and space-group type VIII, 45

$$B_1=\left(\begin{array}{cccc|c}\bar{1}&0&0&0&0\\0&1&0&0&\frac{1}{4}\\0&0&1&0&\frac{1}{4}\\0&0&0&1&\frac{1}{2}\\\hline 0&0&0&0&1\end{array}\right)\qquad B_2=\left(\begin{array}{cccc|c}1&0&0&0&\frac{1}{4}\\0&\bar{1}&0&0&0\\0&0&1&0&\frac{1}{4}\\0&0&0&1&\frac{1}{2}\\\hline 0&0&0&0&1\end{array}\right).$$

With x,y,z,u used as coordinates these matrices give $\bar{x},\frac{1}{4}+y,\frac{1}{4}+z,u$ and $\frac{1}{4}+x,\bar{y},\frac{1}{4}+z,u$ for VIII, 43 or $\bar{x},\frac{1}{4}+y,\frac{1}{4}+z,\frac{1}{2}+u$ and $\frac{1}{4}+x,\bar{y},\frac{1}{4}+z,\frac{1}{2}+u$ for VIII, 45 respectively in the conventional symbols of *International Tables for X-ray Crystallography* (1952).

It is easily verified that the integer matrix

$$T=\left(\begin{array}{cccc|c}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&-2&1&0\\\hline 0&0&0&0&1\end{array}\right)$$

transforms the generators A_i into the generators B_i, that is $B_i=T^{-1}A_iT$. Therefore, by changing the basis a_1,a_2,a_3,a_4 to $b_1=a_1$, $b_2=a_2$, $b_3=a_3-2a_4$, $b_4=a_4$, the group VIII, 45 is obtained from VIII, 43. This shows that VIII, 43 and VIII, 45 describe the same space-group type; they are only referred to different coordinate bases. Exactly the same holds for the pairs X, 99 and X, 101 as well as XI, 45 and XI, 46 with the same T as the transforming matrix. As a result of this the entries VIII 45, X 101, and XI 46 have to be removed from the KB Table 4.

On the other hand, KB list in Table 4, crystal class XV, only 62 space-group types, whereas in the BBNWZ computer determination 63 space-group types in this crystal class were found. Here to the F_6 lattice (U lattice type of Wondratschek, Bülow & Neubüser, 1971) belong three space-group types of which KB list only two. The space group generated by

$$1122=\left(\begin{array}{cccc|c}1&0&0&0&0\\0&1&0&0&0\\0&0&\bar{1}&0&0\\0&0&0&\bar{1}&0\\\hline 0&0&0&0&1\end{array}\right),\qquad 1_{1/2}2_{1/2}1_{1/2}2_{1/2}=\left(\begin{array}{cccc|c}1&0&0&0&\frac{1}{4}\\0&\bar{1}&0&0&\frac{1}{4}\\0&0&1&0&\frac{1}{4}\\0&0&0&\bar{1}&\frac{1}{4}\\\hline 0&0&0&0&1\end{array}\right)$$

$$2_{1/2}1_{1/2}1_{1/2}2_{1/2}=\left(\begin{array}{cccc|c}\bar{1}&0&0&0&\frac{1}{4}\\0&1&0&0&\frac{1}{4}\\0&0&1&0&\frac{1}{4}\\0&0&0&\bar{1}&\frac{1}{4}\\\hline 0&0&0&0&1\end{array}\right)$$

is missing; it should be added under No. 61*a*.

In No. 60 in each set of parallel twofold 'hyper axes' 2211, 2121, 2112, 1212, 1221, and 1122 there occur real rotations. In 61 in three of the six sets of twofold 'hyper axes' there occur real rotations, the other three containing screw rotations only. Finally in 61*a* there are two sets in which real rotations occur, whereas in the other four sets only screw rotations can be found. This shows immediately geometrically that the three space groups considered are non-equivalent (the original derivation by BBNWZ had been an algebraic one).

References

HERMANN, C. (1949). *Acta Cryst.* **2**, 139–145.

International Tables for X-ray Crystallography (1952). Vol. I, 1st ed. Birmingham: Kynoch Press.

KUNTSEVICH, T. S. & BELOW, N. V. (1971). *Kristallografiya* **16**, 5–17, 268–272. [Engl. Trans. *Sov. Phys. Crystallogr.* **16**, 1–8, 221–224.]

WONDRATSCHEK, H., BÜLOW, R. & NEUBÜSER, J. (1971). *Acta Cryst.* A**27**, 523–535.

Acta Cryst. (1976). A**32**, 349

A method of obtaining a stereochemically acceptable protein model which fits a set of atomic coordinates.

By L. F. TEN EYCK, L. H. WEAVER and B. W. MATTHEWS, *Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, Oregon* 97403, *U.S.A.*

(*Received* 2 *September* 1975; *accepted* 12 *September* 1975)

A method is proposed by which a protein model with acceptable stereochemistry can be fitted to a set of atomic coordinates. By expressing all constraints in terms of distances between pairs of atoms it is possible to enforce any desired stereochemistry in a physically realistic manner, and at the same time to substantially reduce computing time. Application of the technique to thermolysin is described. The method has been independently developed and applied to insulin by E. J. Dodson, N. W. Isaacs & J. S. Rollett [*Acta Cryst.* (1976). A**32**, 311–315].

The problem of obtaining a model of a protein which is a best fit to a set of observed coordinates and at the same time adheres to a specified stereochemistry has been discussed by a number of authors. In the methods of Diamond (1966) and Warme, Gō & Scheraga (1972), bond lengths and angles are, in general, fixed, and model fitting is achieved by allowing specified dihedral angles to vary. However, in real polypeptides, all bond lengths and angles may deviate somewhat from their idealized values, and in order to obtain a satisfactory representation of a protein it is desirable that such deviations be allowed. Hermans & McQueen (1974) have suggested that such variations from idealized stereochemistry can be incorporated by a method of 'local change' in which one atom is adjusted at a time in a cyclic process. Freer, Alden, Carter & Kraut (1975) have also used, in the refinement of HiPIP, a cyclic refinement of the model in which deviations in bond lengths and angles from their standard values are minimized.

Recently Dodson, Isaacs & Rollett (1976) have described, and applied to the refinement of insulin, a method of adjusting protein coordinates which seems to be superior to the techniques described above. We have independently developed the same technique, and would like to comment on the method and on its application to thermolysin.

The basic idea is to minimize the function

$$\varphi = W_1 \sum_i w_i(\Delta x_i)^2 + W_2 \sum_{(ij)} w_{ij}(\Delta d_{ij})^2 \qquad (1)$$

where Δx_i is the displacement of atom i from its experimental position and Δd_{ij} is the deviation of the interatomic distance between atoms i and j from an assumed ideal value. The factors w_i and w_{ij} are weights which can be assigned to individual atoms and distances, while W_1 and W_2 are 'global' weights which can be adjusted empirically to give an appropriate balance between adherence to the guide positions and idealized stereochemistry. The important feature of this method is that the constraints on interatomic distances can be used to enforce not only bond lengths but also bond angles, the planarity of groups of atoms, and in fact any desired stereochemistry. Although the procedure is not an energy minimization technique, it could also be used, for example, to constrain the distance between hydrogen-bonded atoms, or to separate pairs of atoms which had formed unacceptably close contacts during structure refinement. Equation (1) is equivalent to a mechanical model in which each atom is attached to its experimental position by a spring and in addition each designated pair of atoms i and j are linked by springs, each of which has a 'stiffness', proportional to w.

The function defined by equation (1) is minimized with respect to the atomic coordinates using one of the many minimization procedures to be found in the literature on numerical analysis (*e.g.* Ralston, 1965). Since all the constraints are expressed as distances between pairs of atoms, the function and all of its first derivatives are very simple to compute.

Our implementation of the method described above differs from that of Dodson *et al.* (1976) in the following respects. (*a*) All constraints are based only on distances between real atoms, whereas Dodson *et al.* introduce dummy atoms to enforce planar groups. The dummy atoms are not necessary provided a class of very rigid constraints is used for appropriate pairs of atoms within the planar groups. (*b*) Our program incorporates a variation of the conjugate gradient procedure due to Fletcher & Reeves (1964) which seems to be very efficient for this problem.

The program has been used to optimize the coordinates of thermolysin (316 amino acids, 2437 non-hydrogen atoms, and 5716 constraints) estimated from a 2·3 Å resolution isomorphous replacement map (Matthews, Weaver & Kester, 1974). The program took about 5 min on a PDP-10 to reduce the r.m.s. deviation from constraints from 0·32 to 0·01 Å. For bond lengths, the maximum deviation from ideal geometry dropped from 0·81 to 0·02 Å, while the deviation in the distance between pairs of non-bonded atoms used to constrain bond angles decreased from a maximum of 0·82 to 0·07 Å. The r.m.s. atomic movement was 0·18 Å. More significant is the fact that on application of the constraints the crystallographic R value *decreased* from 0·402 to 0·392 suggesting that the constrained model was not only superior to the unconstrained one from a stereochemical point of view, but also that it agreed better with the X-ray data.

For comparison, the idealization of the thermolysin coordinates using Diamond's (1966) model-building procedure required about 100 min on a comparable computer (IBM 360/50). In this case the constrained model in which the angle at each alpha C atom was held constant, differed from the measured atomic coordinates by 0·38 Å, on average, and had an R value of 0·425. It is interesting to note that in the case of insulin Dodson *et al.* (1976) also obtained a decrease in the R value on application of the proposed constraint procedure and an increase in R for coordinates derived by Diamond's method. Also, in common with Dodson *et al.*, we have found that satisfactory convergence can be obtained in cases where large adjustments of the model are required (*e.g.* up to 1 Å). We have found it desirable to list, in decreasing order of magnitude, the largest deviations of the model from 'ideal'. Such a listing shows clearly those residues which have the most significant errors in their coordinates.

In summary, we have found the proposed method to work well for thermolysin. Not only is the method much cheaper to apply than existing techniques, but also it allows one to constrain a protein model in such a way that the model agrees as well as desired with a set of guide coordinates, and at the same time has the strain, introduced by the application of constraints, distributed throughout the molecule in a physically reasonable manner. The method clearly improves molecular models at high R values ($\gtrsim 0{\cdot}3$) and can be expected to be applicable throughout the refinement of a protein structure. The assignment of correct weights to the constraints is obviously important and will need careful consideration. For example it is not obvious how one would choose between several structures, differently constrained, with similar R values. Also it is not yet clear how 'soft' the constraints should be at a given stage of refinement.

We thank Drs Dodson, Isaacs and Rollett for sending us a preprint of their manuscript. This work was supported in part by grants from the National Institutes of Health (GM20066, GM15423, GM21967), the National Science Foundation (BMS74-18407) and by the award to B.W.M. of an Alfred P. Sloan Research Fellowship and a PHS Career Development Award (GM70585) from the Institute of General Medical Sciences.

References

DIAMOND, R. (1966). *Acta Cryst.* **21**, 253–266.

DODSON, E. J., ISAACS, N. W. & ROLLETT, J. S. (1976). *Acta Cryst.* **32**, 311–315.

FLETCHER, R. & REEVES, C. M. (1964). *Comput. J.* **7**, 149–154.

FREER, S. T., ALDEN, R. A., CARTER, C. W. JR & KRAUT, J. (1975). *J. Biol. Chem.* **250**, 46–54.

HERMANS, J. JR & MCQUEEN, J. E. JR (1974). *Acta Cryst.* A**30**, 730–739.

MATTHEWS, B. W., WEAVER, L. H. & KESTER, W. R. (1974). *J. Biol. Chem.* **249**, 8030–8040.

RALSTON, A. (1965). *A First Course in Numerical Analysis.* New York: McGraw-Hill.

WARME, P. K., GŌ, N. & SCHERAGA, H. A. (1972). *J. Comput. Phys.* **9**, 303–317.