

## References

- COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.  
 GIACOVAZZO, C. (1983). *Acta Cryst.* **A39**, 585-592.  
 HAUPTMAN, H. (1982). *Acta Cryst.* **A38**, 632-641.  
 HEINERMAN, J. J. L., KRABBENDAM, H., KROON, J. & SPEK, A. L. (1978). *Acta Cryst.* **A34**, 447-450.  
*International Tables for X-ray Crystallography* (1974). Vol. IV. Birmingham: Kynoch Press. (Present distributor D. Reidel, Dordrecht.)  
 KANTERS, J. A., JANSMA, S. & PONTENAGEL, W. M. G. F. (1986). In preparation.  
 KROON, J., SPEK, A. L. & KRABBENDAM, H. (1977). *Acta Cryst.* **A33**, 382-385.  
 MAIN, P., HULL, S. E., LESSINGER, L., GERMAIN, G., DECLERCQ, J.-P. & WOOLFSON, M. M. (1978). *MULTAN78. A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data*. Univs. of York, England, and Louvain, Belgium.  
 OKAYA, Y. & PEPINSKY, R. (1956). *Phys. Rev.* **103**, 1645-1647.  
 PEERDEMAN, A. F. & BIJVOET, J. M. (1956). *Proc. K. Ned. Akad. Wet.* **B59**, 312-313.  
 RAMACHANDRAN, G. N. & RAMAN, S. (1956). *Curr. Sci.* **25**, 348-351.  
 SIEKER, L. C., ADMAN, E. T. & JENSEN, L. H. (1972). *Nature (London)*, **235**, 40-42.

*Acta Cryst.* (1987). **A43**, 60-64

**REDUC: An Automated Procedure for the Determination of Structure-Factor Phases from the Estimated Values of Structure (Sem)invariant Phase Combinations**

BY W. M. G. F. PONTENAGEL,\* H. KRABBENDAM AND J. KROON

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

(Received 11 November 1986; accepted 11 June 1986)

## Abstract

An algorithm is described to determine structure-factor phases from the estimated phases of (sem)invariants of any order, in a mixed mode if desired. The method essentially consists of a reduction of a redundant set of linear equations by successive elimination of unknowns. The main result of the procedure is a set of mutually independent equations in which structure-factor phases are expressed as linear combinations of a limited number of unknowns (among which a suitable set of origin-defining phases must exist), and (sem)invariant phases of which (reliable) estimates are available. Application to an unknown  $P2_1$  structure using triple products only, followed by a tangent refinement, is given. From further tests it appeared that the method is preeminently expedient as a troubleshooter in intricate structure determinations.

## Introduction

For small and medium-sized structures, the number of seminvariant phases ( $\Phi$ 's) which can be determined by direct methods with a reasonable accuracy outpaces by far the number of structure-factor phases ( $\varphi$ 's) to be determined for the calculation of

an interpretable  $E$  map. As seminvariants can actually be seen as linear combinations of structure-factor phases, a matrix inversion may be thought of as an appropriate procedure to determine structure-factor phases. For centrosymmetric problems this approach was first introduced by Cochran & Douglas (1955) and later amended by Vand & Pepinsky (1956). In essence that technique consists of two parts: (i) select a set of mutually independent relations from which the  $\varphi$ 's can be obtained *via* a matrix inversion and (ii) employ the relations that have not been used in a sensible way as a check on the plausibility of the obtained solution.

Fortier, Fronckowiak, Smith, Hauptman & De Titta (1978) and Fortier, De Titta, Fronckowiak, Smith & Hauptman (1979) described a procedure which resembles the matrix-inversion method, but differs from it in the fact that the distinction between the steps (i) and (ii) has been removed. It consists of the reduction of a redundant set of special linear equations by successive elimination of unknowns. Starting with an arbitrary set of  $M$  seminvariants we find the following symbolic representation:

$$\sum_k \varphi_k \equiv \Phi_s \quad (s = 1, 2, \dots, M). \quad (1)$$

Successive elimination of structure-factor phases and repeated substitutions in all previously obtained expressions leads in centrosymmetric space groups

\* Present address: DSM Research, Postbus 18, 6160 MD Geleen, The Netherlands.

to a number of equations of the form

$$\varphi_h \equiv \sum_i \varphi_i + \sum_j \Phi_j. \quad (2)$$

If the process is carried through to its end, all unknown  $\varphi$ 's may be expressed in terms of origin-defining  $\varphi$ 's and a complete set of independent  $\Phi$ 's, provided all necessary phase relations are employed.

Apart from the type (2) expressions, usually a number of so-called identities ( $\sum_j \Phi_j \equiv 0$ ) is obtained as a consequence of the redundancy of the employed set of seminvariants. By their nature, these relations must be exactly fulfilled, so, by substituting the estimated values of the seminvariants, possible errors may be revealed.

Although the basic features of the procedure may also be used in non-centrosymmetric space groups, a number of essential steps must be added to the algorithm to cope with the special problems encountered when phases instead of signs are processed.

In this paper a general description of phase reduction will be given which is suitable for application in any space group, and in addition the successful application of the automated procedure (*REDUC*) to a  $P2_1$  structure will be discussed.

### General description of phase reduction

The algorithm of phase reduction in both centrosymmetric and non-centrosymmetric space groups will be described point by point on the basis of the flow chart in Fig. 1.

[A] Apart from a set of seminvariants, ranked in decreasing order of the reliability of their phase estimates, the procedure requires a priority listing for the relevant structure factors. These priorities indicate the order in which the unknown phases are preferably eliminated, and their selection is discussed later.

[B] Suppose we enter this section at some stage of the procedure. After reading a seminvariant phase relation of type (1), it is checked if one or more of the contributing  $\varphi$ 's was eliminated previously (see [F]). Where possible, the current expressions for the  $\varphi$ 's are substituted leading to

$$\sum_i \alpha_i \varphi_i \equiv \sum_j \beta_j \Phi_j. \quad (3)$$

The coefficients  $\alpha$  and  $\beta$  are reduced modulo 2 if the corresponding phase is restricted to two possible values as a consequence of space-group symmetry.

[C] If all  $\alpha$ 's in (3) appear to be zero, the identity  $\sum_j \beta_j \Phi_j \equiv 0$  is saved. Such a relation expresses the redundancy of the employed set of seminvariant phase relations. The chosen algorithm ensures that any relation that can be expressed exclusively in terms of previously accepted seminvariants will lead to such an identity.

[D] If at least one of the  $\alpha$ 's in (3) is not equal to zero, a so-called primary relation is obtained. In

symbolic form, these relations are exact, but since estimated values for the seminvariants are to be substituted at the end of the process, the variance of a possible phase indication must already be taken into account. An approximate value of the variance, neglecting correlations, may be obtained by summing the individual variances of the estimates for the contributing  $\Phi$ 's. If the result appears to be worse than some preset threshold, the primary relation is rejected and the last-read seminvariant is written to a scratch file. With a certain frequency, depending on a chosen parameter, the scratch file is used as input to the program to try whether previously rejected relations are of better use at that stage of the reduction process. This ensures that the most accurate estimates are used as much as possible in early stages of phase reduction.

[E] If the estimated variance of a primary relation is acceptable, one of the  $\varphi$ 's may be eliminated, provided the corresponding  $|\alpha_i| = 1$ . (Owing to the  $2\pi$  ambiguity in the estimates of the seminvariants, indications for multiples of phases cannot be processed.) The supplied priority listing indicates which of the  $\varphi$ 's is to be preferred. In those cases where only multiple-phase indications can be obtained (all  $|\alpha_i| \geq 2$ ), the last-read seminvariant will be tried again at a later stage.

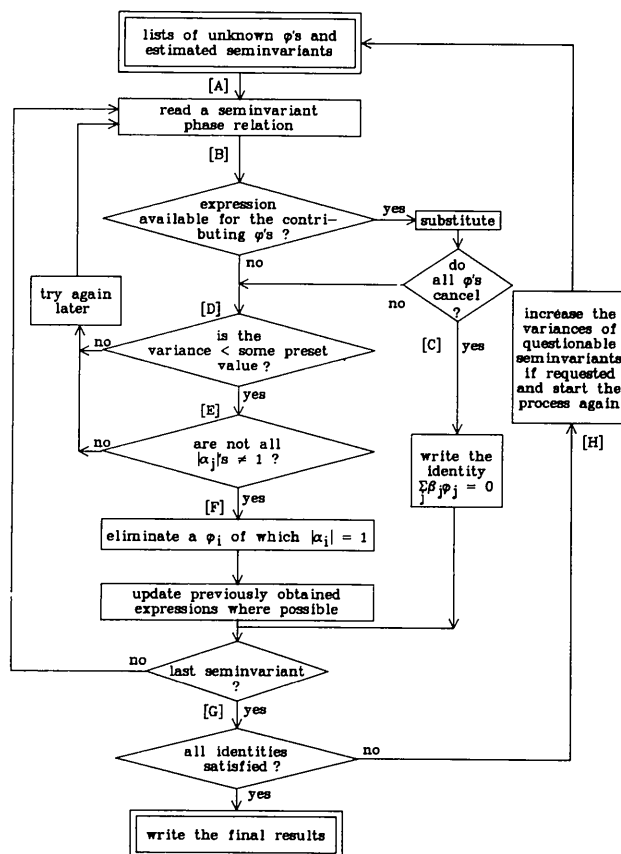


Fig. 1. Flow chart of the reduction procedure.

[F] The current expression

$$\varphi_h \equiv \sum_p \alpha_p \varphi_p + \sum_j \beta_j \Phi_j \quad (4)$$

is substituted in all appropriate previously obtained results and the updated intermediate relations are saved.

[G] If the remaining unused seminvariants do not lead to acceptable primary relations, or to new identities, the elimination process stops. All final expressions of type (4) are printed and the  $\varphi$ 's which are not eliminated ('rest  $\varphi$ 's') are given explicitly. At this point the purely symbolic manipulations are finished and the numerical estimates for the seminvariants are loaded. It will be clear that there should not be sign ambiguity in these numerical estimates, although there is no restriction on their actual value. All identities are screened to reveal possible inconsistencies and for non-centrosymmetric space groups the extent to which an identity is violated is given. In addition, the numerical values of the terms  $\sum_j \beta_j \Phi_j$  are determined together with the corresponding variances. Finally, for each seminvariant the score is given with respect to: (i) its importance for the reduction process [*i.e.* the number of final type (4) expressions in which it appears], (ii) the number of identities satisfied by its original estimate and (iii) the number of identities violated by its estimate.

[H] On the basis of the identities, questionable seminvariants may be allocated. If necessary, the variances of these relations are increased, all derived phase indications are erased and the process is started again. Such a recycling procedure will result in expressions for structure-factor phases in which inconsistently estimated seminvariants are avoided as much as possible.

All through the reduction procedure we are in possession of a variety of phase relations of type (4). It will be clear that these results will lose their practical significance if the coefficients accompanying the not-eliminated  $\varphi$ 's and the estimated  $\Phi$ 's are too large and (or) if the individual expressions contain too many terms. Although the temporary rejection based on the variances of primary relations (part [D] of the flow chart) restrains the elimination process, the repeated updating of expressions in part [F] of the procedure may lead to inefficient phase indications. In this context it is evident that the elimination priority of structure-factor phases must comply with the order in which the seminvariants are supplied. Elimination based on the  $|E|$  values of the contributing structure factors, as suggested by Fortier *et al.* (1978, 1979), seems a reasonable choice if the seminvariants are ranked in decreasing order of reliability.

However, although the not-eliminated  $\varphi$ 's can always define a permitted origin (seminvariants are incapable of relating origin-defining  $\varphi$ 's to each

other), the result will not necessarily be very adequate. As an example, in the space group  $P2_12_12_1$  the procedure may end up with phases of general structure factors, while the phases of projection reflections are known to be much more effective for origin definition. Also it may appear that the resulting origin-defining  $\varphi$ 's do not interact efficiently with the set of eliminated structure-factor phases. [They do not appear frequently in the final type (4) results.] Therefore we usually let the elimination priority of structure-factor phases be determined by a convergence procedure (Germain, Main & Woolfson, 1970) based on all available seminvariants and taking into account the need for an appropriate origin definition. If desired, origin-defining  $\varphi$ 's may be labelled at the outset to prevent them from being eliminated during the reduction procedure.

Normally this will lead to a phase reduction in which the number of not-eliminated  $\varphi$ 's is minimized without introducing intolerable variances in the final results. Apart from origin-defining phases a number of extra phases is expected to remain as unknowns at the end of the process. Based on their frequencies of occurrence in the final type (4) expressions, a few of them are designated to be given different numerical values, thus introducing the multi-solution concept into the phase-reduction procedure. In this way, several sets of trial phases for about 15 to 100 structure factors may be obtained, but since these trial values can only be seen as possible rough estimates for the true phases, a tangent refinement procedure will be necessary, especially for non-centrosymmetric cases. Afterwards, the thus obtained medium-sized starting sets are used for phase expansion to obtain numerical values for those phases which appeared to be unreachable by the reduction process.

### Practical application

The crystal structure of (-)- $\alpha$ -methylbenzylammonium hydrogen *meso*-tartrate,  $C_8H_{12}N^+.C_4H_5O_6^-$ , (Kroon, Duisenberg & Peerdeman, 1984) was solved by the reduction procedure. The relevant data for the compound are: monoclinic, space group  $P2_1$  with  $Z=2$ ;  $a=6.471$ ,  $b=13.74$ ,  $c=7.585$  Å and  $\beta=108.87^\circ$ . Although the phase reduction can handle all kinds of seminvariants simultaneously, the structure was determined with triple-product phase relations only ( $\varphi_h + \varphi_k - \varphi_{h+k} \equiv \Phi_{hk}$ ). All phases  $\Phi_{hk}$  were estimated to be zero in accordance with the Cochran distribution (Cochran, 1955). Consequently, no inconsistencies in the identities are expected but for a few cases in which phase relations, originating from space-group symmetry, are incompatible with the all-zero estimates for the triple-product phases. As an example we may consider  $\sum_1$  relations which can lead to conflicting phase indications for one-phase seminvariants.

After Lp corrections, a scale and temperature factor were obtained from a Wilson plot and all 3570 independent structure factors were normalized. The 225 largest  $E$  values were used to generate 1975 triple products and *via* a convergence procedure, based on these relations only, the elimination priority for the structure-factor phases was determined. The reduction procedure resulted in 200 expressions of type (4) and 282 identities. Two of the identities appeared to be violated, but since in both cases several triple products with a relatively low  $\kappa$  value were involved, no action was taken.

From the frequencies of occurrence in the final expressions it was concluded that only 12 triple-product phases were of ultimate importance. Neither their  $\kappa$  values, nor their scores with regard to the identities led, however, to serious doubts about their original estimates. The other 188 relevant triple products were used only sparsely (less than 20 times) and were therefore not expected to hinder the phase determination, even if the true phases of a number of them would deviate substantially from zero.

From the 25 not-eliminated structure-factor phases, only nine were used more than ten times in the final 200 phase relations. Three of them were set equal to zero to define a permitted origin and three others were given several possible values by quadrant permutation. This resulted in 64 sets of trial values for 87 structure-factor phases. All these sets were put into a refinement procedure based on the tangent formula. In order to make full use of the multi-solution principle, refinements leading to phases which bore no relation to their initial trial values were to be avoided; therefore the modified algorithm suggested by Pontenagel (1984) was used. Next, again using the tangent formula but now with the conventional algorithm, all 64 sets were expanded to the 225 desired phases and the results were allowed to refine. The well known figures of merit ABSFOM, PSIZERO and RESID undoubtedly indicated one set to be the most promising.

The corresponding  $E$  map clearly revealed the *meso*-tartrate ion in the ten highest peaks, but for the rest no significant features were present. One subsequent Fourier synthesis was sufficient to locate the non-H atoms of the counter ion after which the complete structure determination and refinement were straightforward. The final results are published elsewhere (Kroon, Duisenberg & Peerdeman, 1984).

### Discussion

The novel algorithm for the application of phase reduction to non-centrosymmetric problems proves to be capable of determining a structure in a straightforward way. Several other tests on known structures showed that the *REDUC* program works satisfactorily in a variety of space groups,  $\text{CaC}_9\text{H}_{16}\text{O}_9 \cdot 3\text{H}_2\text{O}$ ,

$P1$ ,  $Z=1$  (Kanters, Jansma & Pontenagel, 1986);  $\text{C}_4\text{H}_6\text{O}_6$ ,  $P\bar{1}$ ,  $Z=2$  (Bootsma & Schoone, 1967);  $\text{C}_{17}\text{H}_{21}\text{NSn}$ ,  $P2_12_12_1$ ,  $Z=4$  (van Koten, Jastrzebski, Noltes, Pontenagel, Kroon & Spek, 1978);  $\text{C}_8\text{H}_7\text{O}_3\text{Br}$ ,  $P4_1$ ,  $Z=24$  (Bruins Slot, Krabbendam & Pontenagel, 1987);  $\text{C}_{12}\text{H}_{22}\text{O}_{11} \cdot 2\text{H}_2\text{O}$ ,  $P4_12_12$ ,  $Z=8$  (Kanters & van Bommel, 1987);  $\text{C}_{40}\text{H}_{36}\text{O}_5\text{P}_2 \cdot \text{CH}_2\text{Cl}_2$ ,  $P4_21c$ ,  $Z=4$  (Spek, de Ruiter & Pontenagel, 1982). Therefore *REDUC* can be seen as a convenient alternative to already existing phase-determining procedures. The major advantage of the algorithm is the possibility of including all available phase relations in a natural way; *i.e.* without distinguishing between two-phase seminvariants, triple products, quartets *etc.* The only parameter which determines whether a certain relation is of use to the procedure is the variance of the phase indication which is usually easily obtained from the available statistical formulae. Although no tests have been made up till now with relations other than those of the  $\Sigma_2$  type, we expect much of their incorporation, especially if independent non-zero estimates are available (in those cases the identities are expected to be much more valuable).

Another advantage of the phase reduction procedure is the clear and conveniently arranged book-keeping which is available in all stages of the process. At any time one can get insight into the current situation of the pedigree of phases; this makes *REDUC* a valuable troubleshooter if a routine application does not lead to a satisfactory solution of the phase problem. Especially the straightforward identification of seminvariants which are of crucial importance to the phasing procedure is a valuable tool for localizing weak links.

If desired one can include the phase of the seminvariant in question in the set of variables which are to be given different numerical values *via* a quadrant permutation or *via* a magic integer representation. This approach is certainly realistic since experience has shown that it is always possible to run *REDUC* in such a way that the number of important seminvariants is limited, so it will never be necessary to permute more than a few questionable seminvariants.

In conclusion we stress that *REDUC* is not only a transparent procedure which is capable of treating all kinds of phase relations in a mixed mode, but that it is also easy to introduce non-standard manipulations which may be required to cope with special problems.

### References

- BOOTSMA, G. A. & SCHOONE, J. C. (1967). *Acta Cryst.* **22**, 522-532.
- BRUINS SLOT, H. J., KRABBENDAM, H. & PONTENAGEL, W. M. G. F. (1987). In preparation.
- COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.
- COCHRAN, W. & DOUGLAS, A. S. (1955) *Proc. R. Soc. London Ser. A*, **227**, 486-500.

- FORTIER, S., DE TITTA, G. F., FRONCKOWIAK, M., SMITH, G. D. & HAUPTMAN, H. (1979). *Acta Cryst.* **B35**, 2062-2066.
- FORTIER, S., FRONCKOWIAK, M., SMITH, G. D., HAUPTMAN, H. & DE TITTA, G. (1978). International School of Crystallography, Erice, Italy, Supplement XXXIII-5-20.
- GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* **B26**, 274-285.
- KANTERS, J. A. & VAN BOMMEL, M. (1987). In preparation.
- KANTERS, J. A., JANSMA, S. & PONTENAGEL, W. M. G. F. (1986). To be published.
- KOTEN, G. VAN, JASTRZEBSKI, J. T. B. H., NOLTES, J. G., PONTENAGEL, W. M. G. F., KROON, J. & SPEK, A. L. (1978). *J. Am. Chem. Soc.* **100**, 5021-5028.
- KROON, J., DUISENBERG, A. J. M. & PEERDEMAN, A. F. (1984). *Acta Cryst.* **C40**, 645-647.
- PONTENAGEL, W. M. G. F. (1984). *Acta Cryst.* **A40**, 314-323.
- SPEK, A. L., DE RUITER, E. T. & PONTENAGEL, W. M. G. F. (1982). *Cryst. Struct. Commun.* **11**, 1869-1872.
- VAND, V. & PEPINSKY, R. (1956). *Z. Kristallogr.* **107**, 202-224.

*Acta Cryst.* (1987). **A43**, 64-69

## Isomorphous Replacement in Fiber Diffraction Using Limited Numbers of Heavy-Atom Derivatives

BY KEIICHI NAMBA AND GERALD STUBBS

*Department of Molecular Biology, Vanderbilt University, Nashville, TN 37235, USA*

(Received 17 February 1986; accepted 11 June 1986)

### Abstract

Multi-dimensional isomorphous replacement, analogous to isomorphous replacement in protein crystallography, can be used in fiber diffraction analysis to overcome the problems caused by the cylindrical averaging of the intensity data. Large numbers of heavy-atom derivatives are needed, however. A method is presented by which molecular structures may be determined using data from only one or two derivatives, similar to crystallographic single isomorphous replacement. Partial structure information may also be incorporated. Examples are given using data from oriented gels of tobacco mosaic virus, and possibilities for further applications are discussed.

### Introduction

Many important biological macromolecules, for example actin, myosin, tubulin, flagellin, and the coat proteins of some viruses, naturally form filamentous assemblies and have functions specific to those assemblies. Even in cases where it is possible to crystallize such molecules as monomers or small aggregates, it is important to know the molecular structure of the intact assembly in order to understand the function of the molecule. It is therefore necessary to use the methods of fiber diffraction.

Fiber diffraction from macromolecules has many aspects in common with protein crystallography, but there are also a number of major differences. The most important of these stems from the fact that although the filamentous particles in a fiber diffraction specimen are oriented with their long axes

approximately parallel, they are randomly oriented about those axes. As a consequence, the observed diffraction pattern is the cylindrical average of the diffraction pattern to be expected from one particle (in the absence of interference effects) or from a fully ordered array of particles (in the case of a crystalline fiber). Considerable information is lost in this averaging; for example, the effective number of observable diffraction data for tobacco mosaic virus (TMV) at 3 Å resolution is reduced by a factor about 2.5, and for the bacteriophage Pfl at the same resolution by 1.7 (Makowski, 1982). These factors are much higher for lower-symmetry systems such as microtubules.

The intensity of fiber diffraction can be written (Waser, 1955; Franklin & Klug, 1955)

$$I(R, l) = \sum_n G_{n,l}(R) G_{n,l}^*(R) \quad (1)$$

where  $l$  is the layer-line number,  $R$  is the reciprocal-space radius and  $n$  is the order of the Bessel functions that contribute to the complex Fourier-Bessel structure factor  $G$  (Klug, Crick & Wyckoff, 1958). Equation (1) can be compared with the crystallographic equation

$$I(h, k, l) = F_{hkl} F_{hkl}^*$$

If each  $G$  is known, a Fourier-Bessel transform of the  $G$  terms will give an electron density map, just as a Fourier transform of the  $F$  terms will give an electron density map for a crystal. Instead of two unknowns (the real and imaginary parts of  $F$ ), there are  $2N$  unknowns, where  $N$  is the number of terms ( $G$  terms) contributing to the sum in (1).  $N$  depends on the size and symmetry of the diffracting particle