Acta Cryst. (1974). A**30**, 730

Computer Manipulation of (Macro)molecules with the Method of Local Change*

By Jan Hermans Jr† and John E. McQueen Jr

Department of Biochemistry, University of North Carolina, Chapel Hill, North Carolina 27514, *U.S.A.*

(*Received* 21 *March* 1974; *accepted* 17 *June* 1974)

In computer manipulation of macromolecules bond lengths and bond angles as well as some dihedral angles frequently are held fixed at ideal values observed in small model compounds. Changes of the conformation are then made by internal rotation about chemical bonds. The result of each rotation is the relative motion of large parts of the molecule; this will therefore be referred to as the global method of changing the conformation. The effect of this method is similar to manipulation of a stick model of the molecule. A method is described for manipulating the conformation in which only one atom is moved at a time; hence the name, local method. Each movement is made in order to improve the immediate environment of the atom by decreasing the differences between bond lengths, bond angles and fixed dihedral angles near this atom and their ideal values. Small displacements are calculated and applied for each atom in turn, and this is repeated a number of times for the entire molecule. At the same time, one may require that the position of each atom is not moved too far away from the starting position, so as to give idealization of the starting conformation or model building. Alternatively, inclusion of a term tending to lower the contributions to the intramolecular energy (van der Waals attractive energy, repulsive energy, electrostatic energy) gives energy minimization. A description is given of the progress of the model-building calculation with a fifteen-residue segment of the protein rubredoxin as a test case. The resulting conformation is found to be very close to the best global fit obtainable. This best global fit is obtained by constructing a global fit to the locally fit model and further adjusting this intermediate conformation to improve the agreement with the starting coordinates. A global fit constructed to the original data is found to be inferior. It corresponds to a higher relative minimum of the sum of the squares of the distances between the model coordinates and those to be fitted; the conformation of two side chains is qualitatively different in the two global fits. An example shows how the method is suitable for building trial conformations of chain segments. Finally, advantages of the local method are pointed out which, it is believed, make its use preferable for model building in an interactive computing environment.

Introduction

Purposes of conformational manipulation

Computer building of molecular models and subsequent manipulation of the models according to a pre-established criterion is needed in at least two areas: (*a*) crystallographic structure determination and refinement of large molecules and (*b*) theoretical studies of macromolecular conformation. The criterion used in the former is the agreement with the observed X-ray diffraction pattern, while in the latter the criterion most often used is the conformational energy. In both cases the common practice is to limit the allowed variations of the structure by imposing restraints such as bond lengths, bond angles and some dihedral angles which are not permitted to vary from a set of ideal or canonical values. This is a consequence of the need for a low conformational energy: the energy varies extremely rapidly as a function of the bond angles and bond lengths, but, in general, varying the dihedral angles has a much less dramatic effect on the energy unless the rotations happen to cause severe atomic overlap.

The determination of the structure of a protein starts with the fitting of an idealized model to the electron density map. Constrained refinement, in which this idealized model is changed by internal rotation to improve the fit to the electron density map, has been done successfully by Diamond (1971). More recently, it has been shown that further improvement of the structure also requires repeated recalculation of the map (Watenpaugh, Sieker, Herriott & Jensen, 1971, 1973; Deisenhofer & Steigemann, 1974). Usually, refinement is done by applying small shifts to the atoms of an ideal structure in order to improve the map (or reduce the amplitude of a difference map), then successively calculating a new idealized structure, and a new difference map, and repeating the procedure. Apparently, this approach strikes an acceptable balance between computational effort and progress towards a refined structure.

A least-squares refinement, with as criterion the agreement between calculated and observed intensities of the X-ray reflections, but without imposed conformational constraints, has been done for one protein (Watenpaugh *et al.*, 1971, 1973). This method requires much effort, and may well not converge to a reasonable conformation with less extensive and precise experimental data.

It has recently been shown that the refinement of the crystallographically determined conformation of a protein may be aided by a consideration of the conforma-

* Supported by a research grant from the National Science Foundation (GB-16420).

† Research Career Development Awardee of the National Institutes of Health, U.S. Public Health Service (Grant GM-22015).

tional energy (Birktoft & Blow, 1972; Levitt & Lifson, 1969; Warme & Scheraga, 1974; Levitt, 1974). In addition, the conformational energy is a valuable criterion in the analysis of folding of biopolymers, which for small peptides and regular, helical polymers, has become useful in the prediction of their structure.

Thus, three different procedures are in use in which idealized conformations of the protein are improved. These procedures have a very similar overall approach but a different criterion which determines when there is improvement and when there is not. These criteria are: (1) the agreement between the calculated atomic positions and those positions which are provided on another basis, the experimental or target coordinates, (2) the agreement between the electron density map calculated for the ideal structure and the experimental electron density map and (3) the total energy of the molecule. In addition to these three, another possible criterion function would be the sum of the squares of the difference between observed and calculated intensities of X-ray reflections, a reciprocal-space refinement. As yet, no one to our knowledge has attempted such a calculation on a molecule the size of a protein.

Conformations of protein molecules are known to a rather low precision. Refinement is absolutely essential if one is ever to find adequate detailed answers to the question of how proteins function, as enzymes or otherwise (Watenpaugh *et al.*, 1971; Watson *et al.*, 1963; Branden, Holmes & Kendrew, 1963).

The global method of structure improvement

A number of similar methods and results of conformational refinements have been presented by several authors (Gibson & Scheraga, 1967, 1969; Diamond, 1966, 1971; Warme, Go & Scheraga, 1972). In these methods, the conformation is modified by simultaneously changing a certain specified number of dihedral angles. The amount of change applied to each dihedral angle depends on the magnitude of the derivatives of the criterion function with respect to the dihedral angles. In some methods only the first derivatives need to be calculated, in others both first and second derivatives are used.

When a small rotation is applied about a bond in the middle of the molecule, sizable changes may occur in the relative position of the two halves of the molecule separated by the bond. For this reason, we call this manner of changing the conformation the global method of refinement. This type of refinement proceeds best if the conformations of at least ten to fifteen residues of the molecule are varied simultaneously. This implies that some forty dihedral angles must be considered as variables. On the one hand, this complicates the mathematical treatment; on the other hand, the simultaneous variation of a number of variables may give rapid convergence and is especially powerful whenever atomic positions can only be improved by concerted internal rotations about bonds which are several atoms removed.

The local method of structure refinement

Because of the increasing use of refinement programs, it seemed to us worthwhile to investigate the performance of a different technique of refinement. This technique is based on the rather simple principle that if the environment of every atom is ideal, that is to say if all the bonds leading to every atom have the correct ideal bond lengths and all the angles which these bonds make with other bonds have their ideal values, then, of course, the structure as a whole is an ideal structure. Most atoms of an imperfect structure can individually be moved a little bit in such a way that the difference between bond angles and bond lengths and their ideal values will decrease. Only atoms bonded to one other atom can be moved to a position where bond angles and bond lengths are exactly correct. However, by repeatedly applying this process to all the atoms in turn, it is possible to make the entire structure approach more and more an ideal structure. This article describes a model-building program written using this principle and the results obtained with this program.

Restating the principle of the calculation in algebraic terms, one defines a criterion function, F_i^0, for each atom as follows:

$$F_i^0 = w_l \sum (l-l_0)^2 + w_t \sum (\theta-\theta_0)^2 + w_r \sum (\varrho-\varrho_0)^2 \quad (1)$$

where the three summations are carried out for all the bond lengths, l, all the bond angles, θ, and all the constrained dihedral angles, ϱ, which are affected by the position of the atom i. The parameters w_l, w_t and w_r represent the weights which are to be given to the errors in bond lengths, bond angles and dihedral angles respectively. Shifts of the atomic position are applied on the basis of the first and second derivatives of the criterion function with respect to the coordinates $\mathbf{X}_i$, following the Newton–Raphson method, *i.e.*

$$\mathbf{F}''\Delta\mathbf{X} = -\mathbf{F}' \quad (2)$$

where $\mathbf{F}'(\mathbf{F}'')$ represents the vector (matrix) of first (second) derivatives of F with respect to the components of $\mathbf{X}$, and $\Delta\mathbf{X}$ the shift vector.*

* Our choice of minimization method is somewhat arbitrary. Convergence is slower than one might wish it to be. When several atoms should move to relieve an error, this has to be done in many small steps, since placing any *one* atom (right away) in its ultimate position grossly deforms the structure. An advantage is that the programming is relatively simple. Furthermore, storage requirements and execution time vary linearly with the size of the molecule. Without elaborate tests, we cannot exclude the possibility that some other minimization is more efficient. Several powerful minimization methods consider many variables at a time and work with an array of second derivatives. With the Cartesian coordinates as variables, the second derivative matrix has many vanishing elements. With the use of specialized representations and routines for handling such sparse matrices a 'block' minimization still using the Newton–Raphson method might be rather efficient. However, it is not a trivial undertaking to put this idea into practice, and this may properly be the subject of a separate study.

It is essential that the function which is used not be F_i^0 as defined in equation (1). Rather, we add another term to F_i^0 which depends on how far the atom has moved from its original position. Thus we use

$$F_i = F_i^0 + w_0(\mathbf{X} - \mathbf{X}_0)^2 \tag{3}$$

where $\mathbf{X}_0$ represents the coordinates which the atom had at the start of the calculation, and w_0 a weight determining the importance of the additional term. The shift $\Delta\mathbf{X}$ is calculated for and applied to each atom in turn. This process is applied, atom by atom, for as many cycles through the molecule as are required for convergence. The calculation begins with a fairly large value for w_0 as compared to the values of w_l, w_t and w_r, but after each cycle the weight w_0 is reduced with respect to the other three weights.

The procedure becomes an energy minimization if one adds to the function F_i^0 a term equal to the sum of the energy of interaction between this atom and all atoms to which it is not directly bonded:

$$F_i = F_i^0 + \sum_j E_{ij}\,. \tag{4}$$

Here each term E_{ij} consists of at least three contributions: the van der Waals attractive energy, the repulsive energy and the electrostatic energy. In that case, the weights occurring in equation (1) may be given such values that the terms represent the energy of deformation of bond lengths, bond angles and dihedral angles in the same units in which the energies E_{ij} are calculated. Such a calculation applies the idea of a 'consistent force field', *i.e.* of zero net force on each atom. With this idea, Lifson & Warshel (1968) determined a set of interatomic force constants from spectroscopic and crystal data. Levitt's recent energy refinement of the conformation of lysozyme uses essentially the procedure outlined above (Levitt, 1974).

Calculations

Using the program which we have written, we have made a series of calculations on a segment of rubredoxin in order to find optimum conditions for the model-building procedure. We have also compared the results obtained under the conditions which appeared to us most favorable with results of a model-building program using the global method.

Methods

Our global refinement procedure is similar to that described by Warme, Go & Scheraga (1972) and uses Fletcher–Powell–Davidon minimization, as there described. The initial values of the dihedral angles of the idealized model are those calculated from the given set of coordinates. In the first step of the refinement, the conformation of a piece consisting of the first four residues is allowed to vary until the agreement between calculated and target coordinates for this piece is satisfactory. Next, the conformation of residues three through six is changed in the same manner. In this way the conformation varies four residues at a time until the four-residue piece reaches the end of the molecule. Following this, the conformation of the entire piece of fifteen residues is allowed to vary, both by applying internal rotation and by repositioning and reorienting the molecule.

In both the local and the global procedures the atoms are listed starting with the N of the first residue (the formyl group present on the N-terminal methionine residue was not considered) in such a way that shorter branches are always listed before longer branches (Hermans & Ferro, 1971).

In the global procedure bond lengths and bond angles do not vary while internal rotation is allowed about most chemical bonds. The peptide bond is kept in the planar *trans* conformation and the geometry of ring structures including that in proline is not allowed to vary. In the local method we represent the same constraints by including a minimum but sufficient set of error terms in the criterion function [equation (1)]. This set can be chosen in more than one way. Our set was obtained by having each atom added to the list (except the first three) define one additional bond length, one bond angle and one dihedral angle. When the structure branches, one of the atoms after the branchpoint defines the dihedral angle for rotation about the bond leading to the branchpoint; this angle is constrained only in the peptide bond and in rings. Each other atom connected to the branchpoint defines one fixed dihedral angle indicating by how much the lateral branch is rotated out of the plane of the preceding bond and the main branch. On the basis of these restrictions the error terms in equation (1) are accumulated. (It may be useful to point out that this minimal set is imperfect when the method is used for energy refinement.)

In the local-refinement method the shift of each atom is calculated and applied in the order in which it is listed. We prevent both excessive shifts and shifts in the wrong direction. The latter occur when the determinant of the second-derivatives matrix [*cf.* equation (2)] is negative. In the former case, a smaller shift of the coordinates of the atom is applied, which has components proportional to the components of the calculated shift. In the latter case, a shift is applied of which the components are proportional to the first derivatives of the function F_i, but of opposite sign. The maximum permissible shift and the shifts which are applied in either of the two cases have the same value, δ, obtained according to the following (arbitrary) algorithm:

$$\delta = \delta_0 n/(n + k^2) \tag{5}$$

where δ_0 has a value of 0·2 Å, n is the total number of cycles through the entire molecule which are going to be performed and k is the number of the cycle which is at this time being executed.

The calculations were all carried out using as a test case the first fifteen residues of rubredoxin for a total of 122 atoms; a set of coordinates was kindly provided by Professor L. H. Jensen. These coordinates were obtained by Watenpaugh *et al.* (1973) with several cycles of difference Fourier refinement in which the constraints on the conformation were approximately maintained; this was followed by several cycles of least-squares reciprocal-space refinement. Since no constraints were imposed on the coordinates in the final stages of their refinement, the resulting coordinates are not *particularly* close to those of an idealized structure. Nevertheless, these coordinates clearly fit the steric requirements of the chemical structure of rubredoxin.

One side chain (isoleucine 12) was assumed to be a valine residue in the refinement of Watenpaugh *et al.* We gave the C_δ atom rather bad coordinates by mistake. As will be discussed, the refinement near this 'bad spot' helps in our analysis of the performance of the refinement methods. Most of the residues are in the extended, or β conformation, a few have a conformation which may be called α-helical, and one glycine residue has a conformation which falls in a third area of the Ramachandran plot (Fig. 1). Thus, the set of test data appears to be quite representative of data to which one usually would apply a model-building procedure. As an additional test, we applied the program to a 'preshrunk' set of coordinates obtained from these coordinates of rubredoxin by multiplying the x and y coordinates by a factor of 0·8, leaving the z coordinates as they were before.

In none of our calculations did we take into account that the two sulfur atoms of our 15-residue molecule are part of the FeS_4 cluster of rubredoxin, which has a specific geometry; this imposes further constraints on the model. Both methods of refinement can take these additional constraints into account. The problems arising from the need to provide for ring closure are briefly discussed near the end of this article.

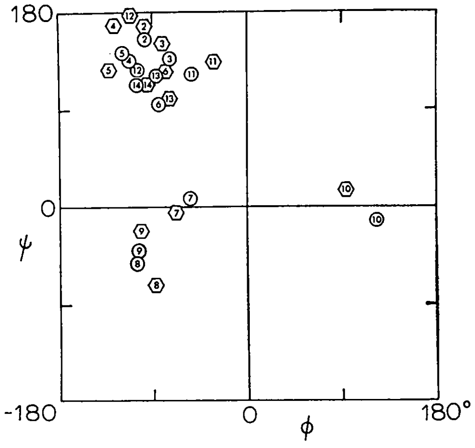

Fig. 1. Representation of the conformation of residues 2 through 14 in a Ramachandran plot. φ and ψ are the dihedral angles for the bonds N–C_α and C_α–C, respectively. Hexagons represent the conformation of the original coordinates, circles represent the conformation resulting after 100 cycles of local refinement. The residues represented by the cluster in the upper left quadrant, are in the more or less extended, or β conformation. Those in the lower left quadrant are in another allowed region of the map, and have conformation roughly that of a residue in an α helix. Residue 10 is glycine.

Test of the refinement method

In determining the shifts to be applied to each atom, different weights are given to the displacements from the original position and to the errors in bond lengths, bond angles and fixed dihedrals [equations (1)–(3)]. We said above that these weights vary throughout the refinement. We have performed some calculations in order to establish that this is in fact the preferable way of doing the calculation and what are desirable values for the weights throughout the calculation. The following arguments can be proposed in order to make a selection from the great many possible choices.

In case the calculation performed is an energy minimization, the weights for deforming bond lengths, bond angles and fixed dihedrals ought to be such that the terms in equation (1) correspond to energies of deformation of these parameters. The three weights would then be very roughly in the ratio 300 to 30 to 15 and this set of weights was in fact used for some of our calculations. However, the calculations which we are here concerned with are model-building calculations. The structure of a model is not fully determined by the constraints (in contrast to what is true when one is doing an energy calculation), since the number of constraints is smaller than the number of independent variables, which is equal to three times the number of atoms.

Thus there appears to be no obstacle to approaching an ideal structure as closely as one wishes, and the choice of the ratios between the three weights should theoretically not affect the final result; however, during the approach to an ideal structure those parameters which have been given the smallest relative weight will deviate most strongly from their ideal values.

On the other hand, one may argue that one has no control over the conformation unless sufficient constraints are imposed to determine the final structure. The conformation which we obtain at the end of the calculation should not only be ideal, but also as close as possible to the starting conformation. Therefore, the presence of the term added in equation (3), which takes into account the distance the atom has moved from its original position, would seem to be essential, serving to keep the calculated structure near the original structure. However, if the displacement from the starting position is given a finite weight in the criterion function F_i, then the final structure obtained will not be perfectly ideal but somewhat deformed because of the 'pull' by the original coordinates. The resolution of this dilemma would appear to be to let the weight for the approach to the original positions be significant in the early cycles, and gradually become very small with respect to the other weights. Then, when the weighted term for the displacement has become negligible, the structure will already be very close to an ideal conformation. From then on, further changes in atomic positions will be very small and gradual.

In Fig. 2 we show graphically the progress, through a number of cycles, of the sums of the squares of the distances moved, of the squares of the errors in bond lengths, bond angles and dihedral angles. Also given are weights and values after twenty (in one case also six) cycles. In the calculations reported in Fig. 2 the weights were set for the first cycle and left unchanged. In another set of calculations, the weights were all set equal to one in the first cycle and allowed to exponentially reach the values shown after twenty cycles. Values after twenty cycles (for one case after ten cycles) are listed in Table 1. In Fig. 2 and Table 1, the order for both weights and total deviations is: displacements, bond lengths, bond angles and dihedral angles.

One notes that when the weight given to the displacement from the original position is relatively large, the structure stabilizes in a few cycles after some initial, quite rapid changes. On the other hand, when the weight for the displacement is small the conformation continues to change gradually after the first initial rapid changes have occurred.

The use of a weight for the error in bond lengths which is much larger than that for the error in bond angles or dihedral angles as in Fig. 2(*d*) and Table 1, set 5, produces a structure with excellent bond lengths. Nevertheless it cannot be considered overall to be as close to an ideal structure as the conformations obtained with weights in a less extreme ratio.

Comparison of Fig. 2(*d*) and Table 1, set 5, shows that the use of exponentially changing weights may produce in twenty cycles a structure which is closer to ideal and also is closer to the original data than that obtained using constant weights. With the changing set of weights, the calculation is more effective and also more efficient. With another set of weights [Fig. 2(*c*) and Table 1, set 4] the calculation with changing weights is the less efficient of the two. However, six cycles of refinement using constant weights move the structure as far as do twenty cycles with changing weights, while the error in the former result is considerably larger.

Structures in which the deviations from the ideal are distributed equally over the bond lengths, bond angles and dihedral angles are obtained if the weights for the errors in these parameters are approximately equal. The choice of the ratio between these weights as one to two to two was made in the expectation that the maximum error in any given bond length (in Å) or bond angle or dihedral angle (in rad) in the entire molecule would then be approximately the same.

As expected, the use of a progressively smaller weight for the displacement from the original position improves the agreement with an ideal structure and causes a greater total displacement from the original structure. Notice, however, that after ten cycles the run reported in Table 1, set 1 has produced the same total movement from the starting structure but less progress has been made to an ideal structure compared with twenty cycles (Table 1, set 3) in which the final weight for the deviation from the starting positions is higher

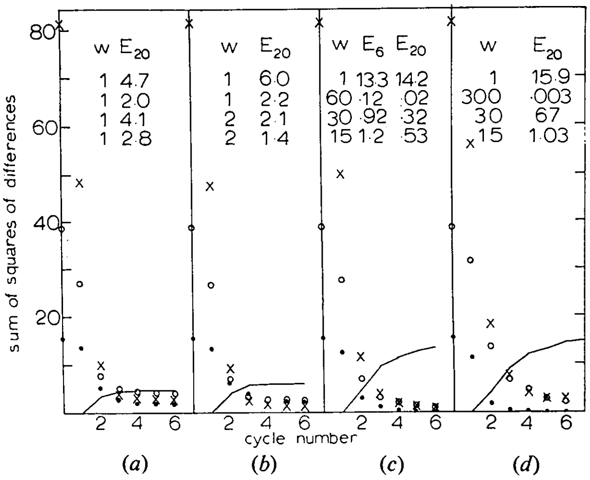

Fig. 2. Progress of the refinement during the first six cycles using various sets of weights which are held constant for all cycles. Plotted are: the sums of the squares of the total distance moved (solid line), of the errors in bond lengths (●), in bond angles (○) and in dihedral angles (×). The weights given to these terms and the values attained after twenty cycles (in one case also at the end of six cycles) are given in tables in each part of the figure, in the order: displacement, error in bond length, bond angles and dihedral angles. Units are Å^2 and rad^2.

Table 1. *Result of twenty cycles of local refinement of the test structure*

Weights start all equal to one and change exponentially with the number of cycles to the values shown (in columns marked *w*). The sums of the displacements, and of errors are given in the columns marked E_{20}. Values are in Å^2 or in rad^2. Column marked E_{10} gives values after 10 cycles. Values in parentheses are root-mean-square values, with angles given in degrees.

	w	E_{10} (r.m.s.)	E_{20} (r.m.s.)	*w*	E_{20} (r.m.s.)	*w*	E_{20} (r.m.s.)	*w*	E_{20} (r.m.s.)	*w*	E_{20} (r.m.s.)
Displacement	1	15·8 (0·36)	20·4 (0·41)	1	14·1 (0·34)	1	15·7 (0·36)	1	13·3 (0·33)	1	13·3 (0·33)
Bond lengths	10^7	0·19 (0·039)	0·08 (0·026)	25	0·21 (0·041)	70	0·15 (0·035)	60	0·037 (0·017)	300	0·003 (0·005)
Bond angles	2 . 10^7	0·45 (3·5)	0·18 (2·2)	50	0·34 (3·0)	140	0·28 (2·7)	30	0·42 (3·4)	30	0·60 (4·0)
Dihedral angles	2 . 10^7	0·21 (2·4)	0·08 (1·5)	50	0·18 (2·2)	140	0·13 (1·9)	15	0·72 (4·4)	15	0·96 (5·3)

relative to the other weights. One concludes that the weight for the displacement from the original positions has in the run of Table 1, set 1 too rapidly become insignificant with respect to the other weights.

The results of twenty cycles of refinement using various sets of weights on the structure with two of the three Cartesian coordinates foreshortened by 20% are given in Table 2 (the weights change gradually in the same manner as they do in the runs reported in Table 1). We see that the same amount of computation on the shrunken molecule produces in each case a structure which has moved farther from the original position and in which, at the same time, the errors in bond lengths, bond angles and dihedral angles are larger than on the non-shrunken molecule. The increase in the error is found to be greatest in the bond lengths, given the same set of weights.

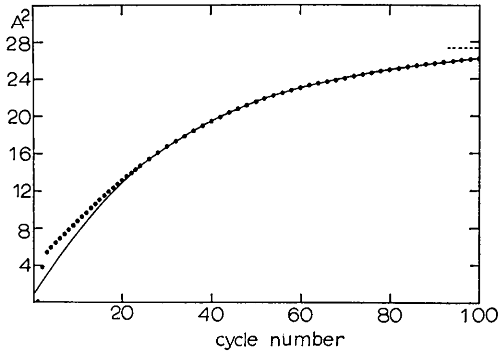

Fig. 3. Progress of the sum of the squares of the displacements from the original positions through a 100-cycle refinement in which the weights varied exponentially with the number of cycles performed.

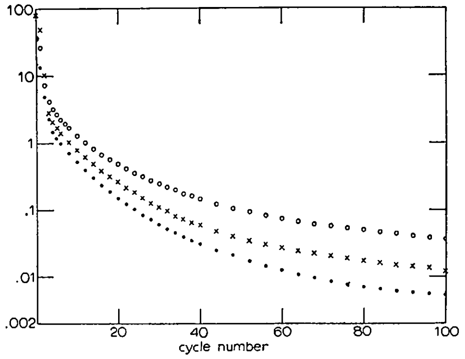

Fig. 4. Progress of the sums of the squares of the errors in bond lengths (●), bond angles (○) and dihedral angles (×), for the same run as the data in Fig. 3. Values are in Å^2 or rad^2.

The runs reported above allow us to select a set of weights and a method for changing them which will produce a reasonably rapidly improving structure which at the same time remains as close as possible to the original starting structure. In the following section we further analyze the progress of such an optimum run and compare the resulting structure with the model fitted to the same data by the global refinement method.

Progress of the refinement and analysis of the results

To illustrate the progress of the refinement we show in Fig. 3 the values of the sum of the displacements from the original coordinates as a function of the number of cycles on a linear scale. In Fig. 4 we show on a logarithmic scale the sums of the squares of the errors in bond lengths, bond angles and dihedral angles. In this calculation the weights for the four contributions to F_i all start equal to one another and the weights change exponentially to final values of 1, 10^7, 2×10^7 and 2×10^7.*

The curve drawn through the points giving the total displacement (Fig. 3) is an exponential curve approaching the limit of 27·2 Å^2 for a very large number of cycles. All the points are very close to this curve over the last 75 cycles of the refinement. During these 75 cycles the sum of the distances moved comes closer to this limit by a factor of ten. Between the third and the 24th cycle the function increases *less* rapidly than the exponential, presumably because at this stage of the refinement the weight given to the distance moved is not yet insignificant with respect to the other weights. In the first two cycles the function changes *more* rapidly than the exponential. In this case all weights are nearly equal to the weight for maintaining the structure close to the original structure and a certain amount of 'slack' is at once taken up.

We now compare the structure resulting from the global refinement calculation with that obtained by the local refinement method. Ideally the two would be identical. However we find that the difference tends to a finite limit. Further analysis shows that at least part of the remaining difference is due to the fact that the two structures correspond to different best fits. There are many ways of building a model which has ideal bond lengths, bond angles and dihedrals and which is

Table 2. *Results of twenty cycles of refinement on the foreshortened conformation*

Cf. legend of Table 1.

	w	E_{20}	(r.m.s.)	w	E_{20}	(r.m.s.)	w	E_{20}	(r.m.s.)
Displacement	1	19·0	(0·39)	1	23·0	(0·43)	1	24·8	(0·45)
Bond lengths	25	1·5	(0·11)	70	1·1	(0·095)	300	0·04	(0·018)
Bond angles	50	0·50	(3·7)	140	0·40	(3·2)	30	2·7	(8·5)
Dihedral angles	50	0·14	(1·9)	140	0·15	(2·0)	15	1·5	(6·4)

* This calculation took 50 s of c.p.u. time on an IBM 370 model 165 for a total of 12200 atom cycles.

also a best fit to a set of target coordinates. In this case one means by best fit that slight perturbations of the model are removed if the perturbed model is resubmitted to the refinement process. In Fig. 5 we show four projections of three side-chains of the molecule: one of the starting structure and three of structures resulting from local and global refinement. For the side chain of isoleucine 12, which has a poor initial geometry, one notes how very close the results of the two fitting procedures are to one another and how far they are from the original positions. The two methods give practically the same result. Just as with isoleucine 12 the initial geometry of the side-chains of lysines 2 and 3 does not correspond well to an idealized geometry, especially near the ends of the chains. However, the two refinement methods produce what appear to be qualitatively different conformations for these side chains.

In order to show that the result of the local refinement indeed corresponds to another possible solution of the global fitting problem, we used the global refinement method to calculate an idealized structure best fitting the result of the local refinement. Subsequently, this conformation was allowed to change further in order to attain the best possible global fit with the original set of coordinates. The resulting coordinates are close to the result of the local refinement procedure and not to the result of the first global refinement calculation. The sum of the squares of the distances between calculated and target coordinates (the criterion function for the global model-building procedure) has apparently more than one minimum. Some of these minima differ sufficiently little in the corresponding values of the independent variables that different refinement methods may end up near different minima.

The sum of the squares of the deviations (Table 3) shows that a direct global fit to the target data is not as good a fit as a global fit to the target data using a global fit to locally refined data as an intermediate conformation. The difference in the quality of the result is greatest where the conformations differ most. The sum of the squares of the displacement for the atoms of residue lysine 2 is 2·71 for the former structure, 2·23 for the latter, while for lysine 3 these numbers are 3·30 and 2·90. These differences account for 0·9 out of a total of 1·5 Å^2 difference between the two structures (see Table 3, penultimate line).

The values of the dihedral angles resulting from different refinements are given in Table 3. The qualitatively different fit of the two lysine side chains is evident from these data. Apart from the difference in lysines 2 and 3, the two variable dihedrals next to the peptide bond between residues 7 and 8 show the largest deviation. It may be that the local fit corresponds to another optimum global fit in this region. If this is the case, then the fact that the local-fitting method tends after many cycles to a structure which is less far removed from the starting structure than any of the structures obtained with the global refinement method indicates that this is a very good global fit, perhaps the best possible. Alternatively, it may be that somehow the local refinement method is incapable of removing the last small errors in bond lengths, bond angles and dihedral angles, even after a great many cycles of calculation. As yet, we have not encountered a case where this is clearly so.

Small errors, particularly in bond angles and fixed dihedral angles, occur without doubt in the true conformation of a folded protein molecule. These are, of course, not really errors but small deviations caused by intramolecular forces. Thus, a perfect model is not a significantly better representation of the true conforma-

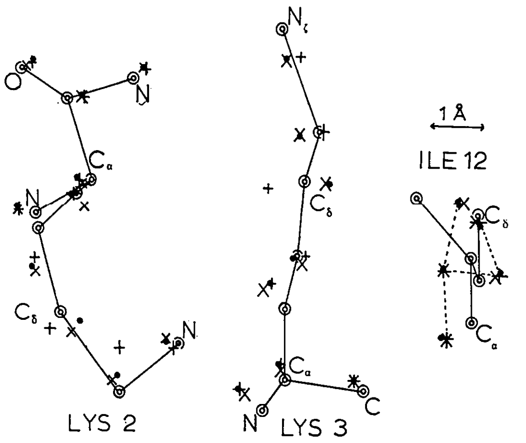

Fig. 5. Projections of the coordinates of the atoms of three side chains before and after refinement. Given are the positions before refinement (◎), after one hundred cycles of local refinement (×), after global refinement of the original structure (+), and a global fit to the starting structure, obtained with as an intermediate conformation, a global fit to the local fit (●).

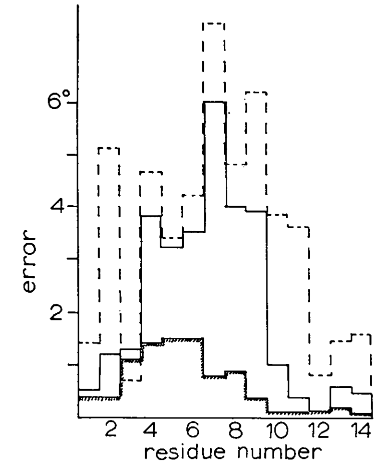

Fig. 6. Distribution of the errors throughout the main chain of the fifteen-residue molecule. The absolute values of the errors in the bond angle at C_{i-1}, N_i and C_i^α were added and are given as a dashed line for the result after 40 cycles of refinement, and as a solid line for 100 cycles of refinement. The solid line with shading represents the error in the dihedral angle for the peptide group after 100 cycles.

Table 3. *Dihedral angles for various fits*

Meaning of columns:

TGT	Original structure
LC20	Local fit obtained in 20 cycles
LC40	Local fit obtained in 40 cycles
LC00	Local fit obtained in 100 cycles
GBL0	Global fit to original structure
GBL3	Global fit to structure LC00
GBL4	Global fit to LC00 refitted to original structure

	TGT	LC20	LC40	LC00	GBL0	GBL3	GBL4
Methionine 1							
χ_1	9	−81	−88	−85	−81	−86	−76
χ_2	−125	−76	−71	−59	−77	−58	−68
χ_3	−21	−82	−91	−88	−78	−92	−81
ψ	6	42	37	29	37	28	36
Lysine 2							
φ	−97	−101	−98	−97	−83	−88	−91
χ_1	−18	−20	−17	−19	−19	−27	−34
χ_2	−73	−125	−130	−132	−60	−122	−117
χ_3	−4	84	81	88	−60	97	104
χ_4	22	−63	−59	−60	92	−71	−77
ψ	166	162	160	156	172	159	162
Lysine 3							
φ	−80	−82	−78	−73	−90	−80	−81
χ_1	−55	−55	−51	−55	−65	−56	−75
χ_2	162	142	145	145	−151	150	144
χ_3	179	113	115	126	−135	114	134
χ_4	−163	−164	−162	−150	139	−152	−135
ψ	152	150	145	138	134	138	132
Tyrosine 4							
φ	−126	−121	−117	−113	−114	−116	−114
χ_1	−110	−78	−80	−80	−77	−78	−78
χ_2	−61	−110	−109	−108	−106	−109	−106
ψ	168	149	145	139	124	130	124
Threonine 5							
φ	−131	−122	−121	−117	−103	−106	−103
χ_1	−2	−35	−33	−29	−32	−26	−32
ψ	127	136	136	141	153	152	152
Cysteine 6							
φ	−83	−76	−77	−84	−96	−97	−96
χ_1	173	173	174	176	179	175	175
ψ	124	107	103	95	92	93	92
Threonine 7							
φ	−69	−68	−63	−57	−55	−57	−56
χ_1	104	68	68	68	75	78	76
ψ	−6	0	2	8	25	37	34
Valine 8							
φ	−88	−97	−102	−106	−118	−131	−128
χ_1	−60	−82	−82	−84	−77	−84	−74
ψ	−72	−48	−48	−52	−66	−65	−63
Cysteine 9							
φ	−102	−107	−103	−104	−93	−94	−97
χ_1	44	75	76	70	65	68	64
ψ	−22	−32	−44	−43	−51	−49	−49
Glycine 10							
φ	94	108	120	125	131	128	126
ψ	16	−2	−8	−13	−10	−11	−3
Tyrosine 11							
φ	−35	−58	−57	−53	−53	−54	−60
χ_1	158	162	164	166	166	166	166
χ_2	102	85	88	89	90	89	90
ψ	136	131	127	123	120	121	119
Isoleucine 12							
φ	−111	−113	−110	−105	−103	−104	−101
χ_1	180	135	139	136	130	136	132
χ_2	180	165	163	162	160	161	158
ψ	178	129	127	129	129	130	129
Tyrosine 13							
φ	−75	−86	−86	−88	−88	−89	−89
χ_1	169	174	173	172	171	171	171
χ_2	86	79	79	79	79	79	79
ψ	99	121	121	122	125	123	124
Aspartic acid 14							
φ	−95	−109	−108	−107	−109	−109	−108
χ_1	−177	−165	−165	−163	−161	−163	−159
χ_2	162	156	156	155	152	156	151
ψ	116	115	115	114	113	114	114
Proline 15							
ψ	−11	−35	−27	−31	−34	−32	−33
Sum of squares of deviations (Å^2)							
	0·0	20·4	24·4	26·0	33·7	33·8*	32·2
Root-mean-square deviation (Å)							
	0·0	0·41	0·44	0·46	0·53	0·53*	0·51

* These values with respect to the original structure, although structure GBL3 was obtained as a fit to structure LC00. The sum of squares of distances moved from coordinates of LC00 is 3·4 Å^2 (0·17 Å r.m.s.).

tion than a model in which bond angles and dihedrals deviate from their ideal values by a few degrees. However, the deviations remaining in models built using the local-refinement procedure do not necessarily bear any relationship to the deviations which occur in the molecule.

After many cycles of local refinement the deviations from an ideal structure are all located in the main chain of the molecule. This is not surprising, since the idealization is straightforward for any atom which is bonded to only one other atom and the surroundings of the next atom in the chain can be made ideal by displacing this atom plus the atom at the end of the chain. In order to make the surroundings of the atom in the middle of the chain ideal, a fairly large number of atoms all along the chain will have to make corresponding small movements. Strain occurring at any one atom is thus distributed over the chain as the refinement progresses until it is either absorbed by a free chain end or absorbed by adjustment of one or more free dihedral angles. This takes more cycles for atoms farther from the chain ends.

In Fig. 6 we show how the errors in bond angles and the error in the frozen dihedral angle for the peptide bond are distributed over the molecule; the errors are concentrated in the middle part of the molecule.

Model building as an end in itself

In the *Introduction* we described how idealization or model building can be used to aid one in the refinement of protein structure. The use of model conformations is of course much wider than that. Model conformations are used in the initial phase of interpretation of the electron density map and at the end of the structure determination in order to give a pleasing and understandable form to the result. Once the structure is known, the three-dimensional model is of great importance in reaching an understanding of the biological function of the protein in terms of its structure and the design of new experiments to test any hypothesis made in this regard. Computed models are not useful in this respect unless they are converted into visual form, either on a computer-driven display or as a three-dimensional model built of solid parts according to the coordinates of the computed model. A general-purpose model-building program should be usable both as one of the several procedures used in refinement and for the purpose of producing a stereochemically reasonable model which can be visually inspected.

A not unusual requirement in model building under the second set of circumstances is to connect two parts of the chain of known conformation with a series of residues in a region where the data are insufficient to determine the conformation. Accordingly, we have investigated the performance of the local fitting method in solving the following problem. Given the conformation of two sections of nearly extended chain forming an antiparallel β structure, it was asked to connect the chain continuously to form a sharp bend. Of a chain of ten alanine residues, we assigned to the first five the coordinates of five residues on one chain and to the other five the coordinates of five residues on the antiparallel chain. Logically, residues 5 and 6 were connected but the bond angles and bond distance near the connection were patently ridiculous. After only twenty cycles of refinement the connection between residues 5 and 6 had closed; shifts of up to 3 Å had occurred in some of the atoms of residues 5 and 6, shifts of at most 0·5 Å in the atoms of residue 4 and shifts of less than 0·1 Å in the atoms of all the other residues. Since the weights used were in the ratio 1 to 60 to 30 to 15 and the largest shift was 3 Å, it is not surprising that the error in bond angle was as large as 0·45 radians near the atoms where the shift was greatest. The errors in bond lengths and dihedral angles were smaller. This conformation would undoubtedly have progressed further to an idealized conformation if the weight for the displacement had been reduced with respect to the other weights.

Alternatively, we could have produced a very similar result by excluding the term dependent on the displacement from the original position for those atoms which are found to move a great deal. Whenever a larger gap is to be bridged with a greater number of residues, and consequently with a larger number of degrees of freedom, more than one qualitatively different conformation may be possible. In this case the choice can be restricted by specifying the target positions of a few selected atoms in the gap.

Projections of the starting conformation and the conformation resulting from the fitting calculation are shown in Fig. 7. A space-filling model of a bend similar to that occurring in the computed structure is easily built. The bend is a type II β bend according to the nomenclature of Venkatachalam (1968). Refinement of the right-hand conformation of Fig. 7 by energy minimization results in further changes, but the conformation remains qualitatively the same. The maximum deviation in any bond angle after energy minimization is 17° (0·29 rad).

Conclusions

The results reported above demonstrate that refinement *via* local changes can replace refinement obtained by internal rotation about covalent chemical bonds. We here review some considerations which may lead to a preference for either method depending on the purpose for which applied.

Quality of final fit

The result of a global refinement is always a perfectly ideal structure. Since it is obtained by function minimization, it will rarely be *exactly* a structure which, when slightly perturbed, will return to itself upon further function minimization. However, the difference can be made small. This calculation will require about the same effort as one needed to obtain a locally refined structure. The latter, while not perfectly ideal, in compensation meets the requirement of being close to the original structure somewhat better.

Much more important is that the local refinement apparently tends to the best distinct solution of the global-fitting problem, while the global method may easily give one a 'best' fit which is farther away from

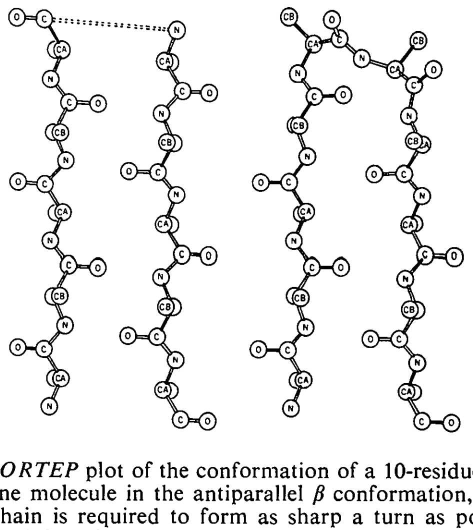

Fig. 7. *ORTEP* plot of the conformation of a 10-residue polyalanine molecule in the antiparallel β conformation, where the chain is required to form as sharp a turn as possible. On the left, before refinement, one covalent bond is present which is several Å long. On the right, the same conformation after only 20 cycles of local fitting.

the target coordinates. Use of both methods in succession (first local, then global) is recommended for obtaining the best of several possible solutions to the global-fitting problem.

Use with cyclic structures

The presence of loops in the structure poses additional constraints to be taken into account in the fitting. In the global calculation the closure, whether approximate or perfect, requires special features and steps (Gibson & Scheraga, 1967, 1969). But if, as in local refinement, only the immediate environment of each atom is considered, ring closure elsewhere is unnoticed and is yet achieved as the refinement progresses. Initial errors are distributed over the atoms of the ring and are eventually 'absorbed' by the free internal rotations.

Simplicity

Both conceptually and algebraically local refinement is a simpler method.* As a result, programming new problems in which the local change of conformation is an element of the calculation is, in our experience, much simpler than if one uses global changes instead. This simplicity is expected to be a great advantage in an interactive environment achieved using a computer-driven display, where the programmer may call for refinement of a conformation under a variety of circumstances (Levinthal, 1966; Levinthal, Barry, Ward & Zwick, 1968; Katz & Levinthal, 1972; Meyer, 1971).

In such a system, both coarse and accurate refinement of a starting set of coordinates will be requested. Clearly a result which is inaccurate by being not altogether ideal but relatively close to the starting coordinates is much preferable to a result which errs by being rather far from the given coordinates but altogether perfect in geometry. The relatively small size of the program module is a further, practical advantage in an interactive system.

Use in energy minimizations

In using the local-refinement method for the minimization of energy the problem of the underdetermination of the final results by the constraints, discussed above, disappears entirely. The starting coordinates do not influence the final result, except insofar as they determine which *qualitatively* different structure out of the many with minimum energy is obtained. The local method automatically imparts flexibility to bonds and bond angles; this can be obtained in the global minimization only at the cost of greatly increasing the number of independent variables.

In the global procedure, only so many variables can be allowed to change simultaneously since both calculation time and needed storage are strongly dependent on their number. As a result, the fitting of a large protein can be done only in pieces. If the criterion for fitting is the deviation from a set of target positions, then the conformation of each piece depends on the conformation of only the adjacent pieces and that not strongly. But in energy minimization each piece interacts through nonbonded interactions with other pieces which may be many residues distant along the chain. As a result, refinement with the global method has to be piecemeal and must be done in several cycles over the pieces. Necessary concerted changes in dihedral angles in chain segments close in space but far distant along the chain cannot be made, even though they are undoubtedly the quicker means to the desired objective and otherwise would give an advantage to the use of the global method.

Therefore, energy minimization by the simpler local method may be preferable (Levitt, 1974). More experience with an energy-refinement procedure by this method will be needed to decide if this is indeed so. This problem is presently being studied by us.

* The essential portion of the program is a module of less than 10.10^3 bytes in system 360, as coded by us. Also required are nine four-byte words for each atom for tree-list matrix and two sets of coordinates.

References

Birktoft, J. J. & Blow, D. M. (1972). *J. Mol. Biol.* **68**, 187–240.

Branden, C. I., Holmes, K. C. & Kendrew, J. C. (1963). *Acta Cryst.* **16**, A175.

Deisenhofer, J. & Steigemann, W. (1974). *Proc. 2nd Int. Res. Conf. Proteinase Inhibitors.* In the Press.

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

Diamond, R. (1971). *Acta Cryst.* A**27**, 436–452.

Gibson, K. D. & Scheraga, H. A. (1967). *Proc. Natl. Acad. Sci. U.S.* **58**, 420–427.

Gibson, K. D. & Scheraga, H. A. (1969). *Proc. Natl. Acad. Sci. U.S.* **63**, 9–15.

Hermans, J. & Ferro, D. (1971). *Biopolymers*, **10**, 1121–1138.

Katz, L. & Levinthal, C. (1972). *Ann. Rev. Biophys. Bioeng.* **1**, 465–504.

Levinthal, C. (1966). *Sci. Amer.* **214** (6), 42–52.

Levinthal, C., Barry, C. D., Ward, S. A. & Zwick, M. (1968). In *Energy Concepts in Computer Graphics*, edited by J. Nievergelt and D. Secrest. New York: Benjamin.

Levitt, M. (1974). *J. Mol. Biol.* **82**, 393–420.

Levitt, M. & Lifson, S. (1969). *J. Mol. Biol.* **46**, 269–279.

Lifson, S. & Warshel, A. (1968). *J. Chem. Phys.* **49**, 5116–5129.

Meyer, E. (1971). *Nature, Lond.* **232**, 255–257.

Venkatachalam, C. M. (1968). *Biopolymers*, **6**, 1425–1436.

Warme, P. K., Go, N. & Scheraga, H. A. (1972). *J. Comput. Phys.* **9**, 303–317.

Warme, P. K. & Scheraga, H. A. (1974). *Biochemistry*, **13**, 757–767.

Watenpaugh, K. D., Sieker, L. C., Herriott, J. R. & Jensen, L. H. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 359–367.

Watenpaugh, K. D., Sieker, L. C., Herriott, J. R. & Jensen, L. H. (1973). *Acta Cryst.* B**29**, 943–956.

Watson, H. C., Kendrew, J. C., Coulter, C. L., Branden, C. I., Phillips, D. C. & Blake, C. F. (1963). *Acta Cryst.* **16**, A81.