Acta Cryst. (1979). A35, 622–626

On the Application of Phase Relationships to Complex Structures. XVII. When *MULTAN* Fails

BY J. P. DECLERCQ AND G. GERMAIN

Laboratoire de Chimie Physique et de Cristallographie, Université de Louvain, Bâtiment Lavoisier, 1 place Louis Pasteur, 1348 Louvain-la-Neuve, Belgium

AND M. M. WOOLFSON

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

(*Received* 15 *January* 1979; *accepted* 16 *February* 1979)

Abstract

Two new multisolution direct-methods procedures are described: *MAGIC*, which employs the magic-integer concept and *YZARC* which refines initially random sets of phases by a least-squares approach. Each procedure produces several sets of phases for a number of reflexions, usually in the range 35–100. These are then extended by the tangent formula but with the constraint that the basis phases are not allowed to change until the final cycle. It is shown that for difficult structures these methods, which deal simultaneously with many phase relationships, may have intrinsic advantages over the *MULTAN* procedure. Examples of their use are given.

Introduction

For the solution of structures containing up to 70 or 80 atoms in the asymmetric unit crystallographers are increasingly relying on the use of multisolution direct-methods computer programs, such as *MULTAN* or *SHELX*, and about one-half of all structures are now solved in this way. However, despite their many successes, these methods are fallible and one structure with, say, 25 independent atoms may be resistant to solution while a 50- or 60-atom problem may be solved in a single computer run.

When a straightforward application of *MULTAN* is unsuccessful then the veteran user may try one or two ruses to persuade it to work; artificially halving or doubling the temperature factor is sometimes efficacious but any other device which causes the program to follow a different path is worth trying.

Recently we have developed two new computer-based procedures which can be used in conjunction with the *MULTAN* 78 package and which have been successful for several structures when *MULTAN* itself has failed. One of these, designated *MAGIC* 78, is closely modelled on that previously described by Declercq, Germain & Woolfson (1975) while the other, *YZARC* 78, uses the random-approach concept developed by Baggio, Woolfson, Declercq & Germain (1978). A brief description of these two procedures, and examples of their application, will now be given.

MAGIC 78

Most of the details of this procedure have already been described by Declercq *et al.* (1975) and they will not be repeated here. Briefly, one uses the magic-integer concept to find a number (typically 50–80) of trial sets of phases for 35–50 reflexions, extends each set by the *FASTAN* routine of *MULTAN* and then follows the *MULTAN* pathway thereafter.

In *MAGIC* 78 there is a small but, we believe, important modification involving the phase-extension process. The original magic-integer-determined phases, from which the phase extension is initiated, are not modified except in the final cycle of the application of the tangent formula. The reasons for this will be discussed later.

The program is very straightforward to apply and the number of user-designated parameters is quite small. The use of suggested standard parameters and default values seems to detract little from the effectiveness of the method.

Some applications of *MAGIC*

For our first example we take a structure (unpublished) which is an extract of coral and was kindly provided by B. Tursch of the Université Libre de Bruxelles. This we have code-named Tursch 5 and the basic data are: formula $C_{15}H_{24}O_4$, $P2_12_12_1$, $Z = 4$, $a = 19{\cdot}625$, $b = 8{\cdot}616$, $c = 8{\cdot}536$ Å.

0567-7394/79/040622-05$01.00

A routine run of *MULTAN* 78 was unsuccessful, giving fragments which could not be developed.

For the *MAGIC* application nine primary reflexions were represented by the elements of {467} (*xyz*) and there were also 24 secondary reflexions. Because the *CONVERGE* routine of *MULTAN* fixed the enantiomorph with a general reflexion (restricted to two of the four possible quadrant values) this run of *MAGIC* took only three origin-defining reflexions of fixed value in the primary set. The method is not sensitive to the lack of definition of enantiomorph and the loss of one possible reflexion from the primary set did not seem to matter.

The 36 reflexions of the combined primary and secondary sets were linked by 20 triple-phase relationships (other than the 24 giving secondary reflexions from primary) and a ψ map was calculated with these. The 80 (default value) highest peaks in the ψ map were translated into trial phases for the 33 reflexions (all, that is, except the origin-defining ones) and these were then refined by a parameter-shift method based on satisfying the available total of 44 relationships to the maximum possible extent.

Each of the 80 phase sets was extended to 200 phases by the use of the tangent formula and the *MULTAN* figures of merit were found. The *E* map corresponding to the highest combined figure of merit (CFOM) was calculated and the form of the line-printer plot, which clearly showed the molecule, is reproduced in Fig. 1. It is interesting to note that of the 80 sets of phases developed by *MAGIC* no less than 10 corresponded to this correct solution. The computer time requirement was quite modest; the stage up to obtaining the 80 sets of 36 phases took 2 min 24 s and the phase extension to 200 phases 5 min 23 s. These times, and all subsequent times quoted, are for an IBM 370/158 computer.

For the second example we take the unpublished structure of 1-(1-hydroxyethyl)hexahelicene, $C_{28}H_{20}O$. The space group is $P2_12_12_1$ with $Z = 4$ and $a = 8{\cdot}541$, $b = 14{\cdot}100$, $c = 15{\cdot}812$ Å.

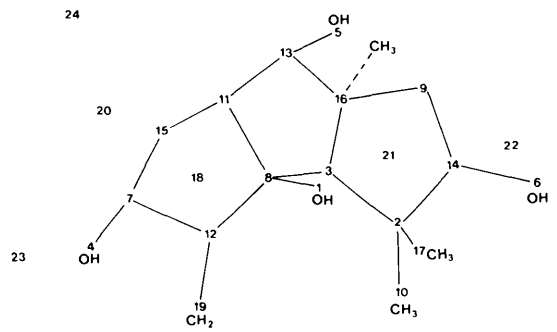

Fig. 1. A reproduction of the *MAGIC* 78 line-printer output for the set of phases with highest CFOM (2·700) for Tursch 5. Full lines show the part of the structure given by the plot and the dashed line indicates a missing atom. Isolated numbers correspond to false peaks. The individual figures of merit were: ABSFOM = 1·065, PSIZERO = 1·296 and RESID = 16·45.

The origin and enantiomorph were fixed by four special reflexions and nine primary reflexions had phases represented by the elements of {457}(*xyz*). There were 21 secondary reflexions and 25 relationships contributing to the ψ map. The 50 highest peaks were selected and after parameter-shift refinement the 50 sets of 34 phases were extended to 290 phases by the tangent formula.

In this case the set of phases with an outstandingly high CFOM (2·775) did not give the solution. On the basis of examining the individual figures of merit, as well as the CFOM, four other phase sets were investigated. One gave the plot shown in Fig. 2. Two atoms are missing but otherwise the molecule is clearly shown.

These two examples are typical of what is now a number of applications of *MAGIC*. No attempt has been made to compare *MAGIC* and *MULTAN* but the general impression has been formed that *MAGIC* is of comparable overall effectiveness. Certainly it can solve some structures which are not solved by a routine run of *MULTAN*.

YZARC 78

In a paper by Baggio *et al.* (1978) it was shown that phases could be refined by applying least-squares techniques to a set of linear equations representing the triple-phase relationships. It was further shown that one could often 'refine' from a completely random set of phases to an essentially correct solution. These results and observations have been incorporated into a systematic procedure for solving crystal structures, the steps of which will now be described.

(1) Assume that a routine run of the *MULTAN* system has failed.

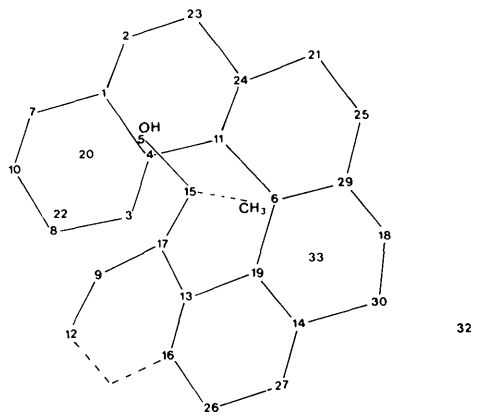

Fig. 2. A reproduction of the *MAGIC* 78 line-printer output for the set of phases giving the structure of 1-(1-hydroxyethyl)hexahelicene. All but two of the atoms in the structure are revealed. The individual figures of merit were: ABSFOM = 1·0554, PSIZERO = 232·1 (old version), RESID = 50·75 and CFOM = 1·960.

(2) The bottom 100 reflexions in the *CONVERGE* map, and the associated relationships, constitute the system which is initially used to obtain trial sets of phases. This system of equations is set up in the matrix form

$$\mathbf{A}\varphi = \mathbf{b} \qquad (1)$$

and the matrix $(\mathbf{A}^T\mathbf{A})^{-1}$ is calculated. The actual number of reflexions may be chosen by the user, with 100 as the default value.

(3) By means of a pseudo-random number generator giving a uniform distribution in the range 0 to 1, M sets of 100 random phases (in cycles) are generated. Each set is refined by the iterative application of

$$\mathbf{n} = \text{integral part } (\mathbf{A}\varphi + \mathbf{c}), \qquad (2)$$

$$\varphi = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{n} + \boldsymbol{\varepsilon} - \mathbf{c}), \qquad (3)$$

where $\mathbf{c}$ is a column vector whose elements are the constant-angle parts of the phase relationships as generated by *MULTAN*. With α as the deviation (in the range 0·5 to −0·5) of each current estimate of a phase relationship from the nearest integer then $\boldsymbol{\varepsilon}$ is the column vector with elements $4\alpha^3$. These $4\alpha^3$ terms provide a convenient way of weighting the individual equations in each cycle without calculating a new inverse matrix each time (Baggio *et al.*, 1978). If there might be a problem in defining the enantiomorph then, after the refinement is completed with $4\alpha^3$, the weighting term is replaced by $\cos^{-1}[I_1(\alpha)/I_0(\alpha)] \times$ (sign α). The default action is to conclude the refinement in this way and it never seems to be harmful.

(4) Extend each set of phases by the tangent formula, only refining the initial phases in the final cycle. Then determine figures of merit and complete as with *MULTAN*.

This program is, if anything, even more automatic than *MAGIC* and is usually run with the default parameters.

Some applications of *YZARC*

Recently *YZARC* has been used to solve the structure of oxa[3.1.1]propellane 4, $C_{27}H_{22}O$ (Szeimies-Seebach, Szeimies, Van Meerssche, Germain & Declercq, 1979), the molecular form of which is illustrated in Fig. 3(*a*). The crystal data are: space group *Cc*, $Z = 4$, $a = 12{\cdot}987$, $b = 10{\cdot}025$, $c = 15{\cdot}095$ Å, $\beta = 95{\cdot}43°$.

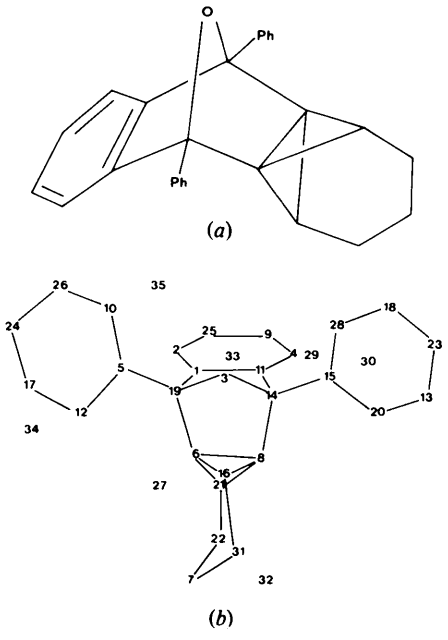

Fig. 3. (*a*) The form of the molecule of oxa[3.1.1]propellane 4. (*b*) The *YZARC* solution with the highest CFOM. Peaks 22 and 31 exactly overlapped on the line-printer output but, for clarity, they have been slightly separated here. The individual figures of merit were: ABSFOM = 1·152, PSIZERO = 1·217, RESID = 18·95 and CFOM = 2·475.

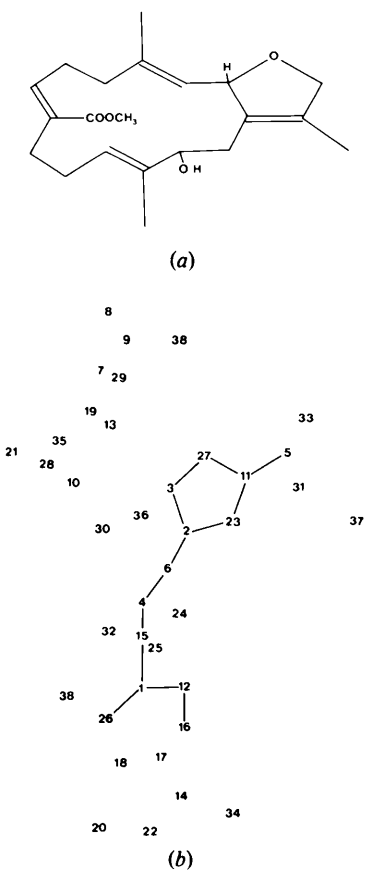

Fig. 4. (*a*) The form of the molecule of sarcoglaucol. (*b*) The *YZARC* solution with the highest CFOM. The 13-atom fragment was eventually developed to the complete structure. The individual figures of merit were: ABSFOM = 0·911, PSIZERO = 189·2 (old version), RESID = 23·61 and CFOM = 2·255.

The bottom of the *CONVERGENCE* map gave 480 relationships relating 100 reflexions. One hundred sets of random phases were generated and refined by the least-squares approach. These were then extended by application of the tangent formula to 212 phases.

The set with the highest CFOM gave the plot shown in Fig. 3(*b*) which revealed the complete structure.

As a second example we shall take another extract from coral, sarcoglaucol (Fig. 4*a*), $C_{21}H_{30}O_4$ (Albericci, Braekman, Daloze, Tursch, Declercq, Germain & Van Meerssche, 1978), space group $P2_1$ with $a = 11{\cdot}256$, $b = 14{\cdot}180$, $c = 6{\cdot}319$ Å, $\beta = 98{\cdot}48°$, $Z = 2$.

A routine run of *MULTAN* 77 (the version available at that time) gave nothing recognizable although attempts were made to develop fragments from some *E* maps. The bottom of the *CONVERGE* map gave 550 relationships for 100 phases. One hundred sets of random phases were generated and refined in 18 min 2 s and then extended to 200 phases by *FASTAN* in another 5 min 3 s. The set with the highest CFOM (2·255) gave the 13-atom fragment shown in Fig. 4(*b*) which was easily extended to reveal the complete structure.

These two examples are typical of several successful applications of *YZARC*. Once again no direct comparison of efficiency with *MULTAN* has been made but we shall see that there are good reasons for thinking that it may have inherent advantages over the *MULTAN* approach.

General comments

It has previously been noted that, because of the use of tangent-formula phase refinement, some structures cannot be solved (Lessinger, 1976; Hull & Irwin, 1978). Even a large set of correct phases fed into tangent-formula refinement can degenerate to a set which will not reveal the structure at all. A common end product of such abortive tangent-formula refinements is a set of phases giving rise to one or two very large peaks or a 'sum of two structures', *i.e.* the structure sought plus its own enantiomorph, which is usually undecipherable.

The underlying processes of *MAGIC* and *YZARC* depend on the simultaneous treatment of a large number of triple-phase relationships. In selecting peaks from a ψ map it is likely that some may well correspond to sets of phases with poor enantiomorph discrimination, *i.e.* which give a sum of two structures, but others, or perhaps only one other, will correspond to almost correct phases with good enantiomorph discrimination. Similarly, in each cycle of refinement by the least-squares method in *YZARC* all phases change simultaneously. This gives much less tendency to switch from an enantiomorph-defining set of phases than does the tangent formula where phases are refined one at a time and a slow drift process is much more likely to take place. Any tendency to enantiomorph loss which may be present with the linear equations can be corrected (Baggio *et al.*, 1978) and the modification of the tangent formula, as recommended by Hull & Irwin (1978), does as much for tangent-formula refinement.

However, what we *have* established is that if tangent-formula refinement is carried out under the condition that a large group of almost correct enantiomorph-defining phases is kept fixed then refinement of the remainder will be consistent with the enantiomorph of the basis set. One final cycle of refinement in which the basis set is allowed to change will relax all the phases to better self consistency without affecting the enantiomorph definition. This is the procedure which we have used in both *MAGIC* and *YZARC*. What has been said about the effect of the fixed basis set in providing an anchor to prevent drift of phases and loss of enantiomorph will equally prevent drift to the one-or-two-large-peaks situation which is common with some space groups.

There is yet another way in which the *MAGIC* and *YZARC* procedures are improvements on *MULTAN*. In the *MULTAN* process one begins with a small starting set of, perhaps, ten reflexions to which phase values are assigned. Then, *via* a convergence map, new phases are developed individually, each determined by a few, initially usually one, two, or three, phase relationships. For complicated structures these phase relationships may sometimes be rather poor and give a completely wrong pattern of *relative* phases. This may not only be due to individual randomly poor relationships; it sometimes seems that phase relationships conspire together in groups to lead one astray. For ergocalciferol, for example, a development of a convergence map by symbolic addition revealed no inconsistencies for a considerable way up the *CONVERGE* map, *i.e.* where there were two or more indications for a new phase they were identical. Nevertheless, the determined relative phases of the initial set were badly in error and subsequent tangent-formula refinement did nothing to rectify the damage. Where this occurs in *MULTAN*, it is true for each and every starting point. *MULTAN* is then not so much a multiple-solution method as a multiple-wrong-solution method!

The simultaneous use of many relationships, as in *MAGIC* and *YZARC*, obviates the danger of repeatedly developing wrong patterns. The statistical behaviour of a large number of relationships is fairly predictable, unlike that of the small number on which one depends at the beginning of the *MULTAN* phase-developing process. Of course, as has been previously mentioned, *MULTAN* can be persuaded to work by various devices such as artificially changing the temperature factor. This merely imposes a different pattern on phase development and may be regarded as

an inefficient and somewhat *ad hoc* way of applying the 'random' principle.

Although *MULTAN*, *MAGIC* and *YZARC* use exactly the same information they do so sufficiently differently for it not to be true that the failure of one method implies that the other two must also fail. The use of any one of them on its own gives a fairly small failure rate but the use of all three, where necessary, should give a very much smaller failure rate. The programs *MAGIC* 78 and *YZARC* 78 (obtained from G. Germain at Louvain-la-Neuve) use all the components of *MULTAN* 78 (obtained from P. Main at York). If the basic *MULTAN* technique is tried and fails then one has available data files with which a *MAGIC* or *YZARC* attempt at solution can immediately be made. It is our belief that the use of this complementary set of techniques will greatly reduce the risk of failure using multisolution direct methods. The way in which *MAGIC* and *YZARC* are used with the *MULTAN* system is shown in Fig. 5.

The future

We are at present attempting to develop algorithms which will enable, say, 5000 sets of 200 phases to be developed with the *YZARC* technique in a reasonable time. There are three approaches under consideration.

(1) We are trying to speed up the phase-refinement algorithm by the use either of steepest-descents or of parameter-shift methods. The potential gains here seem to be limited to perhaps a factor between two and four in time but there are also losses in refinement efficiency which may cancel some of the gain. Another development is the speeding up of the tangent-formula phase-development process. Here we are planning to sacrifice accuracy for speed but still to have sufficient accuracy to recognize a subset of potentially correct sets of phases which may then be redeveloped or refined by the existing slower but more accurate process. Approaches under this head are designed to improve the speed of the method when using a main-frame time-sharing computer where there are practical limits on the amount of CPU time which may be used.

(2) The *YZARC* process will eventually be programed to run on an array processor. This will greatly speed up all the matrix operations and may make the goal of developing 5000 sets of 200 phases feasible in a reasonable time. However, at present, such a development will be of little benefit to the general crystallographic community with limited access to array processors.

(3) We are contemplating the design of a minicomputer version of *YZARC*. A dedicated minicomputer, with its main function as a structure-solving device, could run for 50 h or even more to solve a difficult structure. A machine used solely and efficiently in this way would be quite economical; fully maintained and written off over ten years the cost would be no more than £50, or equivalent, per average structure.

These three lines of development will take some time fully to explore and implement. In the meanwhile *YZARC* and *MAGIC*, as they now exist, form a useful addition to the direct-methods armoury.

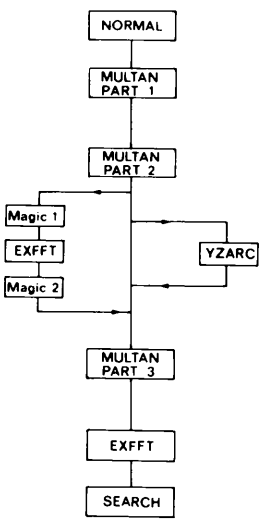

Fig. 5. A flow diagram showing how *MAGIC* and *YZARC* are used in conjunction with *MULTAN*.

We wish to express our gratitude to the Centre de Calcul, Université de Louvain, and also the computing service of the University of York for the provision of facilities for this work. Financial support from the North Atlantic Treaty Organization and the Science Research Council has been crucial to the development of these methods and to the maintenance of a close liaison between our two laboratories. One of us (JPD) is indebted to the Fonds National de la Recherche Scientifique for a Fellowship.

References

ALBERICCI, M., BRAEKMAN, J. C., DALOZE, D., TURSCH, B., DECLERCQ, J. P., GERMAIN, G. & VAN MEERSSCHE, M. (1978). *Bull. Soc. Chim. Belg.* **87**, 487–492.

BAGGIO, R., WOOLFSON, M. M., DECLERCQ, J. P. & GERMAIN, G. (1978). *Acta Cryst.* A**34**, 883–892.

DECLERCQ, J. P., GERMAIN, G. & WOOLFSON, M. M. (1975). *Acta Cryst.* A**31**, 367–372.

HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* A**34**, 863–870.

LESSINGER, L. (1976). *Acta Cryst.* A**32**, 538–550.

SZEIMIES-SEEBACH, U., SZEIMIES, G., VAN MEERSSCHE, M., GERMAIN, G. & DECLERCQ, J. P. (1979). *J. Org. Chem.* Submitted.