

We wish to express our gratitude to Dr C. Tate for his assistance with this project. The support of the Science and Engineering Research Council is also gratefully acknowledged.

References

- BIANCHI, R., PILATI, T. & SIMONETTA, M. (1978). *Acta Cryst.* B34, 2157-2162.
- COLENS, A., DECLERCQ, J.-P., GERMAIN, G., PUTZEYS, J. P. & VAN MEERSSCHE, M. (1974). *Cryst. Struct. Commun.* 3, 119-122.
- DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1988). *Acta Cryst.* A44, 353-357.
- DECLERCQ, J.-P., GERMAIN, G. & KING, G. S. D. (1977). Fourth Eur. Crystallogr. Meet., Oxford. Abstracts A, pp. 279-280.
- DECLERCQ, J.-P., GERMAIN, G. & VAN MEERSSCHE, M. (1972). *Cryst. Struct. Commun.* 1, 13-15.
- GILMORE, C. J. (1984). *J. Appl. Cryst.* 17, 42-46.
- HAUPTMAN, H. (1964). *Acta Cryst.* 17, 1421-1433.
- KARLE, J. (1968). *Acta Cryst.* B24, 182-186.
- KROON, J. & KRABBENDAM, H. (1970). *Acta Cryst.* B26, 312-314.
- MAIN, P. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 97-105. Copenhagen: Munksgaard.
- ODA, T., NAYA, S. & NITTA, I. (1967). *Mem. Osaka Univ.* 16, 19-36.
- OLTHOF, G. J., SINT, L. & SCHENK, H. (1979). *Acta Cryst.* A35, 941-946.
- SZEIMES-SEEBACH, U., HARNISCH, J., SZEIMES, G., VAN MEERSSCHE, M., GERMAIN, G. & DECLERCQ, J.-P. (1978). *Angew. Chem. Int. Ed. Engl.* 17, 848-850.
- THIESSEN, W. E. & BUSING, W. (1974). *Acta Cryst.* A30, 914.
- YAO, J.-X. (1983). *Acta Cryst.* A39, 35-37.

Acta Cryst. (1988). A44, 353-357

On the Application of Phase Relationships to Complex Structures. XXVI. Developments of the Sayre-Equation Tangent Formula

BY T. DEBAERDEMAEKER

*Sektion für Röntgen- und Elektronenbeugung, Universität, Oberer Eselsburg, D-7900 Ulm,
Federal Republic of Germany*

AND C. TATE AND M. M. WOOLFSON

Department of Physics, University of York, York YO1 5DD, England

(Received 8 August 1987; accepted 4 January 1988)

Abstract

The Sayre-equation tangent formula (SETF) develops sets of phases tending to satisfy Sayre's equations for both large and small normalized structure factors. There are two components in the SETF, corresponding to contributions from phase triplets and quartets respectively. The development of objective algorithms for properly weighting these components and for gradually building up the quartet contribution has enabled the SETF, within the procedure SAYTAN, to be incorporated into MULTAN87, the latest version of the package. Examples of tests of MULTAN87 and its use in solving unknown structures are given.

Introduction

In a previous paper Debaerdemaeker, Tate & Woolfson (1985) described the theory of a new tangent formula which had the property of developing phases tending to satisfy a system of Sayre equations. A particular Sayre equation can be written in terms

of normalized structure factors as

$$E(\mathbf{h}) = [f(\mathbf{h})/g(\mathbf{h})V] \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}) \quad (1)$$

where $f(\mathbf{h})$ and $g(\mathbf{h})$ are the scattering factors for atoms and squared-electron-density atoms respectively and V is the volume of the unit cell.

Phases are sought to achieve minimization of

$$R = \sum_{\mathbf{h}} \left| E(\mathbf{h}) - [K/g(\mathbf{h})] \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}) \right|^2 \quad (2)$$

where K is an overall scaling factor which compensates for partial data in the \mathbf{k} summations and the $g(\mathbf{h})$ can be determined on theoretical grounds.

The minimization condition is

$$\partial R / \partial \varphi(\mathbf{l}) = 0 \quad \text{for all } \mathbf{l};$$

application of this to (2), followed by some algebraic manipulation, gives the Sayre-equation tangent formula (SETF)

$$\varphi(\mathbf{l}) = \text{phase of } [t(\mathbf{l}) - 2Kq(\mathbf{l})] \quad (3)$$

where

$$t(\mathbf{l}) = \sum_{\mathbf{h}} [1/g(\mathbf{l}) + 1/g(\mathbf{h}) + 1/g(\mathbf{h}-\mathbf{l})] E(\mathbf{h}) E(\mathbf{l}-\mathbf{h}) \quad (4)$$

and

$$q(\mathbf{l}) = \sum_{\mathbf{k}} [1/g(\mathbf{k})^2] \sum_{\mathbf{h}} E(\mathbf{h}) E(\mathbf{k}-\mathbf{h}) E(\mathbf{l}-\mathbf{k}). \quad (5)$$

A reasonable K is obtained from the condition

$$\partial R / \partial K = 0,$$

leading to

$$K = \frac{\sum_{\mathbf{h}} \sum_{\mathbf{k}} [1/g(\mathbf{h})] E(\bar{\mathbf{h}}) E(\mathbf{k}) E(\mathbf{h}-\mathbf{k})}{\sum_{\mathbf{h}} [1/g(\mathbf{h})^2] \left| \sum_{\mathbf{k}} E(\mathbf{k}) E(\mathbf{h}-\mathbf{k}) \right|^2} \\ = \frac{\sum_{\mathbf{h}} E(\bar{\mathbf{h}}) t(\mathbf{h})}{\sum_{\mathbf{h}} E(\bar{\mathbf{h}}) q(\mathbf{h})}. \quad (6)$$

The system of Sayre equations can include those for which $E(\mathbf{h})$ in (1) is zero or very small. Such equations can be included in the expression for R and so it is possible to use small, ideally zero, E values in the process of determining the phases of the large E 's, in terms of which the structure will be defined. However, experience showed that better results were obtained from the SETF if the terms in the summation of (2) were differently weighted for large and small E 's giving

$$R_w = \sum_{\mathbf{h}} m \left| E(\mathbf{h}) - [K/g(\mathbf{h})] \sum_{\mathbf{k}} E(\mathbf{k}) E(\mathbf{h}-\mathbf{k}) \right|^2 \quad (7)$$

where $m = 1$ for large E 's and some other value for small ones.

Subsequent to the 1985 paper of Debaerdemaeker, Tate & Woolfson the properties of the SETF have been more fully explored and an optimized procedure, applicable to the complete range of structures, has been devised. This has made it possible to include the SETF in the *MULTAN* package and, in the newly created *MULTAN87*, it appears as the *SAYTAN* procedure. Here we describe the developments of the SETF, how it is incorporated in *MULTAN87* and give examples of its performance.

SAYTAN

With the weighting scheme introduced in (7) it is possible to divide the quantity $q(\mathbf{l})$ into two parts so that

$$q(\mathbf{l}) = q^s(\mathbf{l}) + mq^w(\mathbf{l}). \quad (8)$$

The left-hand side and the general term on the right-hand side of (5) have indices corresponding to a

quartet $\mathbf{l}, \bar{\mathbf{h}}, \mathbf{h}-\mathbf{k}$ and $\mathbf{k}-\mathbf{l}$ with a cross term of index \mathbf{k} . The separate components $q^s(\mathbf{l})$ and $q^w(\mathbf{l})$ arise from quartets for which $|E(\mathbf{k})|$ are 'strong' or 'weak' respectively and the m appearing in (8) is the same as that introduced in (7). From the SETF we see that

$$\varphi(\mathbf{l}) = \text{phase of } [t(\mathbf{l}) - 2Kq^s(\mathbf{l}) - 2mKq^w(\mathbf{l})]. \quad (9)$$

For a correct set of phases it can be shown that

$$2Kq^s(\mathbf{l}) = \sum_{\mathbf{h}} [1/g(\mathbf{h}) + 1/g(\mathbf{h}-\mathbf{l})] E(\mathbf{h}) E(\mathbf{l}-\mathbf{h}), \quad (10)$$

so that the SETF then reduces to

$$\varphi(\mathbf{l}) = \text{phase of } \left\{ [1/g(\mathbf{l})] \sum_{\mathbf{h}} E(\mathbf{h}) E(\mathbf{l}-\mathbf{h}) - 2Km \sum_{\mathbf{k}} [1/g(\mathbf{k})^2] \sum_{\mathbf{h}} E(\mathbf{h}) E(\mathbf{k}-\mathbf{h}) E(\mathbf{l}-\mathbf{k}) \right\} \quad (11)$$

and the summation over \mathbf{k} is restricted to terms for which $|E(\mathbf{k})|$ is small (ideally zero).

Strictly the form of (11) only applies for a set of correct phases but, in practice, it is found that this reduced form is almost as effective as the full form (3) in refining random phases. The reduced form also has the advantage that computation is nearly twice as fast.

At the time of reporting the first successful results from *SAYTAN* it was necessary to find a good value for m by trial and error for each structure. Since then, experience has enabled us to incorporate into the program a recipe for the optimization of m . *SAYTAN* is found to work best when the ratio

$$r = \text{NSRPSI} / \text{NSTR} = 0.5, \quad (12)$$

where NSRPSI and NSTR are the number of 'weak' and 'strong' relationships in the \sum_2 list of *MULTAN*. As a first step to achieving this ratio of weak to strong relations *MULTAN87* sets

$$\text{NZRO} = \text{NDET} / 2, \quad (13)$$

where NZRO and NDET are the numbers of weak and strong reflexions respectively, and then, if the resulting r is not very close to 0.5, the program calculates a simple proportional adjustment to NZRO which gives a better r value. Finally the scaling factor for the weak quartets is calculated from

$$m = 5 / (1 + 8r). \quad (14)$$

The weighting factor K given in (6) can also be written as

$$K = T / Q \quad (15)$$

where, from (6), T and Q correspond to a sum of triples and a sum of quartets respectively. In principle, according to the theory, the value of K should be varied from cycle to cycle according to the current value of the ratio T/Q but, in practice, this does not

work. The SETF is often used with initially random phases and the values of T and Q can, by chance, take on a wide range of possible values and so give a very large value to K , which distorts or destroys the phase-refining process. To introduce stability into the system it is found better to increase gradually the influence of the second term of the modified SETF [(11)] as the refinement progresses. This is done by taking

$$K = T_{\text{act}}/Q_{\text{theor}} \quad (16)$$

where T_{act} is the actual sum of triples for the current phase values and Q_{theor} is a theoretical estimate of the sum of quartets. Thus

$$Q_{\text{theor}} = \sum_{\mathbf{h}} \left\langle [1/g(\mathbf{h})^2] \left| \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}) \right|^2 \right\rangle_{\mathbf{h}}. \quad (17)$$

We divide Q_{theor} into two components corresponding to $|E(\mathbf{h})|$ being either 'strong' or 'weak' so that

$$Q_{\text{theor}} = Q_{\text{theor}}^s + Q_{\text{theor}}^w \quad (18)$$

For $|E(\mathbf{h})|$ strong the components in the summation over \mathbf{k} will tend to line up and it is a sufficiently good approximation to write

$$Q_{\text{theor}}^s \approx \sum_{\mathbf{h}} [1/g(\mathbf{h})^2] \left| \left\langle \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}) \right\rangle_{\mathbf{h}} \right|^2. \quad (19)$$

The assumption that individual triples obey the Cochran distribution gives

$$Q_{\text{theor}}^s \approx \sum_{\mathbf{h}} \left\{ [1/g(\mathbf{h})] \sum_{\mathbf{k}} |E(\mathbf{k})E(\mathbf{h}-\mathbf{k})| I_1(\kappa)/I_0(\kappa) \right\}^2 \quad (20)$$

where $\kappa = 2\sigma_3\sigma_2^{-3/2}|E(\mathbf{h})E(\mathbf{k})E(\mathbf{h}-\mathbf{k})|$, as usual.

For $|E(\mathbf{h})|$ small, the quantity

$$\left| \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}) \right|^2$$

is that which appears as part of the PSIZERO figure of merit in *MULTAN*. It is known that for a correct set of phases the value of this quantity is approximately the expectation value for random phases so that

$$Q_{\text{theor}}^w \approx \sum_{\mathbf{h}} [1/g(\mathbf{h})^2] \sum_{\mathbf{k}} |E(\mathbf{k})E(\mathbf{h}-\mathbf{k})|^2. \quad (21)$$

A combination of results (20) and (21) gives the desired estimate of Q_{theor} .

It is the objective algorithms for choosing a suitable number of small E 's and determining the weights m and K which enable *SAYTAN* to be used routinely and hence to be part of a standard package.

Quartet terms have been employed actively in previous direct-methods procedures, for example in

SHELXTL (Sheldrick, 1981), *SIMPEL* (Schenk & Kiers, 1984) and in *MITHRIL* (Gilmore, 1984) but there are significant differences between these procedures and *SAYTAN* in regard to particular characteristics of the quartet contribution and its theoretical basis. In previous schemes, quartets have been invoked on the basis of various probability distributions for four-phase invariants and the resulting phase indications combined with triplet indications through a weighted average. In contrast, SETF is not a probabilistic formula but is derived algebraically from a system of Sayre equations. A set of phases satisfying the SETF minimizes the least-squares residual for the system of Sayre equations and three significant details result from the algebra. First, the optimum weighting for quartets relative to triplets is given algebraically as $-2T/Q$ in (9) and (15). Second, the quartets which are selected belong to the special class which can be constructed from triplets by pairing the latter 'back-to-back' through a common cross term. This offers two practical advantages: such quartets can readily be found from the SIGMA2 list, and their contribution to the tangent formula can be computed economically. Third, these quartets fall into two distinct subsets, those formed with the largest E 's as cross terms and those formed with the smallest E 's as cross terms. This has the effect that the equations tend to develop phase sets which satisfy the smallest magnitudes as well as the largest, and it is reasonable to suppose that this is advantageous in terms of efficient information content, because it is often said that the reflexions carrying most structural information are the very strong and the very weak.

The general structure of *MULTAN87*

MULTAN87 can be used in five modes, the first being the default. These are (i) SWTF with phase permutation; (ii) SWTF in the *RANTAN* mode; (iii) *SAYTAN* with phase permutation; (iv) *SAYTAN* in the *RANTAN* mode; (v) X - Y . SWTF uses the conventional tangent formula but with the statistical weights suggested by Hull & Irwin (1978). The phase permutation approach is that which begins with a small base of phases, some fixing origin and enantiomorph and the others varied from trial to trial by permutation, and builds up a body of phase estimates by a 'bootstrapping' process. The *RANTAN* approach (Yao, 1981) is that for which all phases are given initially random values and some weight, and all terms may be included in the tangent formula from the very beginning of the process. The X - Y method is that described by Debaerdemaeker & Woolfson (1983).

MULTAN87 follows the pattern of previous versions by allowing intervention by the user and the program documentation gives advice about this.

Tests of *MULTAN87*

There exists a set of standard structures, offering a variety of space groups and known for presenting problems to direct methods, which are used to test new direct-methods procedures. A partial list of the tests, with structures referred to by brief code names, and the results obtained is now given.

The number of phase sets generated by *MULTAN87* by default is dependent upon the complexity of the given structure; in the following examples, unless otherwise indicated, 40 trials were made in modes (i) and (iii) and 50 trials in modes (ii), (iv) and (v). It should be mentioned that when *MULTAN87* finds a set of phases with very favourable figures of merit it ceases to generate phase sets and completes the process of presenting the structure.

CORTISONE [Declercq, Germain & Van Meerssche (1972): $C_{21}H_{28}O_5$, $P2_12_12_1$, $Z = 4$]

Solved easily in all five modes.

MUCCAR [Bianchi, Pilati & Simonetta (1978): $C_{13}H_{11}N$, $P1$, $Z = 2$]

Modes (i), (ii) and (v) each revealed less than half the structure after the default numbers of trials indicated above. The results obtained with *SAYTAN* provide an example of the effect, mentioned above, of adjusting NZRO (the number of weak reflexions) to satisfy (12). With the default setting for NZRO [equation (13): $NDET/2 = 156$], mode (iii) gave a perfect result (in the sense that all 28 atoms were revealed in a 'map') on the tenth trial; but after NZRO was reset to the value 240, which had been suggested by the program as an optimum value, the first trial produced a perfect result. In mode (iv), however, the effect of adjusting NZRO was not apparent, because $NZRO = 156$ gave a perfect result for set 3 while $NZRO = 240$ gave a perfect result for set 2.

NORG [Roques, Rossi, Declercq & Germain (1980): $C_{16}H_{20}ClNO_3$, $P1$, $Z = 2$]

The usual numbers of trials in modes (i), (ii) and (v) revealed no useful fragment of the structure whereas 50 trials with *SAYTAN* in mode (iv) produced four solutions showing 40 of the 42 independent atoms. In mode (iii), *SAYTAN* was not so successful; after 40 trials, the best set in terms of the conventional figures of merit was set 24 ($ABSFORM = 0.933$, $PSIZERO = 1.197$, $RESID = 12.59$) but the *E* map showed no convincing part of the structure.

TURSCH 11 [Brackman, Dalozze, Dupont, Tursch, Declercq, Germain & Van Meerssche (1981): $C_{15}H_{24}O_4$, $P2_1$, $Z = 4$]

Nothing useful was obtained from 25 trials in mode (i) and 50 trials in each of modes (ii) and (v). With

SAYTAN in mode (iv), the 18th trial revealed all of the 38 independent atoms. As with *NORG*, *SAYTAN* in mode (iii) was not very effective: nothing useful was obtained from 25 trials.

MUNICH 4 [Szeimies-Seebach, Harnisch, Szeimies, Van Meerssche, Germain & Declercq (1978): $C_{27}H_{22}O$, C_c , $Z = 4$]

The usual default runs in modes (i), (ii) and (v) gave less than half the structure in each case but *SAYTAN* produced a perfect solution for set 4 in mode (iii) and another perfect solution for set 2 in mode (iv).

CINOBUFAGIN [Declercq, Germain & King (1977): $C_{26}H_{27}O_6$, $P2_12_12_1$, $Z = 4$]

This structure is notoriously difficult. Nothing useful was obtained from the usual trials in modes (i), (ii) and (v). The phase-permutation modes (i) and (iii) generate only 12 trials by default for this structure and for *SAYTAN* in mode (iii) there was nothing useful in these 12 sets; but in mode (iv), *SAYTAN* produced a perfect solution for set 4.

AZET [Colens, Declercq, Germain, Putzeys & Van Meerssche (1974): $C_{21}H_{16}ClNO$, $Pca2_1$, $Z = 8$]

The most complete result was obtained in mode (iii) when the best set from 50 trials revealed one of the two independent molecules.

Concluding remarks

There is a great deal of serendipity in the development and application of direct methods. The *X-Y* method (Debaerdemaeker & Woolfson, 1983) has no rational basis, was discovered by making an error and yet seems to be more effective than the standard tangent formula for which there is a strong theoretical basis. In addition, while a comparison of the effectiveness of different direct-method approaches is possible on a statistical basis, it cannot be forecast with certainty which approaches will succeed and which will fail for a particular structure.

There is little doubt on the basis of our present experience that the SETF, as embodied in *SAYTAN*, is the most generally effective tool available in direct methods at present. However, in view of the foregoing, it is desirable to have as many procedures available as possible when it comes down to the practicalities of solving crystal structures. For this reason we intend to re-incorporate in some future version of *MULTAN* other procedures, notably *YZARC* (Baggio, Woolfson, Declercq & Germain, 1978) and *MAGEX* (Zhang & Woolfson, 1982), which were in some previous *MULTAN* releases.

We are grateful to the Science and Engineering Research Council, the European Economic Com-

munity and the Deutsche Forschungsgemeinschaft for their support of this work.

References

- BAGGIO, R., WOOLFSON, M. M., DECLERCQ, J.-P. & GERMAIN, G. (1978). *Acta Cryst.* **A34**, 883-892.
- BIANCHI, R., PILATI, T. & SIMONETTA, M. (1978). *Acta Cryst.* **B34**, 2157-2162.
- BRACKMAN, J. C., DALOZE, D., DUPONT, A., TURSCH, B., DECLERCQ, J.-P., GERMAIN, G. & VAN MEERSSCHE, M. (1981). *Tetrahedron*, **37**, 179-186.
- COLENS, A., DECLERCQ, J.-P., GERMAIN, G., PUTZEYS, J. P. & VAN MEERSSCHE, M. (1974). *Cryst. Struct. Commun.* **3**, 119-122.
- DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* **A41**, 286-290.
- DEBAERDEMAEKER, T. & WOOLFSON, M. M. (1983). *Acta Cryst.* **A39**, 193-196.
- DECLERCQ, J.-P., GERMAIN, G. & KING, G. S. D. (1977). Fourth Eur. Crystallogr. Meet., Oxford. Abstracts A, pp. 279-280.
- DECLERCQ, J.-P., GERMAIN, G. & VAN MEERSSCHE, M. (1972). *Cryst. Struct. Commun.* **1**, 13-15.
- GILMORE, C. J. (1984). *J. Appl. Cryst.* **17**, 42-46.
- HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* **A34**, 863-870.
- ROQUES, R., ROSSI, J. C., DECLERCQ, J.-P. & GERMAIN, G. (1980). *Acta Cryst.* **B36**, 1589-1593.
- SCHENK, H. & KIERS, C. T. (1984). In *Methods and Applications in Crystallographic Computing*, edited by S. R. HALL & T. ASHIDA, pp. 96-105. Oxford: Clarendon Press.
- SHELDRIK, G. M. (1981). *SHELXTL. A System of Computer Programs for the Solution and Refinement of Crystal Structures*. Nicolet Instruments, Madison, Wisconsin, USA.
- SZEIMIES-SEEBACH, U., HARNISCH, J., SZEIMIES, G., VAN MEERSSCHE, M., GERMAIN, G. & DECLERCQ, J.-P. (1978). *Angew. Chem. Int. Ed. Engl.* **17**, 848-850.
- YAO, J.-X. (1981). *Acta Cryst.* **A37**, 642-644.
- ZHANG, S.-H. & WOOLFSON, M. M. (1982). *Acta Cryst.* **A38**, 683-685.

Acta Cryst. (1988). **A44**, 357-368

Refinement at 1.4 Å Resolution of a Model of Erabutoxin b: Treatment of Ordered Solvent and Discrete Disorder

BY JANET L. SMITH*

Department of Biochemistry and Molecular Biophysics and Howard Hughes Medical Institute, Columbia University, New York, NY 10032, USA, and Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375, USA

P. W. R. CORFIELD

Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA

WAYNE A. HENDRICKSON

Department of Biochemistry and Molecular Biophysics and Howard Hughes Medical Institute, Columbia University, New York, NY 10032, USA, and Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375, USA

AND BARBARA W. LOW

Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA

(Received 10 August 1987; accepted 6 January 1988)

Abstract

The latter stages in the refinement of the protein erabutoxin b are described. The crystal structure of the 62-residue protein has been refined to a conventional *R* factor of 0.144 by stereochemically restrained least-squares methods using diffraction data to a limit of 1.4 Å spacings. Emphasis was placed on determining as accurately as possible the solvent

structure and the structures of heterogeneous groups in the protein. The final model includes two conformers for each of seven side chains and for an octapeptide segment. A total of 111 sites for water molecules have been located as well as one sulfate ion with a total of 68 site occupancies. 65 of the solvent sites overlap either with protein atoms belonging to groups in two alternative conformations or with other solvent sites. Dual protein conformers and overlapping solvent sites were both included in the least-squares refinement. Individual thermal and occupancy parameters were refined for solvent molecules. An

* Author to whom correspondence should be addressed at Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA.