

- NILSSON, N. (1957). *Ark. Fys.* **12**, 247–257.  
 REID, J. S. (1983). *Acta Cryst.* **A39**, 1–13.  
 SIMMONS, G. & WANG, H. (1971). *Single Crystal Elastic Constants and Calculated Aggregate Properties*, 2nd ed., p. 28. Cambridge, MA: The MIT Press.  
 SIROTA, N. N. & OLEKHOVICH, N. M. (1961). *Dokl. Acad. Nauk SSSR*, **136**, 879–881.  
 SIROTA, N. N. & OLEKHOVICH, N. M. (1962). *Dokl. Acad. Nauk SSSR*, **143**, 370–372.  
 TALWAR, D. N. & AGARWAL, B. A. (1974). *J. Phys. C*, **7**, 2981–2988.  
 UNO, R., OKANO, T. & YUKINO, Y. (1970). *J. Phys. Soc. Jpn*, **28**, 437–442.  
 VALVODA, V. & JECNY, J. (1978). *Phys. Status Solidi A*, **45**, 269–275.  
 VETELINO, J. F., GAUR, S. P. & MITRA, S. S. (1972). *Phys. Rev. B*, **5**, 2360–2366.  
 ZACHARIASEN, W. H. (1967). *Acta Cryst.* **23**, 558–564.

*Acta Cryst.* (1992). **A48**, 9–11

## Combining *MAGEX* with Random Phases

BY ZHANG SHAO-HUI, CHEN SHU-KANG AND YAO JING-WEN

*Department of Chemistry, Wuhan University, Wuhan 430072, China*

AND LUO BAO-SHENG AND CHEN LIAO-RONG

*Analysis and Testing Center, Wuhan University, Wuhan 430072, China*

(Received 11 March 1991; accepted 20 June 1991)

### Abstract

An algorithm is described which combines the magic-integer concept with that of representing a large number of phases by random numbers. The process, called *MAGEX89*, is more economical than previous magic-integer approaches and in a number of tests its performance compares favourably with those of other commonly used methods of solving small structures.

### Introduction

The magic-integer concept, introduced by White & Woolfson (1975), was utilized in a number of direct-method approaches to structure solution culminating in *MAGEX* (Hull, Viterbo, Woolfson & Zhang, 1981; Zhang & Woolfson, 1982) which was an option in earlier versions of *MULTAN*. In the original *MAGEX* procedure a number,  $q$ , of reflexions had their phases represented (in cycles) by

$$\varphi_j = m_j x \pmod{1}, \quad j = 1, 2, \dots, q, \quad (1)$$

where the  $m_j$  are integers based on the Fibonacci series. The relationships which linked the combination of these reflexions with those defining the origin and enantiomorph ( $O+E$ ) became the foundation for a one-dimensional Fourier series, the  $\psi$  map, with space variable  $x$ . Peaks in the  $\psi$  map, translated into phases by (1), gave plausible starting points for phase extension. There was also a facility for carrying out multiple-pathway phase extension from each starting set of phases by further use of the  $\psi$ -map concept.

Although the general use of *MAGEX* has declined, in favour of the *RANTAN* approach (Yao, 1981) or

*SAYTAN* (Debaerdemaeker, Tate & Woolfson, 1985, 1988) we have continued to develop the idea and would claim that, at least for small structures ( $\leq 100$  independent atoms, say), what we have produced, *MAGEX89*, compares favourably in effectiveness and efficiency with other available methods.

### *MAGEX89*

In the procedure *MAGEX89* we have combined the magic-integer concept with that of representing a large number of phases by random numbers. The algorithm we use to select the starting-set reflexions is that given by Zhang, Luo, Chen & Yao (1989). The  $O+E$  reflexions are chosen to have values of  $E$  and  $\alpha_{\text{est}}$  as large as possible where

$$\alpha(\mathbf{h})_{\text{est}} = \sum_k \kappa(\mathbf{h}, \mathbf{k}) \frac{I_1\{\kappa(\mathbf{h}, \mathbf{k})\}}{I_0\{\kappa(\mathbf{h}, \mathbf{k})\}}, \quad (2)$$

where  $\kappa(\mathbf{h}, \mathbf{k})$  is the usual direct-methods  $\kappa$  value and  $I_1(x)$  and  $I_0(x)$  are modified Bessel functions. A further 5–25 reflexions for the starting set are chosen so as to satisfy the following criteria as well as possible:

(i) these reflexions must be strongly linked with  $O+E$  so that the origin and enantiomorph are strongly defined by the phases of the complete starting set;

(ii) they should have values of  $E$  and  $\alpha_{\text{est}}$  as large as possible.

These extra starting-set reflexions,  $q$  in number, have their phases allocated by (1) with multiple starting sets generated using random values in the range

Table 1. Results of tests with three methods: *S* = SWTR (MULTAN82 version on PDP11/44); *R* = RANTAN (MULTAN82 version on PDP11/44); *M* = MAGEX89

*A/B* indicates that *A* out of *B* atoms were found in the best *E* map. A zero indicates that the default run found no solution.

Structure	Space group and contents	<i>S</i>	<i>A/B</i>		
			<i>R</i>	<i>M</i>	
C <sub>12</sub> H <sub>16</sub> N <sub>2</sub> O <sub>5</sub> S	<i>P</i> 2 <sub>1</sub>	<i>Z</i> = 4	0	27/40	23/40
C <sub>27</sub> H <sub>43</sub> NO <sub>4</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>Z</i> = 4	0	0	22/32
C <sub>16</sub> H <sub>12</sub> ClN	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>Z</i> = 4	17/18	17/18	17/18
C <sub>16</sub> H <sub>11</sub> N <sub>3</sub> O <sub>2</sub>	<i>R</i> 3	<i>Z</i> = 6	21/21	0	21/21
C <sub>12</sub> H <sub>14</sub> N <sub>2</sub> O <sub>4</sub> S	<i>P</i> 1	<i>Z</i> = 2	19/19	0	0
C <sub>14</sub> H <sub>15</sub> BN <sub>4</sub>	<i>P</i> 2 <sub>1</sub> / <i>c</i>	<i>Z</i> = 4	20/20	0	20/20
C <sub>28</sub> H <sub>30</sub> BN <sub>4</sub>	<i>P</i> 2 <sub>1</sub> / <i>c</i>	<i>Z</i> = 4	15/33	0	15/33
C <sub>14</sub> H <sub>28</sub> B <sub>3</sub> FeN	<i>Pnma</i>	<i>Z</i> = 4	6/13	6/13	6/13
C <sub>18</sub> H <sub>18</sub> O <sub>6</sub> ·H <sub>2</sub> O	<i>Cmc</i> 2 <sub>1</sub>	<i>Z</i> = 4	9/15	9/15	11/15

0 to 1 for the variable *x*. The default value of *q* is 15 but can be fixed by the user in the range 5 to 25 and the default number of trials, *i.e.* different values of *x*, equals the number of independent non-H atoms in the structure. A further *N* reflexions with large *E* and  $\alpha_{\text{est}}$  values, where *N* ≈ 300, are chosen by the program. These are allocated random values, a different set for each trial.

With all phases having initial values a process of phase refinement is started using the standard MULTAN weighted tangent formula. Weights are associated with the initial phases as follows: *O* + *E*: weight = 1.0 and the phases are kept fixed; *q* extra starting-set phases: weight = 0.95; other phases: weight = 0.20. At the end of the refinement figures of merit are found and the process continues as for MULTAN.

We have investigated the average error obtained when a number of phases are generated by using a random variable *x*. For example, with *q* = 10 and using 30 trial values of *x* one trial would give a lowest average deviation from true phases of about 50°. To illustrate this, for a set of true phases 46, 78, 160, 315, 130, 35, 245, 348, 4, 88° and with a magic-integer sequence 177, 176, 175, 173, 170, 165, 157, 144, 123, 89 with 30 random values of *x*, one of them gave a mean average error of 52.3°. This process is far less demanding of computer time than calculating a  $\psi$  map. In addition, we have found no advantage in taking equally spaced values of *x* in the range 0 to 1 which is sometimes better and sometimes worse than taking random values. The advantage of using random values is that, if one has been unsuccessful with a first run and wishes to try again, the following batch of trial values of *x* can be completely different from the previous ones.

### Tests of MAGEX89

We have developed MAGEX89 to run on the computer available to us, a PDP 11/44. We also have

Table 2. The best figures of merit found by default runs for the structure C<sub>27</sub>H<sub>43</sub>NO<sub>4</sub>

The MAGEX89 phase set led to the structure solution. The figures of merit are the conventional ones used by MULTAN.

Method	ABSFOM	PSIZERO	RESID
SWTR	0.707	2.340	27.74
RANTAN	0.755	2.294	28.03
MAGEX89	0.835	1.723	22.51

available MULTAN82, which runs on the same machine and contains RANTAN and SWTR, which uses the statistically weighted tangent formula developed by Hull & Irwin (1978). In Table 1 we show the results of applying MAGEX89, RANTAN and SWTR, all run with default parameters, to nine structures and the relative computing times used with these three methods are 91, 100 and 58, respectively. These structures were previously unknown and all eventually refined to *R* factors of less than 0.08 by conventional least-squares methods.

The structure with formula C<sub>27</sub>H<sub>43</sub>NO<sub>4</sub> was solved only by MAGEX89 and we give details of the solution here. As shown in the table the space group is *P*2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> with *Z* = 4. Data were collected on an Enraf-Nonius CAD-4 diffractometer with Mo *K*α radiation,

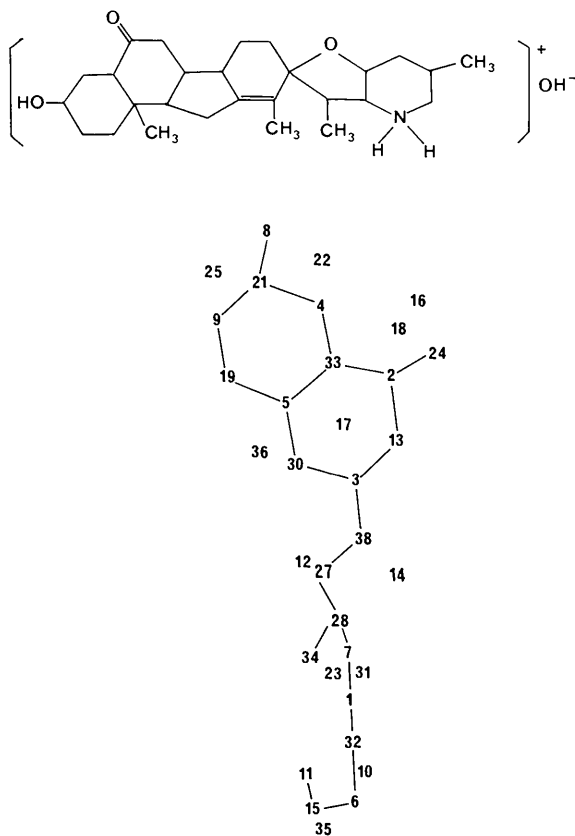


Fig. 1. The finally deduced chemical structure of C<sub>27</sub>H<sub>43</sub>NO<sub>4</sub> and the *E* map found from MAGEX89 which gave the solution.

graphite monochromator and  $\omega/2\theta$  scan mode in the range  $1 \leq \theta \leq 25^\circ$ . Table 2 gives the figures of merit of the best sets for all three methods; that of *MAGEX89* gave an *E* map showing 22 atoms and is reproduced in Fig. 1. Fourier calculations revealed the remaining atoms and least-squares refinement gave a final residual of 0.055 for the observed reflexions.

### Discussion

It will be seen from Table 1 that for the trial structures the *MAGEX89* method performed somewhat better than the other two and the computer resources used by *MAGEX89* were 9% less than those of *RANTAN*. Not too much should be made of that since by modifying parameters all three methods are probably capable of solving all the structures. What we do say is that it is worthwhile having *MAGEX89* available. While any individual method may not succeed for a particular structure, the probability of failure is far lower with many methods available.

The version of *MAGEX89* we have used has been programmed for a PDP11/44 but should be able to

run on most standard personal computers. It can handle all 230 space groups in the standard orientations, including alternative settings, as given in *International Tables for Crystallography* (1987).

We are most grateful to the National Natural Science Foundation of China for its support of this project and to Professor M. M. Woolfson for advice and help with the preparation of this manuscript.

### References

- DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* **A41**, 286-290.  
 DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1988). *Acta Cryst.* **A44**, 353-357.  
 HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* **A34**, 863-870.  
 HULL, S. E., VITERBO, D., WOOLFSON, M. M. & ZHANG, S. H. (1981). *Acta Cryst.* **A37**, 566-572.  
*International Tables for Crystallography* (1987). Vol. A. Dordrecht: Kluwer.  
 WHITE, P. S. & WOOLFSON, M. M. (1975). *Acta Cryst.* **A31**, 53-56.  
 YAO, J. X. (1981). *Acta Cryst.* **A37**, 642-644.  
 ZHANG, S. H., LUO, B. S., CHEN, S. K. & YAO, J. W. (1989). *Acta Phys. Chim. Sin.* **5**, 536-540.  
 ZHANG, S. H. & WOOLFSON, M. M. (1982). *Acta Cryst.* **A38**, 683-685.

*Acta Cryst.* (1992). **A48**, 11-14

## A Method for Multiple Superposition of Structures

BY A. SHAPIRO AND J. D. BOTHA

*Department of Mathematics, Applied Mathematics and Astronomy, University of South Africa, PO Box 392, Pretoria 0001, South Africa*

A. PASTORE

*European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 1022.09, 6900 Heidelberg, Germany*

AND A. M. LESK

*Department of Haematology, University of Cambridge Clinical School, MRC Centre, Hills Road, Cambridge CB2 2QH, England*

(Received 5 April 1991; accepted 15 July 1991)

### Abstract

The study of families of protein structures is important in analysing the results of NMR structure determinations and in investigating mechanisms of molecular evolution at the level of conformation. A method is discussed for finding the transformations that mutually superpose an arbitrary number of structures in the least-squares sense given specified atom-to-atom correspondence.

### 1. Introduction

Superposition has become an important tool for comparing protein structures and for deriving an 'average' structure from a family of conformations of a protein. The problem arises regularly in the determination of the three-dimensional structure of a protein in solution using nuclear magnetic resonance, which typically produces an ensemble of conformations (Wuethrich, 1986). Sets of interproton distances