make it necessary to include components of **k** in directions perpendicular to L_t (here $\frac{1}{2}\mathbf{b}^*$) in the notation for the symmetry group. Therefore we prefer (*a*) to normalize the modulation vector by always choosing it in L_t, (*b*) to express the 'improper' symmetry translation (here **b**) in the lattice symbol. This can be done by using symbols for magnetic Bravais lattices, such as p_{2b} in the case of Fig. 6, *cf.* Opechowsky & Guccione (1965). Indeed the value of s is, with one exception, restricted to $\frac{1}{2}$, so the extra shift, being a binary operation which commutes with all translations, is formally equivalent to time reversal.

The exception is the trigonal system which allows $s = \pm\frac{1}{3}$ for improper translations. A table of Bravais-lattice types including all inequivalent lattices with and without improper symmetry translations, for each of the seven systems from (I) Table 1 is given in Table 1.*

The author gratefully acknowledges various stimulating discussions with Professor A. Janner and Dr T. Janssen (Nijmegen) as well as their reading two earlier versions of this paper.

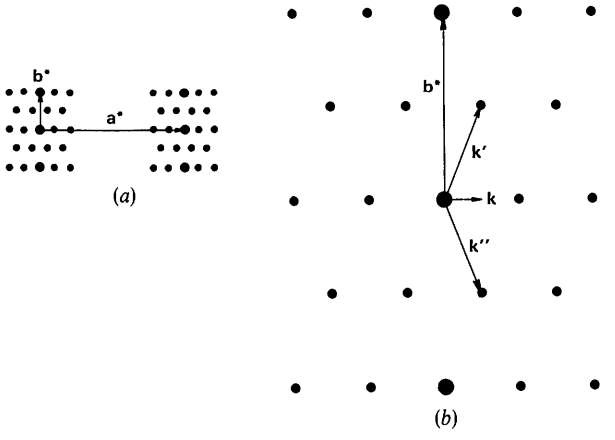

Fig. 7. The reciprocal net of Fig. 6. (*a*) Overall picture; (*b*) surroundings of the origin, magnified.

* Table 1 has been deposited with the British Library Lending Division as Supplementary Publication No. SUP 32323 (3 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 13 White Friars, Chester CH1 1NZ, England.

References

AALST, W. VAN, DEN HOLLANDER, J., PETERSE, W. J. A. M. & DE WOLFF, P. M. (1976). *Acta Cryst.* B**32**, 47–58.
JANNER, A. & JANSSEN, T. (1976). *Phys. Rev.* In the press.
JOHNSON, C. K. & WATSON, C. R. (1976). *J. Chem. Phys.* **64**, 2271–2286.
OPECHOWSKY, W. & GUCCIONE, R. (1965). *Magnetic Symmetry in Magnetism.* IIa, edited by G. T. Rado & H. Suhl. New York: Academic Press.
SCHULZ, H. (1974). *Acta Cryst.* B**30**, 1318–1332.
SHIOZAKI, Y. (1971). *Ferroelectrics*, **2**, 245–260.
WOLFF, P. M. DE (1974). *Acta Cryst.* A**30**, 777–785.

Acta Cryst. (1977). A**33**, 497–499

An Interactive Model-Building Program for Macromolecules

BY A. JACK

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

(*Received* 11 *November* 1976; *accepted* 28 *January* 1977)

A computer model-building program (suitable for interactive use on a small machine) is described. The program fits a polymer chain with idealized stereochemistry to a number of guide coordinates, by rotation about single bonds. It has been used to record accurate atomic coordinates from a skeletal wire model of a protein in a fraction of the time which manual measurement would have taken.

Introduction

A sufficiently large number of protein and nucleic acid structures have now been solved at atomic resolution, that the strategy of structure determination and refinement has become well-established. Much thought is currently being given to automation of the more routine aspects of such work such as X-ray data collection. One step which has not generally been automated is the measurement of atomic coordinates from a skeletal wire model which has been fitted to an electron density map. One approach to this problem is to abandon physical models, and to carry out the fitting by computer, either tracing the polymer chain automatically (Greer, 1976), or building a computer model to fit the density while both are displayed on an interactive graphics terminal (Feldmann, 1976). Nonetheless, a physical model has undeniable advantages for visualizing the structure as a whole, and it seems unlikely that the traditional methods will be totally supplanted.

The computer model-building program described here may be used in conjunction with a model to generate a stereochemically acceptable set of coordinates with a minimum of measurement.

Computer model-building

A preliminary to the refinement of a macromolecular structure is the recording of the coordinates of each

atom, followed by adjustment of these coordinates so that the bond lengths and angles agree with those found from accurate crystallographic studies of small molecules. Two methods of achieving this idealization have been proposed for proteins. The first (Diamond, 1966) constructs a polypeptide chain in which the bond lengths and angles have *exactly* the values found in amino-acid structures. This chain is then folded by rotation about chosen lines (usually single bonds), to fit a set of guide coordinates as well as possible. The second approach applies small shifts to the Cartesian coordinates of each atom individually in an attempt to minimize the potential energy of the molecule (Levitt & Lifson, 1969; Levitt, 1974) and so does not force the resulting coordinates to have *exact* bond lengths and angles. Some variants of this technique have been described (Hermans & McQueen, 1974; Dodson, Isaacs & Rollett, 1976; Ten Eyck, Weaver & Matthews, 1976). For some purposes these energy-based algorithms seem preferable, in that they are flexible and computationally fast, but they are not well-suited to producing a chain with a *given conformation*, since they are concerned with Cartesian coordinates rather than with torsion angles. Diamond (1976) has discussed the relative merits of the different approaches at some length.

During crystallographic refinement it often becomes necessary to change the conformation of part of the polymer chain, leaving bond lengths and bond angles unchanged, and without disturbing the rest of the molecule or breaking the chain continuity. Since conformation is usually defined in terms of torsion angles, a technique such as Diamond's seems the most appropriate. Unfortunately, his program (in its present form) has certain restrictions which make it applicable only to proteins. During the structure refinement of yeast phenylalanine tRNA (Jack, Ladner & Klug, 1976) it became apparent that we needed a similar program able to handle any type of chain whose coordinates can be listed in the way originated by Diamond.

Method of fitting the chain

The program takes as input a list of atomic coordinates in the same format as is used by Diamond's (1966, 1971, 1976) programs, and a library of standard groups, these being Cartesian coordinates of the monomeric units comprising the chain. The model is built sequentially, one residue at a time. The required conformation may be specified in terms of guide coordinates (which are normally approximate atomic positions measured from a model or an electron density map), or in terms of torsion angles, or some combination of both.

The idealization of each residue takes place in two stages (Fig. 1). Stage 1, the initialization, reads the appropriate standard group and then moves through it one angle at a time, at each stage considering only those atoms which will not be affected by subsequent angles. For example, in Fig. 1, if we imagine that atoms *a*, *b* and *c* of the standard group have been fitted to guide points *A*, *B* and *C*, then the initialization will produce that value of α which minimizes $w_d|\mathbf{d}-\mathbf{D}|^2$, where w_d is a weight associated with atom *d*, and **d** and **D** are the position vectors of the two atoms *d* and *D*. β is then taken as the angle which minimizes $w_e|\mathbf{e}-\mathbf{E}|^2 + w_f|\mathbf{f}-\mathbf{F}|^2$. Requests for specific values of torsion angles are satisfied *exactly* at this stage, regardless of the fit to the guide points.

Stage 2 of the idealization is a least-squares minimization of the residual

$$\sum_i w_i|\mathbf{r}_i-\mathbf{r}_i^0|^2+\sum_j w_j|\theta_j-\theta_j^0|^2 \qquad (1)$$

where the summations are over all atoms and torsion angles in the residue. $\mathbf{r}_i^0$ is the target position for atom *i*, $\mathbf{r}_i$ its current position, θ_j^0 is a starting torsion angle produced by stage 1, and θ_j the current value. This last term, with different weights for different angles, allows certain angles to be more flexible than others. Provision is made for fixing angles at given values, so that they are not refined in stage 2, and for decoupling atoms, so that they can 'ride' on other atoms which are being driven towards guide points, without affecting the residual.

When a new residue is to be added to the growing chain, the program writes the previous residue to an output file, and positions the next standard group so that chain continuity is preserved exactly. A 'windup' command causes any remaining unbuilt residues to be copied from the input to the output. Thus the output is always a complete listing of coordinates which may be used as input to a subsequent run.

Interactive use

The program strategy is partly based on Diamond's (1966) work; a similar least-squares technique (without angular terms or continuity constraints between residues) was used by Brown, Takano & Dickerson to refine the structure of cytochrome c (Takano, 1975). The present program, however, has three advantages over these other programs: first, it is not restricted to proteins; second, the weighting of angular terms allows rather more control over the conformations which may

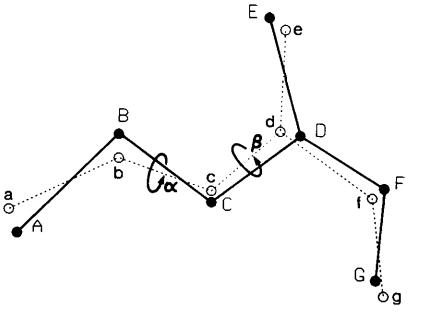

Fig. 1. Diagram to illustrate the method of fitting. The solid atoms joined by solid lines represent the guide coordinates, and the open circles, joined by dashed lines, show the idealized structure.

be adopted, and third, it is designed for interactive use.

This last aspect is particularly important for a program of this type, since ill-chosen input parameters (especially weights) can cause unexpected and undesirable conformations in stage 2. After refining a residue, the program types the values of the torsion angles, and the weighted and unweighted r.m.s. deviations of the atomic positions from the guide points. Positional and angular requests may form part of the input coordinate list [as in version 4 of Diamond's (1971) real-space refinement program], or may be typed at a keyboard while the program is running. If the idealized conformation is not acceptable, the user may reject it, in which case he is invited to change some or all of the target positions, angles or weights, before the program rebuilds the residue.

Recording atomic coordinates

Possibly the most useful application of the program is in recording atomic coordinates from a Kendrew skeletal model built to fit electron density in an optical comparator. Such recording traditionally involves measurement of the x, y and z coordinates of each atom independently with a plumbline or similar device. This process is tedious and error-prone. With the model-building program, the user needs to measure coordinates of only two or three atoms per residue (at the tips of the side chain and the growing backbone), and type these in, together with estimates of the torsion angles. The program then adjusts the angles to fit the guide points as well as possible. Errors of measurement are immediately obvious from the large error of fit, and can be corrected on the spot. The output file may be used directly as input to Diamond's real-space refinement program.

This method has been used to record the atomic coordinates of tyrosyl-tRNA synthetase from *B. Stearothermophilus* (Irwin, Nyborg, Reid & Blow, 1976). An ancillary program was used to generate a 'skeleton' coordinate list from the amino-acid sequence, containing appropriate atom and parameter records, but with all the weights and positional coordinates set to zero. Thus the only terms in the residual (1) arise from guide coordinates and angles which are measured on the model. The torsion angles φ(N–C_α), ψ(C_α–C) and ω(C–N) and the bond angle at C_α were all treated as variables, these last two being made 100 times stiffer than the first two. It proved necessary to measure all the backbone coordinates for the first residue, so as to provide an accurately positioned root for the growing chain. After this, the measurement and idealization were straightforward. In this particular case only backbone and C_β atoms were considered (*i.e.* the protein was treated as polyalanine). For building purposes a residue was deemed to begin and end on a backbone N atom; for each residue the coordinates of the C_β and of the distal N atom were measured, and typed in together with estimates of φ and ψ. Five to ten cycles of refinement generally brought the N atom of the idealized chain to within 0·1–0·3 Å, of its measured position. The idealized coordinates then replaced the dummy coordinates in the input listing, and the updated coordinates were written out.

Implementation

The program is written in ANSI Fortran, with the exception of some short routines which provide free-format keyboard input. It has been used on IBM370/165, PDP11/45 and CTL Modular 1 computers without serious modification. Since the program was designed to run on a number of different machines, no graphical display of the molecule was included, although this should not be too difficult to add. The speed depends on the number of parameters to be refined: for tRNA with 10 parameters and about 20 atoms per residue, one cycle of refinement for one residue takes about 4s on the PDP11, and typically 5 to 15 cycles are needed, depending on the accuracy of the initial model and on the convergence criteria used. For a protein, with 4 parameters and about 10 atoms, the convergence is generally rapid (10 cycles or less), and takes less than 5s on the PDP.

I thank Dr G. Fermi for suggesting the possibility of weighting angular terms in the least-squares refinement, and Dr J. White for his help in installing the program on the Modular 1 machine.

References

DIAMOND, R. (1966). *Acta Cryst.* **21**, 253–266.
DIAMOND, R. (1971). *Acta Cryst.* A**27**, 436–452.
DIAMOND, R. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 291–301. Copenhagen: Munksgaard.
DODSON, E., ISAACS, N. W. & ROLLETT, J. S. (1976). *Acta Cryst.* A**32**, 311–315.
FELDMANN, R. J. (1976). *Advances in Biophysics and Bioengineering*, edited by L. J. MULLINS, pp. 477–507. Annual Reviews Inc., Palo Alto, California, USA.
GREER, J. (1976). *J. Mol. Biol.* **104**, 371–386.
HERMANS, J. & MCQUEEN, J. E. (1974). *Acta Cryst.* A**30**, 730–739.
IRWIN, M. J., NYBORG, J., REID, B. R. & BLOW, D. M. (1976). *J. Mol. Biol.* **105**, 577–586.
JACK, A., LADNER, J. E. & KLUG, A. (1976). *J. Mol. Biol.* **108**, 619–649.
LEVITT, M. (1974). *J. Mol. Biol.* **82**, 393–420.
LEVITT, M. & LIFSON, S. (1969). *J. Mol. Biol.* **46**, 269–279.
TAKANO, T. (1975). Private communication.
TEN EYCK, L. F., WEAVER, L. W. & MATTHEWS, B. W. (1976). *Acta Cryst.* A**32**, 349–350.