

- ERIKSSON, A., HERMANSSON, K., LINDGREN, J. & THOMAS, J. O. (1982). *Acta Cryst.* **A38**, 138-142.
- International Tables for X-ray Crystallography* (1959). Vol. II. Birmingham: Kynoch Press. (Present distributor D. Reidel, Dordrecht.)
- International Tables for X-ray Crystallography* (1974). Vol. IV. Birmingham: Kynoch Press. (Present distributor D. Reidel, Dordrecht.)
- IBERS, J. A. (1959). *Acta Cryst.* **12**, 251-252.
- JOHNSON, C. K. (1969). *Acta Cryst.* **A25**, 187-194.
- JOHNSON, C. K. (1970). *Thermal Neutron Diffraction*, pp. 132-160. Oxford Univ. Press.
- KENDALL, M. G. & STUART, A. (1969). *The Advanced Theory of Statistics*, Vol. I. London: Griffin.
- LANDOLT-BÖRNSTEIN (1976). New Series, Vol. 7. *Structure Data of Free Polyatomic Molecules*. Berlin, Heidelberg, New York: Springer.
- LINNIK, J. W. (1961). *Die Methode der kleinsten Quadrate in moderner Darstellung*. Berlin: VEB Verlag der Wissenschaften.
- MAIR, S. L. & BARNEA, Z. (1975). *Acta Cryst.* **A31**, 201-207.
- OHMS, U., GUTH, H., KUTOGLU, A. & SCHERINGER, C. (1982). *Acta Cryst.* **B38**, 831-834.
- PRYOR, A. W. & SANGER, P. L. (1970). *Acta Cryst.* **A26**, 543-558.
- SCHERINGER, C. (1985). *Acta Cryst.* **A41**, 73-79.
- SCHERINGER, C. (1986a). *Acta Cryst.* Submitted.
- SCHERINGER, C. (1986b). In preparation.
- WHITELEY, B., MOSS, G. & BARNEA, Z. (1978). *Acta Cryst.* **A34**, 130-136.
- WILLIS, B. T. M. (1969). *Acta Cryst.* **A25**, 277-300.
- WILLIS, B. T. M. & PRYOR, A. W. (1975). *Thermal Vibrations in Crystallography*. Cambridge Univ. Press.
- WILSON, A. J. C. (1973). *Acta Cryst.* **B29**, 1488-1490.
- WILSON, A. J. C. (1976). *Acta Cryst.* **A32**, 994-996.
- WILSON, A. J. C. (1979). *Acta Cryst.* **A35**, 122-130.

*Acta Cryst.* (1986). **A42**, 362-368

## Unique Solutions for the SIR and SAS Phase Problems and the Use of Partial Structural Information in Phase Refinement

BY DAVID A. LANGS

*Medical Foundation of Buffalo, Inc., 73 High St, Buffalo, NY 14203, USA*

(Received 12 October 1985; accepted 3 April 1986)

### Abstract

Algebraic formulae are presented which permit a unique phased solution for diffraction data measured from a single isomorphous pair of crystals. Trial calculations performed on an SIR (single isomorphous replacement) data set from an 84-atom structure demonstrate that complete phasing can be achieved from a single chirally positioned replacement atom representing less than one percent of the total scattering power of the derivative structure. Similar phase refinements employing error-free SIR data for 2Zn pig insulin are less remarkable, and converge to an average phase error of 50°. The phase convergence of the formulae can be markedly improved if estimates of the cosine invariants from the SIR data are available [Hauptman (1982). *Acta Cryst.* **A38**, 289-294; Fortier, Moore & Fraser (1985). *Acta Cryst.* **A41**, 571-577]. The precision of these cosine estimates was found not to be critical; modular estimates of +1 or -1 were sufficient to allow the SIR phase refinement of the insulin structure to converge to an average phase error of 6°, which compares favorably with the value of 3° produced if the cosine invariants were known precisely. The derived formulae are also shown to be applicable to single-crystal analyses which utilize one-wavelength anomalous dispersion or partial structural fragments to initiate phasing. Test examples indicate that tangent-formula recycling procedures based on the derived formulae

compare favorably with the traditional tangent-formula methods to exploit partial structure information.

### Introduction

Recent developments in the improvement of translation-function methods (Langs, 1985) have introduced formulae which enable one to determine unambiguously the sine component of translation-function coefficients, given only the real component of these inherently complex-valued quantities. These procedures may be shown to be useful in other crystallographic situations in which it would be advantageous to retrieve the imaginary component of a complex variable for which only the real component is known. Several important applications may be found in structure-determination methods which use either partial structural models, anomalous scattering or isomorphous replacement to initiate phasing.

Crystallographic single isomorphous replacement (SIR) methods possess a known twofold ambiguity in the determination of non-centrosymmetric phases based on the determined positions of the replacement atoms. The resolution of this ambiguity by multiple isomorphous replacement relationships (Green, Ingram & Perutz, 1954; Harker, 1956) is dependent on the ability to prepare additional isomorphous derivatives of the native structure with replacement atoms possessing significantly large scattering powers

substituted at different positions in the crystal. A similar twofold ambiguity arises in the analysis of one-wavelength anomalous scattering data (SAS) (Ramachandran & Raman, 1956), which can be resolved with the availability of additional isomorphous data or anomalous scattering data measured at a different wavelength. It has recently been reported that the SAS ambiguity may in principle be resolved with information provided by an additional normally scattering data set, provided the structure contains only one kind of anomalous scatterer and its structure can be determined (Woolfson, 1984). Early attempts to resolve the SIR and SAS problems used phases which were a weighted average of the two solutions (Ramachandran & Raman, 1956; Blow & Rossmann, 1961), with a slight preference given towards that solution which was closest to the phase of the anomalous scatterer in the SAS case (Peerdeman & Bijvoet, 1956; Ramachandran & Raman, 1956). This latter observation has been successfully exploited to resolve the SAS phase ambiguity even for macromolecular structures (Hendrickson & Teeter, 1981). Other techniques which have used *a posteriori* structural information to resolve the SIR and SAS phase ambiguity include the averaging of non-crystallographic symmetry (Bricogne, 1976) and the leveling of suspected solvent regions that define the envelope of the macromolecule (Wang, 1981). An early attempt to resolve the SAS twofold ambiguity utilizing Friedel-related triple-phase invariants (Kroon, Spek & Krabbendam, 1977) produced formulae which could accurately estimate the sine of the invariant, but estimates of the related cosine components were less reliable in that they were expected, on average, to be positive. Subsequent work on direct methods has focused on obtaining reliable unique estimates for triple-phase invariants employing SIR (Hauptman, 1982*a*; Hauptman, Potter & Weeks, 1982) and SAS (Hauptman, 1982*b*; Giacovazzo, 1983; Karle, 1984) data. The Hauptman (1982*b*) SAS results have recently been used in the first successful *ab initio* direct-methods determination of a protein structure by Furey, Robbins, Clancy, Winge, Wang & Stout (1985). Furthermore, the SIR results have recently been improved to obtain cosine-invariant values of sufficient accuracy to resolve the phase ambiguity (Fortier, Moore & Fraser, 1985). More traditional direct-methods analyses have also introduced a procedure to resolve the SIR and SAS ambiguities (Fan Hai-fu, Han Fu-son, Qian Jin-zi & Yao Jia-xing, 1984), given the structure of the heavy atom or anomalous scatterer, and preliminary tests based on ideal error-free SIR data appear encouraging. Karle (1985), more recently, has shown that the SAS ambiguity may be resolved by an algebraic least-squares procedure which does not require knowledge of the positions of the anomalous scatterers in the structure. The present work describes an efficient

algebraic procedure to resolve the SIR and SAS phase ambiguity, given the structure of the heavy atoms or anomalous scatterers. The procedure can be easily adapted to forcefully utilize cosine-invariant information to obtain crystallographic phases of unprecedented accuracy. The methodology, moreover, can be applied to tangent-formula recycling procedures (Karle, 1968) to extend their range of effectiveness to smaller initial partial structures.

### Analysis

The single isomorphous replacement experiment provides diffraction data for both a native and an isomorphous derivative structure. The success of the method requires that the scattering power of the replacement atoms be sufficiently large to determine their positions by difference Patterson methods. The known quantities determined from the experiment are the amplitudes of the native and derivative structures,  $|FN_h|$  and  $|FD_h|$ , and the phased contribution  $FH_h$  of the heavy replacement atoms to  $FN_h$  to produce  $FD_h$ . The phases of  $FN_h$  and  $FD_h$ ,  $\varphi N_h$  and  $\varphi D_h$ , are unknown.

$$|FD_h| \exp(i\varphi D_h) = |FN_h| \exp(i\varphi N_h) + |FH_h| \exp(i\varphi H_h). \quad (1)$$

The magnitudes of  $FN_h$  and  $FD_h$  are adequate to determine the real component of the phase difference between the known value of  $\varphi H_h$  and either of the other two phases, for example

$$G_h = |FN_h|^2 - |FH_h|^2 - |FD_h|^2 = 2|FD_h FH_h| \cos(\varphi D_h - \varphi H_h). \quad (2)$$

It has previously been considered impossible with the available information to obtain the sign of the corresponding sine component  $S_h$  associated with  $G_h$ , *i.e.*

$$S_h = 2|FD_h FH_h| \sin(\varphi D_h - \varphi H_h), \quad (3)$$

which at best is known in magnitude only as  $|\sin x| = |1 - \cos^2 x|^{1/2}$ . Following the earlier analysis [Langs, 1985, equations (13)-(15)], the signed phase difference between the phase of the known heavy-atom structure and either of its isomorphously related structures may be obtained from the cosine and sine expansions of the phase identity  $(\varphi D_h - \varphi H_h) - \Phi_{h,k} = -(\varphi D_k - \varphi H_k) - (\varphi D_l - \varphi H_l) - \Psi_{h,k}$ , namely,

$$\begin{aligned} G_h \langle A_k A_l \cos \Phi_{h,k} \rangle_k + S_h \langle A_k A_l \sin \Phi_{h,k} \rangle_k &= A_h \langle (G_k G_l - S_k S_l) \cos \Psi_{h,k} \\ &\quad - (G_k S_l + S_k G_l) \sin \Psi_{h,k} \rangle_k, \\ S_h \langle A_k A_l \cos \Phi_{h,k} \rangle_k - G_h \langle A_k A_l \sin \Phi_{h,k} \rangle_k &= -A_h \langle (G_k S_l + S_k G_l) \cos \Psi_{h,k} \\ &\quad + (G_k G_l - S_k S_l) \sin \Psi_{h,k} \rangle_k, \end{aligned} \quad (4)$$

where  $A_h = 2|FD_h FH_h|$ ,  $\Phi_{h,k} = \varphi D_h + \varphi D_k + \varphi D_l$ ,  $\Psi_{h,k} = \varphi H_h + \varphi H_k + \varphi H_l$ , and the vectors  $h+k+l=0$ . All of the variables on the right-hand sides of (4) are known, except the various  $S_{hkl}$ , and these may be obtained on the expected condition that the term  $\langle A_k A_l \sin \Phi_{h,k} \rangle_k$  tends toward zero, such that

$$G_h \approx A_h \langle (G_k G_l - S_k S_l) \cos \Psi_{h,k} - (G_k S_l + S_k G_l) \sin \Psi_{h,k} \rangle_k / \langle A_k A_l \cos \Phi_{h,k} \rangle_k, \quad (5a)$$

$$S_h \approx -A_h \langle (G_k S_l + S_k G_l) \cos \Psi_{h,k} + (G_k G_l - S_k S_l) \sin \Psi_{h,k} \rangle_k / \langle A_k A_l \cos \Phi_{h,k} \rangle_k \quad (5b)$$

and

$$\varphi D_h = \varphi H_h + \tan^{-1}(S_h/G_h). \quad (5c)$$

Note that it is more important to know the sign of the structure-invariant quantity  $\langle A_k A_l \cos \Phi_{h,k} \rangle_k$ , which is expected to be positive, as its magnitude cancels in the tangent expression. Initial estimates of the various  $S_{hkl}$  to insert into the right-hand sides of (5a) and (5b) may be obtained from

$$\begin{aligned} G'_h &= A_h \langle G_k G_l \cos \Psi_{h,k} \rangle_k, \\ S'_h &= -A_h \langle G_k G_l \sin \Psi_{h,k} \rangle_k, \\ S_h &\approx A_h \sin [\tan^{-1}(S'_h/G'_h)]. \end{aligned} \quad (6)$$

A similar expression is readily developed for diffraction data measured at one wavelength for a single crystal containing anomalous scatterers of known or determinable position (Ramachandran & Raman, 1956). Given the Friedel relationships

$$|F_h| \exp(i\varphi_h) = |FN_h| \exp(i\varphi N_h) + |FA_h| \exp(i\varphi A_h)$$

and

$$\begin{aligned} |F_{\bar{h}}| \exp(i\varphi_{\bar{h}}) &= |FN_{\bar{h}}| \exp(i\varphi N_{\bar{h}}) \\ &+ |FA_{\bar{h}}| \exp(i\varphi A_{\bar{h}}), \end{aligned} \quad (7)$$

it has been shown that

$$\begin{aligned} G_h &= |F_{\bar{h}}|^2 - 4|FA_{\bar{h}}|^2 - |F_h|^2 \\ &= 4|F_h FA_{\bar{h}}| \cos(\varphi_h - \varphi A_{\bar{h}}), \\ G_{\bar{h}} &= |F_h|^2 - 4|FA_h|^2 - |F_{\bar{h}}|^2 \\ &= 4|F_{\bar{h}} FA_h| \cos(\varphi_{\bar{h}} + \varphi A_h), \end{aligned} \quad (8)$$

where  $FN_h$  and  $FA_h$  are the normal and the anomalous scattering contributions to  $F_h$ . These known values may be utilized in the same way as were the analogous SIR values of  $G_h$  given by (2). The values of the phases  $\varphi_h$ , corresponding to  $\varphi D_h$  in the SIR example, may be obtained from (5c) as previously described, given that the corresponding values of  $A_h = 4|F_h FA_{\bar{h}}|$ ,  $\Phi_{h,k} = \varphi_h + \varphi_k + \varphi_l$ , and  $\Psi_{h,k} = \varphi A_h + \varphi A_k + \varphi A_l$ .

Finally, tangent-formula recycling procedures (Karle, 1968) provide one of the most successful

methods for completing crystal structures from partial structural information. The equation of interest in these applications is

$$|F_h| \exp(\varphi_h) = |FP_h| \exp(\varphi P_h) + |FR_h| \exp(\varphi R_h), \quad (9)$$

where the values of  $|F_h|$  for the crystal and  $FP_h$  for the partial structure are known and  $|FR_h|$  and  $\varphi R_h$  for the residual structure are unknown. Since the expected value of  $|FR_h|^2$  may be estimated as  $\sum fR_j^2$  over the residual atoms of the structure, it follows that

$$\begin{aligned} G_h &= |F_h|^2 + |FP_h|^2 - \sum fR_j^2 \\ &\approx 2|F_h FP_h| \cos(\varphi_h - \varphi P_h), \end{aligned} \quad (10)$$

which is analogous to an approximation obtained earlier [Langs, 1985, equation (3)]. The probability that the sign of  $G_h$  is the sign of the cosine term on the right-hand side of (10) above was shown to be [Langs, 1985, equation (12)]

$$\begin{aligned} P(G_h) &= \frac{1}{2} + \frac{1}{2} \left[ \sum_{n=0}^{\infty} (-1)^n I_{2n+1}(|\alpha|) I_n(\beta) \right] \\ &\times \left[ \sum_{-\infty}^{\infty} (-1)^n I_{2n}(|\alpha|) I_n(\beta) \right]^{-1}, \end{aligned} \quad (11)$$

where  $\alpha = 2|F_h FP_h|G_h/\sigma^2(x)$ ,  $\beta = |F_h FP_h|/\sigma^2(x)$ , and  $\sigma^2(x) = 2 \sum_{j>k} fR_j^2/fR_k^2$  summed over the residual atoms of the structure. This estimate varies from the usual expected value of  $\cos(\varphi_h - \varphi P_h) = I_1(X)/I_0(X)$ , where  $X = 2|F_h FP_h|/\sum fR_j^2$ , derived by Sim (1960), in that (11) may on occasion indicate a negative sign. This probability criterion may be useful for selecting a starting set of reliable signed  $G_h$  values with which to initiate the previously described phase-refinement procedure, much as one would use the fractional scattering power (Karle, 1968) for selecting a starting set of basis phases for traditional tangent-formula recycling procedures.

### Small-molecule SIR calculations

The crystal structure of isoleucinomycin (Pletnev, Galitskii, Smith, Weeks & Duax, 1980), ILED, was initially used to test (5c) in various hypothetical applications with regard to isomorphous replacement and partial structure phase refinement. Crystal data:  $C_{60}H_{102}N_6O_{18}$ ,  $P2_12_12_1$ ,  $a = 11.516$ ,  $b = 15.705$ ,  $c = 39.310$  Å,  $Z = 4$ . All trial calculations were performed employing  $|E|$  values rather than  $|F|$  values in the calculation of the various  $A_h$  and  $G_h$  used in the refinement procedure. This is expected to eliminate the  $\sin \theta/\lambda$  bias that would result if  $|F|$  moduli were used to calculate the averages over the vector  $k$ . The analyses were restricted to the 500 largest  $|E|$  values which were used to generate 9071 triple relationships as the phasing basis. As a point of information, the average value of  $\langle \cos \Phi_{h,k} \rangle_k$  for the 500 phased terms was observed to be 0.383 from the known structure,

Table 1. *Isomorphous derivative phase refinement for compound ILED for various thresholds of r.m.s. error (% ERR) in the values of  $|ED_h|^2$ ,  $|EH_h|^2$  and  $|EN_h|^2$*

NO ATOMS gives the number of replaceable atoms in the 84-atom derivative for the computation of  $EH_h$  from which the phasing is initiated, and  $\langle \cos \Delta\varphi_0 \rangle$  gives the average cosine of the phase difference between the calculated  $\varphi H_h$  and  $\varphi D_h$  which is known from the solved structure. The fourth column,  $\langle \cos \Delta\varphi_4 \rangle$ , gives the average cosine of the difference between  $\varphi D_h(\text{true}) - \varphi D_h(\text{refined})$ , at the end of four refinement cycles with equation (5), as a measure of the convergence toward the true values. The last column records the number of atom peaks found among the top strongest indicated peaks of the subsequent  $E$  map.

% ERR	NO ATOMS	$\langle \cos \Delta\varphi_0 \rangle$	$\langle \cos \Delta\varphi_4 \rangle$	NO ATOMS/ TOP PEAKS
0	83	0.999	0.999	76/84
0	1	0.137	0.906	72/84
3.5	1	0.137	0.795	52/84, 38/40
7.0	1	0.137	0.678	28/84, 19/40
3.5	2	0.214	0.846	63/84, 37/40
7.0	2	0.214	0.749	47/84, 31/40

all individual averages were positive and only nine terms had  $\langle \cos \Phi_{h,k} \rangle_k$  less than the corresponding value of  $|\langle \sin \Phi_{h,k} \rangle_k|$ .

The isomorphous replacement tests treated the original ILED diffraction data as the derivative structure. The isomorphously replaceable atoms for the various tests were selected from among the 84 non-hydrogen atoms of the molecule, and the corresponding set of native crystallographic amplitudes was generated from the non-replaceable atoms of the structure. The  $G_h$  values corresponding to the 500 strongest  $|E|$  values were calculated and used to obtain initial estimates for  $S_h$  by means of (6). The  $G_h$  values were held fixed while the corresponding  $S_h$  values were refined to convergence by (5b), a process that usually required two or three cycles. The  $S_h$  values input into a particular cycle were replaced by their recalculated values only at the end of the cycle when all 500  $S_h$  had been re-evaluated. At the end of the refinement the phases for the structure were computed using (5c). This global phase-refinement procedure was performed in triplicate, first with error-free data, and then with data sets which incorporated randomly generated Gaussian errors of 3.5 and 7.0% in the magnitudes of the squared amplitudes used in the calculations. A summary of the results is presented in Table 1.

#### Macromolecular SIR calculations

The ILED SIR results indicated that the formula might be successfully applied to larger macromolecular structures, and data generated for a second test structure, 2Zn pig insulin (Blundell, Cutfield, Dodson, Dodson, Hodgkin & Mercola, 1971), were examined. Crystal data:  $R3$ ,  $a = 82.5$ ,  $c = 34.0$  Å,  $Z = 9$ . The asymmetric unit contains the equivalent of 1083 independent non-hydrogen atoms,

protein and water molecules, including atomic sites which exhibit full or partial occupancy. Normalized  $E$  data were generated to 1.9 Å resolution and the 1000 largest  $E$  values were used to generate 64 458 triples for the analysis. The number of these various data compares favorably with those used by Fan Hai-fu *et al.* (1984) in their analysis employing a similar Pb derivative of the insulin structure. The average value for  $\langle \cos \Phi_{h,k} \rangle_k$  for the 1000 phased terms was noted to be 0.074, based on the known structure. Moreover, individual values of  $\langle \cos \Phi_{h,k} \rangle_k$  were calculated to be less than the corresponding value of  $|\langle \sin \Phi_{h,k} \rangle_k|$  for 263 of the 1000 terms, and  $\langle \cos \Phi_{h,k} \rangle_k$  was negative for 74 of these terms.

The above calculations were repeated for SIR data generated for the coordinates of the 2Zn pig insulin which was treated as the native structure. Two derivative structures were obtained by substituting a Pb or S atom for one of the full-occupancy water sites not on the threefold axis. Calculations for each of these native and derivative SIR data sets were performed in triplicate, first assuming that the structure-invariant quantity  $\langle A_k A_l \cos \Phi_{h,k} \rangle_k$  in (5a) and (5b) was unknown but expected to be positive as was done in the ILED example above, second, that the involved cosine values were approximately known as either +1 or -1 (Hauptman, 1982a), and third, that reliable estimates of these cosine values existed for the full range between +1 and -1 (Fortier, Moore & Fraser, 1985). The accuracy of the phase-refinement procedure can be enhanced by multiplying the right- and left-hand sides of (4) by the  $\cos \Phi_{h,k}$  estimate for each of the triple contributors such that the resultant  $\langle A_k A_l \cos^2 \Phi_{h,k} \rangle_k$  term is required to be maximal and positive relative to the  $\sin \Phi_{h,k}$  term. Following on from this the denominators of (5a) and (5b) should be divided by this  $\langle A_k A_l \cos^2 \Phi_{h,k} \rangle_k$  term. These results are given in Table 2.

#### Small-molecule tangent-formula recycling calculations

Finally, tangent-formula recycling procedures were used in three test applications employing known molecular fragments of the ILED structure comprising 20, ten and five atoms. The calculations were performed using the same largest 500  $|E|$  values and 9071 triples for ILED as above. Basis set phases were selected using the fractional scattering power criteria, phases from the molecular fragment being accepted if  $|F_{\text{cal}}|/|F_{\text{obs}}| > p$ . The order of the phase extension was determined as that phase which had the largest sum  $\alpha$  (Karle & Karle, 1966; Germain, Main & Woolfson, 1970) computed from the basis set and extended phases. The phase refinement was completed in four cycles in each trial. The first two cycles extended and refined phases for the remaining 40 and 80% of the data while the phases of the basis set were held fixed. The basis set was allowed to refine in the

Table 2. Comparison of SIR phase refinements for the hypothetical Pb and S heavy-atom derivatives of 2Zn pig insulin as indicated by DERV in the first column

The three refinement conditions concerning the expectation value of  $\cos \Phi_{h,k}$  in equation (5) are:  $U$  = unitary weights representing the unknown but expected positive value of the cosine invariant,  $M$  = modular weights representing the sign or mode of the true value of the invariant (+1 or -1), and  $D$  = the true value of the cosine invariant determined in the range of +1 to -1, as indicated in the second column under TYPE. The letter  $K$  in the third row indicates that the true known values of  $S_h$ , as well as  $G_h$ , were used to initiate phasing and test the stability of equation (5).  $\langle \text{COS } \Delta\varphi_0 \rangle$  and  $\langle \text{COS } \Delta\varphi_1 \rangle$  have the same meaning as in Table 1.  $NR$  and  $NI$  give the corresponding number of  $G_h$  and  $S_h$  for which equation (5) indicates the wrong sign at the end of the first and fourth refinement cycles.

DERV	TYPE	$\langle \text{COS } \Delta\varphi_0 \rangle$	$\langle \text{COS } \Delta\varphi_1 \rangle$	$NR$	$NI$	$\langle \text{COS } \Delta\varphi_2 \rangle$	$NR$	$NI$
Pb	$U$	0.125	0.654	339	376	0.574	187	404
S	$U$	0.050	0.654	296	343	0.568	213	386
Pb	$K, U$	0.125	0.750	228	227	0.669	172	304
Pb	$M$	0.125	0.914	46	80	0.995	4	25
S	$M$	0.050	0.918	47	72	0.995	2	31
Pb	$D$	0.125	0.916	39	71	0.998	4	18
S	$D$	0.050	0.920	43	72	0.998	5	16

last two cycles which determined phase at the end of the convergence map.

Several protocols were examined in utilizing (5c) for molecular fragment phase recycling with regard to the ILED structure. The traditional refinement scheme of selecting a basis set of reliably accurate phases from which to determine the others by phase extension and refinement did not produce significantly different results from the global refinement procedure, as described above, but with the important added provision that both  $G_h$  and  $S_h$  were allowed to refine. The recycling methods were found to converge to essentially the same phase solution whether one initiated the phasing with as few as 100  $G_h$  values, selected as having the highest probability values by (11), or indiscriminantly used the full set of 500 as was done in the earlier SIR tests. Global refinement procedures were used in the subsequent molecular-fragment phase recycling tests, largely because they eliminated the need to generate a convergence mapping. A comparison between these results and those produced by traditional tangent-formula recycling methods is presented in Table 3.

### Discussion

The results in Table 1 for the ILED structure clearly indicate that the various SIR phase refinements have resolved the initial SIR phase ambiguity. These trials suggest that this method should be successful for moderately large structures of 100 or more atoms, employing diffraction data of average accuracy, provided that the replaceable atoms comprise at least 1% of the scattering power of the derivative structure and are not centrosymmetrically related. In the latter case it would be impossible to obtain non-zero estimates of  $S_h$  and refine away from the values of the phases of the replaceable atoms.

The results of the trials for 2Zn pig insulin presented in Table 2 illustrate the limitations of (5c) in dealing with considerably larger macromolecular structures, in which cases the value of the

Table 3. Comparison of partial structure phase refinements for compound ILED

The first column cites the number of atoms in the molecular fragment; the number of spurious atoms in the model is indicated by an  $s$ . Columns two through five summarize the results produced by tangent-formula recycling, and columns six and seven the results produced by equation (5). The fractional scattering power ( $P$ ) and starting number of basis phases (NO  $\varphi$ 's) selected for the recycling procedure are given in columns two and three. The remaining columns record the number of atomic positions found among the given number of top largest peaks of the  $E$  maps synthesized from the various refined phase sets. The last three rows summarize fragment recycling efforts which selected coordinates from among the strongest peaks in the  $E$  map produced by the five-atom model. The numbers in parentheses record the number of correct and spurious atoms from the input model which are found among the number of top peaks indicated.

AT/84	$P$	Tangent-formula recycling		Equation (5)
		NO $\varphi$ 's	NO ATOMS/ TOP PEAKS	NO ATOMS/ TOP PEAKS
20	0.20	200	67/84, 39/40	69/84, 40/40
10	0.20	97	54/84, 33/40	42/84, 30/40
5	0.20	32	5/20, 2/10	11/20, 10/12
	0.15	74	5/20, 2/10	
	0.10	176	6/20, 6/10	
10+2s				47/84, 32/40
11+9s				(10, 4s)/20
14+16s				(14, 1s)/15

$\langle A_k A_l \cos \Phi_{h,k} \rangle_k$  term is expected to be less reliably large and positive relative to the corresponding  $\sin \Phi_{h,k}$  term. The entries in the first two rows of Table 2 show that the first cycle of refinement has reduced the phase error from about 85 to 50° as indicated by the cosine values in the second and third columns of the table. Additional refinement cycles demonstrate that (5c) does not further converge on the phase solution. Although some improvement may be seen in reducing the number of contrary signs,  $NR$ , indicated for  $G_h$ , which is *a priori* known, the number of wrong-sign indications,  $NI$ , for  $S_h$  is seen to increase. Row three of the table summarizes the results of a refinement initiated with the true values of  $S_h$  and  $G_h$ , which is also seen to diverge. It should

be stated that Fan Hai-fu *et al.* (1984) were able to achieve a significantly lower average phase error of  $25^\circ$  for a similar insulin test example using probability formulae which associate the SIR difference phases with the crystallographic amplitudes,  $|ED_h|$ , rather than the structure products  $2|ED_hEH_h|$  as indicated in (2) and (3) of this work. Although it has been possible to verify this  $25^\circ$  accuracy with our insulin test structure, we caution that the success of these methods with real data sets may be overly dependent on how accurately one can estimate  $\cos(\varphi D_h - \varphi H_h)$  from (2) for those terms for which  $EH_h$  tends to be small. The analyses presented in this paper have purposefully avoided separating the structure product amplitudes from the cosine and  $|\text{sine}|$  of the SIR phase difference for this very reason.

Stable convergent SIR phase refinements for the insulin data can be demonstrated for (5c), but require one to obtain reliable estimates of the  $\cos \Phi_{h,k}$  invariants as have been provided by Hauptman (1982a) and Fortier *et al.* (1985) and incorporate these in the refinement as described above. The last four rows of Table 2 clearly show the unprecedented accuracy which may be obtained utilizing either modular (+1 or -1) or true values of these cosine invariants. The SIR phase refinement using modular cosine estimates is seen to converge to an average phase error of less than  $6^\circ$ , demonstrating that precise estimates for these invariants are not as critical as was shown for earlier tangent-formula refinements employing three-phase invariants of moderate accuracy (Weeks, Potter, Smith, Hauptman & Fortier, 1984). The results produced by the error-free Pb and S data sets do not appear to be significantly different in spite of the large differences in scattering powers of the replacement atoms, but should show a phasing advantage for the Pb derivative given experimentally measured diffraction data. No attempt was made to simulate real SIR data. It should be cautioned that errors encountered by lack of isomorphism, inadequate modeling of the heavy-atom structure, and the measurement of the small but crucial differences in the diffraction data in which the structural information resides will to some degree adversely affect the usefulness of the refinement procedure described in this section.

The partial structure recycling trials summarized in Table 3 indicate that (5c) yields results which compare favorably with those produced by traditional tangent-formula recycling methods. The  $P(G_h)$  criterion was not particularly useful for selecting starting basis sets of  $G_h$  values for fragments which represented less than a quarter of the structure, as very few  $G_h$  terms had probabilities which exceeded 0.75. Phase refinements that employed low-threshold acceptance values for  $P(G_h)$  in order to select meaningfully sized basis sets consequently produced results which were not significantly different from

those obtained by initiating the refinement using the full set of  $G_h$  values. Traditional tangent-formula recycling methods are routinely successful in determining structures from fragments comprising 10% of the structure, which appears to be well below the range of usefulness of the  $P(G_h)$  criteria.

One feature of the global phase refinements provided by (5c) is that they tend to produce phases which strongly preserve the structure of the trial model as the strongest peaks in subsequent  $E$  maps. The entries for the traditional tangent formula recycling of the five-atom fragment in Table 3 indicate that the positions of the five atoms of the input fragment are to be found among the top 20 peaks of the  $E$  map, but only two of the five atoms are among the top ten peaks for the first two trials using 32 and 74 starting phases. A larger basis set of 176 phases produces a solution in which the five atoms are found in the top ten peaks, but only one additional valid peak is to be found in the top 20 peaks to the map. The results from (5c) appear to be better in that ten valid peaks, including the fragment, are to be found among the top 12 peaks in the map. Furthermore, (5c) appears to be able to converge on the correct phase solution even when the input structure has numerous spurious atoms. The top 12 peaks, which include two spurious atoms, can be recycled to produce a solution which is as good as that provided by ten correct atoms. If the top 20 peaks from the  $E$  map phased by the five-atom fragment are selected for recycling, these define 11 correct atoms and nine spurious atoms as indicated in the table. Phase recycling produces an  $E$  map which appears to be no better than that produced by the original five-atom model in that 12 correct atoms are found among the top 20 peaks. However, a comparison of the 20-atom input model with the resultant  $E$  map reveals that ten of the eleven correct atoms but only four of the nine spurious atoms reappear among these top 20 peaks, and this information can be used to improve the model for the next cycle of refinement. The results from recycling the top 30 peaks from the five-atom refinement cycle are more remarkable in that, although the input model contains 14 correct and 16 incorrect atoms, the 14 correct atoms are found among the 15 largest peaks of the subsequent synthesis. These examples indicate that, although (5c) has a strong tendency to produce solutions which preserve the structure of the input model, the refinement procedure is rather forgiving in that the spurious atoms of the model tend to produce weaker peaks than those produced by correct atoms.

Refined coordinates for the 2Zn pig insulin structure were obtained from Dr G. D. Smith. Dr Suzanne Fortier is acknowledged for kindly providing a pre-publication copy of her paper. This work was

supported in part by NIH grant HL32303 from the National Heart, Lung and Blood Institute.

#### References

- BLOW, D. M. & ROSSMANN, M. G. (1961). *Acta Cryst.* **14**, 1195-1202.
- BLUNDELL, T. L., CUTFIELD, J. F., DODSON, E. J., DODSON, G. G., HODGKIN, D. C. & MERCOLA, D. A. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 233-241.
- BRICOGNE, G. (1976). *Acta Cryst.* **A32**, 832-847.
- FAN HAI-FU, HAN FU-SUN, QIAN JIN-ZI & YAO JIA-XING (1984). *Acta Cryst.* **A40**, 489-495.
- FORTIER, S., MOORE, N. J. & FRASER, M. E. (1985). *Acta Cryst.* **A41**, 571-577.
- FUREY, W., ROBBINS, A. H., CLANCY, L. L., WINGE, D. R., WANG, B. C. & STOUT, C. D. (1985). *Science*, **231**, 704-710.
- GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* **B26**, 274-285.
- GIACOVAZZO, C. (1983). *Acta Cryst.* **A39**, 585-592.
- GREEN, D. W., INGRAM, V. W. & PERUTZ, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287-307.
- HARKER, D. (1956). *Acta Cryst.* **9**, 1-9.
- HAUPTMAN, H. (1982a). *Acta Cryst.* **A38**, 289-294.
- HAUPTMAN, H. (1982b). *Acta Cryst.* **A38**, 632-641.
- HAUPTMAN, H., POTTER, S. & WEEKS, C. M. (1982). *Acta Cryst.* **A38**, 294-300.
- HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **290**, 107-113.
- KARLE, J. (1968). *Acta Cryst.* **B24**, 182-186.
- KARLE, J. (1984). *Acta Cryst.* **A40**, 4-11.
- KARLE, J. (1985). *Acta Cryst.* **A41**, 387-394.
- KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849-859.
- KROON, J., SPEK, A. L. & KRABBENDAM, H. (1977). *Acta Cryst.* **A33**, 382-385.
- LANGS, D. A. (1985). *Acta Cryst.* **A41**, 578-582.
- PEERDEMAN, A. F. & BIJVOET, J. M. (1956). *Acta Cryst.* **9**, 1012-1015.
- PLETNEV, V. Z., GALITSKII, N. M., SMITH, G. D., WEEKS, C. M. & DUAX, W. L. (1980). *Biopolymers* **19**, 1517-1534.
- RAMACHANDRAN, G. N. & RAMAN, S. (1956). *Curr. Sci.* **25**, 348-351.
- SIM, G. A. (1960). *Acta Cryst.* **13**, 511-512.
- WANG, B. C. (1981). *Acta Cryst.* **A37**, C11.
- WEEKS, C. M., POTTER, S. A., SMITH, G. D., HAUPTMAN, H. & FORTIER, S. (1984). *Am. Crystallogr. Assoc. Meet.* 20-25 May 1984, Lexington, Kentucky, USA. Abstr. Q2.
- WOOLFSON, M. M. (1984). *Acta Cryst.* **A40**, 32-34.

*Acta Cryst.* (1986). **A42**, 368-380

## Space-Group Analyses of Thin Precipitates by Different Convergent-Beam Electron Diffraction Procedures

BY J. M. HOWE

*Department of Metallurgical Engineering and Materials Science, Carnegie-Mellon University, Pittsburgh, PA 15213, USA*

M. SARIKAYA

*Department of Materials Science and Engineering, University of Washington, Seattle, WA 98195, USA*

AND R. GRONSKY

*Materials and Molecular Research Division, Lawrence Berkeley Laboratory, and Department of Materials Science and Mineral Engineering, University of California, Berkeley, CA 94720, USA*

(Received 23 August 1985; accepted 7 April 1986)

#### Abstract

Convergent-beam electron diffraction point- and space-group analyses were performed on thin  $\gamma'$  precipitate plates, which had been extracted from an Al-15 wt% Ag alloy aged for either 30 or 120 min at 623 K. Although the space groups of precipitates in both samples were determined to be  $P6_3/mmc$ , it is shown that different results can be obtained, depending on the method of convergent-beam electron diffraction analysis that is employed. Comparative analyses using a pure  $\alpha$ -titanium standard demonstrate that the limited thickness of the plate-shaped precipitates is responsible for the variable results,

suggesting a preferred method for point- and space-group determination of thin particles.

#### 1. Introduction

Howe & Gronsky (1985) recently demonstrated that symmetry determinations performed on thin specimens by convergent-beam electron diffraction (CBED) may reflect the limited thickness of the specimen along the electron-beam direction, rather than the actual space group of the material. This is a particularly important effect in materials science, where many of the microconstituents which strongly influence the properties of engineering materials are