \gamma_{246} + \alpha_1 \alpha_3 \gamma_{345} + \alpha_2 \alpha_3 \gamma_{123} \\
 &\quad - \alpha_1 \gamma_{456} - \alpha_2 \gamma_{126} - \alpha_3 \gamma_{135} + \gamma_{156}), \\
 \beta_{22} &= \left[ \prod_i (1 - \alpha_i^2) \right]^{-1} (-\alpha_1 \alpha_2 \alpha_3 \gamma_{135} \\
 &\quad + \alpha_1 \alpha_2 \gamma_{156} + \alpha_1 \alpha_3 \gamma_{123} + \alpha_2 \alpha_3 \gamma_{345} \\
 &\quad - \alpha_1 \gamma_{126} - \alpha_2 \gamma_{456} - \alpha_3 \gamma_{234} + \gamma_{246}),
 \end{aligned}$$

$$\beta_{23} = \left[ \prod_i (1 - \alpha_i^2) \right]^{-1} (-\alpha_1 \alpha_2 \alpha_3 \gamma_{126} + \alpha_1 \alpha_2 \gamma_{123} + \alpha_1 \alpha_3 \gamma_{156} + \alpha_2 \alpha_3 \gamma_{246} - \alpha_1 \gamma_{135} - \alpha_2 \gamma_{234} - \alpha_3 \gamma_{456} + \gamma_{345}),$$

$$\beta_3 = \left[ \prod_i (1 - \alpha_i^2) \right]^{-1} (-\alpha_1 \alpha_2 \alpha_3 \gamma_{123} + \alpha_1 \alpha_2 \gamma_{126} + \alpha_1 \alpha_3 \gamma_{135} + \alpha_2 \alpha_3 \gamma_{234} - \alpha_1 \gamma_{156} - \alpha_2 \gamma_{246} - \alpha_3 \gamma_{345} + \gamma_{456}),$$

$$\beta_{0i} = \alpha_i / (1 - \alpha_i^2) \quad \text{for } i = 1, 2, 3.$$

In their turn

$$\alpha_1 = \alpha_{11}(\mathbf{h}) / [\alpha_{20}(\mathbf{h}) \alpha_{02}(\mathbf{h})]^{1/2}$$

$$\alpha_2 = \alpha_{11}(\mathbf{k}) / [\alpha_{20}(\mathbf{k}) \alpha_{02}(\mathbf{k})]^{1/2}$$

$$\alpha_3 = \alpha_{11}(\mathbf{l}) / [\alpha_{20}(\mathbf{l}) \alpha_{02}(\mathbf{l})]^{1/2}$$

where

$$\alpha_{20}(\mathbf{h}) = \sum_{j=1}^N f_j^2(\mathbf{h}) = \Sigma_1,$$

$$\alpha_{20}(\mathbf{k}) = \sum_{j=1}^N f_j^2(\mathbf{k}) = \Sigma_2,$$

$$\alpha_{20}(\mathbf{l}) = \sum_{j=1}^N f_j^2(\mathbf{l}) = \Sigma_3,$$

$$\alpha_{02}(\mathbf{h}) = \sum_{j=1}^N g_j^2(\mathbf{h}) = \Sigma'_1,$$

$$\alpha_{02}(\mathbf{k}) = \sum_{j=1}^N g_j^2(\mathbf{k}) = \Sigma'_2,$$

$$\alpha_{02}(\mathbf{l}) = \sum_{j=1}^N g_j^2(\mathbf{l}) = \Sigma'_3,$$

$$\alpha_{11}(\mathbf{h}) = \sum_{j=1}^N f_j(\mathbf{h}) g_j(\mathbf{h}),$$

$$\alpha_{11}(\mathbf{k}) = \sum_{j=1}^N f_j(\mathbf{k}) g_j(\mathbf{k}),$$

$$\alpha_{11}(\mathbf{l}) = \sum_{j=1}^N f_j(\mathbf{l}) g_j(\mathbf{l}).$$

Also

$$\gamma_{126} = (\Sigma_1 \Sigma_2 \Sigma_3)^{-1/2} \sum_{j=1}^N f_j(\mathbf{h}) f_j(\mathbf{k}) g_j(\mathbf{l}),$$

$$\gamma_{135} = (\Sigma_1 \Sigma'_2 \Sigma_3)^{-1/2} \sum_{j=1}^N f_j(\mathbf{h}) g_j(\mathbf{k}) f_j(\mathbf{l}),$$

$$\gamma_{156} = (\Sigma_1 \Sigma_2 \Sigma'_3)^{-1/2} \sum_{j=1}^N f_j(\mathbf{h}) g_j(\mathbf{k}) g_j(\mathbf{l}),$$

$$\gamma_{234} = (\Sigma'_1 \Sigma_2 \Sigma_3)^{-1/2} \sum_{j=1}^N g_j(\mathbf{h}) f_j(\mathbf{k}) f_j(\mathbf{l}),$$

Table 1. Correspondence between parameters occurring in Hauptman's and in our distribution

$$\alpha \rightarrow \alpha_1, \alpha_2, \alpha_3$$

$$\beta \rightarrow \beta_{01}, \beta_{02}, \beta_{03}$$

$$\beta_0 \rightarrow \beta_0$$

$$\beta_1 \rightarrow \beta_{13}, \beta_{12}, \beta_{11}$$

$$\beta_2 \rightarrow \beta_{21}, \beta_{22}, \beta_{23}$$

$$\beta_3 \rightarrow \beta_3$$

$$\gamma_{345} = (\Sigma'_1 \Sigma'_2 \Sigma_3)^{-1/2} \sum_{j=1}^N g_j(\mathbf{h}) g_j(\mathbf{k}) f_j(\mathbf{l}),$$

$$\gamma_{246} = (\Sigma'_1 \Sigma_2 \Sigma'_3)^{-1/2} \sum_{j=1}^N g_j(\mathbf{h}) f_j(\mathbf{k}) g_j(\mathbf{l}),$$

$$\gamma_{456} = (\Sigma'_1 \Sigma'_2 \Sigma'_3)^{-1/2} \sum_{j=1}^N g_j(\mathbf{h}) g_j(\mathbf{k}) g_j(\mathbf{l}),$$

$$\gamma_{123} = (\Sigma_1 \Sigma_2 \Sigma_3)^{-1/2} \sum_{j=1}^N f_j(\mathbf{h}) f_j(\mathbf{k}) f_j(\mathbf{l}).$$

If (3) is compared with (1) it may be observed that:

(a) A one-to-many correspondence exists between  $\alpha$  and  $\beta_i, i = 0, 1, 2, 3$  in (1), and parameters occurring in (3). Parameters playing a similar (but not identical role) lie on the same line in Table 1.

(b)  $\alpha_i$  and  $\beta_{ij}$  parameters occurring in our distribution are numerically markedly different from the corresponding parameters of Hauptman. This also holds for terms  $\beta_0$  and  $\beta_3$ , in spite of the fact that the same symbols are used in (1) and (3).

(c) To the eight parameters  $\gamma_{ijl}$  only four distinct parameters (i.e.  $\alpha_{30}, \alpha_{21}, \alpha_{12}, \alpha_{03}$ ) correspond in Hauptman's distribution.

From (3) the following conditional probability distribution may be derived:

$$P(\omega_1 | R_1, R_2, R_3, S_1, S_2, S_3)$$

$$\approx [2\pi I_0(A_1)]^{-1} \exp[A_1 \cos \omega_1] \quad (4)$$

where

$$A_1 = [2\beta_0 R_1 R_2 R_3 + (2\beta_{11} S_1 R_2 R_3) T_1 + (2\beta_{12} R_1 S_2 R_3) T_2 + (2\beta_{13} R_1 R_2 S_3) T_3 + 2\beta_{23} S_1 S_2 R_3 T_1 T_2 + (2\beta_{22} S_1 R_2 S_3) T_1 T_3 + 2\beta_{21} R_1 S_2 S_3 T_2 T_3 + 2\beta_3 S_1 S_2 S_3 T_1 T_2 T_3]$$

where

$$T_i = I_1(2\beta_{0i} R_i S_i) / I_0(2\beta_{0i} R_i S_i)$$

is the ratio of two modified Bessel functions. Equation (4) may usefully be compared with (H3.14).

Similar expressions for  $\omega_i, i = 2, \dots, 8$  may be found in a recursive way. Adopting the practical rule described by Fortier, Weeks & Hauptman (1984), we

get

$$A_i = [\tau_0(2\beta_0 R_1 R_2 R_3) + \tau_{11}(2\beta_{11} S_1 R_2 R_3) \\ + \tau_{12}(2\beta_{12} R_1 S_2 R_3) + \tau_{13}(2\beta_{13} R_1 R_2 S_3) \\ + \tau_{23}(2\beta_{23} S_1 S_2 R_3) + \tau_{22}(2\beta_{22} S_1 R_2 S_3) \\ + \tau_{21}(2\beta_{21} R_1 S_2 S_3) + \tau_3(2\beta_3 S_1 S_2 S_3)] \quad (5)$$

where  $\tau = C_1 C_2 C_3$  is obtained by comparing the  $i$ th structure factor associated with the coefficient of  $\tau$  with the  $i$ th structure factor associated with the invariant. If they are of the same type, *i.e.* both  $R$  or both  $S$ , then  $C_i = 1$  for  $i = 1, 2, 3$ , otherwise  $C_i = T_i$ .

### 3. The case of a native protein heavy-atom derivative

Formulas (3) and (1) are completely general and include as special cases native protein and heavy-atom isomorphous derivatives as well as X-ray and neutron diffraction data. We explicitly treat here the case of a native protein heavy-atom derivative because:

- (a) it frequently occurs;
- (b) the parameters of the distribution (3) are then rather simple;
- (c) it allows an easier interpretation of the formula in terms of diffraction experiments.

The following notation is introduced:

$$\sum_{H1} = \sum_H f_j^2(\mathbf{h}), \quad \sum_{H2} = \sum_H f_j^2(\mathbf{k}), \quad \sum_{H3} = \sum_H f_j^2(\mathbf{l}), \\ \gamma_P = \left[ \sum_P f_j(\mathbf{h}) f_j(\mathbf{k}) f_j(\mathbf{l}) \right] / \left( \sum_1 \sum_2 \sum_3 \right)^{1/2}$$

and

$$\gamma_H = \left[ \sum_H f_j(\mathbf{h}) f_j(\mathbf{k}) f_j(\mathbf{l}) \right] / \left( \sum_1 \sum_2 \sum_3 \right)^{1/2},$$

where summations over  $P$  and over  $H$  state that free indices vary over protein atoms and over heavy atoms respectively. In its turn  $\gamma_P$  may be approximated by the well known parameter  $[\sigma_3/\sigma_2^{3/2}]_P$ , where  $\sigma_n = \sum Z_j^n$ , and  $Z_j$  is the atomic number of the  $j$ th atom. Furthermore

$$\alpha_{20}(\mathbf{h}) = \alpha_{11}(\mathbf{h}) = \sum_1, \quad \alpha_{02}(\mathbf{h}) = \sum_1 + \sum_{H1} = \sum_1' \\ \alpha_{20}(\mathbf{k}) = \alpha_{11}(\mathbf{k}) = \sum_2, \quad \alpha_{02}(\mathbf{k}) = \sum_2 + \sum_{H2} = \sum_2' \\ \alpha_{20}(\mathbf{l}) = \alpha_{11}(\mathbf{l}) = \sum_3, \quad \alpha_{03}(\mathbf{l}) = \sum_3 + \sum_{H3} = \sum_3' \\ \alpha_1 = \left( \sum_1 / \sum_1' \right)^{1/2}, \quad \alpha_2 = \left( \sum_2 / \sum_2' \right)^{1/2},$$

$$\alpha_3 = \left( \sum_3 / \sum_3' \right)^{1/2}$$

$$\beta_{01} = \left( \sum_1 \sum_1' \right)^{1/2} / \sum_{H1},$$

$$\beta_{02} = \left( \sum_2 \sum_2' \right)^{1/2} / \sum_{H2},$$

$$\beta_{03} = \left( \sum_3 \sum_3' \right)^{1/2} / \sum_{H3},$$

$$\varepsilon_1 = \prod_i (1 - \alpha_i), \quad \varepsilon_2 = \prod_i (1 - \alpha_i^2),$$

$$\gamma_{234} / \alpha_1 = \gamma_{135} / \alpha_2 = \gamma_{126} / \alpha_3 =$$

$$\gamma_{345} / \alpha_1 \alpha_2 = \gamma_{246} / \alpha_1 \alpha_3 = \gamma_{156} / \alpha_2 \alpha_3 = \gamma_{123} = \gamma_P,$$

$$\gamma_{456} = (\gamma_P + \gamma_H) \alpha_1 \alpha_2 \alpha_3,$$

$$\beta_0 = (\varepsilon_2 \gamma_P - \alpha_1^2 \alpha_2^2 \alpha_3^2 \gamma_H) / \varepsilon_2,$$

$$\beta_{11} = \alpha_1 \alpha_2^2 \alpha_3^2 \gamma_H / \varepsilon_2,$$

$$\beta_{12} = \alpha_1^2 \alpha_2 \alpha_3^2 \gamma_H / \varepsilon_2,$$

$$\beta_{13} = \alpha_1^2 \alpha_2^2 \alpha_3 \gamma_H / \varepsilon_2,$$

$$\beta_{21} = -\alpha_1^2 \alpha_2 \alpha_3 \gamma_H / \varepsilon_2,$$

$$\beta_{22} = -\alpha_1 \alpha_2^2 \alpha_3 \gamma_H / \varepsilon_2,$$

$$\beta_{23} = -\alpha_1 \alpha_2 \alpha_3^2 \gamma_H / \varepsilon_2,$$

$$\beta_3 = \alpha_1 \alpha_2 \alpha_3 \gamma_H / \varepsilon_2.$$

Let us now suppose that the terms  $(2\beta_{0i} R_i S_i)$  are sufficiently large to justify the assumption  $T_i = 1$  for  $i = 1, 2, 3$ . Then (5) may be approximated by

$$A_i = 2\gamma_P R_1 R_2 R_3 + (2\gamma_H / \varepsilon_2) \alpha_1 \alpha_2 \alpha_3 (S_1 - \alpha_1 R_1) \\ \times (S_2 - \alpha_2 R_2) (S_3 - \alpha_3 R_3). \quad (6)$$

According to (6): (a) every invariant  $\omega_i$  for  $i = 1, \dots, 8$  is estimated with about the same reliability; (b) the first component of  $A_i$ ,  $2\gamma_P R_1 R_2 R_3$ , is nothing else but a Cochran (1955)-type contribution which is usually negligible for proteins; (c) the second component of  $A_i$  depends on the parameter  $2\gamma_H \alpha_1 \alpha_2 \alpha_3 / \varepsilon_2$ . Since

$$1 / \varepsilon_2 = \sum_1' \sum_2' \sum_3' / \sum_{H1} \sum_{H2} \sum_{H3}$$

then

$$2\gamma_H \alpha_1 \alpha_2 \alpha_3 / \varepsilon_2 \\ = \frac{2 \sum_H f_j(\mathbf{h}) f_j(\mathbf{k}) f_j(\mathbf{l})}{\left( \sum_{H1} \sum_{H2} \sum_{H3} \right)^{1/2}} \left( \frac{\sum_1' \sum_2' \sum_3'}{\sum_{H1} \sum_{H2} \sum_{H3}} \right)^{1/2} \\ \approx 2[\sigma_3 / \sigma_2^{3/2}]_H \left( \frac{\sum_1' \sum_2' \sum_3'}{\sum_{H1} \sum_{H2} \sum_{H3}} \right)^{1/2}. \quad (7)$$

Thus, according to Fortier, Weeks & Hauptman (1984), high reliability estimates will be available if differences in the diffraction intensities between the native protein and the derivative can be observed in correspondence with a small number of heavy atoms and large ratios  $(\sum_i' / \sum_{Hi})$ ;

(d) a simple algebraic rule for estimating the values of triplet phase invariants (0 or 180°) was presented by Karle (1983): a triplet invariant is expected to be positive or negative according to the sign of  $(F_{d1} - F_{p1})(F_{d2} - F_{p2})(F_{d3} - F_{p3})$ , where  $F_p$  and  $F_d$  are

Table 2. Triplets arranged according to Hauptman's equation (H3.12)

For each interval  $(A_i - A_{i+1})$ , the number of triplets (NR) with  $A_i < A_0 < A_{i+1}$  is given; % ( $\times 100$ ) is the percentage of triplets having cosine sign equal to the predicted one;  $\langle |\Phi_3| \rangle$  and  $\langle |\Phi_3 - \pi| \rangle$  may be considered as estimates of the average phase errors for positive and negative estimated triplets respectively (in parentheses the expected absolute value of the phase error for the  $A_0$  interval considered).

| Intervals for $A_0$<br>( $A_i - A_{i+1}$ ) | Positive estimated triplets |      |                            | Negative estimated triplets |      |                                  |
|--|-----------------------------|------|----------------------------|-----------------------------|------|----------------------------------|
|  | NR                          | %    | $\langle  \Phi_3  \rangle$ | NR                          | %    | $\langle  \Phi_3 - \pi  \rangle$ |
| < 0.8                                      | 353                         | 72.8 | 67                         | 180                         | 27.2 | 106                              |
| 0.8- 1.2                                   | 401                         | 79.3 | 58 (63)                    | 146                         | 36.3 | 100 (63)                         |
| 1.2- 1.6                                   | 494                         | 76.7 | 58 (55)                    | 149                         | 28.9 | 102 (55)                         |
| 1.6- 2.0                                   | 818                         | 78.6 | 55 (49)                    | 113                         | 44.2 | 91 (49)                          |
| 2.0- 2.4                                   | 1032                        | 81.1 | 52 (43)                    | 97                          | 52.6 | 86 (43)                          |
| 2.4- 3.0                                   | 1686                        | 84.9 | 48 (38)                    | 123                         | 56.1 | 86 (38)                          |
| 3.0- 3.6                                   | 1741                        | 87.6 | 45 (34)                    | 112                         | 58.9 | 82 (34)                          |
| 3.6- 4.2                                   | 1526                        | 91.1 | 40 (31)                    | 97                          | 64.9 | 77 (31)                          |
| 4.2- 4.8                                   | 1302                        | 90.8 | 40 (29)                    | 58                          | 74.1 | 71 (29)                          |
| 4.8- 5.5                                   | 1244                        | 93.3 | 35 (27)                    | 76                          | 71.1 | 68 (27)                          |
| 5.5- 6.5                                   | 1392                        | 91.7 | 39 (24)                    | 77                          | 64.9 | 75 (24)                          |
| 6.5- 9.0                                   | 2345                        | 93.2 | 36 (21)                    | 125                         | 80.0 | 58 (21)                          |
| 9.0-15.0                                   | 2469                        | 95.9 | 34 (17)                    | 120                         | 90.8 | 49 (17)                          |
| 15.0-20.0                                  | 1249                        | 99.4 | 28 (14)                    | 52                          | 96.2 | 35 (14)                          |

structure-factor moduli for protein and derivative respectively. A related result may be obtained from (6) by partially expressing it in terms of  $F_d$  and  $F_p$ :

$$A = 2[\sigma_3/\sigma_2^{3/2}]_P R_1 R_2 R_3 + 2 \sum_H f_j(\mathbf{h}) f_j(\mathbf{k}) f_j(\mathbf{l}) (F_{d1} - F_{p1}) \times (F_{d2} - F_{p2})(F_{d3} - F_{p3}) / \sum_{H1} \sum_{H2} \sum_{H3},$$

or,

$$A = 2[\sigma_3/\sigma_2^{3/2}]_P R_1 R_2 R_3 + 2[\sigma_3/\sigma_2^{3/2}]_H \Delta_1 \Delta_2 \Delta_3,$$

where

$$\Delta = (F_d - F_p) / (\sum_H)^{1/2}$$

is a pseudo-normalized difference (with respect to the heavy-atom structure).

Thus the signs for a triplet cosine predicted by Karle's rule and by (6) coincide provide the Cochran-type term may be neglected. Furthermore, from a probabilistic point of view, (6) suggests that large reliability values do not depend on  $(F_{d1} - F_{p1}) \times (F_{d2} - F_{p2})(F_{d3} - F_{p3})$  but on the product  $\Delta_1 \Delta_2 \Delta_3$ .

#### 4. Experimental results

We have generated a P1 random structure consisting of 30 carbon atoms and a heavy-atom derivative by adding one Na atom: then the ratio  $\sum_H / \sum'$  ( $\approx 0.10$ ) is not too different from values occurring in real cases. Calculated structure factors were used up to  $[(\sin \theta)/\lambda]_{\max} = 0.4 \text{ \AA}^{-1}$ : 20 000 triplets of type  $\omega_1 = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}+\mathbf{k}}$  were found among the 1500 strongest  $E$  magnitudes ( $E_{\min} = 1.17$ , while  $G$  could take any

Table 3. Triplets arranged according to formula (6)

For each interval  $(A_i - A_{i+1})$ , the number of triplets (NR) having  $A_i < A < A_{i+1}$  is given; % ( $\times 100$ ) is the percentage of triplets having cosine sign equal to the predicted one;  $\langle |\Phi_3| \rangle$  and  $\langle |\Phi_3 - \pi| \rangle$  may be considered as estimates of the average phase errors for positive and negative estimated triplets respectively (in parentheses the expected absolute value of the phase error for the  $A$  interval considered).

| Intervals for $A$<br>( $A_i - A_{i+1}$ ) | Positive estimated triplets |       |                            | Negative estimated triplets |       |                                  |
|--|-----------------------------|-------|----------------------------|-----------------------------|-------|----------------------------------|
|  | NR                          | %     | $\langle  \Phi_3  \rangle$ | NR                          | %     | $\langle  \Phi_3 - \pi  \rangle$ |
| < 0.8                                    | 1385                        | 59.9  | 85                         | 578                         | 71.1  | 68                               |
| 0.8- 1.2                                 | 2012                        | 64.2  | 76 (63)                    | 370                         | 77.0  | 60 (63)                          |
| 1.2- 1.6                                 | 3599                        | 71.2  | 66 (55)                    | 306                         | 84.6  | 53 (55)                          |
| 1.6- 2.0                                 | 5331                        | 75.4  | 60 (49)                    | 243                         | 93.8  | 45 (49)                          |
| 2.0- 2.4                                 | 6038                        | 79.6  | 56 (43)                    | 159                         | 96.9  | 40 (43)                          |
| 2.4- 3.0                                 | 8456                        | 83.0  | 51 (38)                    | 146                         | 99.3  | 35 (38)                          |
| 3.0- 3.6                                 | 7220                        | 86.5  | 46 (34)                    | 85                          | 100.0 | 28 (34)                          |
| 3.6- 4.2                                 | 5741                        | 89.4  | 43 (31)                    | 46                          | 100.0 | 23 (31)                          |
| 4.2- 4.8                                 | 4617                        | 91.9  | 40 (29)                    | 13                          | 100.0 | 15 (29)                          |
| 4.8- 5.5                                 | 4098                        | 94.7  | 36 (27)                    | 2                           | 100.0 | 25 (27)                          |
| 5.5- 6.5                                 | 4029                        | 96.0  | 35 (24)                    | 0                           |       |                                  |
| 6.6- 9.0                                 | 4013                        | 98.6  | 29 (21)                    | 0                           |       |                                  |
| 9.0-15.0                                 | 624                         | 100.0 | 23 (17)                    | 0                           |       |                                  |
| 15.0-20.0                                | 3                           | 100.0 | 19 (14)                    | 0                           |       |                                  |

value), and triplet phase cosines were estimated according to (H3.12) and our formula (6). Tables 2 and 3 show the outcome. It may be observed that:

(1) Triplet reliabilities are rather overestimated by Hauptman's approach (see also Fortier, Weeks & Hauptman, 1984). According to Table 2, 1249 positive estimated triplets occur with  $A > 20$ : their average error  $\langle |\Phi_3| \rangle$  is rather large ( $\langle |\Phi_3| \rangle \sim 28^\circ$ ). The situation is worse for negative estimated triplets: 52 triplets occur with  $|A| > 15$  and  $\langle |\Phi_3 - \pi| \rangle \approx 35^\circ$ .

Overestimation of triplet reliability is reduced if (6) is used: indeed only three positive estimated triplets occur with  $A > 20$ , to which an average error  $\langle |\Phi_3| \rangle = 19^\circ$  corresponds: furthermore, no negative estimated triplets occur with  $|A| > 5.5$ . Such an effect has to be expected because of the non-negligible effect of Cochran's contribution.

(2) Estimation of negative triplets according to (H3.12) is rather poor. The majority of the triplets which are estimated as negative with  $|A| < 2.0$  is actually positive. The larger efficiency of (6) in picking up negative triplets is clearly shown in Table 3.

#### 5. Concluding remarks

The basic ideas proposed by Hauptman for the estimation of triplet invariants given six magnitudes have been further developed. The new probabilistic formulas seem statistically more robust than previous results, and are expected to be of interest in macromolecular crystallography.

ZCD is indebted to CNR (Consiglio Nazionale delle Ricerche) and to ICTP (International Centre

for Theoretical Physics, Programme for Training and Research in Italian Laboratories, Trieste, Italy) for financial support during a stay at the University of Bari.

#### References

- CASCARANO, G., GIACOVAZZO, C. & LUIĆ, M. (1985). *Acta Cryst.* **A41**, 544-551.  
 CASCARANO, G., GIACOVAZZO, C. & LUIĆ, M. (1987). *Acta Cryst.* **A43**, 14-22.

- COCHRAN, W. (1955). *Acta Cryst.* **8**, 433-438.  
 FORTIER, S., WEEKS, C. M. & HAUPTMAN, H. (1984). *Acta Cryst.* **A40**, 544-548.  
 GIACOVAZZO, C. (1977). *Acta Cryst.* **A33**, 50-54.  
 GIACOVAZZO, C. (1980). *Direct Methods in Crystallography*. London: Academic Press.  
 GIACOVAZZO, C. (1983). *Acta Cryst.* **A39**, 685-692.  
 HAUPTMAN, H. (1982). *Acta Cryst.* **A38**, 289-294.  
 HAUPTMAN, H., POTTER, S. & WEEKS, C. M. (1982). *Acta Cryst.* **A38**, 294-300.  
 KARLE, J. (1983). *Acta Cryst.* **A39**, 800-805.

*Acta Cryst.* (1988). **A44**, 51-61

## Multiple Scattering Theory for Fast Electrons in Single Crystals and Kikuchi Patterns

BY S. L. DUDAREV AND M. I. RYAZANOV

*Moscow Engineering Physics Institute, Moscow 115409, USSR*

(Received 8 May 1987; accepted 11 August 1987)

### Abstract

By solving the kinetic equation for the density matrix of fast electrons the intensity distribution of the diffuse background and the Kikuchi lines and bands is found. It is shown that the Kikuchi pattern contrast depends on the angle of deflection of the scattered particles and the crystal thickness. For a thin crystal the expressions obtained coincide with the usual results found in the single inelastic scattering approximation. The theory takes into account both the dynamical diffraction and multiple inelastic scattering of electrons and gives a simple interpretation of a variety of contrast effects observed in thick single crystals.

### 1. Introduction

As is well known the diffraction of inelastically scattered electrons in crystals leads to Kikuchi line and band formation (Tomas & Goringe, 1979; Reimer, 1984). Kainuma (1955), Fujimoto & Kainuma (1963), Okamoto, Ichinokawa & Ohtsuki (1971) and Ohtsuki (1983) calculated the intensity distribution in Kikuchi patterns using the single inelastic scattering approximation. This approximation is only valid for very thin crystals (the thickness should be smaller than the mean free path with respect to inelastic collisions). In the case of a thick crystal one needs to take into account multiple inelastic scattering. As was pointed out by Hall (1970), Ishida (1970, 1971) and Chukhovskii, Alexanjan & Pinsker (1973), the contrast reversal of Kikuchi bands can be explained by means of absorption of inelastically scattered electrons. This procedure, however, does not conserve the total prob-

ability and leads to an exponential decrease of the scattered electron density (Thomas & Humphreys, 1970; Serneels, Van Roost & Knuyt, 1982). This result is at variance with experimental observation (Uyeda & Nonoyama, 1967, 1968) of the Kikuchi pattern in thick crystals (where the thickness is larger than the mean free path with respect to inelastic collisions). The application of an iterative method to the problem of electron multiple scattering has been reported by Høier (1973). This approach leads to a result in the form of an infinite series with many difficulties for quantitative evaluation.

The most consistent treatment of this problem is the use of the quantum kinetic equation for a single-particle density matrix (Blum, 1981). This formulation of multiple scattering theory was first used by Migdal (1955) for the case of random space-distribution scatterers. Kagan & Kononets (1970, 1973, 1974) in their theory of proton channelling generalized this method for the case of a crystalline medium. Bird & Buxton (1980) discussed the application of similar equations to electron diffraction.

Most of the known analytical solutions of the quantum kinetic equation in single crystals were obtained in the diffusive approximation (Kagan & Kononets, 1973). This approximation does not work for 10-1000 keV electrons, *i.e.* in the energy range used for observations by transmission electron microscopy.

For the two-beam case the quantum kinetic equation was solved by Dudarev & Ryazanov (1985) without using the diffusive approximation.

Nevertheless, the two-beam approximation fails in many cases for fast electrons, especially if they move along high-symmetry directions. For this reason, to