ACTA CHEMICA SCANDINAVICA 25 (1971) 1040–1044

A Simple Procedure for Solving Structure Factor Sign Relationships With a Small Computer, Giving a Limited Number of Alternative Solutions

ROLF NORRESTAM

Institute of Inorganic and Physical Chemistry, University of Stockholm, S-104 05 Stockholm, Sweden

A procedure for solving structure factor sign relationships is described. Special regard has been paid to the problem of selecting a proper basis set to limit the number of alternative solutions, and also to the use of simple multiplicity criteria for estimating the quality of sign indications. The flow of a program for an IBM 1800 computer, using the procedure, is described in some detail.

A brief description of the results obtained for six crystal structures solved by the use of this program is given.

In recent years, it has been shown by numerous examples that the solution of triple product sign relationships among the structure factors in the centrosymmetrical case is a very efficient method of structure solution. The necessary computations are fairly simple and consist, apart from calculations of probabilities, mostly of logical operations and simple integer arithmetic. Hence, a fairly small computer suffices, especially if the probability criteria, usually used to estimate the quality of sign indications, could be replaced by some simpler calculated criteria which do not demand, *i.a.*, time-consuming floating-point arithmetic.

The present program was written for an IBM 1800, cycle time 1.4 μs, no hardware floating point, 24 K 16-bit memory. Parts of the memory are occupied by programs for on line control of different processes, *e.g.* a film scanner and a diffractometer, so that the memory available for non-process calculations is about 14 K. Thus, the non-process programs must be fairly small and must avoid floating point arithmetic as far as possible. However, two disks are available for intermediate storage of data and subprograms during the execution of non-process programs.

The operating system presently used is the IBM Time Sharing Executive System.[1] The program has been written in IBM 1800 Basic Fortran IV language, except for two very simple bit-handling subroutines, written in IBM 1800 Assembler language.

GENERAL DESCRIPTION OF THE PROCEDURE

Input to the program consists of E-values and triple product sign relationships. The rather lengthy routine calculations necessary for obtaining triple product sign relations have usually been performed on a larger computer (IBM 360/75), using a program (GSR) written by C. T. Grainger. Usually, only E-values larger than about 1.5 and triple product sign relations, among them with probabilities greater than about 0.95 have been used. It was hoped that by using only such selected data, multiplicity criteria for sign indications could be used by the phasing program instead of the more time-consuming calculations of probability criteria. The phasing program does not use symbolic addition (*cf.* Karle and Karle [2]) with one basis set, consisting of origin and symbol defining reflections to generate multiple solutions, from unresolved symbols and/or allowance for the failure of symbolic equations (*cf.* Germain and Woolfson [3]). Rather it uses multiple basis sets, consisting of origin defining reflections and all possible sign permutations for a limited number of reflections with unknown signs, to give all the corresponding solutions. The obvious advantage of this method compared with the symbolic addition method is that especially in the initial, and most important, stages, the multiplicities (and thus the probabilities) of sign indications are either increased or eventual contradictions are easier to recognize, since no unknown symbols are involved. A disadvantage is the need for more than one complete calculation to obtain all possible solutions. However, the computer time needed per calculation is very small even with a fairly small modern computer. The problem of reducing the number of necessary calculations is of course still important, and therefore the procedure described below for selecting a proper basis set with a rather small number of reflections with unknown signs has been adopted. In the six sign determinations (*cf.* Table 1) performed so far, using this procedure, not more than two calculations per sign determination have been necessary (*i.e.* not more than one unknown sign has been included in each basis set).

SELECTION OF A PROPER BASIS SET

The initial step of the procedure for selecting a proper basis set consists of examining the triple relations, to search for explicit sign indications for intensity invariants (*e.g.* sign relations of the type $\text{sign}(E_{2h})\cdot\text{sign}(E_h)\cdot\text{sign}(E_h)\approx +$). The signs determined by at least two such relations, without contradictions, are then substituted into the triple relations, and new sign indications thus obtained are substituted into the sign relations, *etc.* When no more sign indications are found, the frequencies of the reflections in the remaining sign relations are calculated. The reflection with the highest frequency is selected, and in this preliminary calculation to choose a proper basis set, it is assigned a symbol, say A, which is then substituted into the sign relations. All new sign or symbol indications, chosen as those with the highest frequency in each case, are then used to further reduce the remaining sign relations, *etc.* When no more new sign or symbol indications are found, the reflection occurring most frequently in the remaining sign relations is assigned a new symbol, say B. In this way, successive new symbols are introduced, until either the most frequent reflection in the remaining relations has a frequency less than the average frequency, calculated from the original triple relations, or all reflections have assignments.

Since the primary object of this small initial calculation is the selection of a proper basis set, no attempt is made to determine relations among symbols. Among the symbol defining reflections, a proper origin-defining set is selected, and assigned fixed signs. The rest of the symbol-defining reflections necessary for the complete sign-determination will then have variable signs. As a rule, only one reflection with unknown sign needs to be included in the basis set, to generate sign indications for most of the remaining reflections. Basis sets with different sign permutations for the reflections with variable signs are then used to generate the corresponding alternative solutions of the triple product sign relationships.

GENERATION OF SOLUTIONS OF SIGN RELATIONS

The solution procedure for each basis set is divided into three steps, initialization, expansion of basis set, and finally completion.

The first step, initialization, yields a preliminary set of signs for the reflections. First a search for explicit sign indications for intensity invariants among the triple relations is performed. All signs determined by at least two such relations and without contradictions are then used, together with the signs given in the basis set to reduce the triple relations as far as possible. This done, all newly-determined signs are used to reduce the triple relations, and so on. The step is finished when no more signs are determined.

In the next step, an expansion of basis set is performed in steps to reduce the propagation of errors. The signs determined with a multiplicity greater than or equal to a given number m in the previous step are selected and added to the original basis set, while the rest of the signs are considered undetermined. The multiplicity of a sign indication for a reflection is calculated as the number of indications of the most frequent sign minus the number of indications of the opposite sign. This extended basis set is then used as described above, to reduce triple relations and so on, except that a new sign is approved only if its multiplicity during the solution procedure becomes equal to or exceeds the number m. When no more signs are approved, the value of m is decreased by one, and a new, more extensive basis set is formed, and the above procedure is repeated. Thus, a stepwise expansion of the basis set is performed by reducing the value of m, until m reaches the value 2.

The last step is performed to complete the determination of signs and to check the self-consistency of the solution obtained. In this step, one reflection at a time is examined, and all the sign indications for this reflection obtainable from all those reflections whose signs have been previously determined with a multiplicity greater than or equal to 2 are calculated. To check the self-consistency of the solution, the examination of indicated signs also includes the reflections of the original basis set. For each reflection, the number of indications of both the most frequent sign and of the opposite sign (number of contradictions) is calculated. Finally, all the contradictions

Table 1.

Compound No.	Composition of asymmetric unit	Space group	Number of E-values	$E_{\min}$	Number of sign relations	$P_{\min}$	Number of unknown signs in basis set	Number of signs determined	Number of erroneous signs
1	$C_{14}H_{14}N_4O_2$	$P2_1/c$	233	1.50	997	0.95	1	205	0
2	$C_{16}H_{18}N_4O_3$	»	250	1.57	929	0.95	1	238	5
3	$C_{12}H_{14}N_8O_7$	»	250	1.31	958	0.91	1	228	0
4	$C_{22}H_{22}BrN_4O_5$	»	243	1.70	934	0.985	0	236	0
5	$C_{17}H_{21}ClN_4O_6$	»	250	1.25	776	0.90	0	235	37
6	$C_{24}H_{36}Cl_3Cu_3N_{12}O_{12}$	*Pnma*	174	1.40	907	0.975	0	170	–

Compound No. 1: 3-Methyl-lumiflavin.
2: 3,7,8,10-Tetramethyl-5-acetyl-1,5-dihydroalloxazine.
3: Di-lumazin trihydrate.
4: Hydroquinone lumiflavinium hydrobromide.
5: *O,N*-Diethyl-lumiflavinium perchlorate.
6: Copper(I)tetraacetonitrile perchlorate.

and the sum of signed E-values for all reflections with determined signs are calculated to facilitate the choice of the best solution.

The output of the program, apart from listings, consists of a punched card deck of signed E-values for input to our Fourier programs.

APPLICATIONS OF THE PROCEDURE

In Table 1, a brief description of the results obtained for the first six structures solved by the use of this program is given. More detailed descriptions of the solutions and of the structures will shortly be published elsewhere. The first four applications (*cf.* Table 1) are quite successful. However, in the fifth example, the number of erroneously assigned signs is rather high. The reason for this is probably that the reflections used for phasing included 11 reflections which through errors in intensity measurements were assigned high values of E, averaging 3.8, whereas their true values are each much less than 1.

The sixth example, copper(I)tetraacetonitrile perchlorate, is in fact non-centrosymmetric (space group $Pn2_1a$), but the copper and chlorine positions could be approximately described by using the centrosymmetric space group *Pnma*. In this case, the calculations were performed to obtain approximate positions for the six independent heavy atoms.

A procedure similar to the one described here, but for solving non-centrosymmetric triple phase relations, has been developed and programmed for the IBM 1800 by the present author. A description will be published in the near future.

Acknowledgements. The author wishes to express gratitude to Professor Peder Kierkegaard for his continuous interest in this work. The author is indebted to Dr. Don Koenig for his correction of the English of this paper.

A research fellowship from the *Stiftelsen Bengt Lundqvists Minne* is gratefully acknowledged.

This investigation has been performed with financial support from the *Tri-Centennial Fund of the Bank of Sweden*, and from the *Swedish Natural Science Research Council*. Permission for the use of the computer IBM 360/75 was granted by the *Computer Division of the National Swedish Rationalization Agency*.

REFERENCES

1. *IBM 1800 Time-sharing Executive System*, Order No. GC26-3703-1 (1970).
2. Karle, J. and Karle, I. L. *Acta Cryst.* 21 (1966) 849.
3. Germain, G. and Woolfson, M. M. *Acta Cryst.* **B 24** (1968) 91.

Received September 2, 1970.