References

ABAD-ZAPATERO, C., ABDEL-MEGUID, S. F., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I. R., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1980). *Nature* (*London*), **286**, 33–39.

ARGOS, P., FORD, G. C. & ROSSMANN, M. G. (1975). *Acta Cryst.* A**31**, 499–566.

BANTZ, D. & ZWICK, M. (1974). *Acta Cryst.* A**30**, 257–260.

BARRETT, A. N. & ZWICK, M. (1971). *Acta Cryst.* A**27**, 6–11.

BIESECKER, G., HARRIS, J. I., THIERRY, J. C., WALKER, J. E. & WONACOTT, A. J. (1977). *Nature* (*London*), **266**, 328–333.

BLOOMER, A. C., CHAMPNESS, J. N., BRICOGNE, G., STADEN, R. & KLUG, A. (1978). *Nature* (*London*), **276**, 362–368.

BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.

COLLINS, D. M., BRICE, M. D., LA COUR, T. F. M. & LEGG, M. J. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, K. HUML and B. SEDLACEK, pp. 330–335. Copenhagen: Munksgaard.

COLLINS, D. M., COTTON, F. A., HAZEN, E. E. JR, MEYER, E. F. JR & MORIMOTO, C. N. (1975). *Science*, **190**, 1047–1052.

FLETTERICK, R. J. & STEITZ, T. A. (1976). *Acta Cryst.* A**32**, 125–132.

HARRISON, S. C., OLSON, A. J., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). *Nature* (*London*), **276**, 368–373.

HENDRICKSON, W. A., KLIPPENSTEIN, G. L. & WARD, K. B. (1975). *Proc. Natl Acad. Sci. USA*, **72**, 2160–2164.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136–143.

HOPPE, W. & GASSMANN, J. (1968). *Acta Cryst.* B**24**, 97–107.

JOHNSON, C. D., ADOLPH, K., ROSA, J. J., HALL, M. D. & SIGLER, P. B. (1970). *Nature* (*London*), **226**, 1246–1247.

PODJARNY, A. D., SCHEVITZ, R. W., SUSSMAN, J. & SIGLER, P. B. (1981). In preparation.

PODJARNY, A. D. & SUSSMAN, J. (1981). In preparation.

SCHEVITZ, R. W., PODJARNY, A. D., KRISHNAMACHARI, N., HUGHES, J. J. & SIGLER, P. B. (1979). *Nature* (*London*), **278**, 188–190.

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

SPRINZL, M., GRÜTER, F. & GAUSS, D. H. (1978). *Nucleic Acids Res.* Special Supplement r15.

SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S.-H. (1977). *Acta Cryst.* A**33**, 800–804.

WARD, K. B., HENDRICKSON, W. A. & KLIPPENSTEIN, G. L. (1975). *Nature* (*London*), **257**, 818–821.

ZWICK, M., BANTZ, D. & HUGHES, J. (1976). *Ultramicroscopy*, **1**, 275–277.

Acta Cryst. (1981). A**37**, 677–684

An Application of the Theory of Representations to the Estimation of the One-Phase Seminvariants of First Rank

BY M. C. BURLA AND A. NUNZI

Istituto di Mineralogia dell'Università, 06100 *Perugia, Italy*

C. GIACOVAZZO

Istituto di Mineralogia, Università, 70121 *Bari, Italy*

AND G. POLIDORI

Centro Calcolo Elettronico dell'Università, 06100 *Perugia, Italy*

(*Received* 20 *March* 1980; *accepted* 3 *March* 1981)

Abstract

A procedure has been devised which uses both the estimates of the one-phase seminvariants of first rank *via* their second representations and the estimates of a special class of two-phase seminvariants *via* their first representations in order to obtain accurate estimates of the one-phase seminvariants.

1. Introduction

Probabilistic theories for the estimation of the one-phase seminvariants of first rank were supplied by various authors (*e.g.* Hauptman & Karle, 1953; Cochran & Woolfson, 1954; Naya, Nitta & Oda, 1964; Weeks & Hauptman, 1970).

The *MULTAN* procedure (Germain, Main &

0567-7394/81/050677-08$01.00

Woolfson, 1971) applies the $\sum_1$ formula to find the probable signs of centric one-phase seminvariants: each sign determination can be incorporated into the starting set in order to make the phase determination process more efficient. Owing to the frequency of failures, a cautious use of one-phase seminvariants is suggested. In particular: (i) an appropriate weighting scheme is associated with $\sum_1$ relationships in such a way that, when one is incorrect, the tangent formula refinement can handle the situation and lead to a correct set of phases; (ii) high sign probability values are required for the default use of the $\sum_1$ relationships.

These problems suggested to Karle (1970) and to Overbeek & Schenk (1976) that the passive use of one-phase seminvariants is the most convenient way of exploiting their phase information. In particular, Karle deemed that $\sum_1$ relationships should be used at the end of a symbolic addition procedure in order to decide among the possible alternatives for a symbol. Overbeek & Schenk (1976) formulated a criterion based on $\sum_1$ relationships which can be used to select the correct $\sum_2$ solution in multisolution procedures. Joint probability distributions involving sets of structure factors larger than those exploited by the $\sum_1$ relationships were studied by various authors in order to obtain more accurate estimates of the one-phase seminvariants. For example, $P(E_{2\mathbf{h}}, E_{\mathbf{k}}, E_{\mathbf{h}+\mathbf{k}})$ was studied by Hauptman & Karle (1953) and Cochran & Woolfson (1955); $P(E_{2\mathbf{h}}, E_{\mathbf{h}}, E_{\mathbf{k}}, E_{\mathbf{h}+\mathbf{k}})$ by Cochran (1954) and by Hauptman & Karle (1957); $P(E_{2\mathbf{h}}, E_{\mathbf{h}}, E_{\mathbf{k}}, E_{\mathbf{h}+\mathbf{k}}, E_{2\mathbf{h}+\mathbf{k}})$ and $P(E_{2\mathbf{h}}, E_{\mathbf{h}}, E_{\mathbf{k}}, E_{\mathbf{h}-\mathbf{k}}, E_{2\mathbf{h}-\mathbf{k}}, E_{\mathbf{h}+\mathbf{k}})$ by Giacovazzo (1976, 1975).

More recently, the theory of representations (Giacovazzo, 1977) has given new insights into probabilistic methods for obtaining accurate estimates of one-phase seminvariants of first rank. In accordance with this theory a one-phase seminvariant of first rank $\varphi_{\mathbf{H}}$ may be *represented* (*via* its first representations) by means of triplet and quintet invariants, the value of which differ from $\varphi_{\mathbf{H}}$ by constants which arise because of translational symmetry. In particular:
(i) The triplet invariants

$$\psi_1 = \varphi_{\mathbf{H}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}\mathbf{R}_n},$$

where the vectors $\mathbf{h}$ and the matrices $\mathbf{R}_n$ can be found by solving the equation

$$\mathbf{H} = \mathbf{h}(\mathbf{I} - \mathbf{R}_n). \qquad (1)$$

The solution of (1) may be obtained by application of the generalized inverse matrices as described by Giacovazzo (1980*a*). The collection $\{\psi\}_1$ of the triplets ψ_1 constitutes the first representation of $\varphi_{\mathbf{H}}$. The estimate of $\varphi_{\mathbf{H}}$ *via* its first representation coincides with that provided by the $\sum_1$ relationships.
(ii) The quintet invariants

$$\psi_2 = \varphi_{\mathbf{H}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}\mathbf{R}_n} - \varphi_{\mathbf{k}\mathbf{R}_j} + \varphi_{\mathbf{k}\mathbf{R}_j}, \quad j = 1, \ldots, r,$$

where $\mathbf{h}$ varies over the set of vectors defined in $\{\psi\}_1$, $\mathbf{R}_n$ over the subsets of matrices which satisfy (1) and $\mathbf{R}_j$ varies over the subset of the r matrices not related by the centre of symmetry. That is, $r = m/2$ or $r = m$ (m is the order of the space group) according to whether the space group is centrosymmetric or not. $\mathbf{k}$ is a free vector which sweeps over the asymmetric region of the reciprocal space.

Probabilistic formulae estimating $\varphi_{\mathbf{H}}$ *via* its first and second representations were secured by Giacovazzo (1978). Burla, Nunzi, Polidori, Busetta & Giacovazzo (1980) tested these formulae on various structures of different complexity, covering the symmetry classes $\bar{1}$, 2, $2/m$, 222 and $mm2$. The estimates of one-phase seminvariants *via* the second representation were found to be considerably more accurate than the corresponding estimates *via* the first representation. In particular, the number of seminvariants estimated with high probability value was always larger than for the $\sum_1$ estimates, and, as a rule, they are correctly estimated. These results suggest again the active use of one-phase seminvariants in the direct procedures.

We are thoroughly conscious of the great importance of knowing a number of seminvariant phases at the beginning of a multisolution procedure. Then: (i) the phase determination process is faster and more accurate; (ii) the necessity of introducing a large number of variable phases in the starting set is reduced. We have therefore devoted our efforts to improving further on the estimates of the one-phase seminvariants. In this paper we describe a new procedure which uses some recent advances in the theory of representations in order to obtain, in all the space groups up to orthorhombic, some very accurate estimates of the one-phase seminvariants *via* the estimation of special two-phase seminvariants.

In § 2 we show how the concept of *generalized first representation* recently introduced by Giacovazzo (1980*a*) may be applied with some modifications to the estimation of the two-phase seminvariants of first rank. In § 3 it is shown how the estimates of the two-phase seminvariants can be improved by probabilistic relations involving three or four two-phase seminvariants which form tripoles or quadrupoles. In §§ 4 and 5 an automatic procedure is described which leads to the accurate estimate of the one-phase seminvariants. In § 6 the practical applications of the procedure to some test structures are described and conclusions are drawn.

2. An application of the concept of generalized phasing shell

We recall the concept of a generalized first phasing shell only with respect to the two-phase seminvariants of first rank. Let $\Phi = \varphi_{\mathbf{u}} + \varphi_{\mathbf{v}}$ be a two-phase seminvariant in which $\varphi_{\mathbf{u}}$ and $\varphi_{\mathbf{v}}$ are one-phase

seminvariants. If $\{B\}_1$, $\{B'\}_1$, $\{B''\}_1$ are the first phasing shells of Φ, φ_u and φ_v respectively, then the set theoretical union

$$\{B\}_1^g = \{B\}_1 \cup \{B'\}_1 \cup \{B''\}_1 \qquad (2)$$

is said to be the generalized first phasing shell of Φ. It was shown by Giacovazzo (1980*a*) that the estimate of Φ *via* $\{B\}_1^g$ is expected to be more accurate in the statistical sense than *via* $\{B\}_1$. When Φ is estimated *via* the set $\{B\}_1$, then we say that Φ is estimated *via* its first representation. A probabilistic theory which estimates Φ *via* $\{B\}_1$ was secured by Giacovazzo (1979): the practical aspects of this theory together with extensive applications were described by Giacovazzo, Spagna, Vicković & Viterbo (1979). When Φ is estimated *via* $\{B\}_1^g$ then we say that it is estimated *via* its *generalized first representation.*

We give two practical examples in order to clarify the situation. In $P\bar{1}$ let $\Phi = \varphi_{842} + \varphi_{266}$. Then the first representation of Φ is the collection of the two quartets

$$\psi_1 = \varphi_{842} + \varphi_{266} - \varphi_{554} - \varphi_{554},$$
$$\psi_1' = \varphi_{842} - \varphi_{266} - \varphi_{3\bar{1}\bar{2}} - \varphi_{3\bar{1}\bar{2}}.$$

The first phasing shell $\{B\}_1$ contains all the basis and the cross magnitudes of the two quartets:

$$\{B\}_1 = \{R_{842}, R_{266}, R_{554}, R_{3\bar{1}\bar{2}}, R_{10,10,8}, R_{6\bar{2}\bar{4}}\},$$

where the R's symbolize the $|E|$ magnitudes. Since φ_{842} and φ_{266} are one-phase seminvariants, they depend, in their first representation, on the first phasing shells

$$\{B'\}_1 = \{R_{842}, R_{421}\},$$
$$\{B''\}_1 = \{R_{266}, R_{133}\},$$

respectively. Therefore, according to (2),

$$\{B\}_1^g = \{R_{842}, R_{266}, R_{554}, R_{3\bar{1}\bar{2}}, R_{10,10,8}, R_{6\bar{2}\bar{4}}, R_{421}, R_{133}\}.$$

As a second numerical example, in $P2_1$, let $\Phi = \varphi_{802} + \varphi_{204}$. Then the first representation of Φ is the collection of quartets

$$\psi_1 = \varphi_{802} + \varphi_{204} - \varphi_{5k3} + \varphi_{\bar{5}k\bar{3}},$$
$$\psi_1' = \varphi_{802} + \varphi_{\bar{2}0\bar{4}} - \varphi_{3k\bar{1}} + \varphi_{\bar{3}k\bar{1}},$$

where k is a free index. The first phasing shell $\{B\}_1$ is then

$$\{B\}_1 = \{R_{802}, R_{204}, R_{5k3}, R_{3k\bar{1}}, R_{10,0,6}, R_{60\bar{2}}\}.$$

Since φ_{802} and φ_{204} are one-phase seminvariants, they depend, in their first representations, on the phasing shells

$$\{B'\}_1 = \{R_{802}, R_{4k1}\}$$

and

$$\{B''\}_1 = \{R_{204}, R_{1k2}\},$$

respectively. Therefore, according to (2)

$$\{B\}_1^g = \{R_{802}, R_{204}, R_{5k3}, R_{3k\bar{1}}, R_{10,0,6}, R_{60\bar{2}}, R_{4k1}, R_{1k2}\}.$$

So far no probabilistic formula estimating Φ *via* its generalized first representation is available. Thus we have used an approximate mathematical approach based on the reasonable assumption that the estimates of Φ, φ_u and φ_v *via* their mere first representations are independent of one another.

Let P^+, P_u^+ and P_v^+ be the sign probabilities for Φ, φ_u and φ_v as given by their first representations. From P_u^+ and P_v^+ the following sign probability arises for $\varphi_u + \varphi_v$:

$$P_0^+ = P_u^+ P_v^+ + (1 - P_u^+)(1 - P_v^+). \qquad (3)$$

P_0^+ is then composed with P^+ in order to give the overall sign probability (Woolfson, 1961).

$$P_\Phi^+ = \frac{P_0^+ P^+}{P_0^+ P^+ + (1 - P_0^+)(1 - P^+)}. \qquad (4)$$

The value of (4) may be considered a reasonable estimation of Φ *via* its generalized first representation.

We have tested the efficiency of (4), with respect to P^+ as calculated by Giacovazzo, Spagna, Vicković & Viterbo (1979), on seven known structures of different complexity. Table 1 shows the reference and the most relevant features of the test structures, the results of which are discussed in this paper. Our results proved that P_Φ^+ is always more reliable than P^+. That convinced us that better estimates of Φ can be obtained by a slight modification of the procedure.

We now introduced in (3) not the sign probabilities calculated *via* the first representations of φ_u and φ_v but those arising from their second representations. Since the latter proved to be more reliable than the former (Burla *et al.*, 1980), better P_0^+ values and, consequently, more accurate values of P_Φ^+ can be expected. We give in Tables 2 and 3 the outcome for TOXE ($N = 104$, $P2_12_12_1$) and AZET ($N = 192$, $Pca2_1$). In the tables the phase seminvariants are estimated *via* their

Table 1. *Abbreviations, references, space groups and formulae for the seven test structures*

	References	Space group	Formula	Z
RIBO	James & Stevens (1977)	$P2_1$	$C_{13}H_{18}O_9$	4
HEPTA	Beurskens, Beurskens & van den Hark (1976)	$P2_1$	$C_{30}H_{18}$	4
METHOX	Hanson & Nordman (1975)	$P2_1/c$	$C_{18}H_{22}O_2$	4
TOLY	Brufani, Cellai, Cerrini, Fedeli & Vaciago (1978)	$P2_12_12_1$	$C_{37}H_{43}NO_{13}$	4
TOXE	Cerrini, Fedeli, Gavuzzo & Mazza (1975)	$P2_12_12_1$	$C_{21}H_{35}O_5$	4
KARLE	Karle, Karle & Estlin (1967)	$P2_12_12_1$	$C_{12}H_{13}NO_4$	4
AZET	Colens, Declercq, Germain, Putzeys & Van Meerssche (1974)	$Pca2_1$	$C_{21}H_{16}ClNO$	8

Table 2. *TOXE: Number of two-phase seminvariants (n) and percentage of correct relations (%) calculated via their first representation (columns A), equation (4) (columns B) and equation (10) (columns C)*

The two-phase seminvariants are constructed starting from the one-phase seminvariants with $|E| \geq 1{\cdot}55$.

	A		*B*		*C*	
ARG	%	*n*	%	*n*	%	*n*
0·3	84·2	38	93·4	61	97·4	117
0·6	100	16	100	46	100	106
0·8	100	10	100	42	100	102
1·0	100	6	100	36	100	93
1·2	100	5	100	30	100	79
1·4	100	1	100	24	100	78
1·6	100	1	100	18	100	77
2·0	100	1	100	10	100	77
2·5			100	4	100	76
3·0			100	1	100	76

Table 3. *AZET: Number of two-phase seminvariants (n) and percentage of correct relations (%) calculated via their first representation (columns A), equation (4) (columns B) and equation (10) (columns C)*

The two-phase seminvariants are constructed starting from the one-phase seminvariants with $|E| \geq 1{\cdot}30$.

	A		*B*		*C*	
ARG	%	*n*	%	*n*	%	*n*
0·3	83·9	31	90·7	43	97·5	79
0·6	100	12	100	30	100	70
0·8	100	9	100	26	100	62
1·0	100	4	100	22	100	53
1·2	100	4	100	18	100	49
1·4	100	4	100	17	100	48
1·6	100	4	100	16	100	48
2·0	100	4	100	11	100	47
2·5	100	3	100	8	100	47
3·0	100	3	100	8	100	47

first representations by means of equation (36) of Giacovazzo *et al.* (1979):

$$P^+ \simeq 0{\cdot}5 + 0{\cdot}5 \tanh \text{ARG}.$$

In order easily to compare the accuracy of P^+ with that of P_Φ^+ we also calculated for any value P_Φ^+ the corresponding value of ARG. Therefore Tables 2 and 3 show an analysis of the number of corrected sign indications for the relations having the value of |ARG| above the limits shown in the first columns. The higher reliability of (4) is clearly shown by the tables.

3. Improved estimates of Φ by means of multipoles

A basis of representations theory (Giacovazzo, 1977) is the phase interrelationship principle according to which if $\Phi_1, \Phi_2, \ldots, \Phi_n$ are n seminvariants for which $\sum_{i=1}^n \Phi_i = 0$ and if $\{R\}_j$, $j = 1, \ldots, n$, are n subsets of magnitudes for which the following conditional distributions may be calculated

$$P(\Phi_1 | \{R\}_1), \tag{5.1}$$

$$P(\Phi_2 | \{R\}_2), \tag{5.2}$$

$$\vdots \qquad\qquad \vdots$$

$$P(\Phi_n | \{R\}_n), \tag{5.n}$$

then the estimate of Φ *via*

$$\{R\}_U = \{R\}_1 \cup \{R\}_2 \cup \ldots \cup \{R\}_n \tag{6}$$

is in principle more accurate in the statistical sense than that given by (5.1).

We apply this principle to tripoles of type

$$\Phi \equiv \Phi_1 = \varphi_u + \varphi_v \tag{7.1}$$

$$\Phi_2 = -\varphi_u + \varphi_k \tag{7.2}$$

$$\Phi_3 = -\varphi_k - \varphi_v. \tag{7.3}$$

In our context Φ_1, Φ_2, Φ_3 are estimated according to § 2. In other words, $\{R\}_1$ is the subset of magnitudes contained either in the first phasing shell of Φ_1 or in the second phasing shells of φ_u and φ_v; $\{R\}_2$ is the subset, *etc.*

So far no probabilistic formula estimating Φ *via* $\{R\}_u$ is available. However, if the estimates of Φ, Φ_2 and Φ_3 are assumed independent of one another, then the following mathematical approach may be used. Let P_Φ^+, $P_{\Phi_2}^+$, $P_{\Phi_3}^+$ be the sign probabilities for Φ, Φ_2, Φ_3 calculated according to § 2. Then the additional sign probability

$$P_{\Phi,\mathbf{k}}^+ = P_{\Phi_2}^+ P_{\Phi_3}^+ + (1 - P_{\Phi_2}^+)(1 - P_{\Phi_3}^+)$$

arises for Φ from the mere estimates of Φ_2 and Φ_3. Since $\mathbf{k}$ is a free vector in (7), more tripoles may be found for a given Φ. The various $P_{\Phi,\mathbf{k}_j}^+$ can be multiplied together to give the overall probability (Woolfson, 1961)

$$P'^+_\Phi = \left(1 + \frac{\prod_j P^-_{\Phi,\mathbf{k}_j}}{\prod_j P^+_{\Phi,\mathbf{k}_j}}\right)^{-1}.$$

In its turn P'^+_Φ may be composed with P_Φ^+ by means of

$$P^+_{\Phi,t} = \frac{P'^+_\Phi P^+_\Phi}{P'^+_\Phi P^+_\Phi + (1 - P'^+_\Phi)(1 - P^+_\Phi)}.$$

In addition to the tripoles, the quadrupoles

$$\Phi \equiv \Phi_1 = \varphi_u + \varphi_v \tag{8.1}$$

$$\Phi_2 = -\varphi_u + \varphi_l \tag{8.2}$$

$$\Phi_3 = -\varphi_l + \varphi_p \tag{8.3}$$

$$\Phi_4 = -\varphi_p - \varphi_v \tag{8.4}$$

can also be exploited in order to improve the estimate

of Φ. Let us assume that the estimates of $\Phi \equiv \Phi_1$, Φ_2, Φ_3, Φ_4 are independent of one another. We denote by $P^{\pm}_{\Phi}$, $P^{\pm}_{\Phi_2}$, $P^{\pm}_{\Phi_3}$, $P^{\pm}_{\Phi_4}$ the sign probabilities for Φ, Φ_2, Φ_3, Φ_4 calculated according to § 2. The mere knowledge of $P^{\pm}_{\Phi_2}$, $P^{\pm}_{\Phi_3}$, $P^{\pm}_{\Phi_4}$ gives for Φ the additional sign probability

$$P^{\pm}_{\Phi,\mathbf{k},\mathbf{l}} = 4\,P^{\pm}_{\Phi_2}\,P^{\pm}_{\Phi_3}\,P^{\pm}_{\Phi_4} - 2\,P^{\pm}_{\Phi_2}\,P^{\pm}_{\Phi_3} - 2\,P^{\pm}_{\Phi_2}\,P^{\pm}_{\Phi_4} - 2\,P^{\pm}_{\Phi_3}\,P^{+}_{\Phi_4} + P^{+}_{\Phi_2} + P^{+}_{\Phi_3} + P^{+}_{\Phi_4}.$$

Since $\mathbf{k}$ and $\mathbf{l}$ are free vectors in (8), more quadrupoles may be found for a given Φ. The various $P^{\pm}_{\Phi,\mathbf{k}_i,\mathbf{l}_j}$ may be multiplied together to give the overall probability

$$\left(1 + \frac{\prod_{i,j} P^{-}_{\Phi,\mathbf{k}_i,\mathbf{l}_j}}{\prod_{i,j} P^{+}_{\Phi,\mathbf{k}_i,\mathbf{l}_j}}\right)^{-1}.$$

In our approach we also assume any $P^{\pm}_{\Phi,\mathbf{k}_j}$ to be independent of any $P^{\pm}_{\Phi,\mathbf{k}_i,\mathbf{l}_j}$. Therefore the formula

$$P'^{+}_{\Phi} = \left(1 + \frac{\prod_{r} P^{-}_{r}}{\prod_{r} P^{+}_{r}}\right)^{-1} \tag{9}$$

may be used, where $P^{\pm}_{r}$ symbolizes $P^{\pm}_{\Phi,\mathbf{k}_j}$ and $P^{\pm}_{\Phi,\mathbf{k}_i,\mathbf{l}_j}$.

In its turn P'^{+}_{Φ} is composed with $P^{\pm}_{\Phi}$ by means of

$$P^{\pm}_{\Phi,t} = \frac{P'^{+}_{\Phi}\,P^{\pm}_{\Phi}}{P'^{+}_{\Phi}\,P^{\pm}_{\Phi} + (1 - P'^{+}_{\Phi})(1 - P^{\pm}_{\Phi})}. \tag{10}$$

Equation (10) is really more reliable than (4). We give in Tables 2 and 3 the number and the percentage of the correctly estimated two-phase seminvariants according to (10) (columns *C*), compared with the corresponding figures obtained *via* P^{+} and $P^{\pm}_{\Phi}$.

4. The estimation of the one-phase seminvariants

We are now able to obtain, from the improved estimates of the two-phase seminvariants, improved estimates of the one-phase seminvariants. In fact, if $P^{\pm}_{\Phi,t}$ [as given by (10)] and $P^{+}_{\mathbf{v}}$ (as calculated *via* the second representation of $\varphi_{\mathbf{v}}$) are known, then the additional sign probability

$$P'^{+}_{\mathbf{u}} = P^{\pm}_{\Phi,t}\,P^{+}_{\mathbf{v}} + (1 - P^{\pm}_{\Phi,t})(1 - P^{\pm}_{\Phi,t}) \tag{11}$$

arises for $\Phi_{\mathbf{u}}$. Since any $\Phi_{\mathbf{u}}$ may enter in several two-phase seminvariants, several probabilities such as (11) may be calculated which may be combined by

$$P^{0+}_{\mathbf{u}} = \left(1 + \frac{\prod_{j} P'^{-}_{\mathbf{u},j}}{\prod_{j} P^{+}_{\mathbf{u},j}}\right)^{-1}. \tag{12}$$

$P'^{\pm}_{\mathbf{u},j}$ are the sign probabilities [according to (11)] arising from the jth two-phase seminvariant containing $\Phi_{\mathbf{u}}$.

5. The procedure

The basic steps of our procedure are:

(*a*) Calculation of the one-phase seminvariants *via* their second representation for the reflections with the largest E values. $P^{+}_{\mathbf{u}}$ is the outcome of this step for any reflection $E_{\mathbf{u}}$.

(*b*) Calculation of the two-phase seminvariants (constituted by one-phase seminvariants) *via* their first representation. P^{+} (see § 2) is the outcome of this step for $\Phi = \varphi_{\mathbf{u}} + \varphi_{\mathbf{v}}$.

(*c*) Calculation of $P^{\pm}_{\Phi}$ according to (4). A subset is defined (subset $\{M_2\}$) which contains the most reliable two-phase seminvariants (in our tests those with $|P^{\pm}_{\Phi} - 0{\cdot}5| \geq 0{\cdot}35$, but a variable threshold may in general be used).

(*d*) $P^{\pm}_{\Phi,t}$ [according to (10)] is calculated. In this step Φ and all the other two-phase seminvariants involved in the multipoles belong to $\{M_2\}$. This restriction aims to avoid the probability of obtaining, for any $\Phi \in \{M_2\}$ a wrong probability caused by wrong estimates of some unreliable seminvariants.

(*e*) $P^{\pm}_{\Phi,t}$ is calculated for any Φ. The sign indication arising from any multipole is used in (9) only if all the seminvariants are sufficiently reliable. In our tests no seminvariant is used for which $|P^{\pm}_{\Phi} - 0{\cdot}5| < 0{\cdot}30$.

(*f*) A subset $\{M_1\}$ is defined which contains the most reliable one-phase seminvariants (*i.e.* $|P^{+}_{\mathbf{u}} - 0{\cdot}5| \geq 0{\cdot}35$). Then $P^{0+}_{\mathbf{u}}$ [according to (12)] is calculated for any $\varphi_{\mathbf{u}} \in \{M_1\}$ *via* the most reliable two-phase seminvariants $\varphi_{\mathbf{u}} + \varphi_{\mathbf{v}}$ for which $\varphi_{\mathbf{v}} \in \{M_1\}$. These restrictions aim to avoid a wrong estimation of $\varphi_{\mathbf{u}}$ caused by previous wrong estimates of the various $\varphi_{\mathbf{v}}$ and $\varphi_{\mathbf{u}} + \varphi_{\mathbf{v}}$.

(*g*) The other one-phase seminvariants $\varphi_{\mathbf{u}} \notin \{M_1\}$ are estimated according to (12), but always the $\varphi_{\mathbf{v}}$'s $\in \{M_1\}$. This step is cyclic because the one-phase seminvariants can belong to $\{M_1\}$ or not, according to their current probability value. The procedure stops when no new one-phase seminvariant goes in $\{M_1\}$. In this step all the available two-phase seminvariants are used to assign the new $\varphi_{\mathbf{u}}$'s.

6. Practical applications and conclusions

The estimates of the one-phase seminvariants with $|E|$ larger than a given threshold are given in Tables 4–10 for each test structure. For each seminvariant the sign probabilities are shown calculated *via*: (*a*) the $\sum_1$ relationships as used by the *MULTAN* system (Germain, Main & Woolfson, 1971); (*b*) the second representation as described by Burla *et al.* (1980); (*c*)

Table 4. *RIBO: indices, $|E|$'s and actual signs of the one-phase seminvariants with $|E| > 1.3$ together with the various probability values*

The asterisks indicate incorrectly evaluated signs, and the character – the unpredicted signs.

h	k	l	E	$\sum_1$ P^+	Second representation P^+	Present procedure P^+
0	0	2	−1·58	0·40	0·26	0·0
4	0	$\bar{2}$	−1·79	0·42	0·28	0·0
10	0	$\bar{2}$	−2·17	0·59*	0·28	0·0
18	0	$\bar{10}$	1·96	–	0·86	1·0
4	0	10	−3·05	0·31	0·13	0·0
14	0	$\bar{6}$	2·29	0·62	0·88	1·0
10	0	$\bar{8}$	1·38	0·49*	0·91	1·0
18	0	$\bar{8}$	−2·43	0·21	0·09	0·0
10	0	$\bar{10}$	−1·39	0·33	0·04	0·0
8	0	$\bar{6}$	1·37	0·57	0·82	1·0
2	0	$\bar{6}$	1·70	0·32*	0·52	0·90
4	0	8	1·50	0·49*	0·39*	0·85
22	0	$\bar{8}$	1·55	0·53	0·37*	0·85
12	0	$\bar{6}$	−2·13	0·58*	0·52*	0·21
6	0	10	1·80	0·60	0·46*	0·78
22	0	$\bar{6}$	−1·37	0·44	0·37	0·23
8	0	$\bar{8}$	1·35	0·56	0·57	0·75
16	0	0	1·42	0·48*	0·33*	0·74
16	0	$\bar{2}$	−1·70	0·36	0·34	0·65*
2	0	0	−2·01	0·43	0·30	0·47

Table 5. *HEPTA: indices, $|E|$'s and actual signs of the one-phase seminvariants with $|E| > 1.3$ together with the various probability values*

The asterisks indicate incorrectly evaluated signs.

h	k	l	E	$\sum_1$ P^+	Second representation P^+	Present procedure P^+
2	0	0	1·42	0·67	0·98	1·0
8	0	0	1·33	0·68	0·97	1·0
10	0	$\bar{6}$	1·36	0·65	0·95	1·0
4	0	0	1·44	0·64	0·89	1·0
8	0	$\bar{4}$	−1·36	0·30	0·12	0·0
16	0	2	−1·61	0·22	0·03	0·0
8	0	$\bar{6}$	−2·62	0·46	0·09	0·0
6	0	2	−1·51	0·23	0·01	0·0
2	0	$\bar{10}$	−1·60	0·69*	0·81*	0·27
14	0	$\bar{6}$	1·80	0·61	0·71	0·57
2	0	8	2·24	0·68	0·77	0·53

the procedure described in the present paper. The lists of seminvariants are in decreasing order of reliability, calculated according to the new procedure.

In Table 4 the outcome for RIBO ($N = 88$, $P2_1$) is shown. The most reliable sign according to the $\sum_1$ formula is that of $E_{18,0,\bar{8}}$, the only one with $|P^+ - 0.5| \geq 0.28$. Seven good signs estimated *via* the second representation (all correct) have $|P^+ - 0.5| \geq 0.28$ and three of them (*i.e.* $E_{10,0,\bar{8}}$, $E_{10,0,\bar{10}}$ and $E_{18,0,\bar{8}}$) are so reliable that they may be used in an active way. The present procedure defines a set of ten signs (all correct) suitable for active use in direct procedures.

In Table 5 the results of our calculations for HEPTA ($N = 120$, $P2_1$) are shown. The $\sum_1$ formula gives probability values which are substantially correct but numerically poor, for ten of the 11 one-phase seminvariants. Our procedure gives eight very reliable sign estimates suitable for active use, and all 11 signs are correct.

For METHOX ($N = 80$, $P2_1/c$) 44 one-phase seminvariants were found with $|E| \geq 1.60$. We only give in Table 6 the outcome for the most reliable 15 seminvariants. It is remarkable that seven seminvariants in the list have unpredictable sign according to the $\sum_1$ relationship. The situation is strongly improved when the second representation is used. The procedure here described defines a subset of four very reliable seminvariants which can be actively used in a direct procedure.

Table 7 shows the 14 estimates of the one-phase seminvariants with $|E| \geq 1.30$ for KARLE ($N = 68$, $P2_12_12$). The ordinary $\sum_1$ formula is unable to predict the signs of five seminvariants, whereas the second representation correctly estimates 12 of the 14 seminvariants. By means of our procedure a subset of seven seminvariants with very high reliability is correctly defined.

The results for TOXE ($N = 104$, $P2_12_12_1$) are shown in Table 8. The $\sum_1$ formula gives correct estimates of 19 of the 21 seminvariants in the table, however the numerical values of the sign probabilities are poor. Our procedure defines a subset of 17 seminvariants with very high reliability, all of them correctly estimated.

Table 6. *METHOX: indices, $|E|$'s and actual signs of the more reliable one-phase seminvariants among the 44 with $|E| > 1.6$*

The asterisks indicate the incorrect signs and the character – indicates the unpredicted signs.

h	k	l	E	$\sum_1$ P^+	Second representation P^+	Present procedure P^+
10	0	2	−2·31	–	0·11	0·01
6	0	8	1·85	0·68	0·88	0·99
6	0	16	−2·12	0·19	0·13	0·01
2	2	10	1·75	0·73	0·93	0·96
0	0	34	2·00	0·76	0·81	0·81
$\bar{2}$	2	18	1·65	–	0·78	0·78
$\bar{10}$	0	4	1·72	0·66	0·78	0·78
6	0	2	−1·95	–	0·26	0·26
6	4	12	2·10	0·71	0·74	0·74
$\bar{2}$	4	4	1·98	–	0·71	0·71
$\bar{10}$	0	2	2·81	0·68	0·70	0·70
6	0	18	2·15	–	0·68	0·68
$\bar{4}$	6	12	−1·75	0·67*	0·64*	0·64*
$\bar{2}$	0	22	1·71	–	0·36*	0·36*
2	6	4	1·66	–	0·63	0·63

Table 7. *KARLE: indices, $|E|$'s and actual signs of the one-phase seminvariants with $|E| > 1{\cdot}3$ together with the various probability values*

The asterisks indicate incorrectly evaluated signs, and the character – the unpredicted signs.

h k l	E	$\sum_1$ P^+	Second representation P^+	Present procedure P^+
4 0 12	−1·53	0·38	0·07	0·0
0 4 12	−1·46	0·39	0·09	0·0
0 4 4	−1·51	0·51*	0·11	0·0
4 2 0	−2·54	0·11	0·01	0·0
0 4 14	1·95	0·75	0·99	1·0
4 0 0	1·81	–	0·91	1·0
0 0 12	−2·42	–	0·44	0·05
4 0 4	−1·87	–	0·39	0·15
0 6 16	1·76	–	0·59	0·64
0 2 8	−2·05	0·55*	0·68*	0·41
0 6 10	−1·35	0·36	0·74*	0·43
0 6 18	−1·82	0·31	0·23	0·46
0 2 10	1·33	–	0·67	0·54
0 4 8	2·22	0·56	0·60	0·46*

Table 8. *TOXE: indices, $|E|$'s and actual signs of the one-phase seminvariants with $|E| > 1{\cdot}55$ together with the various probability values*

The asterisks indicate incorrectly evaluated signs, and the character – the unpredicted signs.

h k l	E	$\sum_1$ P^+	Second representation P^+	Present procedure P^+
4 0 14	−1·62	0·38	0·06	0·0
2 0 0	−2·00	0·21	0·01	0·0
0 8 6	−1·60	0·16	0·01	0·0
6 0 2	2·89	0·91	0·99	1·0
8 0 0	2·14	0·92	0·98	1·0
4 16 0	2·36	0·77	0·96	1·0
4 0 2	−1·86	0·13	0·06	0·0
0 6 2	−1·83	0·52*	0·00	0·0
8 4 0	1·97	0·56	0·93	1·0
6 0 10	1·91	0·79	0·93	1·0
0 0 2	−1·63	0·41	0·08	0·0
8 0 2	−1·99	0·22	0·12	0·0
6 6 0	−2·08	0·44	0·40	0·0
0 6 10	−1·77	–	0·34	0·0
2 10 0	−1·78	0·37	0·71*	0·0
8 6 0	1·64	0·52	0·64	0·99
0 10 6	−1·74	0·49	0·62*	0·08
0 14 4	1·70	0·59	0·57	0·79
10 8 0	1·57	0·45*	0·23*	0·57
4 0 6	−1·87	0·45	0·28	0·56*
10 0 14	1·58	0·51	0·36*	0·55

In Table 9 the outcome for the more reliable 13 one-phase seminvariants among the 22 with $|E| \geq 1{\cdot}55$ is shown for TOLY ($N = 208$, $P2_12_12_1$). The number of incorrect estimates is four for the $\sum_1$ relationship and two for the present method: only one *via* the $\sum_1$ formula and five *via* our procedure have a sufficient reliability for active use.

Table 9. *TOLY: indices, $|E|$'s and actual signs of the more reliable one-phase seminvariants among the 22 with $|E| > 1{\cdot}55$*

The asterisk indicates an incorrect sign.

h k l	E	$\sum_1$ P^+	Second representation P^+	Present procedure P^+
0 0 12	3·38	0·81	0·97	1·0
8 0 12	−1·93	0·18	0·04	0·0
8 0 2	−1·78	0·30	0·11	0·0
0 14 0	2·03	0·79	0·99	1·0
8 0 0	−3·19	0·00	0·00	0·0
0 8 2	−2·76	0·40	0·32	0·24
0 16 16	1·62	0·49*	0·18*	0·68
2 2 0	−1·93	0·51*	0·82*	0·33
10 0 6	1·59	0·65	0·76	0·64
0 8 14	−1·91	0·46	0·37	0·37
0 2 8	−1·93	0·53*	0·58*	0·62*
0 20 6	−1·55	0·64*	0·56*	0·61*
4 8 0	−2·87	0·50	0·64*	0·40

Table 10. *AZET: indices, $|E|$'s and actual signs of the one-phase seminvariants with $|E| > 1{\cdot}3$ together with the various probability values*

The asterisks indicate incorrectly evaluated signs, and the character – the unpredicted signs.

k k l	E	$\sum_1$ P^+	Second representation P^+	Present procedure P^+
8 4 0	2·31	0·72	0·93	1·0
30 0 0	1·55	0·78	0·98	1·0
14 0 0	−1·56	0·27	0·00	0·0
0 8 0	1·38	0·97	1·00	1·0
0 6 0	−2·67	0·00	0·00	0·0
0 4 0	3·46	1·00	1·00	1·0
24 4 0	1·39	0·64	0·91	1·0
4 4 0	1·31	0·62	0·85	1·0
24 6 0	1·61	–	0·43*	1·0
28 4 0	2·20	–	0·60	1·0
8 2 0	−2·78	–	0·33	0·0
12 6 0	−1·45	0·56*	0·41	0·0
12 2 0	−2·36	0·44	0·26	0·0
30 2 0	−1·56	0·40	0·18	0·0
22 6 0	1·71	0·68	0·85	0·46*

The results for AZET ($N = 192$, $Pca2_1$) in Table 10 prove again the usefulness of the method. 14 one-phase seminvariants with $|E| \geq 1{\cdot}30$ are correctly estimated with very high reliability.

The efficiency of the proposed procedure is now well documented. Our calculations show that it is possible to have quite good estimates of some one-phase seminvariants which can be actively used in the process of phase determination. The computer time requirement is quite modest: typical times are 30–40 s per structure by an IBM 370/158.

Besides the merits, Tables 4–10 also show some limits:

(*a*) The present approach postulates the mutual statistical independence of the various phase relationships which concur to define φ_u. This is not strictly true: the consequence is that the accuracy of the sign indication may often be overestimated.

(*b*) The procedure may fail on some occasions. For example, a wrong estimate of φ_u is possible when: (i) it is wrongly estimated *via* its second representation with a very high probability value; luckily that occurs rarely if $|E_u|$ is sufficiently large; (ii) a large percentage of two-phase seminvariants $\varphi_u + \varphi_v$ is wrongly estimated with high probability values. This case is not frequent either.

There are several ways for improving the present situation: *e.g.* (i) to improve the estimates of the one-phase seminvariants by application of the concept of generalized representations (Giacovazzo, 1980*b*,*c*); (ii) to obtain improved estimates of the two-phase seminvariants, *e.g. via* their second representations; (iii) by application of the three-phase seminvariants.

The work in these fields is in progress.

References

BEURSKENS, T., BEURSKENS, G. & VAN DEN HARK, TH. E. M. (1976). *Cryst. Struct. Commun.* **5**, 241–246.

BRUFANI, M., CELLAI, L., CERRINI, S., FEDELI, W. & VACIAGO, A. (1978). *Mol. Pharmacol.* **14**, 693–703.

BURLA, M. C., NUNZI, A., POLIDORI, G., BUSETTA, B. & GIACOVAZZO, C. (1980). *Acta Cryst.* A**36**, 573–578.

CERRINI, S., FEDELI, W., GAVUZZO, E. & MAZZA, F. (1975). *Gazz. Chim. Ital.* **105**, 651–655.

COCHRAN, W. (1954). *Acta Cryst.* **7**, 581–583.

COCHRAN, W. & WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 450–451.

COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1–12.

COLENS, A., DECLERCQ, J. P., GERMAIN, G., PUTZEYS, J. P. & VAN MEERSSCHE, M. (1974). *Cryst. Struct. Commun.* **3**, 119–122.

GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1971). *Acta Cryst.* A**27**, 368–376.

GIACOVAZZO, C. (1975). *Acta Cryst.* A**31**, 602–609.

GIACOVAZZO, C. (1976). *Acta Cryst.* A**32**, 958–967.

GIACOVAZZO, C. (1977). *Acta Cryst.* A**33**, 933–944.

GIACOVAZZO, C. (1978). *Acta Cryst.* A**34**, 562–574.

GIACOVAZZO, C. (1979). *Acta Cryst.* A**35**, 296–305.

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

GIACOVAZZO, C. (1980*b*). *Acta Cryst.* A**36**, 704–711.

GIACOVAZZO, C. (1980*c*). *Acta Cryst.* A**36**, 711–715.

GIACOVAZZO, C., SPAGNA, R., VICKOVIĆ, I. & VITERBO, D. (1979). *Acta Cryst.* A**35**, 401–412.

HANSON, J. C. & NORDMAN, C. E. (1975). *Acta Cryst.* B**31**, 493–501.

HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem. 1. The Centrosymmetric Crystal.* ACA Monograph No. 3. Pittsburg: Polycrystal Book Service.

HAUPTMAN, H. & KARLE, J. (1957). *Acta Cryst.* **10**, 267–270.

JAMES, V. J. & STEVENS, J. D. (1977). *Cryst. Struct. Commun.* **6**, 241–246.

KARLE, I. L., KARLE, J. & ESTLIN, J. A. (1967). *Acta Cryst.* **23**, 494–500.

KARLE, J. (1970). In *Crystallographic Computing*, p. 25. Copenhagen: Munksgaard.

NAYA, S., NITTA, I. & ODA, T. (1964). *Acta Cryst.* **17**, 421–433.

OVERBEEK, A. R. & SCHENK, H. (1976). *Proc. K. Ned. Akad. Wet.* B**79**, 341–343.

WEEKS, C. M. & HAUPTMAN, H. (1970). *Z. Kristallogr.* **131**, 437–442.

WOOLFSON, M. M. (1961). *Direct Methods in Crystallography*. Oxford: Clarendon Press.

Acta Cryst. (1981). A**37**, 684–689

DELCRI, an Enantiomorph-Specific Figure of Merit

BY GER J. OLTHOF AND HENK SCHENK

Laboratory for Crystallography, University of Amsterdam, Nieuwe Achtergracht 166, *Amsterdam, The Netherlands*

(*Received* 17 *July* 1980; *accepted* 3 *March* 1981)

Abstract

A figure of merit DELCRI, with estimates Δ_3 for the absolute values of the triplet phase sums φ_3, is described for the selection of numerical values of symbols used in a symbolic addition procedure. From tests with a number of structures crystallizing in polar space groups this figure of merit was found to enable the selection of enantiomorph-specific phase sets.

0567-7394/81/050684-06$01.00