ATTI ACCADEMIA NAZIONALE DEI LINCEI

CLASSE SCIENZE FISICHE MATEMATICHE NATURALI

RENDICONTI

RICCARDO BIANCHI, RICCARDO DESTRO, ANGELO GAVEZZOTTI

A set of programs for the application of direct methods in Crystallography

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8, Vol. **52** (1972), n.5, p. 764–769.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1972_8_52_5_764_0>

Strutturistica chimica. — *A set of programs for the application of direct methods in Crystallography* (*). Nota di RICCARDO BIANCHI, RICCARDO DESTRO e ANGELO GAVEZZOTTI, presentata (**) dal Corrisp. M. SIMONETTA.

RIASSUNTO. — La serie di tre programmi di calcolo (SIG2–SYMBAD–TANFOR) descritti in questa Nota affronta la soluzione del problema della fase in cristallografia. In questi programmi, i « metodi diretti » vengono applicati a tutti i gruppi spaziali senza restrizioni. Ad una sintesi generale del problema segue una descrizione dei tre stadi in cui è stata suddivisa la sua soluzione. L'ultimo di questi stadi (applicazione della « formula della tangente ») è usualmente impiegato solo in strutture non centrosimmetriche.

INTRODUCTION

The purpose of this paper is to describe a set of three programs for handling the problem of phase determination by direct methods in crystallography. In recent years, many Authors have developed similar programs [1], [2], [3] with remarkable success for centrosymmetric, and, with some restrictions, for non-centrosymmetric structures. Two years ago, when the main features of our programs have been established, the need was still felt for a general purpose program which could handle space groups other than those of triclinic, monoclinic, and orthorhombic systems; and since most of the programs to day available still suffer from this restriction, we give here an outline of our straightforward method for solving this problem, in a suitable form for medium-size computer use. On the other hand, our philosophy has been such as to leave the user the task of taking the most sensitive logical decisions, while automatizing the standard part of the process. Under this respect, we feel that a balance is needed between automatization and human judgement.

OUTLINE OF THE PROCEDURE

A synthesis of the general formulation of direct methods theory for phase determination in crystallography can be found in a paper by J. Karle and I. L. Karle [4].

According to this theory, the phase of a reflexion can be estimated as

$$\varphi_{\boldsymbol{h}} = \langle \varphi_{\boldsymbol{k}} + \varphi_{\boldsymbol{h}-\boldsymbol{k}} \rangle_{\boldsymbol{k}}$$

where $\boldsymbol{h}$, $\boldsymbol{k}$, and $\boldsymbol{h} - \boldsymbol{k}$ are reciprocal lattice vectors.

(*) Work done at the Institute of Physical Chemistry–University of Milan, Milan, Italy.

(**) Nella seduta del 13 maggio 1972.

A more accurate formula, known as the "tangent formula", is

$$\tan \varphi_h = \frac{\sum_k |E_k \, E_{h-k}| \sin (\varphi_k + \varphi_{h-k})}{\sum_k |E_k \, E_{h-k}| \cos (\varphi_k + \varphi_{h-k})} = \frac{S_h}{C_h}$$

where E_k, E_{h-k} are normalized structure factors. These formulae, usually applied to reflexions with E's of largest value, require the generation of reflexions with equivalent indices, and consequently the calculation of their phase relationship with the original ones, according to the space-group symmetry laws. This stage of the process, which is often the limiting one for many analogous programs, is here performed in a way that under many respects is similar to an early treatment by Patterson [5] recently emphasized by Brown [6]. If $\mathbf{M}_1, \mathbf{M}_2, \cdots, \mathbf{M}_n$ are the space-group rotation matrices, and $\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_n$ their associated translation vectors, so that a general equivalent position is represented by

$$\boldsymbol{x}'_i = \mathbf{M}_i \boldsymbol{x} + \boldsymbol{t}_i$$

it can easily be shown that the equivalent indices can be generated as

$$\boldsymbol{h}'_i = \boldsymbol{h}\mathbf{M}_i$$

$\boldsymbol{h}'_i$ and $\boldsymbol{h}$ being row reciprocal space vectors. The phase difference between $E_{\boldsymbol{h}'_i}$ and $E_{\boldsymbol{h}}$ is given by

$$\varphi_{\boldsymbol{h}'_i} - \varphi_{\boldsymbol{h}} = \Delta\varphi_i = -2\pi (\boldsymbol{h}\boldsymbol{t}_i) .$$

These equations afford an easy route to a computer-suitable solution of the problem of equivalent reflexions. Program SIG2, the first in our chain of programs, can do these calculations, keeping the amount of needed fast core within reasonable limits, and providing intermediate paper and tape output to be transmitted to successive programs.

The next step, that is, the choice of origin-defining and symbolic phases, is in our present philosophy left to the user's judgement, although we are planning to introduce automatic criteria to select the more appropriate starting set of phases. This is done with remarkable success for instance by Bob Long's program [1] for centrosymmetric structures, and by MULTAN program [2] for non-centrosymmetric ones. Our intermediate program SYMBAD, in its present version, performs in a first approach a symbolic addition procedure to determine a given number of phases, recording all the relationships among symbols and known phases. A remarkable feature of this procedure is that probable values for symbols can be obtained, thus limiting the number of trial computations by the tangent formula. Other programs, as the previously mentioned ones, give more reliable information about reflexions to be chosen as symbols but no information at all about the value of their phases before the complete phase-determining procedure. The last program, TANFOR, applies the tangent formula to determine unknown phases and to refine known ones, until internal consistency is reached. In the following sections, the three programs will be described in detail.

DESCRIPTION OF THE PROGRAMS

a) *SIG2 program.*

Input requirements for SIG2 are: *a*) independent normalized structure factors, arranged in decreasing order, *b*) a single parameter card, and *c*) symmetry information, that is, space-group rotation matrices and translation vectors, as obtainable from International Tables [7].

Equivalent reflexions are generated by the previously outlined method; it is important to observe that not only are phase relationships obtained, but also restrictions on the phase value are detected, and stored in coded form, for special reflexions.

The maximum number of input reflexions can be determined as a function of the size of the main data array (NSIZE) and the number of equivalent positions in the space-group (NUASIM) by the formula

$$\text{max no. of reflexions} = \frac{\text{NSIZE}}{4 + \text{A}\cdot\text{NUASIM}}$$

where A is 5 for centrosymmetric, and 10 for non-centrosymmetric space-groups. In our present version, NSIZE = 18000; if desired, this parameter can be varied by minor modifications, carefully specified on the program listing.

The paper output consists of the interactions list, in the following fashion: the reflexion whose phase is to be determined is called $h_1 k_1 l_1$; the phase relationship is given by

$$\varphi_{h_1 k_1 l_1} = \varphi_{h_2 k_2 l_2} + \varphi_{h_3 k_3 l_3}$$

where

$$h_1 - h_2 = h_3 \quad , \quad k_1 - k_2 = k_3 \quad , \quad l_1 - l_2 = l_3 \, .$$

The phases for $h_2 k_2 l_2$ and $h_3 k_3 l_3$ are always given in terms of the initial set of E's. For example, if $h_3 k_3 l_3$ is -2, -4, -3 and the original input reflexion is 2 , 4 , 3 then 2 , 4 , 3 will be printed and $\varphi_{h_3 k_3 l_3} = -\text{F}$, where F is the phase of the 2 , 4 , 3 reflexion. For each interaction, the probability value is computed according to the formulae

$$\text{P}_{\boldsymbol{h}} = \frac{1}{2} + \frac{1}{2} \tanh \{ \sigma_3 \, \sigma_2^{-3/2} \, | \text{E}_{\boldsymbol{h}} \, \text{E}_{\boldsymbol{k}} \, \text{E}_{\boldsymbol{h}-\boldsymbol{k}} | \}$$

$$\text{K}_{\boldsymbol{h},\boldsymbol{k}} = 2 \, \sigma_3 \, \sigma_2^{-3/2} \, | \text{E}_{\boldsymbol{h}} \, \text{E}_{\boldsymbol{k}} \, \text{E}_{\boldsymbol{h}-\boldsymbol{k}} |$$

with

$$\sigma_n = \sum_{1}^{\text{N}}{}_{j} \, \text{Z}_j^n$$

where P applies for centrosymmetric, and K for non-centrosymmetric space-groups. The essential information is also recorded on magnetic tape for successive use.

b) *SYMBAD program.*

This program has been devised rather like a trial-and-error routine than a decision-making automatization of the problem of the starting set of phases. Many origin assignments and symbol choices can be tested in turn for internal consistency; the program output includes an alphabetic display of phase determinations in terms of symbols, messages and information about detected inconsistencies, deriving both from special phases limitations or from contradictory assignments. A list of the relationships among symbols is also included, from whose analysis the choice of the starting set can be submitted to useful criticism. No numeric criteria of convergence or figures of merit are given; though we are planning improvements of this part of the process, at this stage only the user's experience can decide, while the dull work of manual addition is avoided.

Beside phase- and origin-assigning cards, input to SYMBAD consists only of a few key integers; all other needed information is fed to the program by the output tape of SIG2. Two options are available, that is, newly assigned phases can be used at once or only in a subsequent cycle. Up to three different definitions for a given phase are memorized, while the actual value of the phase, as used to determine other phases, is the most probable one, in the sense of the largest P or K. Up to 200 relationships among symbols are memorized and displayed at the end of each cycle.

Program stops after a given number of cycles or after determining a given number of phases.

c) *TANFOR program.*

Starting from a set of known phases, this program determines a certain number of unknown phases, which are then refined together with the starting set. This procedure is repeated until all phases are determined and refined. The whole process is performed by the tangent formula, where the E's can be given a weight calculated by any of the three formulae

$$W = 1$$

$$W = \tanh \frac{1}{2} \{1 + \cos \Delta\varphi\} \frac{\alpha_h}{2}$$

$$W = \tanh \{\exp(-C[\Delta\varphi]^2)\} \frac{\alpha_h}{2}$$

$\Delta\varphi$ is the change in phase in two subsequent refinement cycles, and C is an input constant; α_h is defined by

$$\alpha_h = \left\{ \left[\sum_k K_{h,k} \cos(\varphi_k + \varphi_{h-k}) \right]^2 + \left[\sum_k K_{h,k} \sin(\varphi_k + \varphi_{h-k}) \right]^2 \right\}^{1/2}$$

where $K_{h,k}$ is the formerly defined parameter.

During the phase determination and refinement process, a phase assignment is accepted only if two reliability criteria are contemporaneously satisfied, that is, α_h and T_h greater than an assigned value; α_h is the above defined quantity, while T_h is [3], [8],

$$T_h = [S_h^2 + C_h^2]^{1/2} \Big/ \sum_k |E_k E_{h-k}| .$$

Note that α_h is related to the variance of the phase angle [4], [9] and T_h represents an internal consistency index, being equal to a maximum value of 1 when no contradictions are present in the assignment of the phase under consideration. For a single relationship, T_h is obviously equal to 1, and α_h is a more reliable criterion; with many relationships, α_h is usually large and T_h is the more reliable criterion for acceptance or rejection.

Two figures of merit are calculated at the end of each refinement cycle [3], [8], [9]:

$$Q = \frac{\sum_h E_h (1 - T_h)}{\sum_h E_h}$$

$$S_w = \sum_h E_h^2 (S_h^2 + C_h^2) .$$

When comparing two sets of refined phases in order to determine which one is most likely correct, the one with the smaller Q value, or the one with the largest S_w value may be chosen. Both Q and S_w are calculated including only the accepted reflexions. Q is much more a function of the number of reflexions than S_w is; it is therefore usually advisable to make the choice on S_w rather than on Q.

CONCLUDING REMARKS

From the above description it appears that our programs are mainly devoted to non-centrosymmetric structures. However, it is worth while to point out that in at least one case a highly disordered centrosymmetric structure, that could not be solved by the usual formulae for centrosymmetric space groups, was readily determined making use of the tangent formula. In our opinion, this is due to the accurate criteria of acceptance or rejection during the phase determination process.

Various non-centrosymmetric structures have been solved by this set of programs in two years of use. A good example is the structure of 9-ethylguanine, which crystallizes in space group $P4_12_12$ with 208 non-hydrogen atoms in the unit cell. For this structure, 207 reflexions have been introduced in SIG2 program, which produced 15300 relationships in 2′ 30″ on an UNIVAC 1106 computer. About 15″ are taken by SYMBAD program for the determination of 50 symbolic phases, working only on E's greater than 2.1, again on UNIVAC 1106. The expansion and refinement of the 207 phases by TANFOR takes 1′ 45″ on the same computer.

Acknowledgement. – This work has been financially supported by the Consiglio Nazionale delle Ricerche.

REFERENCES

[1] R. E. LONG, Doctoral Thesis, University of California, Los Angeles (1965).
[2] G. GERMAIN, P. MAIN and M. M. WOOLFSON, «Acta Cryst.», *A27*, 368 (1971).
[3] M. G. B. DREW, D. H. TEMPLETON and A. ZALKIN, «Acta Cryst.», *B25*, 261 (1969).
[4] J. KARLE and I. L. KARLE, «Acta Cryst.», *21*, 849 (1966).
[5] A. L. PATTERSON, in *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, R. Pepinsky, Editor, State College Pennsylvania, X-ray Crystal Analysis Laboratory (1952).
[6] G. M. BROWN, «Acta Cryst.», *B27*, 1675 (1971).
[7] *International Tables for X-ray Crystallography*, Birmingham, Kynoch Press (1962).
[8] M. G. B. DREW, «Acta Cryst.», *B25*, 1320 (1969).
[9] G. GERMAIN, P. MAIN and M. M. WOOLFSON, «Acta Cryst.», *B26*, 274 (1970).