

*Acta Cryst.* (1971) **A27**, 681

**A program for choosing starting sets to be used in the tangent formula structure solution method.** By W. D. S. MOTHERWELL and N. W. ISAACS, *University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England*

(Received 16 July 1971)

The program described here has been used successfully to choose starting sets for 12 crystal structures (5 previously undetermined). The number of sets of phases to be calculated was not greater than 32 for any structure.

### Introduction

One of the most successful applications of direct methods for solution of crystal structures is the multiple-solution approach of Germain & Woolfson (1968). This method is critically dependent upon the choice of starting set, particularly for larger structures ( $N > 150$  atoms *per* unit cell). This problem has been discussed by Germain, Main & Woolfson (1970), who have written a program to select starting sets based on the elimination of weakest interactions. It is hoped that the generation of new phases from such starting sets will follow the pathway of strongest interactions, and thus be most likely to lead to a correct solution. We wish to describe here an alternative approach in which the starting set is chosen to give the greatest number of new phases in the first steps of expansion, using interactions stronger than a specified level. We have found with this method that we are able to control to some extent the number of symbol reflexions required, and thus keep the number of permutations to a minimum. We follow the same terminology as given in the references.

### Description of the program

- (1) A list of about 200–250 reflexions, sorted by decreasing  $|E|$  magnitude, are input. About 2000–3000 non-redundant triple product phase relationships are computed and stored.
- (2) Any reflexions with 'known' phases are now indicated, *e.g.* origin-defining or derived by the  $\Sigma_1$  formula (Karle & Karle, 1966).
- (3) Each of the first  $n$  unknown reflexions is postulated as a potential member of the starting set.
- (4) The reflexions whose phases may be determined (subject to certain criteria) from this potential starting set are found by cyclic application of the  $\Sigma_2$  formula.
- (5) When all  $n$  reflexions have been examined the one which generated most new phases becomes a member of the starting set.
- (6) The search is repeated as from (3), until  $> 95\%$  of the reflexions can be generated.

For each newly generated reflexion the quantity  $\langle \alpha^2 \rangle^{1/2}$  is calculated following Germain, Main & Woolfson (1970). This is the expected value of  $\alpha$  in the absence of phase information, calculated from the triple product magnitudes of the contributors, and is related to the standard deviation of the phase calculated by the tangent formula (Karle & Karle, 1966). The reflexion is accepted if the following criteria are satisfied:

- (a)  $\langle \alpha^2 \rangle^{1/2} > \alpha_{\min}$  (input value)
- (b) If generation is by a single triple product relationship

all three  $|E|$  must be greater than  $E_{\min}$  (input value). This reduces the effect of unreliable single indications of phase.

It usually turns out that in choosing symbols the same 'family' of reflexions may be generated by any of several members of the family. In these cases the choice is made of the generating member of highest  $|E|$ . The program does not distinguish between zonal and general reflexions so that a two or one dimensional reflexion might be chosen as a symbol. However the user may choose any other generating symbol from within the same family without affecting the general pattern of expansion from the starting set. The actual path followed during the numerical tangent formula calculation will not be identical with the one followed by this program because  $\alpha$  may be lower than  $\langle \alpha^2 \rangle^{1/2}$  expected for any given step.

We normally start investigations on an unknown structure by examining the first 30 reflexions with  $\alpha_{\min} = 2.5$  (a standard deviation in phase of  $43^\circ$ ), and  $E_{\min} = 2.0$ . Because of computing expenses we wish to limit our number of permutations to 32, which may mean (typically) that no more than 3 symbolic reflexions should be chosen. If the program chooses only one or two symbols it may be forced to choose a set of three by raising either  $\alpha_{\min}$  or  $E_{\min}$ . Similarly if 5 symbols were chosen we might reduce  $\alpha_{\min}$  to, say, 2.0 so that only 3 were chosen.

### Discussion of results

The results for the crystal structure of 20- $\alpha$ -hydroxyprogesterone [Table 2 compound (*k*)] are used to illustrate some features of the program (see Table 1). In the tetragonal space group  $P4_1$  two reflexions are required to define the origin (Karle & Hauptman, 1956). A suitable pair of reflexions were picked by hand (Table 1) and input to the program as 'known'. The program chose two symbols using  $\alpha_{\min}$  of 3.0 and  $E_{\min}$  2.0. One symbol was allowed 4 values and the other only 2 values to define the enantiomorph. Of the 8 sets computed using an iterative tangent formula program (Larson & Motherwell, 1969) the correct structure was clearly indicated by the lowest  $R$  value between  $|E_{\text{obs}}|$  and  $|E_{\text{calc}}|$ ,  $R_{\text{Karle}}$  of 23% (Karle & Karle, 1966).

Table 1. Starting set chosen by the program for 20- $\alpha$ -hydroxyprogesterone

<i>h</i>	<i>k</i>	<i>l</i>	$ E $	Starting values of phase (degrees)
9	2	0	2.96	0 } origin
3	7	1	3.02	0 } defining
9	3	5	2.09	45, 135
2	4	3	2.39	45, 135, 225, 315

Table 2. Structures solved using this program

Compound	Formula	Space group	<i>N</i>	Reflexions input	Number of known phases	Number of symbols chosen	Number of phases sets calculated
(a)	C <sub>10</sub> H <sub>10</sub> O <sub>2</sub>	<i>P</i> $\bar{1}$	24	281	3	3	8
(b)	C <sub>33</sub> H <sub>33</sub> O <sub>8</sub> P	<i>C</i> 2/ <i>c</i>	336	275	4	3	8
(c)	C <sub>29</sub> H <sub>47</sub> O <sub>2</sub> Cl	<i>P</i> 2 <sub>1</sub>	64	250	4	3	32
*(d)	C <sub>12</sub> H <sub>15</sub> NO <sub>2</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	60	100	1	4	32
(e)	C <sub>19</sub> H <sub>28</sub> O <sub>2</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	84	178	4	3	32
*(f)	C <sub>21</sub> H <sub>30</sub> O <sub>3</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	96	151	4	3	32
(g)	C <sub>15</sub> H <sub>21</sub> N <sub>5</sub> O <sub>5</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	100	150	3	4	32
*(h)	C <sub>21</sub> H <sub>24</sub> O <sub>9</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	120	187	4	2	16
(i)	C <sub>31</sub> H <sub>52</sub> O <sub>4</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	140	198	7	2	16
(j)	C <sub>15</sub> H <sub>24</sub> ClPO <sub>3</sub>	<i>P</i> 4 <sub>1</sub>	80	162	3	3	32
*(k)	C <sub>21</sub> H <sub>32</sub> O <sub>2</sub>	<i>P</i> 4 <sub>1</sub>	92	143	2	2	16
*(l)	C <sub>15</sub> H <sub>22</sub> O	<i>P</i> 3 <sub>2</sub>	48	155	1	3	32

\* Unknown structure solved using this program.

- (a) *cis-trans-cis*-Cyclodeca-2,4,8-triene-1,6-dione. Kennard, Wampler, Coppola, Motherwell, Watson & Larson (1971).  
 (b) Tri-*p*-totylphosphine dimethyl acetylenedicarboxylate adduct. Kennard, Motherwell & Coppola (1971).  
 (c) Iso-cholesteryl chloroacetate. Harrison, Hodgkin, Maslen & Motherwell (1971).  
 (d) Bellendine. Motherwell, Isaacs, Kennard, Bick, Bremner & Gillard (1971).  
 (e) Epitestosterone. O. Kennard *et al.* (to be published).  
 (f) 3 $\beta$ ,5-Cyclo-6-keto-androstane-17-acetate. R. C. Pettersen *et al.* (to be published).  
 (g) 2'-*O*-Tetrahydropyranyladenosine. Kennard, Larson, Motherwell, Coppola, Griffin & Reese (1971).  
 (h) Melampodin. Neidle & Rogers (1971).  
 (i) Methyl shoreate. O. Kennard *et al.* (to be published).  
 (j) Isocyperene glycol chlorophosphoridate. Dreyfus, Thierry, Weiss, Kennard, Motherwell, Coppola & Watson (1969).  
 (k) 20- $\alpha$ -Hydroxyprogesterone. O. Kennard *et al.* (to be published).  
 (l) (—)-Aristolone. F. H. Allen *et al.* (to be published).

Table 2 summarizes the results for 12 structures, 5 of which were solved using this program and the others by previous manual choice of starting sets. Some points of interest are

- (i) The program has been successful over a considerable range of space groups (all symmetry information is input on a single data card).  
 (ii) The number of symbols chosen was acceptably low in every case, and in some, *e.g.* compounds (a) and (i), the number chosen was less than the number of symbols used in the original solution. Several compounds had given considerable difficulty in the original attempts *viz.* (a), (c), (f), (h), (i), (l).  
 (iii) It is not necessary to define the origin in the conventional manner. For example the compound (l) did not have reflexions of high  $|E|$  and the correct parity to define the origin. Only one reflexion was entered as 'known' and by limiting the values given to the three symbols chosen, the origin and enantiomorph was effectively defined. A similar procedure was followed for compound (d) where all symbols chosen were three-dimensional reflexions.  
 (iv) In each case the correct set of phases was found to be the one with lowest  $R_{\text{Karle}}$  in the range 15–25%.

We have attempted unsuccessfully to apply the program to three large structures ( $N > 280$ ). Because the probability of correct phase calculation with a given relationship decreases as  $N$  increases it appears that an unacceptably large number of symbols are needed to give a correct solution. However within our present limit of 32 permutations the procedure seems reliable for structures of the size presented in Table 1.

The program is written in FORTRAN IV for an IBM 360/44 computer with a 32K word store.

We are grateful to Dr S. Neidle for use of his data for compound (h), and to the Office for Scientific and Technical Information for financial support.

#### References

- DREYFUS, H., THIERRY, J. C., WEISS, R., KENNARD, O., MOTHERWELL, W. D. S., COPPOLA, J. C., & WATSON, D. G. (1969). *Tetrahedron Letters*, **43**, 3757.  
 GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* **B26**, 274.  
 GERMAIN, G. & WOOLFSON, M. M. (1968). *Acta Cryst.* **B24**, 91.  
 HARRISON, H. R., HODGKIN, D. G., MASLEN, E. H. & MOTHERWELL, W. D. S. (1971). *J. Chem. Soc. (C)*, p. 1275.  
 KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635.  
 KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849.  
 KENNARD, O., LARSON, A. C., MOTHERWELL, W. D. S., COPPOLA, J. C., GRIFFIN, B. E. & REESE, C. B. (1971). *J. Chem. Soc. (C)*. In the press.  
 KENNARD, O., MOTHERWELL, W. D. S. & COPPOLA, J. C. (1971). *J. Chem. Soc. (C)*, p. 2461.  
 KENNARD, O., WAMPLER, D. L., COPPOLA, J. C., MOTHERWELL, W. D. S., WATSON, D. G. & LARSON, A. C. (1971). *Acta Cryst.* **B27**, 1116.  
 LARSON, A. C. & MOTHERWELL, W. D. S. (1969). *Program Manual*. Crystallographic Data Centre, Cambridge, England.  
 MOTHERWELL, W. D. S., ISAACS, N. W., KENNARD, O., BICK, I. R. C., BREMNER, J. B. & GILLARD, J. (1971). *Chem. Comm.* p. 133.  
 NEIDLE, S. & RODGERS, D. (1971). *Chem. Comm.* To be published.