independent non-hydrogen atoms and the highest spurious peak was 35th in order of height. Similarly, for TPH, with 270 E's and all the Σ_2's, the 'best' map from the 56 sets of phases generated gave all 39 atoms of the structure. In this case, the highest spurious peak was 39th. The performance of the individual figures of merit for both of these structures is worth reporting. For RR, the best set of phases was first in ranking order on residual, second on ψ_0 and 15th on $\Sigma\alpha$, while the best set for TPH was first on residual, first on ψ_0 and 18th on $\Sigma\alpha$ (out of 56). Since $\Sigma\alpha$ is a measure of the internal consistency of the Σ_2 relationships, it is clear that the correct phases do not correspond to the most consistent set of Σ_2's and, in fact, are a long way from this situation. As the phase-determination process relies initially upon the assumption that $\varphi_{\bar{h}} + \varphi_{h-k} + \varphi_k = 0$ for all h and k, this could be the reason why these two structures were previously rather difficult to solve. Copies of this latest version of $MULTAN$ are available from the author upon request.

Only the starting phases which are not restricted by space-group symmetry are given values by magic-integer permutation. Phases which can take on only two values because of symmetry are still given those two values in different phase combinations as before. For n phases assigned to magic integers, the number of sets produced is

$$N_1 = 2^{n+2} - 4. \qquad (18)$$

This number is obviously modified by the requirements of enantiomorph and origin fixing in the usual way. If the Fibonacci sequences were used, the number of phase sets produced would be

$$N_2 = \frac{8}{\sqrt{5}} \left\{ \left[\frac{1 + \sqrt{5}}{2} \right]^{(n+1)} - \left[\frac{1 - \sqrt{5}}{2} \right]^{(n+1)} \right\} - 4. \qquad (19)$$

It is clearly of interest to determine how large an error can be tolerated in the starting phases in $MULTAN$, since any permitted increase in error results in a large decrease in the number of phase sets which must be generated. In general, when the phase sets are generated from a magic integer sequence based on an integer-geometric progression of limiting ratio r, the number of sets produced for n phases will be a function of r^n.

References

DECLERCQ, J.-P., GERMAIN, G. & HENKE, H. (1973). *Cryst. Struct. Commun.* 2, 405–409.
DECLERCQ, J.-P., GERMAIN, G. & WOOLFSON, M. M. (1975). *Acta Cryst.* A31, 367–372.
HAUPTMAN, H. & DUAX, W. L. (1972). *Acta Cryst.* A28, 393–395.
HOEKSTRA, A., VOS, A., BRAUN, P. B. & HORNSTRA, J. (1975). *Acta Cryst.* B31, 1708–1715.
KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* 3, 181–187.
LESSINGER, L. (1976). *Acta Cryst.* A32, 538–550.
MAIN, P. (1977). *Acta Cryst.* A33, 750–757.
MARGULIS, T. N. & LESSINGER, L. (1977). Am. Crystallogr. Assoc. Winter Meet., Asilomar, California, Abstract JQ5. Also submitted to *Acta Cryst.*
WHITE, P. S. & WOOLFSON, M. M. (1975). *Acta Cryst.* A31, 53–56.

On the Application of Phase Relationships to Complex Structures. XIII.* The Choice of Reflexions

BY S. E. HULL

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

Alternative methods of choosing the reflexions for use in direct methods are discussed. The original *CONVERGE* method, as implemented in the *MULTAN* system of computer programs, is compared with several variants and an iterative procedure for maximizing the ratio of triple-phase invariants to reflexions is described.

Introduction

Woolfson (1977) has outlined a computer program package for direct phase determination which, in time,

may replace tangent-formula-based programs such as *MULTAN* (Main, Woolfson, Lessinger, Germain & Declercq, 1974). At the heart of *MULTAN* is the *CONVERGE* procedure (Germain, Main & Woolfson, 1970) which determines the reflexions to be used in the starting set (origin and enantiomorph-defining

---

* Part XII: Main (1978).

reflexions and permuted phases) and, to a large extent, the order in which the remaining phases will be determined. The algorithm currently incorporated in *MULTAN* is designed to produce an ordered list of phases suitable for step-by-step phase extension and ideally all reflexions (except those in the starting set) should be strongly related by triple-phase invariants (TPI's) to reflexions with previously determined phases. This paper discusses and compares several variations on the *CONVERGE* procedure which may be of use in *MULTAN* and its probable successor (to be called *MAGLIN*). The most important difference between the two programs is that *MAGLIN* does not depend on the separate determination of individual phases, but on the simultaneous refinement of a relatively large number of phases (perhaps up to 150) by a least-squares process. In order that phases can be refined efficiently and correctly by least squares it is desirable to have a large ratio of TPI's to reflexions. The second part of this paper describes a method of maximizing this ratio by choosing a subset of reflexions from a larger set.

Convergence procedures

The *CONVERGE* process in *MULTAN* (Germain, Main & Woolfson, 1970) involves calculating the estimated value of α_h for each reflexion in the initial set from:

$$\langle \alpha_h{}^2 \rangle = \sum_h K^2_{hh'}$$

$$+ 2 \sum_{\substack{h' \\ h' \neq h''}} \sum_{\substack{h'' \\ h' \neq h''}} K_{hh'} K_{hh''} \frac{I_1(K_{hh'})}{I_0(K_{hh'})} \frac{I_1(K_{hh''})}{I_0(K_{hh''})}$$

where $K_{hh'}$ is the weight of the relationship $\varphi_h + \varphi_{h'} + \varphi_{h-h'} \simeq 0$. At each stage the reflexion with the lowest α_h is eliminated and the α_h's are updated. This is repeated until only the reflexions required in the starting set, including the origin and enantiomorph-defining reflexions, are left. A good convergence list should have the following characteristics: (1) There will be few reflexions eliminated with zero α_h. If many reflexions are eliminated at an early stage of the procedure with unacceptably low α_h, then it is advisable either to use more TPI's or fewer reflexions (Lessinger, 1976). Any reflexions which are eliminated with zero α towards the end of the procedure should be taken into the starting set; if they are not, phase development *may* be impossible. (2) There should be no places in the list where eliminated reflexions have low α_h or very few TPI's involving non-eliminated reflexions; this is particularly important in the middle to late stages of convergence, where a wrongly determined phase may not easily be corrected by refinement. (3) The final group of reflexions elimi-

nated should not be biased towards any particular index group. (4) The phase-determination pathway should not allow the formation of 'islands', *i.e.* groups of reflexions which are strongly linked among themselves, but not with other reflexions.

The procedure outlined above usually meets all these conditions reasonably well and relatively few structures are difficult to solve because of a poor convergence list.

In order to see if the *CONVERGE* process could be modified to produce a set of reflexions suitable for use in the least-squares procedure of phase development (*i.e.* as large a number of TPI's for a given number of reflexions as possible), several tests were made with alternative elimination criteria (α_h). The results of these tests for ergocalciferol (Hull, Leban, Main, White & Woolfson, 1976) and for 3,3-dimethyl-4,5,9,10,11,12-hexacarboxymethyltetracyclo[7.2.1.0^{2,4}.0^{2,8}]dodeca-5,7,10-triene (RR) (Declercq, Germain & Henke, 1973) are shown in Tables 1 and 2 respectively. Ergocalciferol (space group $P2_12_12_1$, $C_{28}H_{44}O$, $Z = 8$) was chosen because the difficulties experienced in solving the structure were attributed largely to the poor ordering of the reflexions by *CONVERGE*. In the tests made with this structure all 4431 TPI's for the 300 reflexions with the largest $|E|$ values were used. The RR compound (space group $Pn2_1a$, $C_{26}H_{28}O_{12}$, $Z = 4$) was used as an example because it is moderately difficult to solve by *MULTAN* (Lessinger, 1976) and is typical of the medium sized structures which are solved routinely by direct methods. For this structure all 2301 TPI's involving the 250 reflexions with the largest $|E|$ values were used.

The α_h used in the various cases shown in Tables 1 and 2 are:

(1) as in *MULTAN*; the formula given above is approximated by making

$$K'_{hh'} = K_{hh'} \frac{I_1(K_{hh'})}{I_0(K_{hh'})}$$

and then putting

$$\alpha_h = \sum_{h'} K_{hh'};$$

Table 1. *Convergence-procedure results for ergocalciferol, showing the number of TPI's for N reflexions in cases 1 to 7*

The different procedures employed in each case are explained in the text.

N	20	40	60	80	100	120	140
Case							
1	37	161	317	464	672	914	1201
2	45	160	336	521	711	971	1242
3	42	161	335	485	708	966	1242
4	49	163	335	551	743	978	1294
5	49	163	336	553	726	974	1291
6	49	165	337	553	737	984	1294
7	49	165	337	553	743	984	1294

Table 2. *Convergence-procedure results for the RR compound in the same format as Table 1*

N	20	40	60	80	100	120	140
Case							
1	46	135	254	346	456	609	838
2	52	144	260	425	527	684	878
3	49	134	260	395	487	645	867
4	43	125	244	386	556	722	921
5	37	124	260	386	549	702	907
6	35	138	262	391	554	706	917
7	40	138	262	391	554	706	917

$$(2) \qquad \alpha_h = \sum_{h'} K_{hh'};$$

(3) α_h = the number of TPI's involving reflexion **h** (*i.e.* $K_{hh'}$ = 1); in the event of several reflexions having the same value of α_h, the reflexion with the lowest value of $|E|$ is eliminated;

$$(4) \qquad \alpha_h = \sum_{h'} K_{hh'} - \sum_{h'} K_{h''h'}$$

where **h''** is the reflexion with the lowest value of $\sum_{h'} K_{hh'}$ after reflexion **h** has been eliminated;

(5) as (4), but with $K_{hh'} = 1$ for all TPI's;

$$(6) \qquad K_{hh'} = 1.0, \, \alpha_h = 2\sum_{h'} K_{hh'} - \sum_{h'} K_{h''h'}$$

(7) as (3), but if several reflexions have the same value of α_h, then the criterion used in (5) is used to decide which to eliminate.

Tables 1 and 2 show that the largest jump in the number of TPI's for a given number of reflexions occurs when the $I_1(K)/I_0(K)$ modification is removed from α_h (case 1 to case 2). Application of this factor has the effect of lowering the relative weight of the weak TPI's and therefore reflexions associated with many TPI's become relatively less important than those associated with a few strong TPI's, and are eliminated at an earlier stage of the CONVERGE process. The modification by the factor $I_1(K)/I_0(K)$ is necessary if phases are to be determined one at a time, but not in methods which determine or refine many phases simultaneously, and where it is an advantage for the phases to be required to satisfy many TPI's at once, even if some of these would be contradictory indications in the step-by-step methods. For the same number of reflexions case 2 gives up to 10% more TPI's than case 1. Further improvement is obtained for ergocalciferol by weighting all TPI's equally (case 4); this results in an increase of up to 6% over case 2. However, for the RR compound the trend is not so clear; for systems of less than 100 reflexions case 2 has more TPI's than case 4.

Cases 3, 5, 6 and 7 are attempts to maximize the interaction between the reflexions remaining as elimination takes place by requiring that the reflexion eliminated be the one that leaves the remaining set with

the greatest minimum elimination criterion (α_h). For both test structures case 3 shows that this is not a good criterion when the TPI's are weighted by $K_{hh'}$. With $K_{hh'} = 1$, case 5 is less good than case 4 for both structures, but case 6 is similar to case 4 for ergocalciferol, for which case 7 has the maximum number of TPI's at all stages and suggests that the 'greatest minimum' test may best be used to decide between reflexions which are involved in equal numbers of TPI's. The case 7 results for the RR compound are less good than for case 4 when more than 100 reflexions are considered. This anomalous result is due to the distribution of weak links in the convergence processes.

More detailed examination of cases 1, 2 and 3 for ergocalciferol shows that there are five points in each list where a reflexion is eliminated with no TPI's; however in cases 1 and 3 one of these breaks occurs a considerable way along the phase-development pathway (after 72 and 73 reflexions respectively). In case 2 the weakest link at this stage is at reflexion 78 which has only one TPI. The next reflexion to be determined after the weak link has one TPI (case 1), three TPI's (case 2) and one TPI (case 3), so it seems that the second convergence method may present the most effective phase-determining pathway for this structure. Cases 4 to 7 do not have any extremely weak links at this point of the phase development, but do have reflexions with fewer TPI's between the 90th and 100th phase to be determined.

For the RR compound case 1 has six reflexions eliminated with no TPI's, cases 2, 4 and 7 have four such reflexions and cases 3, 5 and 6 have five. However the first break in the convergence listing occurs with only 36 reflexions remaining in case 4, whereas in the other cases it occurs with between 74 and 88 reflexions remaining. This accounts for the low ratios of TPI's to reflexions for small reflexion sets in case 4, and for the relatively high ratios in this case for larger sets of reflexions.

An iterative method of optimizing the ratio of TPI's to reflexions

Since the least-squares phase-refinement method described by Woolfson (1977) develops phases in blocks rather than singly, it is not necessary to specify the exact order of phase development by use of a CONVERGE-type process. An alternative way of maximizing the ratio of TPI's to reflexions by an iterative process has been devised and is now described.

The first approximation to the optimum order of the reflexions is obtained by listing them in order of the number of TPI's in which each is involved, from best to worst connected. Then for each reflexion the following quantity is calculated:

$$A_h = \sum_{h'} f(N_{h'}) f(N_{h-h'})$$

where the summation is over all the TPI's involving reflexion **h**, N_h is the position of reflexion **h** in the sorted list of reflexions and $f(N)$ is some function of N which decreases with increasing N. The reflexions are then resorted on A_h, and the process repeated, either for a set number of cycles or until the sorted reflexion order ceases to change significantly.

The idea behind this method is to produce a well connected set of reflexions by giving large weights to TPI's involving strongly interacting and small weights to those involving weakly interacting reflexions. The exact form of the function $f(N)$ is critical; empirical tests suggest that it should have a sigmoidal form, with reflexions of low N having weights near unity and those of high N having weights close to zero: in the intermediate range the weights should vary rapidly with N. The function should be continuous; use of a non-continuous function gives rise to oscillation between the iterations and the reflexion order may never stabilize. Fig. 1 shows several members of a family of curves which have such properties and have been found to be suitable as weighting functions. These curves are given by:

$$f(N) = \left[e^{-\lambda} \sum_{s=0}^{n-1} \frac{\lambda^s}{s!} \right]^3$$

where λ is a linear function of N, n is an integer and the expression in square brackets is the Poisson distribution formula. Some results obtained for ergocalciferol and the RR compound are shown in Tables 3 and 4 for various values of n, taking $\lambda = 20N/N_R$, where N_R is the total number of reflexions. In the tests 20 cycles of iteration were performed, although in most cases the reflexion order ceased to change significantly in about half that number of cycles.

Tables 3 and 4 show that the number of reflexions (N_m) for which the number of TPI's is a maximum depends on the value of n. Although the choice of n
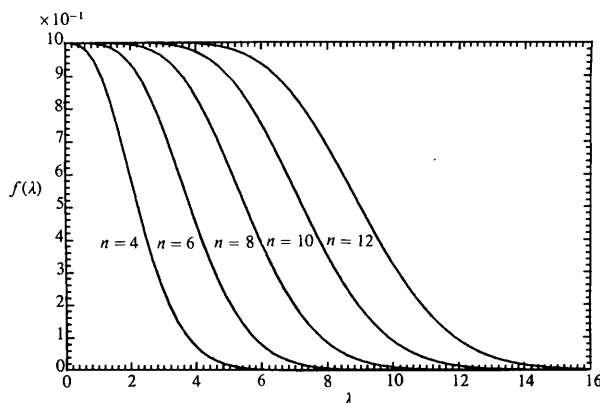

Fig. 1. A plot of the function $f(\lambda) = \left[e^{-\lambda} \sum_{s=0}^{n-1} \frac{\lambda^s}{s!} \right]^3$ for several values of n.

Table 3. *Results of the iterative procedure for ergocalciferol for various values of n, showing the number of TPI's obtained for N reflexions*

The last line shows the number of TPI's obtained for N reflexions by ordering the reflexions on the number of TPI's in which each is involved (*i.e.* after the first stage of the iterative procedure).

N	20	40	60	80	100	120	140
n							
4	45	170	320	492	732	965	1191
5	41	170	338	518	722	977	1195
6	36	151	337	544	800	1025	1232
7	33	147	330	556	805	1043	1253
8	23	142	323	559	807	1049	1253
9	21	137	312	551	803	1049	1263
10	20	138	307	548	803	1051	1277
11	26	126	278	469	709	1000	1304
12	25	116	250	449	709	993	1303
Ordered on number of TPI's	14	71	170	337	587	820	1167

Table 4. *Results of the iterative procedure for the RR compound in the same format as Table 3*

N	20	40	60	80	100	120	140
n							
4	46	142	254	394	459	648	920
5	33	131	249	394	549	754	914
6	25	123	261	405	595	761	923
7	20	117	261	424	604	763	923
8	23	114	260	427	602	762	925
9	20	112	262	431	605	781	928
10	22	109	257	427	608	788	928
11	22	111	253	428	602	747	917
12	23	95	214	379	563	749	923
Ordered on number of TPI's	6	55	155	270	426	610	852

is not critical it is obvious that the best results for a particular N_m are obtained by making N_m close to the point of maximum slope of $f(N)$. The number of TPI's obtained for most values of N_m improved considerably over any of the *CONVERGE* algorithms given above (Tables 1 and 2), and the method is probably the most suitable to use in selecting reflexions for use in phase-determining or refinement techniques which depend on having a large number of well connected reflexions.

One danger of algorithms which pick out sets of well connected reflexions is that either the reflexions will be biased towards certain index groups or that they will occur in unconnected 'islands'. The first of these problems was encountered for both ergocalciferol and the RR compound when the standard *CONVERGE* procedure was used and the iterative procedure gave similar results in these two tests. An additional test was carried out (Table 5) for cortisone (Declercq, Germain & Van Meerssche, 1972) because it has been found that methods designed to produce 'good' Karle–Hauptman matrices (Main, 1975), which necessarily

Table 5. *Results of the iterative procedure for cortisone in the same format as Table 3*

N	20	40	60	80	100	120	140
n							
4	54	222	426	637	846	1143	1539
5	50	222	445	661	886	1197	1530
6	45	216	444	664	926	1241	1596
7	37	184	395	644	916	1228	1507
8	35	166	385	644	940	1239	1528
9	33	150	374	642	943	1240	1536
10	39	152	356	604	899	1235	1584
11	28	136	336	563	885	1245	1612
12	25	125	315	549	874	1240	1619
Ordered on number of TPI's	21	78	167	381	672	962	1398

contain a set of well linked reflexions, lead to sets of reflexions heavily biased towards reflexions with $l = 0$ (modulo 3) (Irwin, 1977). The sets given in Table 5 have been examined in detail, and in the set which gives the best connected set of 20 reflexions ($n = 4$) all 20 reflexions were found to have $l = 0$, 3 or 6. For the set ($n = 8$) having the best connected set of 100 reflexions, this bias is still present, but is not so strong that it is likely to hinder any phase-determining process. This indicates that a well connected set of reflexions does not necessarily contain reflexions from only a few index groups.

The time taken for the iterative process is not excessive, although somewhat longer than for the standard *CONVERGE* procedures (cases 1, 2 and 3 above). Typically, one cycle of iteration, that is calculation of A_h, counting the number of TPI's and sorting, takes less than 1·5 s on a DEC System 10 computer for the structures given here as examples, compared with about 4 s for the complete *CONVERGE* process. The *CONVERGE* processes which involve looking at the set of reflexions left after a reflexion has been eliminated (cases 4, 5, 6 and 7) are slower, and may take up to 30 s computer time.

Conclusions

The alternative *CONVERGE* procedures described in this paper probably do not have any advantage over the standard procedure used in *MULTAN* when phases are to be developed sequentially, one at a time, but may be worth trying for some difficult structures if the *CONVERGE* list produced by *MULTAN* does not satisfy the criteria given above. If phases are to be determined by a method which benefits from a high ratio of relationships to reflexions, then the iterative process described in the last section is a good way of choosing reflexions. The methods which may benefit from this iterative algorithm include the least-squares procedure mentioned previously and the magic integers method (Declercq, Germain & Woolfson, 1975). The algorithm could also be of use in cases where, for large structures, it was necessary to choose a limited number of reflexions to be used in phase determination from the complete set of reflexions.

References

Declercq, J.-P., Germain, G. & Henke, H. (1973). *Cryst. Struct. Commun.* **2**, 405–409.
Declercq, J.-P., Germain, G. & Van Meerssche, M. (1972). *Cryst. Struct. Commun.* **1**, 13–15.
Declercq, J.-P., Germain, G. & Woolfson, M. M. (1975). *Acta Cryst.* A**31**, 367–372.
Germain, G., Main, P. & Woolfson, M. M. (1970). *Acta Cryst.* B**26**, 274–285.
Hull, S. E., Leban, I., Main, P., White, P. S. & Woolfson, M. M. (1976). *Acta Cryst.* B**32**, 2374–2381.
Irwin, M. J. (1977). Private communication.
Lessinger, L. (1976). *Acta Cryst.* A**32**, 538–550.
Main, P. (1975). *Crystallographic Computing Techniques*, pp. 165–175. Copenhagen: Munksgaard.
Main, P. (1978). *Acta Cryst.* A**34**, 31–38.
Main, P., Woolfson, M. M., Lessinger, L., Germain, G. & Declercq, J.-P. (1974). *MULTAN 74. A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data*. Univs. York, England, and Louvain, Belgium.
Woolfson, M. M. (1977). *Acta Cryst.* A**33**, 219–225.