**Abstract**

A program is presented that solves crystal structures completely using advanced Fourier methods starting from a small known fraction (about 8%) of the structure. The program uses a statistical test based on the \( R_2 \) factor to check the correctness of the model obtained so far. The results obtained with four test structures are given.

**Introduction**

Modern direct-method programs or Patterson search methods sometimes yield only inaccurate and/or incomplete models. Extension of the partial models is not always easy. If the trial positions are not sufficiently accurate or the model contains too many incorrectly placed atoms even direct methods applied to difference structure factors will fail. Our purpose was to develop a procedure that needs only marginal information to obtain the complete structure automatically. To achieve this we used advanced Fourier methods (Sim, 1959; Main, 1979) in combination with a test based on the \( R_2 \) factor (Van Havere & Lenstra, 1983a, b).

\[
R_2 = \sum_h \left[ \frac{1}{E_0} \left( \frac{\|E|}{E_0} - 1 \right)^2 \right]^2 \sum_h \left[ \frac{1}{E} \right]^4;
\]

where \( h \) is a reciprocal-lattice vector, \( E_0 \) is the observed normalized structure factor, \( E \) is the calculated normalized structure factor, \( n = \) number of known atoms in the unit cell and \( N = \) total number of atoms in the unit cell.

\[
\eta_1 = \frac{\eta}{\eta_0}; \quad \eta_2 = \sum_{j=1}^{n} f_j^2; \quad \eta_3 = \sum_{j=1}^{n} f_j^2.
\]

**Procedure**

The input of the program consists of the overall thermal parameter \( (B_{av}) \), the scale factor, the number of atoms of each kind in the unit cell and the positional parameters of the known atoms. On the assumption that the atoms put in are correctly placed, \( R_2 \) can be estimated by (2) or (2a):

\[
\langle R_2 \rangle \langle E_0 \rangle = \sum_h \left[ \frac{1}{E_0} \left( \frac{\|E|}{E_0} - 1 \right)^2 \right]^2 \sum_h \left[ \frac{1}{E} \right]^4;
\]

for non-centrosymmetric space groups (Van Havere & Lenstra, 1983b). and:

\[
\langle R_2 \rangle \langle E_0 \rangle = \sum_h \left[ \frac{1}{E_0} \left( \frac{\|E|}{E_0} - 1 \right)^2 \right]^2 \sum_h \left[ \frac{1}{E} \right]^4;
\]

for centrosymmetric space groups (Van Havere & Lenstra, 1983b), where \( r \) is the number of correctly placed atoms in the unit cell (we assume \( r = n \)).

If \( R_2 \) is equal to or lower than the estimated value, the input model is supposed to be essentially correct. If not, \( R_2 \) is calculated leaving out each atom in turn. Atoms causing \( R_2 \) to decrease on their omission are deleted from the model. The revised model obtained after this selection is used to calculate structure factors and Fourier coefficients of a map to be constructed. If only a very small part of the structure is known, the coefficients

\[
x_{gen} = \frac{\|E\|^2 E}{N}
\]

are used (Ramachandran & Raman, 1959).

The best map can be obtained from coefficients

\[
X = 2\|E\| \left[ \frac{\|E\|}{\|E\|} - \eta_1 - \eta_2 - \eta_3 \right]
\]

(Woolfson, 1956), \( \phi = \) the phase of \( E \).

\[
X = 2\|E\| \left[ \frac{\|E\|}{\|E\|} - \eta_1 - \eta_2 - \eta_3 \right]
\]

(Main, 1979).

In the map about 1-5 times as many peaks as were input – or obtained from a previous cycle – are searched for. Note that the weighted Fourier suppresses wrong atoms. During the iterations only three types of atom are used as possible scatterers (i.e., the carbon-like, the chlorine-like and the metals). To each new peak the scattering factor giving rise to minimal \( R_2 \) is assigned.

The model then obtained is treated as was the original input. The procedure is repeated until all atoms are found and the \( R_2 \) is below a threshold of about 0-2-0-3. We are currently investigating the possibility of estimating the final \( R_2 \), given small errors in the positional parameters of the model (Van de Mieroop & Lenstra, 1978), in order to obtain an independent convergence criterion.

When a complete model is obtained the atomic numbers are calculated from the set of equations

\[
0021-8898/84/050364-03S01.50 \quad \text{© 1984 International Union of Crystallography}
\]
\[
R_j = \left[1 - \sum_{h,k} (\eta^2|E|_h^2)/(\eta^2|E|_h^2) \right]^{1/2} 
\]

for \(j = 1,2,\ldots,N\) (3)

For a derivation we refer to the Appendix.

**Remarks**

According to Lenstra (personal communication), application of the \(R_2\) criterion requires atomic resolution. Thus, **AUTOFOUR** cannot be expected to work properly for protein structures, although the Fourier methods used in the program are common in this branch of crystallography.

If we encounter disordered structures, the program will not be able to assign the right scattering-factor tables to the disordered atoms because a proper estimate of the occupancy is not available at the time.

Sim's weighting scheme can only work properly if a good estimate of both \(B_{ov}\) and the scale factor is available. We intend to implement proper statistics for the estimation of these quantities (Subramanian & Hall, 1982; French & Wilson, 1978).

If we denote a model with \(r\) right and \(w\) wrong atoms as \(M(r,w)\ (r + w \leq N)\), we can conveniently describe the two fundamentally different methods for obtaining the structure. The first method starts from a complete but completely wrong model \(M(0,N)\) and tries to improve it. This is called the random-atom method (Booth, 1947, 1949). The second strategy, starting from an incomplete but correct model \(M(n,0)\), \(n \leq N\), is called the zero-atom method (Lenstra, 1974). It has been found by Van Havere & Lenstra (1983b) that the second strategy is best combined with the \(R_2\) selection.

If we compare **AUTOFOUR** with a procedure proposed by Simonov (1978) it is seen that **AUTOFOUR** tries to follow the zero-atom method as well as possible, whereas Simonov's procedure is more akin to the random-atom method (i.e. **AUTOFOUR** starts from \(M(0,0)\) and expands the model via \(M(n+m,0)\) as well as possible, while Simonov starts from \(M(n,0)\) and expands the model via \(M(n+m,N-n-m)\)). Another difference between the procedures is the way in which the Fourier map is interpreted. Simonov's program tries to assign 'atomic numbers' to the peaks by multiplying some kind of standard scattering power by a weight proportional to the peak height, where **AUTOFOUR** determines the atomic numbers by minimizing the \(R_2\).

Essential improvements on Simonov's method are the use of Sim's weighting scheme and the \(R_2\) selection. Fourier methods based on Sim's scheme are far more powerful in suppressing wrong atoms than the conventional ones.

It is our experience that only the combination of weighted Fouriers and the \(R_2\) selection enables the procedure to cope with a sizeable percentage of erroneous input, an essential property of the algorithm sought after.

An interesting procedure has already been presented by Sheldrick (1982). The strategy is similar but Sheldrick's procedure differs from **AUTOFOUR** by the selection criterion. Where **AUTOFOUR** uses \(R_2\), Sheldrick selects new peaks by minimizing \(R_E\):

\[
R_E = \left[1 - \sum_{h} |E|_h^2 / \sum_{h} |E|_h^2 \right] {\sum_{h} |E|_h^2}^{1/2} 
\]

Since no independent estimate, \(<R_2|E>|\), is available, Sheldrick minimizes the agreement index each cycle. It should also be noted that Sheldrick uses only high \(E\) values during the iterations. It is shown by Van Havere & Lenstra (1983b, c) that omission of reflexions decreases significantly the resolving power of selection criteria based on intensities.

**Test structures**

The program can be used for various types of structure. We have tested the procedure on the equal-atom structures of toxisterol \(C_1\) (de Kok, Boomsma & Romers, 1976) and methyl 3,4-O-isopropylidene-2,6-di-O(2,3,4,6-teta-O-acetyl-\(\beta\)-d-galactopyranosyl)-x-d-galactopyranoside (hereafter TRIGAL) (Hoogendorp, de Kok & Romers, 1983), both \(P2_1\). Furthermore, we have used the program in the structure determinations of bis[triaquabis(4-ethyl-1,2,4-triazole-\(N^1\)-iron(II)-\(N^2\)N\(^2\)N\(^2\)N\(^2\)]iron(II) hexakis(trifluoromethanesulfonate) (hereafter FETRIFL) (Vos, 1983) and hexakis(1-methyltetrazole)copper(II) bis(tetrafluoroborate) (hereafter CUIMTZ).

FETRIFL crystallizes in space group \(P31c\) and CUIMTZ in \(P2_1\)/\(n\). We have compared the results of **AUTOFOUR** with those of **DIRDIF** and Simonov's procedure. **DIRDIF** is a program which applies direct methods to difference structure factors (van den Hark, Prick & Beurskens, 1976). It is our experience that **DIRDIF** does not work when the model contains too many wrongly placed atoms. We were unable to solve the structures of toxisterol \(C_1\) and TRIGAL using **DIRDIF**, while **AUTOFOUR** was quite successful in determining all positions automatically using the same input models. These were obtained by using **DETER** (Vermin & de Graaff, 1982) as conventional direct methods failed for these structures. Simonov's procedure also failed for these structures. Furthermore, the structure of CUIMTZ could not be solved using Simonov's program. **AUTOFOUR** is able to solve even those structures where little \(a\) priori information is available.

In our experience about 8% of the scattering power should be placed more or less correctly. In the case of CUIMTZ we obtained one position for a copper atom from the Patterson. The rest was done automatically by **AUTOFOUR**. This structure could also be solved using **DIRDIF**. The structure of FETRIFL could be obtained using only one of the two known iron positions \((1, 3, 2)\) by all three programs.

The authors wish to thank Professor Dr A. T. H. Lenstra for his helpful suggestions and Dr W. K. L. Van Havere for permission to use his results before publication.

**APPENDIX**

We assume that the correct set of scattering-factor tables is the one that minimizes the \(R_2\) factor, thus

\[
\partial R_2/\partial Z_j = 0 \quad \text{for } j = 1,2,\ldots,N. \quad (A.1) 
\]

In order to make full use of the prior information we slightly modify the formalism according to

\[
\partial R^{(1)}_2/\partial Z_j = \partial R^{(2)}_2/\partial Z_j. \quad (A.2) 
\]

where \(\Delta Z_j = \text{the shift in } Z_j \text{ and } R^{(1)}_2 \text{ is the } R_2 \text{ factor after and}
An enhanced interactive PLUTO78 for molecular display. By K. M. CRENNELL and G. M. CRISP, Computing Division, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX12 OQX. England

(Received 12 January 1984; accepted 30 May 1984)

Abstract

The well known molecular drawing program, PLUTO78, has been made more widely available in the UK by mounting it on the minicomputers of the Joint Academic Network (JANET) wide-area network. Enhancements include interaction, faster plotting, use of colour and simulated half tone, easy access to centrally maintained high-quality-output devices, and the provision of comprehensive documentation in the form of on-line help and a user manual.

1. Introduction

Many small university groups need to display molecular structures but lack either programming effort to mount simple interactive display programs or capital for a more sophisticated molecular graphics unit. Some have therefore used PLUTO78 (Motherwell, 1978) and ORTEP (Johnson, 1976) in batch mode to obtain the plots necessary for research and teaching. ORTEP's chief advantage over PLUTO78 is that it can draw thermal ellipsoids, but it cannot do space-filling plots and its command language is difficult to understand. In cases where thermal ellipsoids are needed, the PLORTEP program (Bandel & Sussman, 1983) can be used to convert PLUTO78 commands to ORTEP ones. The lack of interaction and the consequent delay in obtaining plots has restricted the use of PLUTO78 to the production of final plots for publication. To provide useful molecular plots on simple graphics terminals, PLUTO78 has been implemented as a limited interactive package (limited in the sense that selective erasure of parts of the current plot is not supported, and interaction is solely through the keyboard), on the wide-area Joint Academic Network (JANET; Wells, 1984).

Another benefit of connection to JANET is the ability to interrogate the Cambridge database of organic structures interactively, via the Crystal Structure Search and Retrieval (1981) program maintained at Edinburgh by Daresbury Laboratory. Selected datasets may easily be transferred across the network to the user's local site and displayed with PLUTO78.