J. Mol. Biol. (1983) **170, 723-764**

Protein Folding by Restrained Energy Minimization and Molecular Dynamics

MICHAEL LEVITT

Department of Chemical Physics
Weizmann Institute of Science
Rehovot, Israel

(Received *1* ***November*** *1982, and* ***in revised form*** *20 June 1982)*

Native-like folded conformations of bovine pancreatic trypsin inhibitor protein are calculated by searching for conformations with the lowest possible potential energy.

Twenty-five random starting structures are subjected to soft-atom restrained energy minimization with respect to both the torsion angles and the atomic Cartesian co-ordinates. The restraints used to limit the search include the three disulphide bridges and the **16** main-chain hydrogen bonds that define the native secondary structure. The potential energy functions used are detailed and include terms that allow bond stretching, bond angle bending, bond twisting, van der Waals' forces and hydrogen bonds. Novel features of the methods used include soft-atoms to make restrained energy minimization work, writhing numbers to classify chain threadings, and molecular dynamics followed by energy minimization to anneal the conformations and reduce their energies further. Conformations are analysed using writhing numbers, torsion angle distributions, hydrogen bonds and accessible surface areas.

"The resulting conformations are very diverse in their chain threadings. energies and root-mean-square deviations from the X-ray structure. There is a relationship between the root-mean-square deviation and the energy, in that the lowest energy conformations are also closest' to the X-ray structure. The best conformation calculated here has a root-mean-square deviation of only 3 Å and shows the same special threading found in the X-ray structure.

The methods introduced here have wide ranging applications: they can be used to build models of protein conformations that have low energy values and obey a wide variety of restraints.

1. Introduction

Calculating the native folded conformation of a protein molecule from the amino acid sequence is a most challenging intellectual problem. In principle, such a calculation is possible in that it has been shown experimentally that the amino acid sequence specifies the three-dimensional arrangement of atoms in the native conformation (Sela *et al., 1957).*

The basic method to be used in this calculation of native structure became obvious after the X-ray analysis of the myoglobin crystal (Kendrew ***et al.,*** 1960).

0022–2836/83/3 10723–42 $03.00/0

The native protein conformation was found to be stabilized by the same types of forces that stabilize small organic molecules. In particular, bond lengths, bond angles and double-bond torsion angles are close to standard values, the atoms interact to form a close-packed interior, and almost all internal hydrogen bonding groups are paired to form good hydrogen bonds. The method used to calculate the native conformation of small organic molecules might, therefore, be applied to a protein. In this approach, the forces between atoms are represented by simple empirical functions calibrated on known properties of small molecules, and the native conformation is found by moving the atoms to give an equilibrium structure (Hendrickson, 196 1; Lifson & Warshel,' 1968).

This process is exactly equivalent to moving all the atoms until the potential energy function has a minimum value and there is no net force on any atom. As such, the method shares the basic deficiencies of all minimization methods, namely, there are usually many different minimum energy conformations. From thermodynamic considerations, the predominant conformation at any temperature will be the one that has the lowest free energy. As a first approximation, choosing conformations with the lowest potential energy may be adequate. Thus, the native conformation of a protein could perhaps be calculated by finding many different equilibrium arrangements of the atoms and choosing the arrangement with the lowest value of the potential energy.

The method is not guaranteed to work as the native protein conformation may not have a lower potential energy than all other conformations for two reasons. (I) Some other conformation with a higher potential energy may be entropically favoured and have a lower free energv. (2) The native conformation may be the conformation reached most easily during the folding process; other conformations with lower free energy would never be encountered due to the huge number of possible conformations and the rapidity with which real proteins fold (Levinthal, 1968). Problem (1) can be solved by calculating the free energy in the vicinity of each equilibrium structure. Problem (2) is much more fundamental and would make the very difficult calculation of the native conformation much more difficult.

For the present, we follow accepted practice and adopt the working assumption that the native structure does have the lowest possible value of the potential energy (Levitt, 1982). The problem is then reduced to a search for the set of atomic co-ordinates that have this lowest energy value. Such a search has two phases: (1) the generation of trial structures, and (2) the refinement of these trial structures to make them more like the native conformation. The number of trial structures needed depends on the power of the refinement. If any trial structure could be refined to become the conformation with lowest energy, only one trial structure would be needed.

In the past, all attempts to calculate native conformation have followed this approach. Almost all have been tested on the same small protein, bovine pancreatic trypsin inhibitor, whose X-ray structure is known (Huber et *al.*, 1971). This means the methods can be evaluated on the basis of the root-mean-square deviation of the calculated conformation from the X-ray conformation. One series of calculations tried to avoid the multitude of false

minima by using a simplified potential energy function. The atoms of each side-chain were approximated by a single spherical group and r.m.s.† deviations of 5·3 Å to 6·7 Å were obtained (Levitt & Warshel, **1975;** Levitt: 1976). Subsequently, the simple energy terms were replaced by sets of constraints on the positions of C^{α} or C^{β} atoms, and r.m.s. deviations ranging from 3·8 Å to 6·0 Å were obtained (Kuntz *et al.,* 1976,1979). In these studies, the energy value was not always lowest for the conformation with the lowest r.m.s. (Levitt, **1976).** The energy value could not, therefore, be used to find that conformation. When atoms are treated as spherical groups, the detailed interactions that seem to be responsible for the stability of the X-ray structure are omitted or greatly weakened. As a result, the native conformation is no longer especially stable.

Energy minimization that includes the details of all interatomic interaction is much more expensive computationally and has been tested far less extensively than for simplified interactions. When such energy minimization was applied to a small peptide fragment of a protein, there were many different low energy conformations (Gibson & Scheraga, **1969).** When the method was used on BPTI, the resulting structures bore little resemblance to the X-ray structure and no r.m.s. deviation was given (Burgess & Scheraga, **1975).** Recently, closer agreement, to the X-ray structure (r.m.s. deviation of 4·4 A) has been obtained bv energy minimization that started from a conformation model built manually to "have the same loop structure as the native protein (Meirovitch & Scheraga, **1981).**

Any computational search for the native conformation must satisfy the following criteria. (1) The conformations with the lowest r.m.s. deviation must have the lowest energy, enabling energy to be used to select the best conformations. (2) A wide variety of trail structures must be generated so that a native-like conformation is not excluded. (3) The refinement technique must be powerful enough to bring the trail structures closer to the X-ray structure.

The present studv uses such a scheme to calculate the native conformation of BPTI. The potential energy is represented in atomic detail so as to be more discriminating than with the simplified interactions. The vast number of possible chain foldings is reduced to manageable proportions by using restraints that make all calculated structures have the same secondary structure and disulphide bridge pairings as native BPTI. A random set of 25 trial structures are generated by a new method of restrained energy minimization that employs soft-atoms. These trial structures are then refined by a powerful combination of energy minimization and molecular dynamics that avoids local minima, and changes the conformation by up to 2 A to become more like the native structure. All the conformations are analysed in terms of energy, torsion angles: hydrogen bonds and atomic accessibilities. In spite of the common constraints, there are big differences between the calculated conformations which have r.m.s. deviations ranging from 3·0 to 6·7 A. The conformations with the lowest energies always have the lowest r.m.s. deviations. The value of the potential energy can therefore be used to select the best conformation, whose r.m.s. deviation of 3 A is lower than obtained in all previous studies.

† Abbreviations used: r.m.s., root-mean-square; BPTI. bovine pancreatic trypsin inhibitor.

2. Methods to Generate Conformations

Two classes of methods are needed for the present study: (1) methods used to generate low energy conformations that satisfy restraints, and (2) methods used to analyse the generated conformations. These methods are presented in this and the next section emphasizing original developments.

(a) *Possible approaches*

(i) *Distance geometry*

Previous methods used to generate protein conformations subject to external restraints have emphasized the fit to the restraints rather then the energy of the resulting conformation (Kuntz *et al.*, 1976,1979). The method of distance geometry (Mackay, 1974; Crippen, 1977) works from individually specified lower and upper bounds on the distances between all pairs of atoms; energetic considerations are represented as distance constraints. For example, bounds are set to hold covalently bonded atoms to the equilibrium bond length separation or to keep non-bonded atoms from getting closer than the sum of the van der Waals' radii.

In spite of its elegance and simplicity, the distance geometry method has two shortcomings. (1) Computer storage requirements increase as n^2 (actually $9n^2/2$ words for n atoms), explaining why the method has so far been used only to generate simplified protein conformations in which each residue is represented as one "effective" atom (Havel *et al.*, 1979). (2) Proper representation of a general interatomic potential by distance constraints is impossible due to the presence of non-quadratic r^{-6} and r^{-12} non-bonded interactions.

(ii) *Soft-atom restrained energy minimization*

Here a different more direct approach is taken. Methods- used previously to minimize the total potential energy of a macromolecule are modified to include terms that restrain selected quantities to target values (interatomic distances, angles, distances from the molecular centroid etc.). At first sight this approach seems straightforward and offers the combined advantages that the energy can be represented properly and that any number of very general restraints can be used (not just distances). However, in preliminary attempts at such calculations, energy minimization made little progress when either there were many restraints or the restrained variables were initially far from their target values. This failure occurred with a variable metric minimization method, VA09D from the Harwell Subroutine Library (Fletcher, 1970) known for its robustness and rapid convergence.

"Soft-atoms" were introduced in an attempt to improve the performance of restrained energy minimization. For such atoms, the van der Waals' interaction is modified so that the infinitely high energy that results from atoms overlapping is replaced by a high but finite value (see Fig. 1). This modification of the van der Waals' in&action removes all the infinities from the potential and even allows atoms to pass through each other. With this improvement. any number of

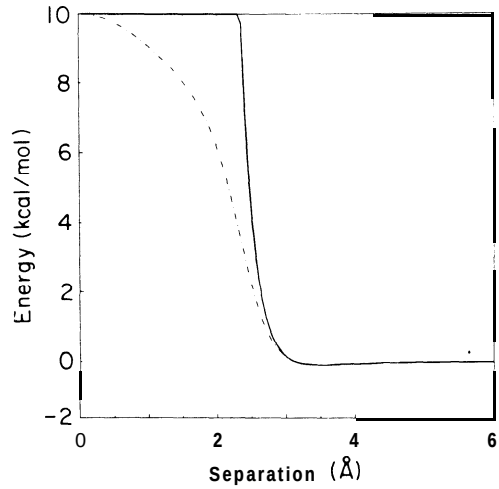


FIG. 1. Comparison of the soft-atom van der Waals' interaction energy of a pair of nitrogen atoms (broken line) with the corresponding normal-atom van der Waals' interaction energy (continuous line). The normal potential has the form $U_{vd}(r) = A/r^{12} - B/r^6$ for the atom pair at a distance r apart. The soft-atom potential has the form:

$$U_{vd}^{soft} = \{A/r^{12} - B/r^6\}/\{(A/r^{12})(1+0{\cdot}1r^2)/h+1\}.$$

When r is greater than $(2A/B)^{1/6}$, the separation at which the energy U_{vd} is minimum. U_{vd}^{soft} is very like U_{vd}. As r gets smaller, U_{vd}^{soft} varies like $h(1-0{\cdot}1r^2)$, and $U_{vd}^{soft} = h$ at $r = 0$ when 2 atoms are completely overlapped. In the present study h is taken as 10 kcal/mol; tests with h = 3 and 30 kcal/mol gave similar results.

restraints can be added to the potential energy function (over **1200** were tried) without affecting the rapid convergence of energy minimization.

(b) *Energy minimization in torsion angle space*

(i) *The molecule studied*

The method outlined above was tested on BPTI, a molecule whose conformation has been determined by X-ray diffraction studies (Huber et *al.*, 197 **1**; Deisenhofer & Steigemann, **1975).** This same protein has been used in most previous folding calculations (Burgess & Scheraga, **1975;** Levitt & Warshel, **1975;** Kuntz et *al.*, 1976,1979; Levitt, **1976;** Goel & Yčas, **1979;** Robson & Osguthorpe, **1979;** Meirovitch & Scheraga, 1981), facilitating comparative evaluation of the results. BPTI protein has 58 amino acid residues, 892 atoms, 454 non-hydrogen atoms and 208 ϕ, ψ, and χ single-bond torsion angles. Omitting all hydrogen atoms makes it difficult to represent hydrogen bonds realistically, so the 61 hydrogen atoms on peptide and amide groups are included giving a total of 515 atoms. (The eight hydrogen atoms on hydroxyl groups, which are free to rotate, are not included to avoid the difficulty of initial positioning and the additional single bond torsion angles.)

(ii) *Starting conformations*

Different low energy constrained conformations are generated by starting the minimization from a variety of random chain conformations. These sets of atomic co-ordinates of BPTI are built with standard bond lengths, bond angles and

double-bond torsion angles. Every χ_1 torsion angle is set to $-60°$, the most common value found in real globular protein conformations (Janin et *al.,* 1978). whereas all other χ angles are set to 180°, giving a standard *trans* conform&ion. The initial (ϕ, ψ) va l ues of each residue depend on whether the residue is assigned to a-helix or P-sheet secondary structure (see Table 1). For a-helix residues, ϕ is chosen randomly to be in the range -65" and $-55°$, and ψ is $-45°$ to $-35°$. For the P-sheet residues, ϕ is -130" to $-110°$ and ψ is 140° to 160". For the other residues, the initial ϕ and ψ values are each within 60" of $(\phi, \psi) = (-90°, 90°)$, giving extended chain structure.

The particular starting conformation depends on the random number generator, which is, in turn, controlled by specifying a starting number known as the "seed". For simplicity, seed values of 1 to 25 are used to generate 25 different random starting conformations. These conformations (see Fig. 2) are not at all compact and contain rod-like segments (the a-helices and P-sheet strands) separated by less regular regions.

(iii) *Restraints*

The restraints used to guide the energy minimization include the torsion angles for the regions assigned to a-helices and P-strands, the 16 main-chain hydrogen

TABLE 1

Secondary structure, hydrogen bond and disulphide bond restraints

Secondary structure		Residues and atom pairs‡
3_{10}-Helix	2–5†	2(H). . .5(O), 3(H). . .6(O)
β-Strand	17-23	18(H). . .35(O), 35(H). . .18(O),
		20(H). . .33(O), 33(H). . .20(O)
β-Strand	30–36	22(H) . . . 31(O), 31(H). . .22(O),
		24(H) . . . 29(O)
β-Strand	**4 4 4 6**	21(H . . . 45(O), 45(H). . .21(O)
a-Helix	**47-55**	47(H) . . . 51(O), 48(H). . .52(O),
		49(H) . . . 53(O), 50(H). . .54(O),
		51(H) . . . 55(O)
S-S bridges		5(S^γ)–55(S^γ),
		14(S^γ)–38(S^γ),
		30(S^γ)–51(S^γ)

† These residues in secondary structure are given appropriate initial (ϕ, ψ) values (see section 2(b)(v)). These values are enforced during the torsion angle minimization by the extra energy term:

$$K_{\phi\psi}\{(\phi-\phi_{\text{init}})^2 + (\psi-\psi_{\text{init}})^2\}, \tag{17}$$

where $K_{\phi\psi}$ = 100 kcal/mol per radian^2. The (ϕ, ψ) angles of other residues are not affected.

‡ The atom pairs in the 16 hydrogen bonds and 3 disulphide bonds are brought together by the extra energy term (see eqn (1)):

$$K_d\{[(d-d_0)^2 + d_c^2]^{1/2} - d_c\}, \tag{18}$$

where K_d = 100 kcal/mol per Å^2, and the target value, d_0, is **2** Å for the O . . . H separation in hydrogen bonds and 3 Å for the S-S separation in S-S bridges. The constant d_c (taken as **2** A) ensures that for $|d - d_0|$ greater than d_c, the constraint energy varies linearly with $|d - d_0|$, whereas for $|d - d_0|$ less than d_c, it varies quadratically. Without this device, the excessively high restraint energies that occur when ***d*** is far from its target value would dominate the energy minimization.

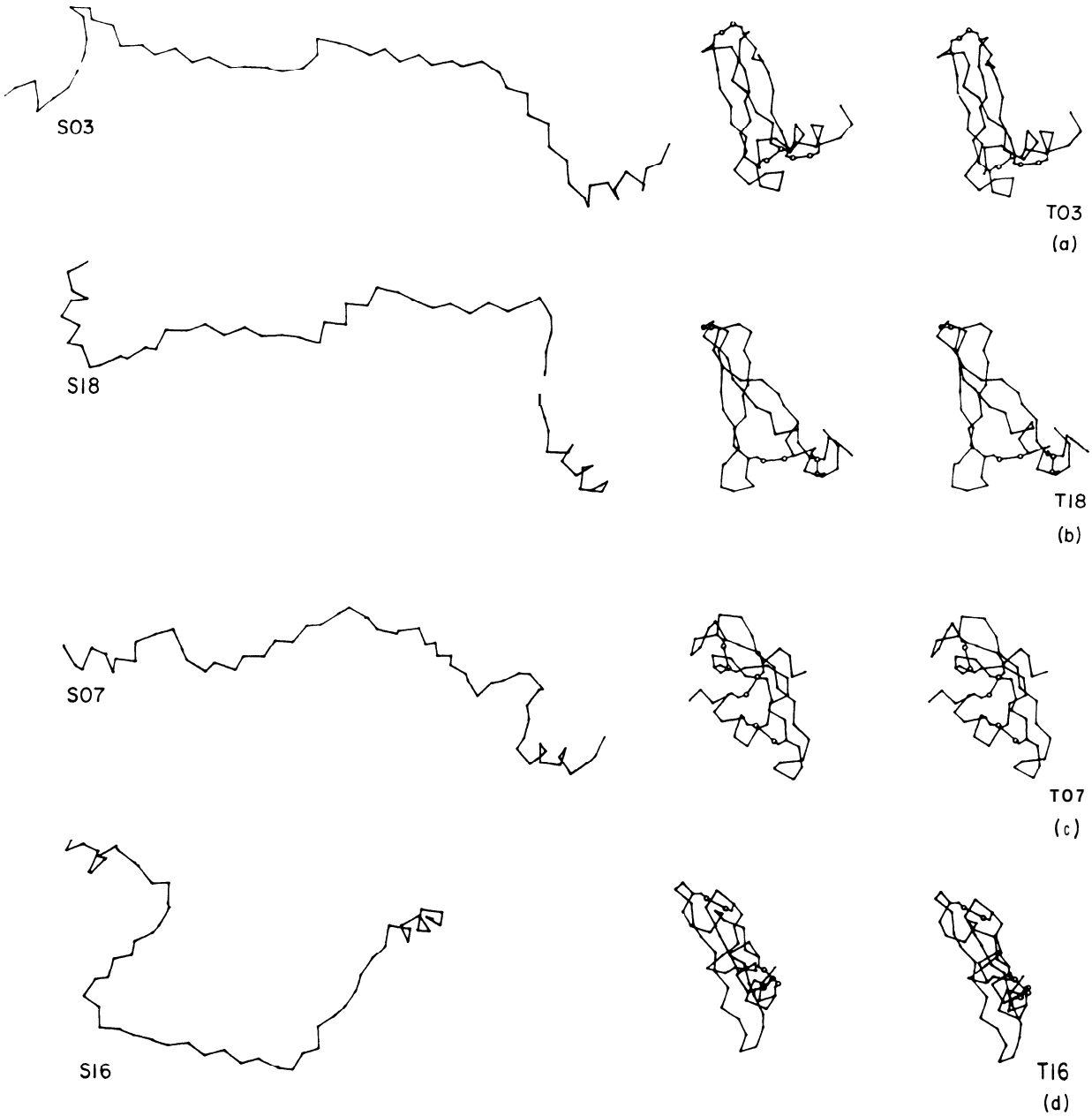


FIG. 2. Four of the starting conformations together with the conformations that. result from torsion angle minimization. The conformations shown are (a) SO3 and T03, (b) S18 and T18, (c) SO7 and T07, (d) S16 and T16. Two conformations, SO3 and S18, lead to the 2 best conformations generated here: A03 and Al8 both have the lowest energy and closest fit to the X-ray co-ordinates. The 2 other conformations, SO7 and S16, lead to the worst conformations generated here: Cl6 has the highest energy and CO7 is the worst fit to the X-ray co-ordinates. Although each calculated conformation consists of all 515 atoms corresponding to the BPTI model used here, the drawings only show the virtual bonds between adjacent a-carbons, and also, in the case of the minimized conformations, the 6 cystine residues forming the 3 disulphide bridges (residues 5 : 55, 14 : 38, 30 : 51). These and all other molecular conformations shown in this paper were drawn using Dr S. Motherwell's program "PLUTO"; I am grateful to him for having developed such a powerful and useful tool.

bonds defining this secondary structure, and the three disulphide bonds (see Table 1). The hydrogen bonding groups in the a-helices are close together in the starting conformation, whereas those in the P-strands are not. These long-range hydrogen bonds are formed during the energy minimization as the restraints take

effect. These restraints are equivalent to having prior knowledge of the secondary structure, including the specific pairing of residues in P-sheets, and the disulphide bridges.

(iv) ***Choice of variables***

The potential energy of this molecule is minimized with respect to the 208 single-bond torsion angles (ϕ, ψ and χ) (53 ϕ angles, excluding those of Arg1, Pro2, Pro8, Pro9, and Pro13; 58 ψ angles; 97 χ angles). Although minimization with respect to the 1545 atomic Cartesian co-ordinates is used at a later stage of this work, torsion angles are preferred for minimization from the open starting conformations. The smaller number of torsion angle variables allows use of the very efficient variable metric minimization method. Because bond lengths, bond angles, and many torsion angles are rigid in torsion angle minimization, the initially good standard geometry cannot be distorted by the large forces that may arise from close contacts or restraints.

(v) *The **energy function** and first **derivatives***

The complete expression for the potential energy (in kcal/mol; 1 kcal = 4·184 kJ) for restrained energy minimization with respect to single-bond torsion angles is:

$$U_{\mathrm{TOT}} = \sum_{\text{side-chains}} K_\chi[1+\cos(n\chi_i)] + \sum_{\text{secondary structure}} K_{\phi\psi}\{(\phi_i-\phi_0)^2 + (\psi_i-\psi_0)^2\} + \sum_{19\ \text{restraints}} K_d\{[(d_i-d_0)^2+d_c^2]^{1/2}-d_c\} + \sum_{\text{atom pairs}} \{[A/r_{ij}^{12}-B/r_{ij}^6-C]/[(A/r_{ij}^{12})(1+0{\cdot}1r_{ij}^2)/h+1]\}S(r_{ij}). \quad (1)$$

The first term represents the preference that certain side-chain 'torsion angles have for the staggered conformation; values for the energy parameters K_χ and n are given in Table 2. The second term restrains those (ϕ, ψ) angles assigned to secondary structure to their initial values (see Table 1). The third term restrains distances d_i to their target values d_0 and is also described more fully in Table 1.

The final term represents non-bonded interactions, namely the van der Waals' and hydrogen bond interactions between pairs of atoms that are not close neighbours along the chemical structure (separated bv at least 4 bonds). The energy parameters A and B given in Table 2 for each type of atom were derived by fitting the crystal dimensions and sublimation energies in a series of hydrocarbon, amide and amino acid crystals (Levitt & Lifson, unpublished work). The parameter C($= A/6^{12} - B/6^6$) ensures that the non-bonded energy is zero at the cutoff separation of 6 A. For r greater than 6 Å, non-bonded interactions are neglected. The function $S(r)$ is a polynomial step function (Levitt, 1976) that makes the energy smoothly differentiable in spite of the truncation at 6 A. The second bracketed term makes the atoms soft and the parameter h is taken as 10 kcal/mol (see Fig. 1 for details).

No hydrogen bond function is used,. but the attraction of hvdrogen bonding groups is represented by special van der Waals' parameters.' The interaction

TABLE 2

Torsion and non-bonded energy parameters

Type of interaction	Energy parameters	
Torsion †	K_χ (kcal/mol)	n
ϕ	0·0	0
ψ	0·0	0
χ	1·4	3
χ_2 (aromatic)	0·1	6
χ (acid or amide)	0·1	6

Non-bond‡	*A*	*B*	ε	r°
H . . . H§‖	290	1·08	0·00100	2·8525
O . . . O	145,834	328	0·18479	3·1005
V . . . V	417,000	1383	1·14630	2·9067
N . . . N	3,952,850	2556	0·41315	3·8171
M . . . M	3,952,850	2556	0·41315	3·8171
C . . . C	3,075,695	953	0·07382	4·3150
A . . . A	1,200,965	425	0·03763	4·2202
O . . . H	2913	241	5·0	1·7
V . . . H	2913	241	5·0	1·7
O . . . V	417,000	1383	1·14630	2·9067
V . . . N	417,000	1383	1·14630	2·9067
M . . . O¶	417,000	0	0·02486	4·0
M . . . V	417,000	0	0·02486	4·0

† The torsion angle energy contribution has the form:

$$K_\chi[1 + \cos(n\chi)] \text{ kcal/mol.} \tag{19}$$

‡ These 4 energy parameters are not independent. The ***A*** and ***B*** parameters are used in the energy function (eqn (1)), while ε and r° are given for comparative purposes: $\varepsilon = B^2/4A$, $r^\circ = (2A/B)^{1/6}$. The potential has a minimum value of $-\varepsilon$ at separation r°.

§ The interaction for atom pairs that are not listed uses the geometric mean of the ***A*** and ***B*** values of pairs involved, for example, $A_{\ldots X} = (A_{\ldots O} A_{\ldots X})^{1/2}$.

‖ The atom types are defined as follows: H, hydrogen atom attached to peptide or amide nitrogen; V, extended oxygen atom representing the entire hydroxyl group; O, oxygen atom in a carbonyl or carboxyl group; N, nitrogen atom in peptide or amide group; C, extended carbon atom including 1, 2 or 3 bonded hydrogen atoms; A, carbon atom in carbonyl or carboxyl group; M, extended nitrogen atom including 1, 2 or 3 bonded hydrogen atoms in lysine or arginine side-chains.

¶ The M . . . O and M . . . V interactions were made purely repulsive to prevent these side-chains from folding back onto the protein surface; ε is the energy value at the nominal separation of 4 A.

between H and O atoms has a minimum of depth 5 kcal/mol at a separation of 1·7 Å representing the N-H. . . O hydrogen bonds. For the hydroxyl groups (denoted as extended V atoms) there is a weaker but significant interaction (depth 1·14 kcal/mol at 2·9 A) with other hydroxyl groups, O atoms and N atoms. This represents the fact that a hydroxyl group can be a hydrogen bond acceptor or donor. The NH, NH_2 and NH_3 groups in side-chains (extended M atoms) are not allowed to make hydrogen bonds with other protein atoms as they are expected to interact strongly with the surrounding solvent.

The expressions for first derivatives of the energy function with respect to the 208 single-bond torsion angle variables are needed for efficient energy minimization. The first two terms of the potential energy function depend

explicitly on the torsion angle variables and can be differentiated directly. The third and fourth terms depend explicitly on the distances between pairs of atoms (d_i and r_{ij}) and must be differentiated in stages:

$$U_{\mathrm{TOT}} = \sum_i U_1(\phi_i) + \sum_{i,j} U_2(r_{ij}), \tag{2}$$

where ϕ_i denotes a ϕ, ψ or χ torsion angle and r_{ij} denotes either *a* non-bonded or restrained interatomic distance. Differentiation with **respect to** ϕ_k gives:

$$\partial U_{\mathrm{TOT}}/\partial\phi_k = \partial U_1(\phi_k)/\partial\phi_k + \sum_{i,j} [\partial U_2(r_{ij})/\partial r_{ij}] \cdot [\partial r_{ij}/\partial\phi_k]. \tag{3}$$

Because there are many (i, j) pairs and because $\partial r_{ij}/\partial\phi_k$ is non-zero for any torsion angle k along the chain joining atoms i and j, the transformation in the second term is very inefficient. The correct method is to use the Cartesian co-ordinates $\mathbf{r}_n$. The second term then becomes:

$$\sum_{i,j}\sum_n [\partial U_2(r_{ij})/\partial r_{ij}][\partial r_{ij}/\partial \mathbf{r}_n][\partial \mathbf{r}_n/\partial\phi_k] = \sum_n \left\{\sum_{i,j} [\partial U_2(r_{ij})/\partial r_{ij}] \cdot [\partial r_{ij}/\partial \mathbf{r}_n]\right\} \cdot [\partial \mathbf{r}_n/\partial\phi_k]. \tag{4}$$

Because $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, $\partial r_{ij}/\partial \mathbf{r}_n = 0$ unless $n = i$ or j, making the summation over (i, j) very sparse. When the efficient formulation is used (eqn (4)), the derivative calculation takes less time than the energy calculation alone. When the inefficient formulation is used in an otherwise highly optimized energy minimization program specially written for a super-fast array processor (Pottle et *al.*, 1980), the derivative calculation takes 170 times longer than the calculation of the energy alone.

Because of its high efficiency, the above scheme is described in more detail as follows. (1) Generate the Cartesian co-ordinates of starting conformations using standard geometry and randomly chosen torsion angle values. (2) Calculate the non-bonded energy for each distance r_{ij} and at the same time calculate the contribution of this term to the Cartesian first derivative $\partial U_2/\partial \mathbf{r}_n$. (3) Transform the completed Cartesian first derivative vector to torsion angle space using:

$$\begin{aligned}\partial U_2/\partial\phi_k &= \sum_n [\partial \mathrm{U}_2/\partial \mathbf{r}_n] \cdot [\partial \mathbf{r}_n/\partial\phi_k] \\ &= \sum_n [\partial \mathrm{U}_2/\partial \mathbf{r}_n] \cdot [\mathbf{n}_{\phi k} \times (\mathbf{r}_n - \mathbf{r}_k)],\end{aligned} \tag{5}$$

where $\mathbf{n}_{\phi k}$ is the unit vector along the bond about which ϕ_k operates and $\mathbf{r}_k$ is the position of the atom at the end of this bond. Both $\mathbf{n}_{\phi k}$ and $\mathbf{r}_k$ are calculated in the current Cartesian co-ordinate system. (4) Calculate the torsion angle energy terms and their contributions to the total energy and **first derivatives** in **torsion** angle co-ordinates. (5) Use the total energy and first derivatives to get changes in the torsion angle values $\Delta\phi_i$. (6) Rotate by $\Delta\phi_i$ about the relevant bonds of the current co-ordinates to generate a new set of Cartesian co-ordinates and repeat the process from step (2). (For reasons of accuracy, all rotations are actually referred back to the initial Cartesian co-ordinates.) When changing the torsion

angles, it is assumed that the a-carbon of the middle residue (residue 29 in BPTI) is fixed; this is more efficient than fixing an atom at one end of the chain.

(vi) *Energy minimization*

The robust, very well tested minimization routine VA09D (Fletcher, 1970) is used to minimize the energy by simultaneously changing all 208 torsion angle variables. On each iteration, the method uses the energy and first derivative values from previous steps to approximate the inverse of the second derivative matrix. This matrix, known as the metric of the energy surface, describes the local curvature and is used to calculate the change in conformation that gets to a local energy minimum. By using "soft-atoms", a smoothly truncated energy function (eqn (1)), and double precision arithmetic (16 significant decimal places), the minimization procedure reaches a precisely defined minimum energy conformation.

More specifically, generation of a minimum energy conformation for BPTI requires between 849 and 1430 energy and derivative evaluations. The angles of these final conformations are accurate to 0·00001 radians and no component of the energy first derivative vector (the torsional forces or couples) exceeds 0·00001 kcal/mol per radian. Fewer energy and derivative evaluations would have given essentially the same final conformation but the small value of the final forces are a reassurance against errors. A major factor contributing to the efficiency of Fletcher's (1970) VA09D minimization routine is the fact that most iterations (over 95%) require only one evaluation of the energy and its first derivatives. Other variable metric minimizers (Davidon, 1959; Fletcher & Powell, 1963) require at least two evaluations per iteration.

(c) Energy minimization and molecular dynamics in Cartesian space

(i) *Advantages of Cartesian co-ordinates*

The soft-atom energy minimization in torsion angle space described above generates a stereochemically acceptable conformation, which also satisfies the imposed restraints. Additional energy minimization with respect to all the atomic Cartesian co-ordinates is necessary for the following reasons. (1) Bond lengths, bond angles and double-bond torsion angles are not able to deviate from standard values in torsion angle minimization. (2) Molecular dynamics, needed to anneal the conformations by moving groups over small energy barriers, requires Cartesian co-ordinates as the equations of motion in torsion angle space are too complicated. (3) The energy function in torsion angle space (eqn (1)) uses physically unrealistic soft-atoms introduced only to allow restrained minimization, omits a directional hydrogen bonding term, and lacks a special (ϕ, ψ) energy term needed to give realistic (ϕ, ψ) angle distributions.

All these deficiencies are overcome by using Cartesian co-ordinates with the same energy function used in studies of proline isomerization (Levitt, 1981*a*) and hydrogen bond dynamics (Levitt, 1981*b*). This potential behaved well in these studies, was better than other potentials in a series of comparative energy

minimizations from the X-ray structure (Levitt, 1980), and has been described in detail (Levitt, 19833).

(ii) *The energy function*

The energy function is given here to allow comparison with the function used in torsion angle space:

$$
\begin{aligned}
U_{\mathrm{T\ \ O}} = {} & \sum_{\text{bonds}} K_b(b_i - b_0)^2 + \sum_{\text{bond angles}} K_\theta(\theta_i - \theta_0)^2 \\
& + \sum_{\text{torsion angles}} K_\chi[1 + \cos(n\chi_i + \delta)] + \sum_{(\phi,\psi)\text{ pairs}} F(\phi_i, \psi_i) \\
& + \sum_{\text{non-bonded pairs}} A/r_{ij}^{12} - B/r_{ij}^{6} \\
& + \sum_{\text{O}\cdots\text{H pairs}} (A/r_{ij}^{12} - B/r_{ij}^{6})\, e^{-\theta^2/\sigma^2} + (A'/r^{12} - B'/r^{6})(1 - e^{-\theta^2/\sigma^2}) \\
& + \sum_{16\ \text{restraints}} K_d(d_i - d_0)^2.
\end{aligned}
$$

The description and values of the parameters are given elsewhere (Levitt, 19833). Differences from the torsion space energy function include: (1) bond lengths and bond angles can deviate from standard values. (2) All torsion angles can change and a special potential is used for the (ϕ, ψ) angle pairs. (3) Van der Waals' interactions are not softened. No smoothing potential is needed as atom pairs in range are listed and used for 100 iterations. (4) The hydrogen bonding interaction depends on the O . . . N-H angle, θ. (5) The 16 hydrogen bond restraints are still imposed but the restraint energy is a simple quadratic function. The disulphide bonds are treated like any other bonds and no restraint is used on the (ϕ, ψ) angles in regions of secondary structure.

(iii) *Energy minimization*

Analytical first derivatives of the energy function are used in energy minimization with respect to all 1545 Cartesian co-ordinates of the 515 atoms in the molecule. The conjugate gradient minimization routine (Hestenes & Steifel, 1952; Fletcher & Reeves, 1964) is used as the variable matrix method would require memory space for 1,194,285 numbers (1545 x 1546/2), which is impractical at present. Conjugate gradients minimization requires space for only 1545 x 3 numbers Here the minimization is continued for 3000 energy evaluations, at the end of which the r.m.s. force (the first derivative of the energy) is less than 0·05 kcal/mol per Å and the energy charge over the last 100 evaluations is less than 1 kcal/mol.

(iv) *Molecular dynamics*

Molecular dynamics methods use the values of the energy and its analytical first derivative to simulate the motion of the atoms in the presence of thermal energy. Besides yielding a wealth of information about the amplitudes and rates of atomic motion on the picosecond time-scale (McCammon *et al.*, 1977), molecular dynamics has been shown to "anneal" conformations (Levitt, *1983a).* This annealing process

removes unfavourable interactions that remain after energy minimization, using the thermal energy to hop over small energetic barriers. Subsequent energy minimization is then able to reach a lower energy value. Dynamic annealing is used here on selected conformations in an attempt to get to the lowest possible minimum energy values.

The iterative solution of the equations of motion that is the basis of molecular dynamics is done as described (Levitt, 19836); the annealing dynamics is continued for 30 picoseconds (15,000 energy and derivative evaluations with a time-step of 2×10^{-15} s). The energy function used for dynamics is almost the same as that used for minimization (eqn (6)). In particular, the restraint on the 16 hydrogen bonds is included. The A and B energy parameters used for dynamics are slightly different from those used for minimization in an attempt to correct for thermal expansion. In fact, such correction is probably unnecessary but the modified parameters are used to be consistent with previous dynamics simulations (Levitt, 1980,1981*b*,1983*a*,*b*).

3. Methods to Analyse Conformations

Analysing one protein conformation is not straightforward; here about 30 conformations have to be analysed and compared. For this task, existing methods have been extended and new methods introduced.

(a) Comparing atomic co-ordinates

The differences between two sets of co-ordinates i and j is quantified using the r.m.s. deviation of the inter-C^{α} distances as follows:

$$\mathrm{Ad}_{\alpha}^{ij} = \left[\frac{1}{N_{kl}} \sum_{k>l} \sum_{l} (d_{kl}^{i} - d_{kl}^{j})^2\right]^{1/2}, \tag{7}$$

where d_{kl}^{i} is the distance between m-carbon atoms k and l in conformation i, d_{kl}^{j} is the corresponding distance in conformation j, and N_{kl} is the number of terms in the summation (58 x 57/2 = 1653 for BPTI). Δd_{α}^{ij} does not require any superposition of the co-ordinates and has been used in most previous comparisons of protein conformations. Here this r.m.s. deviation is used most often to compare conformation i with the X-ray conformation X and the deviation Δd_{α}^{Xi} is then referred to simply as the r.m.s. deviation.

The distance deviation given above is the same for a conformation and its mirror-image (Cohen & Sternberg, 1980) and two other deviations are also used here:

$$\begin{aligned} \Delta r_{\mathrm{A}}^{ij} &= \left[\frac{1}{N_{\mathrm{A}}} \sum_{\text{all atoms } k} (\mathbf{r}_k^i - \mathbf{r}_k^j)^2\right]^{1/2} \\ \Delta r_{\alpha}^{ij} &= \left[\frac{1}{N_{\alpha}} \sum_{(\alpha \text{ atoms } k} (\mathbf{r}_k^i - \mathbf{r}_k^j)^2\right]^{1/2}, \end{aligned} \tag{8}$$

where $\mathbf{r}_k^i$ is the position of the kth atom in conformation i, $\mathbf{r}_k^j$ is the position of the

corresponding atom in conformation j, N_A is the number of atoms, and N_α is the number of a-carbon atoms. This deviation does depend on the relative orientation of the two sets of co-ordinates which must be superimposed using matrix algebra (Kabsch, 1976). All the **454** non-hydrogen atoms are included when calculating the best superposition even if only the a-carbon positions are used in subsequent calculation of Δr_α^{ij}. Projections of the conformation space containing several conformations are obtained as before (Levitt, 1983c) using Δd_α^{ij}.

(b) *Torsion angles, hydrogen bonds and solvent accessibility*

(i) *Torsion angles*

The (ϕ, ψ) torsion angles of different conformations were compared as described above for Cartesian co-ordinates but this measure was not found useful. Instead, the common scheme of plotting the ϕ angle of residue i against the corresponding ψ angle value is used.

(ii) *Hydrogen bonds*

Hvdrogen bonds are **a** very convenient index for describing and comparing conformations. They are few in number, show the spatial proximitv of groups that may be distant along the chain, and play an important role in stabilizing the native conformation. A computer program is used to find hydrogen bonds automatically according to the following criteria: **(1)** the H . . . *acceptor* separation must be less than 2·4 Å; **(2)** the *donor-H . . . acceptor* angle must be within 35" of linearity. When the hydrogen atom is not explicitly included (hydroxyl groups and NH, NH, and NH, groups in arginine and lysine side-chains), it is added with standard geometry so as to make as good a hydrogen bond as possible. Hydrogen bonds in different conformations are compared by collating the lists automatically.

(iii) *Solvent accessibility*

The area of individual atoms accessible to the solvent is calculated using my own implementation of the Lee & Richards **(1971)** algorithm. The radius of the solvent atom is taken as 1·4 A, while the solvent exclusion radii of the different types of atoms are taken as: 0 Å (H), 1·4 Å (0 or V), 1·65 Å **(N),** 1·87 Å **(C),** 1·76 Å **(A),** 1·85 Å (S) and 1·65 Å (M). This same program and set of radii have been used extensively in previous solvent accessibility calculations (Chothia, **1976).** The accessible areas of classes of atoms, individual residues and entire conformations are obtained by summing up the accessible areas of individual atoms.

(c) *Writhing numbers and loop threading*

The line joining adjacent a-carbon atoms along the protein backbone traces out a curve in three-dimensional space. The variation of twist and curvature along the curve has been analysed by formulae of differential geometry (Rackovsky & Scheraga, **1980).** Another property of a space curve is the writhing number (Fuller, 1971), which is not a local measure but depends on the overall shape of

the curve. Although the writhing number has been used before to analyse random-coil conformations of DNA (Le Bret, 1979; Benham, 1978), it has not been applied to proteins. Here a simple formula for the writhing number is developed and shown to be able to distinguish between different chain threadings.

The writhing number is conventionally calculated by the following steps (Fuller, **1971). (1)** Project the conformation onto a plane perpendicular to the particular viewing direction $\boldsymbol{\omega}$. (2) Score each region of self-overlap in projection as + 1 or **− 1** depending on its handedness (see Fig. 3(a) and eqn. **(14))** and sum the scores to get the directional writhing number. (3) Repeat steps **(1)** and (2) for all view directions, $\boldsymbol{\omega}$, and calculate the writhing number as the mean value of the directional writhing numbers.

Mathematically the writhing number is then defined as:

$$W = \frac{1}{4\pi}\int \mathrm{d}\boldsymbol{\omega}\;\; W(\boldsymbol{\omega}) = \frac{1}{4\pi}\int \mathrm{d}\boldsymbol{\omega}\;\sum_{i,j>i}\delta_{ij}(\boldsymbol{\omega}), \tag{9}$$

where $\delta_{ij} = 0$ if the chain segments i and j do not overlap when viewed along direction $\boldsymbol{\omega}$, $\delta_{ij} = \mathbf{1}$ if they overlap in a right-handed sense, and $\delta_{ij} = -\mathbf{1}$ if they overlap in a left-handed sense. This formula is not suited to efficient computation as we must average over many different directions $\boldsymbol{\omega}$ to get an accurate result.

Reversing the order of integration and summation gives a much better formula:

$$W = \sum_{i,j>i}^{n}\frac{1}{4\pi}\int \mathrm{d}\boldsymbol{\omega}\;\; \delta_{ij}(\boldsymbol{\omega}) \;= \sum_{i,j>i}^{n} W_{ij}, \tag{10}$$

where W_{ij} is the fraction of viewing directions along which line segments i and j are seen to overlap; it is positive or negative depending on the handedness of the overlap.

An analytical formula for W_{ij} can be obtained as follows. The directions of overlap of segment i and j are defined by the lines of sight $\mathbf{R}_{ij} = \mathbf{R}_i - \mathbf{R}_j$, where $\mathbf{R}_j$ is any point on segment i and $\mathbf{R}_j$ is any point on segmentj. The limiting values of $\mathbf{R}_{ij}$ occur when one end of segment i is seen in projection to touch one end of segment j. The solid angle Ω_{ij}, in which all $\mathbf{R}_{ij}$ directions of overlap lie, is defined by the four limiting directions (see Fig. 3(b)). Because segments i and j overlap when viewed along either $\mathbf{R}_{ij}$ or $-\mathbf{R}_{ij}$, $W_{ij} = 2\Omega_{ij}/4\pi$.

The solid angle Ω_{ij} is calculated from the angles A, B, C and D of the spherical quadrilateral using the standard formula of spherical triangles:

$$\Omega_{ij} = \mathrm{A} + \mathrm{B} + \mathrm{C} + \mathrm{D} - 2\pi \tag{11}$$

The angles A, B, C and D are calculated from the vectors **a, b, c** and d (see Fig. 3), which are directed to the poles of the great circles forming the sides of the spherical quadrilateral:

$$\begin{aligned}
\mathbf{a} &= (\mathbf{r}_i - \mathbf{r}_{j+1}) \times (\mathbf{r}_i - \mathbf{r}_j)\\
\mathbf{b} &= (\mathbf{r}_i - \mathbf{r}_{j+1}) \times (\mathbf{r}_{i+1} - \mathbf{r}_{j+1})\\
\mathbf{c} &= (\mathbf{r}_{i+1} - \mathbf{r}_j) \times (\mathbf{r}_{i+1} - \mathbf{r}_{j+1})\\
\mathbf{d} &= (\mathbf{r}_{i+1} - \mathbf{r}_j) \times (\mathbf{r}_i - \mathbf{r}_j)
\end{aligned} \tag{12}$$

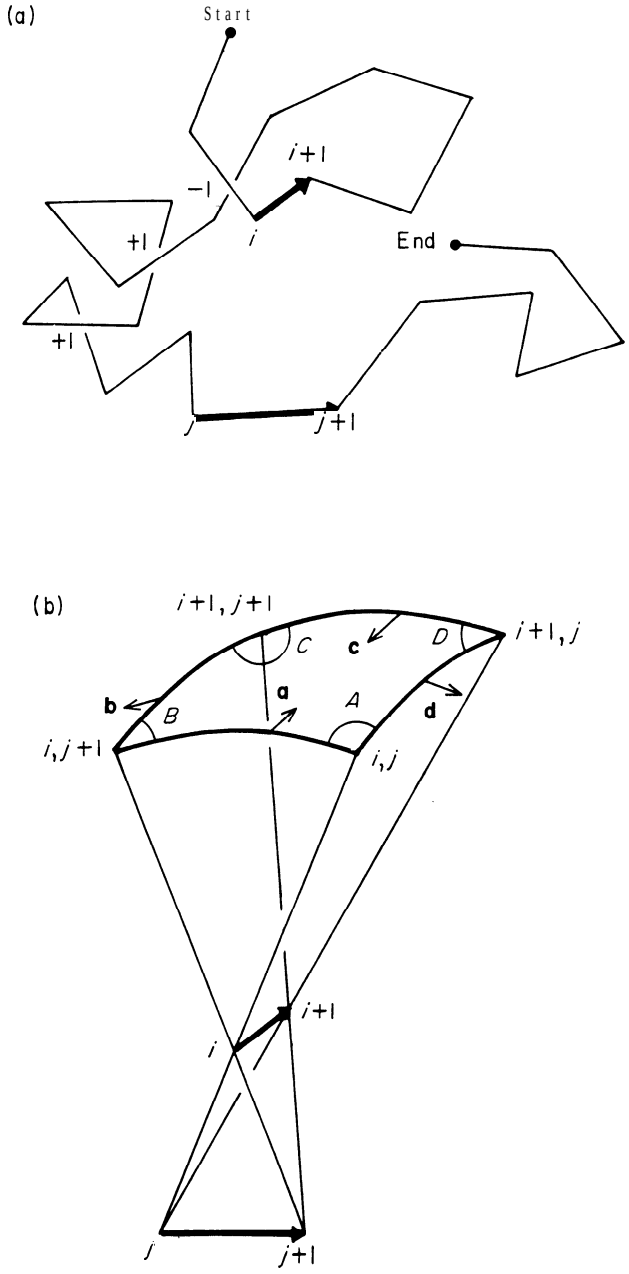


FIG. 3. (a) The total writhing number W can be calculated as a sum of the probabilities, W_{ij}, that segments (i, i + 1) and (j, j + 1) overlap when the structure is viewed from a random direction. (b) W_{ij} is calculated from the positions of the segments. The 2 segments are seen to overlap if the system is viewed from a direction that passes through the spherical quadilateral shown. Consult the text for the equations that give the area of the quadrilateral in terms of angles, A, ***B,*** C and D and pole unit vectors a, b, c and d.

and

$$\begin{aligned} \mathbf{A} &= \cos^{-1}(\mathbf{a} \cdot \mathbf{d}/|\mathbf{a}||\mathbf{d}|) \\ B &= \cos^{-1}(\mathbf{b} \cdot \mathbf{a}/|\mathbf{b}||\mathbf{a}|) \\ &= \cos^{-1} \quad \mathbf{b}/|\mathbf{c}||\mathbf{b}|) \\ D &= \cos^{-1}(\mathbf{d} \cdot \mathbf{c}/|\mathbf{d}||\mathbf{c}|), \end{aligned} \tag{13}$$

where $\mathbf{r}_i$ is the vector of atomic Cartesian co-ordinates of the ith a-carbon at the

beginning of line segment i. The sign of W_{ij} is the same as that of the vector triple product:

$$[(\mathbf{r}_{i+1}-\mathbf{r}_i)\ \mathbf{x}\ (\mathbf{r}_{j+1}-\mathbf{r}_j)]\ .(\mathbf{r}_i-\mathbf{r}_j). \tag{14}$$

The quantity W_{ij} is the contribution of line segments i and j to the total writhing number. Mapping W_{ij} for all (i, j) pairs shows which parts of the chain writhe most. Plotting $\sum_j W_{ij}$ against i shows the total writhing contributions of line segment i.

An approximate formula for W_{ij} is:

$$W'_{ij} = 2(\mathbf{S}_i\ \mathbf{x}\ \ldots\ \mathbf{r}_{ij}/4\pi|\mathbf{r}_{ij}|^3, \tag{15}$$

where $\mathbf{S}_i = (\mathbf{r}_{i+1}-\mathbf{r}_i)$ is the vector along line segment i, and:

$$\mathbf{r}_{ij} = \tfrac{1}{2}(\mathbf{r}_{i+1}+\mathbf{r}_i)-\tfrac{1}{2}(\mathbf{r}_{j+1}+\mathbf{r}_j) \tag{16}$$

is the vector between the midpoints of segments i andj. The approximate formula has a simple interpretation. The numerator is the area of the region delimited by the limiting viewing directions projected into the plane perpendicular to the line joining the segment centres; the denominator is simply the surface area of a sphere of radius $|\mathbf{r}_{ij}|$. This formula is accurate to 10% so long as $|\mathbf{r}_{ij}| > |\mathbf{S}_i|$. For proteins $|\mathbf{S}_i| = 3{\cdot}8$ Å and $|\mathbf{r}_{ij}|$ is large enough for the error to be small. Nevertheless the accurate formula is used here.

(d) *Computing requirements*

The central processing unit times required by one step of torsion angle minimization, Cartesian minimization or Cartesian dynamics are given in Table 3.

TABLE 3

Computer requirements† for a single step on ***BPTI*** *protein*

Program section	CPU time (s)	Percentage of total	Dependence
Torsion angle minimization			
Energy value	1·5	***68***	n
First derivatives	0·4	18	n
Variable metric step	0·3	14	n^2
Total	2·2	100	
Cartesian minimization ***or dynamics***			
Energy value	0·5	***62***	n
First derivatives	0·2	***25***	n
Conjugate gradient or dynamics step	0·1	13	n
Total	0·8	100	

† All calculations were done on an IBM 370/165 computer with the fast-multiply option. The programs were written in a MORTRAN, a rationalized extension of FORTRAN used double precision (64 bit) floating point variables. and were compiled with the FORTRAN Q optimizing compiler OPT = 2. The IBM 370/165 is rated at about 2·5 x 10^6 instructions/s (mips); a typical multiplication as coded in FORTRAN takes 1·4 μs. CPU, central processing unit.

Most time is spent in calculating the energy. Although very similar potentials are used for torsion and Cartesian co-ordinates, the energy calculation in torsion space takes longer. This is because of the different schemes used to find which pairs of atoms (i, j) are close enough in space to be included in the calculation. In torsion space, where substantial conformational changes can occur in one iteration of the minimizer, all (i, j) pairs are scanned on every step (this is done in an efficient way by not checking distances between atoms in residues whose centroids are very far apart). In Cartesian space, where the conformation is expected to change more slowlv, a list of (i, j) pairs close enough to interact is made once and used for about **100** iterations (the precise number depends on the initial value of the energy and the progress of the minimization).

Calculation of derivatives with respect to the 208 torsion angles increases the computer time per step by only 30%. In a torsion angle minimization of B PTI employing **186** variables (Pottle et *al.*, 1980), the derivative calculation increases the computer time per step by 1 7,000% (170-fold). The much greater efficiency of the present calculation results from proper factorization of the derivative calculation (see section 2(b)(v), above).

The additional time required to calculate a change in conformation bv either of the two minimization methods or the dynamics method is less than 15%. In comparing these times it is important to remember that there are **208** variables in torsion space and **1545** in Cartesian space. Because the variable matrix algorithm involves multiplication of vectors by matrices, the time will increase as n^2 (for n variables). In Cartesian space the variable metric step would take $(1545/208)^2$ times longer than in torsion space (the minimization step would then take 17 s or 9 times more than the energy and derivative calculation).

The time required by all the other steps increases linearly with the number of variables n, and the methods could be applied to much larger systems. The memory requirement of the different program sections (see Table 3) has the same dependence on n as the central processing unit time. For BPTI, the programs now run in 400,000 bytes of memory. Because modern computers have verv large memories (up to 20,000,000 bytes), memory requirements will not limit the size of molecule that can be studied.

4. Results

(a) *Generation of conformations*

(i) *M inimixation*

Almost **100** different conformations of BPTI protein are generated in this studv; the convention used to name conformations simply and uniquely is given in Table 4.

Each of the **25** different starting conformations **(SO1** to S25) obtained with an initial random number of **1** to 25 (the "seed") is minimized to convergence with respect to the 208 single-bond torsion angles, using soft-atoms and restraints on the torsion angles in secondary structure, and the lengths of **16** hydrogen and three disulphide bonds (Table **1).** Figure **2** shows four of these starting

TABLE 4

Convention used to name conformations

Name	Meaning
1st letter	
S	Starting
T	Torsion space energy minimized
C	Cartesian space energy minimized
A	Annealed by Cartesian space molecular dynamics and energy minimization
2nd letter	
X	X-ray co-ordinates (Deisenhofer & Steigemann, 1975)
i	The 25 starting conformations are labelled SO1 through S25 and were generated by using *i* as random number seed in the generation of initial torsion angles

conformations and the corresponding structures after torsion angle minimization. At the start of the minimization, the energy is very high as the pairs of atoms that are forced to come together to form the hydrogen and the disulphide bonds are initially very far apart. The use of soft-atoms allows the program to deal with these high strain energies and generate a roughly folded structure that satisfies the restraints after only 30 steps. Many more cycles of minimization (between 500 and 1500) are required, however, to improve the packing of the side-chains and converge to a true energy minimum.

Although the starting conformations are similar in character, the compact folded structures show a surprising range of energy and r.m.s. deviation values. The energies range from − 109 kcal/mol to 1008 kcal/mol and the r.m.s. deviations range from 3·5 Å to 6·0 A. A plot of the energy against the deviation (Fig. 4(a)) shows no obvious trend: the five conformations with lowest energies (T20, T18, T22, T17 and T03) have r.m.s. deviations of 4·9 A, 4·5 A, 3·7 A, 4·5 A and 3·7 A. Such a diversity of conformations was unexpected as each conformation has been minimized in the same way and fits the same set of constraints.

Next, each of the torsion angle minimized conformations (TO1 to T25) is subjected to further energy minimization with respect to the 1545 atomic Cartesian co-ordinates using the more complete energy function (eqn (6)). The total energies and r.m.s. deviations of the resulting minimum energy conformations (CO1 to C25) are shown in Figure 4(b) together with the corresponding properties of the X-ray (X) and minimized X-ray conformation (CX). Although there is still a large spread in energy and r.m.s. deviations, there is a clearer trend than after torsion angle minimization (Fig. 4(a)). In particular, the four conformations of lowest energy (C03, C17, C18 and C22) are also those that have low r.m.s. deviations from the X-ray co-ordinates. None of the 25 C*i* conformations has an energy that is as low as the energy of the X-ray minimum (CX); the lowest in energy (C18) has an energy value that is still 51 kcal/mol above that of CX.

There are nine conformations whose energies are less than 100 kcal/mol above the. energy of the CX conformation. The energy contributions and r.m.s. deviations of these conformations are listed in Table 5. The bond length, bond

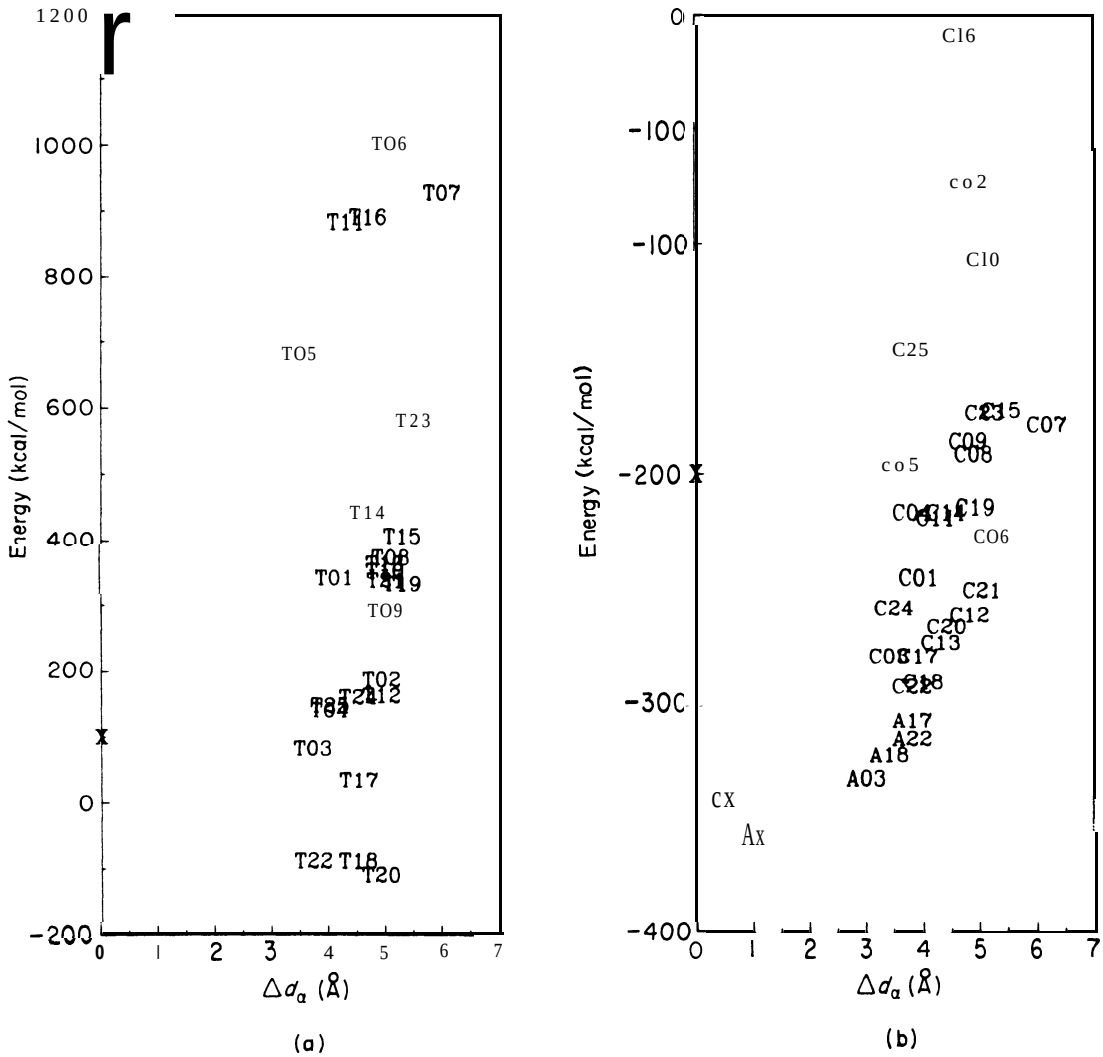


FIG. 4. Minimum energy values and r.m.s. a-carbon distance deviations of conformations generated here. (a) The 25 T conformations obtained by energy minimization with respect to the 208 single-bond torsion angles, and (b) the 26 C conformations obtained by energy minimization with respect, to the 1545 atomic Cartesian co-ordinates (including CX), together with the 5 conformations produced by dynamic annealing (AX., A03, A17, Al8 and A22).

angle and van der Waals' energy contributions show a much smaller variation than the torsion angle and hydrogen bond contributions. The van der Waals' and hydrogen bond energy contributions of conformations C03, C21 and C22 are almost as favourable as those in the X-ray minimum CX. Much of the difference between the energy of the X-ray minimum and the other minima comes from the torsion angle energy. This suggests that many torsion angles have been forced to unfavourable values to satisfy both the requirements of a close-packed interior and the 16 hydrogen bond restraints.

(ii) *Dynamic annealing*

The four lowest energy conformations (C03, C17, Cl8 and C22) are subjected to 30 picoseconds of molecular dynamics that are followed by a second pass of Cartesian space energy minimization to give annealed conformations A03, Al 7,

TABLE 5

Energy contributions of the nine lowest energy C conformations

Conformation	Energy contribution (kcal/mol) Total	Bond	Angle	Torsion	van der Waals'	H bond	r.m.s. deviation (A)
CX	-342	3	33	-27	-228	-123	0·5
c03	— 280	4	39	6	-221	-109	3·4
C12	-262	4	35	2	— 205	-99	4·8
C13	-274	4	40	7	-218'	108	4·3
C17	— 280	3	34	- 1	-218	-98	3·9
C18	-291	3	29	- 3	-216	-105	3·9
C20	-267	4	35	7	-202	-112	4·4
c21	-251	5	51	22	-222	— 107	5·0
c22	-293	5	43	10	— 229	-122	3·8
C24	-259	4	42	16	-211	-110	3·5

Al8 and A22. The CX conformation is also annealed in this way to give conformation AX. Table 6 gives the relative energy contributions and r.m.s. deviations of these ten conformations. In every case, annealing leads to a conformation of lower total energy, due mainly to more favourable hydrogen bonding and less torsion angle strain. The energy values of the folded conformation (A*i*) are, however, always greater than those of both the CX and AX conformation derived from X-ray co-ordinates. Annealing changes these four

TABLE 6

Relative energy contributions? before and after annealing

Conformation	Energy contribution (kcal/mol) Total	Bond	Angle	Torsion	van der Waals'	H bond	r.m.s. (Å) Δd_α^{prev}‡‡	Ad,"	Δr_A^X	R_g§
After annealing										
AX (absolute)	-359	3	33	-34	— 230	-132	0·9	1·0	0·8	10·9
A03	26	1	9	19	- 1	- 3	1·5	3·0	5·2	11·0
Al7	51	1	8	33	8	1	1·5	3·8	8·4	11·4
A18	36	1	4	16	4	11	1·6	3·4	7·5	11·4
A22	43	1	7	24	6	5	1·9	3·9	6·7	11·1
Before annealing										
CX	17	0	0	6	2	9	0·5	0·5	0·6	11·0
c03	79	1	6	40	9	23	0·9	3·4	5·6	11·4
C17	79	0	1	33	12	34	1·6	3·9	8·6	12·0
C18	68	0	- 4	31	14	27	1·5	3·9	8·3	12·1
c22	66	2	10	44	1	10	1·2	3·8	5·9	11·7

† The energy contributions of all conformations except AX are relative to those of the AX conformation.

‡ For the A series, prev refers to the corresponding C conformation e.g. for A03, Δd_α^{prev} is measured to C03. For the C series prev refers to the corresponding T conformation.

§ R_g is the radius of gyration calculated as $R_g = [1/N \sum (\mathbf{r}_i - \bar{\mathbf{r}})^2]^{1/2}$, where $\mathbf{r}_i$ is the ith Cartesian position vector, $\bar{\mathbf{r}}$ is the position of the molecular centroid and the summation extends over all non-hydrogen atoms. For the X-ray conformation, R_g = 10·96 Å.

conformations by 1·5 Å to 1·9 Å r.m.s. The change caused by the initial Cartesian co-ordinate minimization is comparable at 0·9 Å to 1·6 Å. Application of the same refinement techniques to the X-ray co-ordinates causes smaller changes of 0·9 Å and 0·5 Å, respectively.

For conformations A03 and A18, the decrease in total energy caused by dynamic annealing is also accompanied by a significant decrease in the r.m.s. deviation from the X-ray co-ordinates (3·4 Å to 3·0 Å and 3·9 A to 3·4 A). The energies of the annealed conformations increase monotonically with r.m.s. deviation (see Fig. 4(b)). This same trend is observed for those C conformations that have the lowest energy value for a particular value of the r.m.s. deviation (C22, C18, C13, C12, C21 and C07). To a first approximation the lowest energy that one can get for any A or C conformation varies quadratically with:

$$E_{\min} = -360 + 3{\cdot}9\,(\text{r.m.s.})^2. \qquad (17)$$

This indicates that the only way to get to a very low energy value is to become more similar to the X-ray co-ordinates. The existence of conformations at each r.m.s. value with higher energies can be explained by trapping in high energy local minima. The fact that the equilibrium bond length and bond angle values are taken from the X-ray structure (Levitt, 19833) does not cause a lower total energy at conformation X; all minimized conformations have very low bond and bond angle energies (see Table 6).

The five conformations that have the lowest energies after torsion space minimization (T03, T17, T18, T20 and T22) include four of the five conformations that, have the lowest energy after Cartesian space minimization (C03, C13, C17, C18 and *C22:* see Fig. 4). Applying Cartesian co-ordinate minimization only to those T conformations with low energies would save computer time and allow many more starting conformations to be generated. It was hoped that the energy value after a small number of steps of torsion angle minimization would also provide a selection criteria. Unfortunately, only one of the five conformations with lowest energy after 300 steps of torsion angle minimization is among the five best conformations after Cartesian co-ordinate minimization.

(b) *The diversity of conformations*

(i) *Comparing* co-ordinates

All the Cartesian space minimized conformations *(26* C and 5 A) are obtained using the same potential energy function. All these conformations are forced to have the 16 hydrogen bonds found in the native secondary structure. The diversity of these conformations is still considerable *as* evidenced by the range of r.m.s. deviations from the X-ray structure. This diversity is investigated further by calculating the r.m.s. deviation between all pairs of conformations. The closest pair of conformations (Cl7 and C18) are 2·3 A apart, while the furthest pair (CO2 and C07) are 8·5 Å apart

A two-dimensional representation of the conformational space is shown in Figure 5 for the 32 X, C and A conformations. In the two-dimensional representation, the distance between any pair of conformations is a measure of the

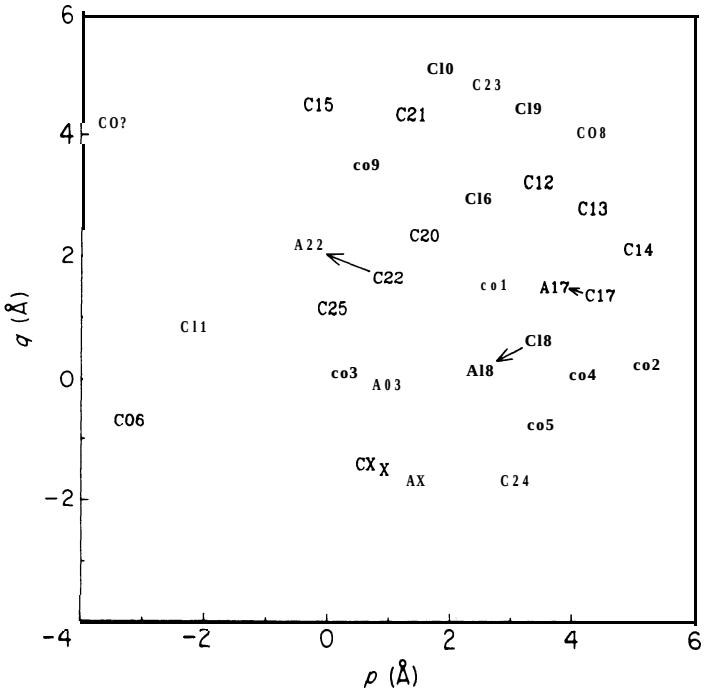


FIG. 5. A 2-dimensional representation of the 1545-dimensional conformational space containing X and the 30 C and A minimum energy conformations. In the projection, the distance on paper between any pair of conformations approximates the actual r.m.s. deviation (Δd_α^{ij}) between them (Levitt, 1983*c*). Notice how conformations C02, C06, CO7 and CO8 are least, similar to each other and map the conformational extremes obtained here.

actual r.m.s. deviation between that pair. All but three conformations cluster in a patch that is about 6 Å (r.m.s.) across. The native conformations X, CX, and AX are to one side of the patch and are close to it (3 Å from A03 to X). The mean r.m.s. deviation between any pair of C conformations is 4·3 Å. Conformations C02, C06, CO7 and CO8 are most different from one another and define the extremes of conformation generated by the constrained energy minimization (see Fig. 6).

It is of interest to estimate how many different C*i* conformations could be generated. Let us assume that any two structures with r.m.s. deviation less than 1 A are identical (this is the deviation between X and AX). The distribution of r.m.s. deviation for all pairs of C conformations is approximately a Gaussian distribution with a mean of 4·3 Å and a standard deviation of 1·2 A. This gives a probability of 0·01 that the r.m.s. deviation is less than 1 A. If a space containing n conformations is sampled by choosing m conformations completely at random then the probability that they are all different is:

$$P_{\text{diff}} = n(n-1)(n-2)\ldots(n-m)/n^m \approx \{(n-m/2)/n\}^m \approx 1-m^2/2n. \qquad (18)$$

The probability that at least two choices are identical is then simply $P_{\text{ident}} = 1-P_{\text{diff}}$. In the space of C conformations, P_{ident} is estimated above to be 0·01 and m is 25, giving:

$$n = m^2/2\,P_{\text{ident}} = 31{,}250. \qquad (19)$$

Thus, we estimate there are about 30,000 different minimum energy conformations

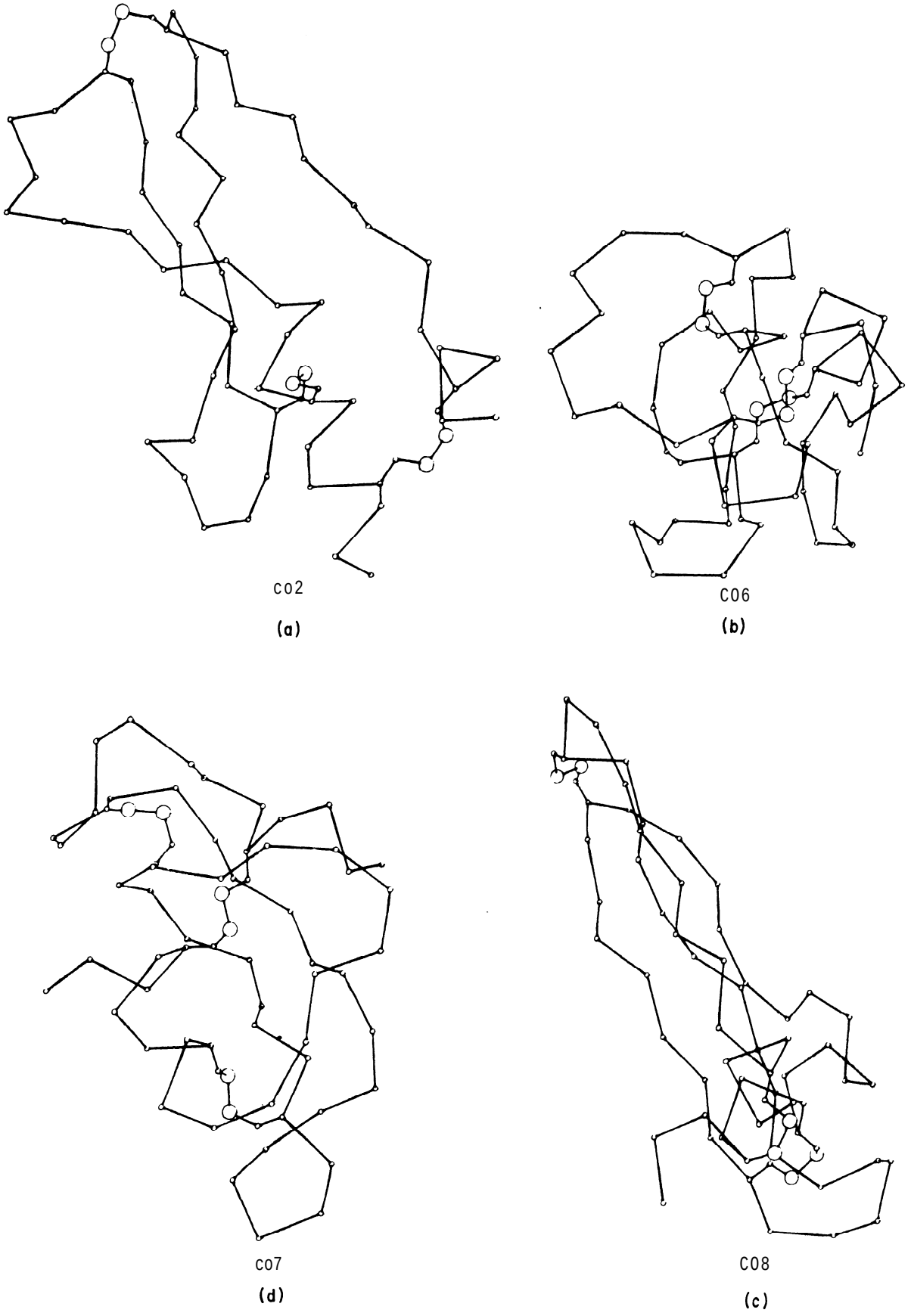


FIG. 6. Drawing of the chain path in the 4 most, different conformations (a) C02, (b) C06, (c) CO7 and (d) C08 that define the extremes of the conformation space shown in Fig. 5. Note how CO6 is almost spherical, C08 is long and thin, CO2 is expanded and CO7 has the 3 S-S bridges distributed evenly over the volume of the molecule.

that can be generated by the present method of constrained energy minimization. While this is a large number, it is very much smaller than the astronomically large number of possible conformations of the BPTI polypeptide chain.

(ii) *Comparing chain threading*

In native BPTI there is an unusual threading with the chain segment consisting of residues 1 to 30 actually passing through the loop formed by the disulphide bond between cystine 30 and 51. None of the BPTI conformations generated by previous studies shows this native threading in published stereo drawings (Levitt & Warshel, 1975; Levitt, 1976; Kuntz *et al.*, 1976,1979; Hagler & Honig, 1978; Goel & Yčas, 1979; Robson & Osguthorpe, 1979). The r.m.s. deviation used above to distinguish folded conformations is not very sensitive to the chain topology or threading. Two structures that seem very similar as measured by the r.m.s. deviation may actually have different chain threadings. Threading can be seen by looking at a stereo drawing of the chain fold (Fig. 7); it can also be detected by using the writhing number of the chain. The plot of energy against writhing number for the C conformations (see Fig. 8) shows that W varies from $-1{\cdot}3$ turns to $3{\cdot}1$ turns. Most structures have $W = 2$ turns, which is significantly different from the value of the X-ray conformation ($W = 0{\cdot}2$). Four conformations (C03, C18, C20 and C22) have $W < 0$ and all these also have low energy values. Inspection of the stereo drawings of these conformations shows that they all have the correct threading, of residues 1 to 25 through the 30-51 loop. This same threading is also found in CO2 and C25, which have W values of $0{\cdot}4$ and $0{\cdot}6$ turns, respectively, but all 19 other conformations, which have W greater than 1 turn, are not threaded. Clearly, the writhing number is a reliable measure of the chain threading and can be used to classify chain folds. Conformations Cl7 and C18, which are the closest pair of C conformations (r.m.s. deviation of $2{\cdot}3$ A, see Fig. 5), have $W = 2{\cdot}1$ and $W = -0{\cdot}2$ turns, respectively: in order to get from conformation Cl7 to Cl8 along the shortest path, the chain would have to pass through itself near the 14 : 38 disulphide bond (see Fig. 7(c) and (d)).

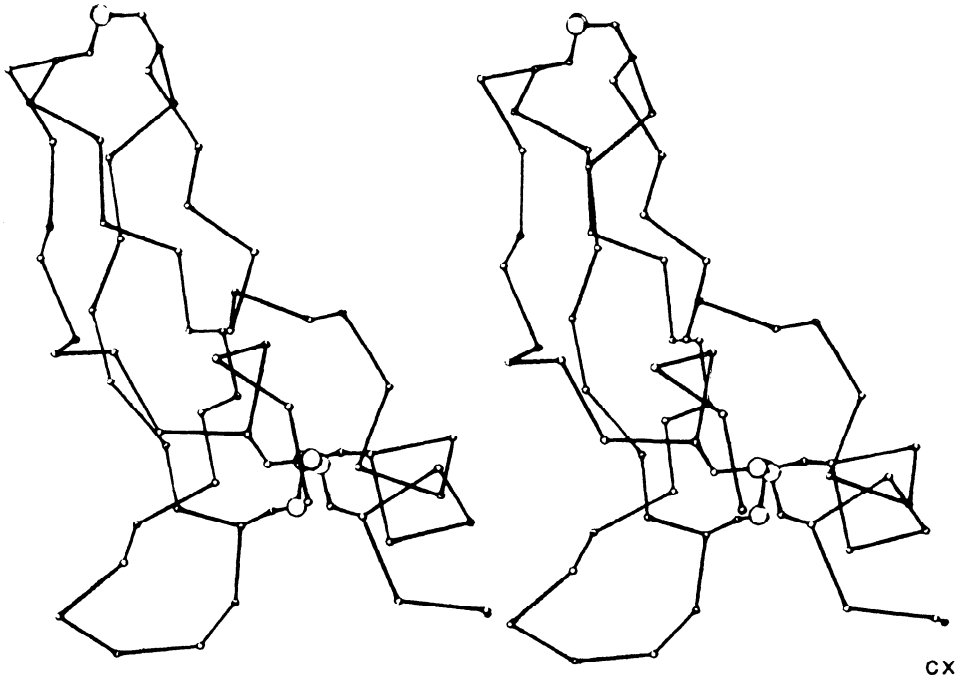
(c) Analysis of the lowest energy conformations

In this section we focus attention on the four C conformations with lowest energies (C03, Cl 7, Cl8 and C22), the corresponding annealed conformations (A03, A17, Al8 and A22) and the X-ray conformations CX and AX.

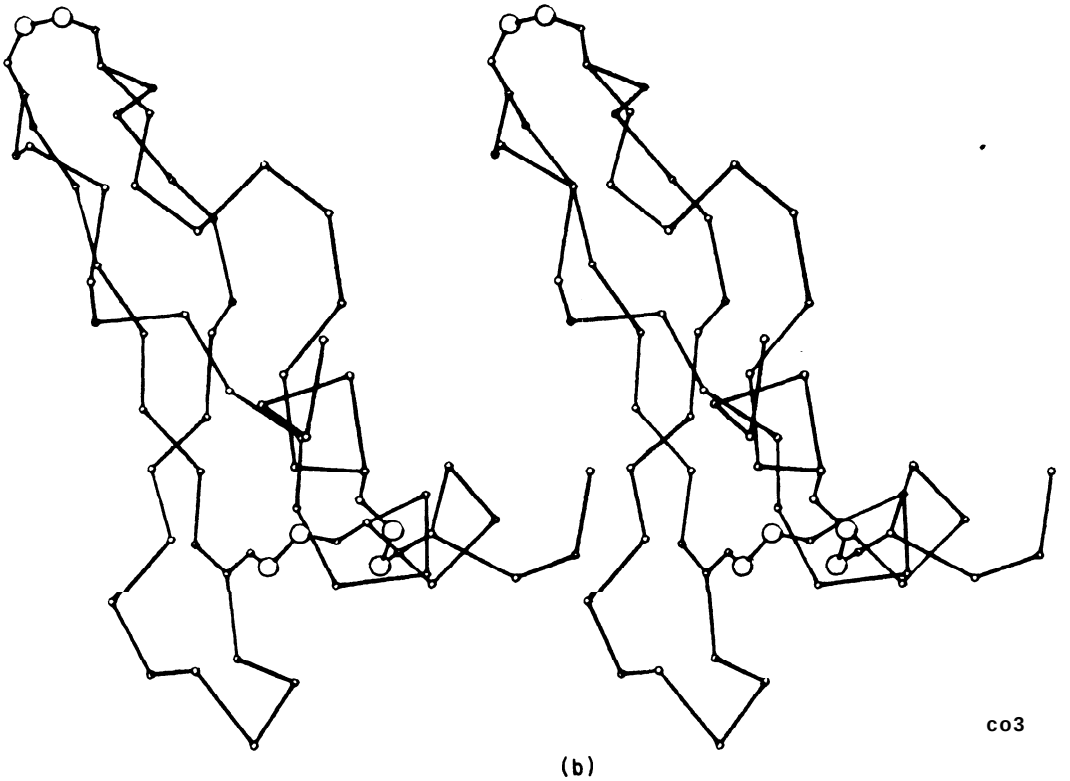
(i) *Energetics*

The four annealed conformations are extremely well-stabilized. Their total energies are between 26 and 51 kcal/mol above that of AX (Table 6), with most of this difference in the torsion angle energy term. The van der Waals' and hydrogen bond energies of A03 are actually more favourable than in AX, the annealed X-ray conformation.

In view of the relatively high torsion angle strain, the distribution of the (ϕ,ψ) angles was examined and compared with those of the X-ray conformation (Fig. 9). All the distributions have similar clusters of points about the cc-helical and β-sheet



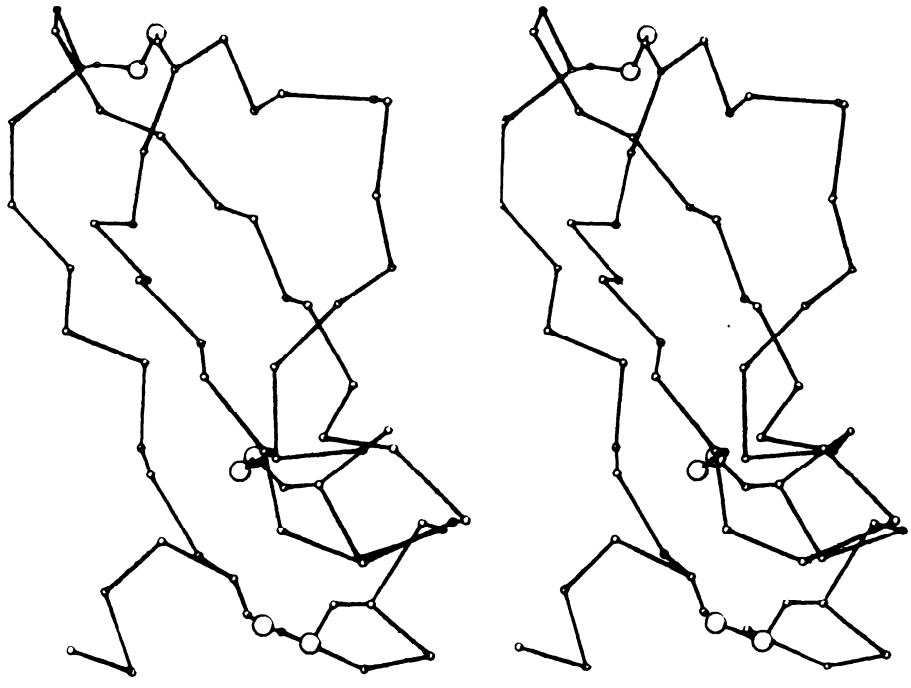
(a)



(b)

FIG. 7.

regions. In the folded conformations A03, A17, Al8 and A22 there are many more non-glycine residues that fall outside these regions. Those conformations with more of the abnormal conformations (Al7 and A22) have higher torsion angle energies (see Table 6).



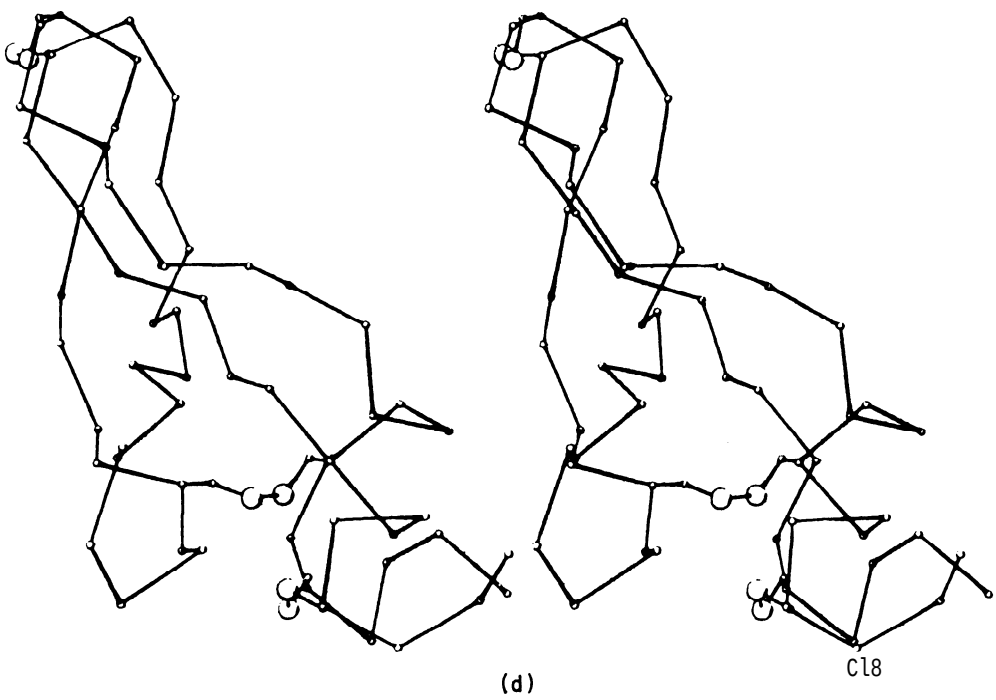


FIG. 7.

(ii) ***Chain folding***

The chain fold in the A conformations is shown in the stereo drawings of Figure 7 and the schematic drawings of Figure 10. Although the folded A conformations have the native secondary structure hydrogen bonds. are of low

energy and very similar to the X-ray conformation (r.m.s. $< 3{\cdot}8$ A), they still show a surprisingly large diversity of chain foldings. Conformations A03 and **A18** have the native threading, **Al7** is unthreaded and A22 has a double threading. The native conformation, AX, has a more regular chain fold in that the chain segments between the turns are less bent or kinked (see Fig. 7).

(iii) ***Hydrogen bonding***

All the conformations generated here must have the 16 main-chain hydrogen bonds enforced by the restraints. Each conformation has other hydrogen bonds that arise as a natural consequence of the strongly attractive 0 . . . HN interaction (eqns **(1)** and ***(6)).*** The total number of hydrogen bonds in the five A conformations (see Table 7) varies from **31** in AX to 39 in A03; all the A*i* conformations have more hydrogen bonds than AX. In spite of the large number of hydrogen bonds formed in each A conformation, there are few common hydrogen bonds. A collated list of the 36 variable hydrogen bonds, which are not formed in all the A conformations, shows that most (27) are formed in only one of the conformations. Thus, although each conformation has the same set of **16** restrained hydrogen bonds, most of the additional hydrogen bonds (between 6 and 15 in number) are unique to that **conformation.**

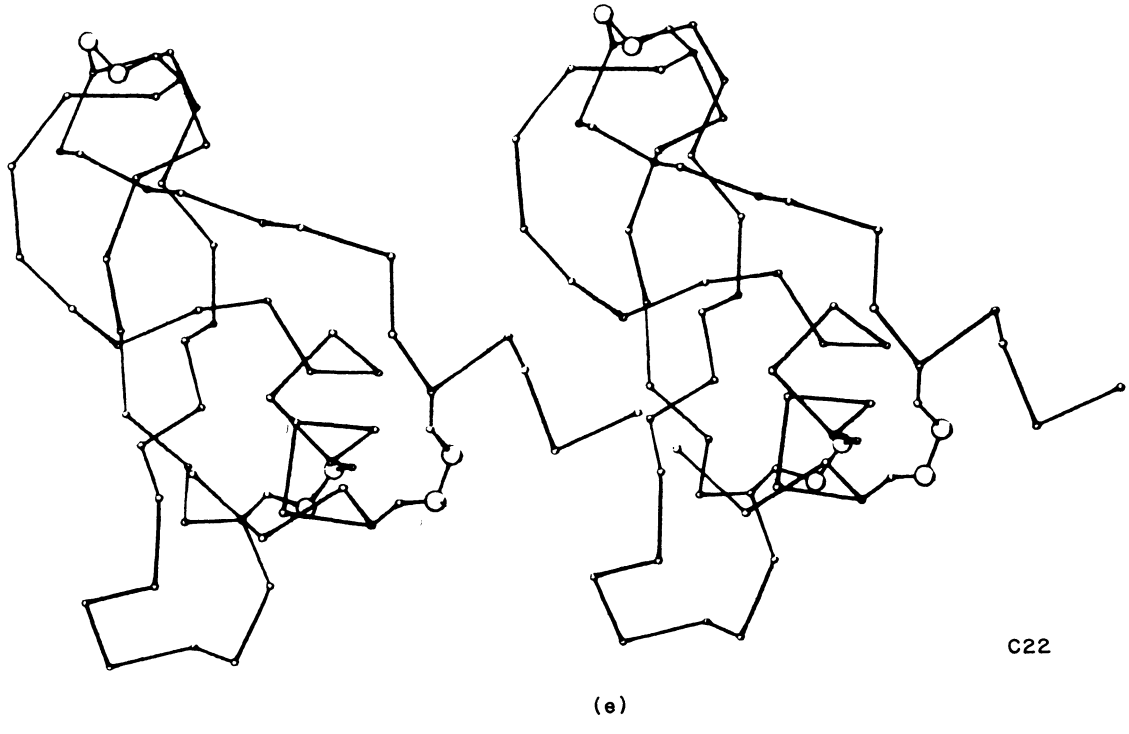


FIG. 7. Stereoscopic drawings showing the main-chain fold and disulphide bridges in the 5 lowest energy C conformations: (a) CX, (b) C03, (c) C17, (d) Cl8 and (e) C22. The a-carbons are joined by virtual bonds and the atoms of the 6 cysteine residues are included to show the 3 disulphide bonds between residues 5 : 55, 14 : 38 and 30 : 51.

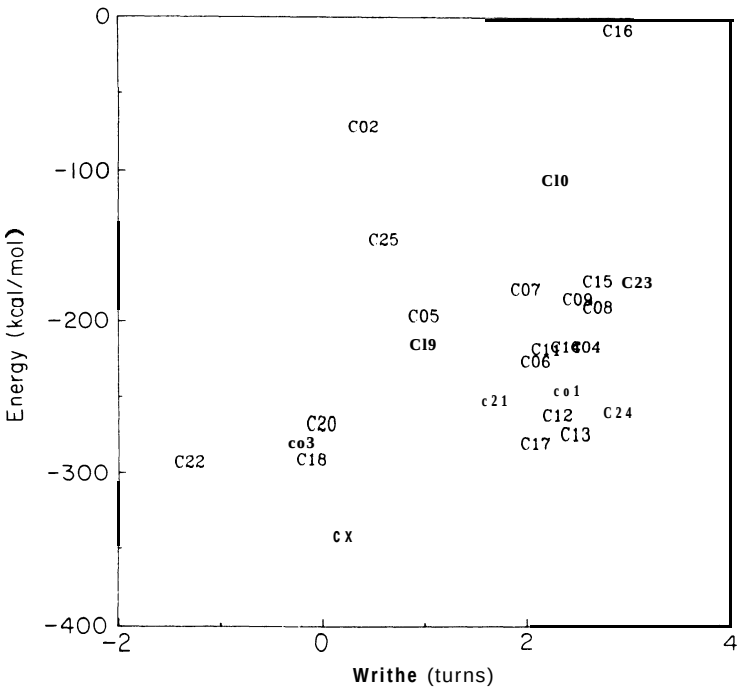


FIG. 8. The minimum energy values and writhing numbers (W) of the native conformation (X) and the 25 Cartesian minimized conformations (C). All those conformations with W values less than 0 are threaded in the same way as the X-ray structure (see Fig. 10) and also have low energy values (conformations C22, Cl8 and C03). Some of the wrongly threaded conformations with W greater than 1·0 also have low energy values (conformations Cl3 and C17).

Closer analysis of the location of these variable hydrogen bonds shows that most **(25)** occur in four regions of the molecule: H1, the N-terminal 3_{10}-helix (residues 1 to 5); B1, the open end of the P-hairpin (residues 10 to 15 pairing with residues **34** to 39); B2, the closed end of the P-hairpin (residues **24** to **29): H2,** the C-terminal a-helix (residues 46 to 58). The numbers of bonds in each region are two in H 1, ten in B1, six in **B2,** and seven in H2 (11 are in other regions). The variable hydrogen bonds in the two helical regions result in the elongation of the 3_{10}-helix **(1, 0..** .4, N), the elongation of the a-helix (46, 0. . .50, N and 53, O . . .57, N), and the addition of 3_{10} and 5_{16} bonds to the a-helix (47, O . .. **50,** N; 52, O . . . **57,** N and 53, O . . .58, N). The variable hydrogen bonds in the B-hairpin are more irregular. None results in a regular extension of the hairpin (such hydrogen bonds would be **14,** H. . .39, O; 14,0. . .39, H; **16,** H.. .37, O; **16,0.** . .37, H; 24, O . . .29, H; **26,** H. . .27, O and 26, O . . .27, H). Instead there are other pairings of residues in these regions. In the native conformation (X) the extension of the b-hairpin is also irregular.

(iv) Accessible surface area

The accessible surface area of atoms, residues and the entire molecule provide additional means of describing the conformation with particular regard to the interaction with the surrounding solvent (see Table 8). Each of the calculated A*i*

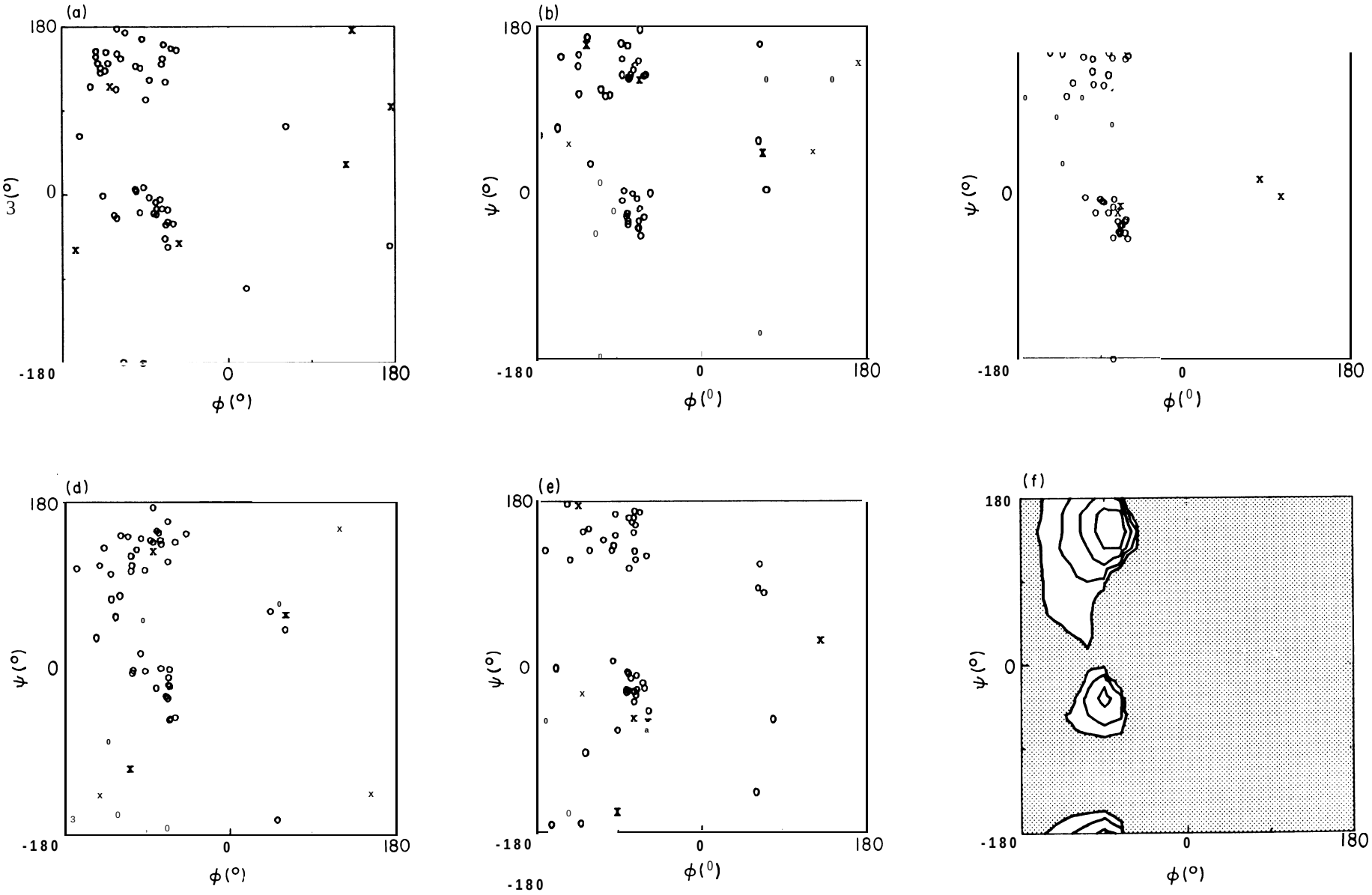


FIG. 9. The distribution of the (ϕ, ψ) angles of the X-ray and 4 annealed conformations (a) A03, (b) A17, (c) X, (d) A18 and (e) A22. The (ϕ, ψ) energy contours at 1 kcal/mol intervals are shown in (f) for alanine dipeptide (CCONHC(C)CONHC), calculated with the energy function used for the Cartesian minimization.

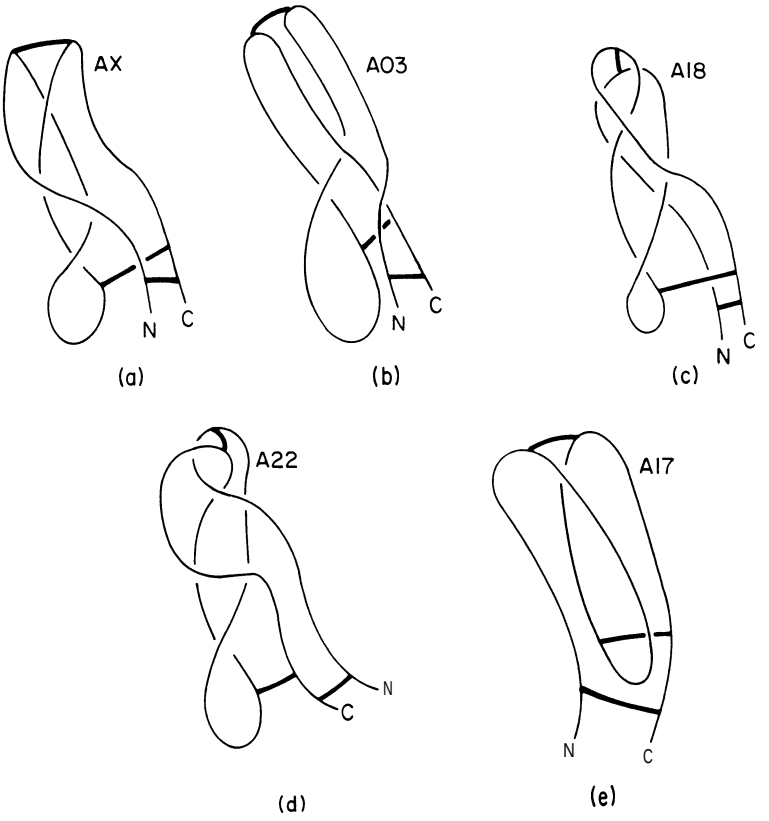


FIG. 10. Schematic drawing of the chain path in the 5 annealed A conformations: (a) AX, (b) A03, (c) A18, (d) A22 and (e) A17. The total energies of these conformations are -359, -333, -3323, -316 and −308 kcal/mol, respectively, and the r.m.s. deviations from X are 1·0, 3·0, 3·4, 3·9 and 3·8 A, respectively. The writhing numbers are 0·2, 0, 0, − 1·1 and 2·3 turns, respectively. The loop between residues 30 and 51 is threaded once in (a), (b) and (c), twice in (d) and not at all in (e), explaining the trend seen in the writhing numbers.

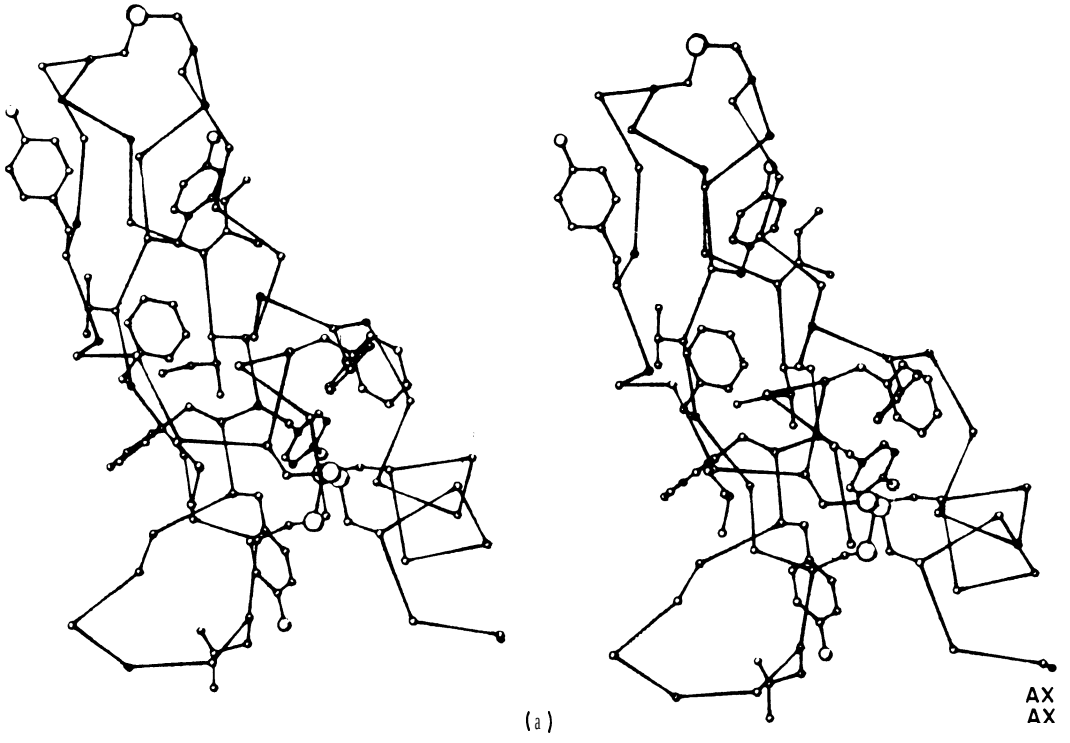
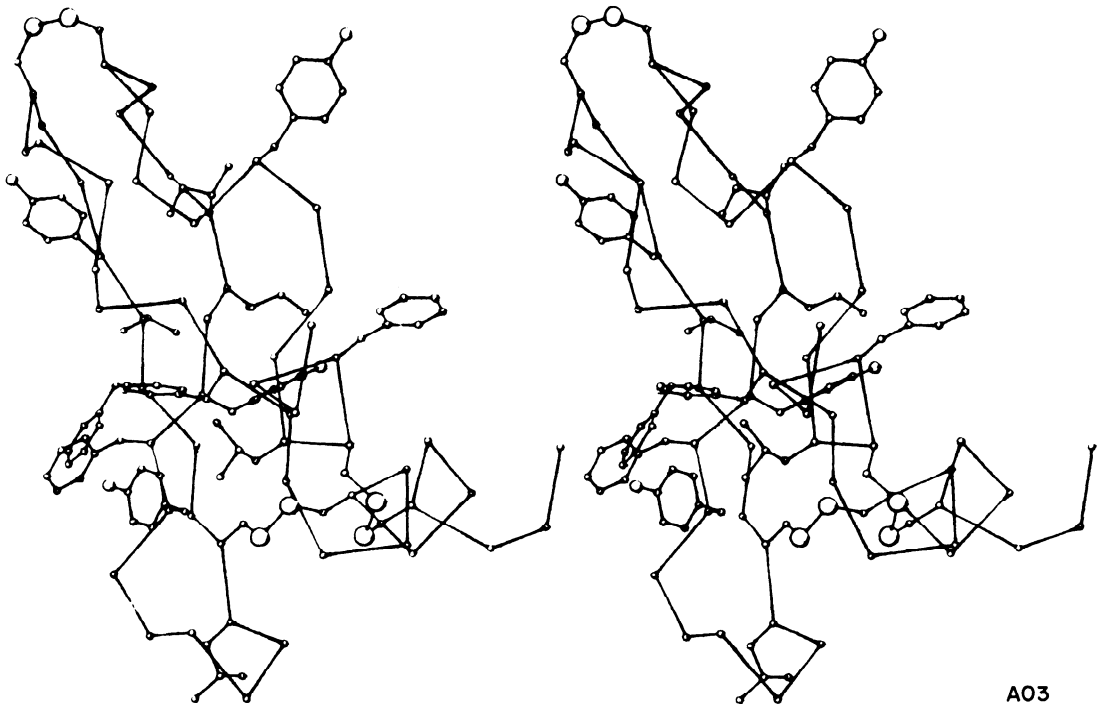
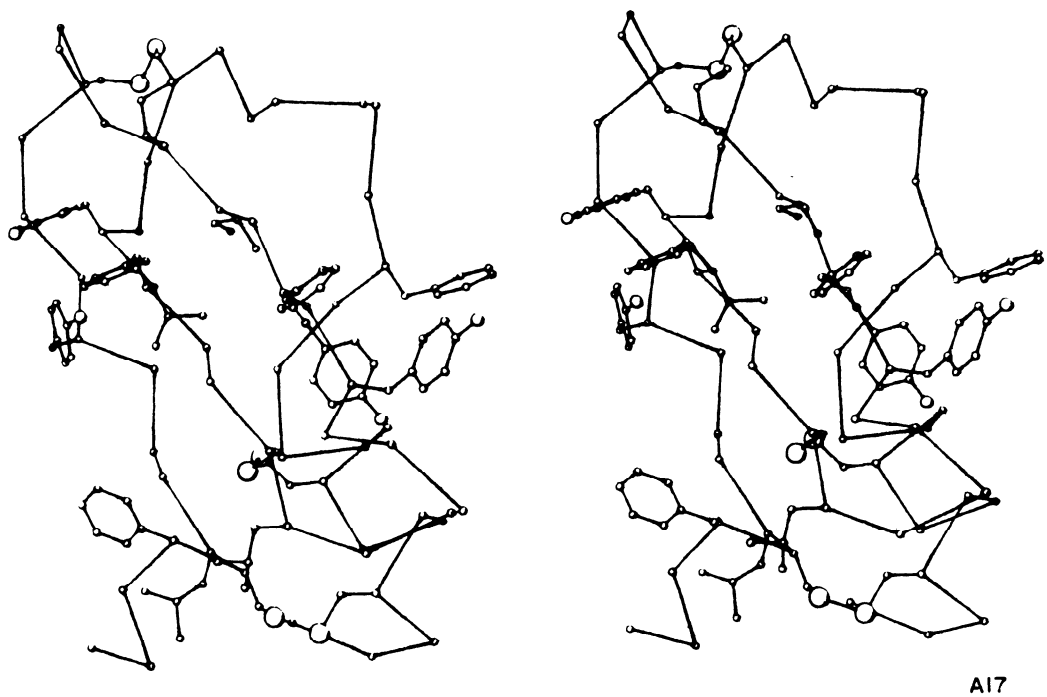


FIG. 11.



(b)



(c)

FIG. 11.

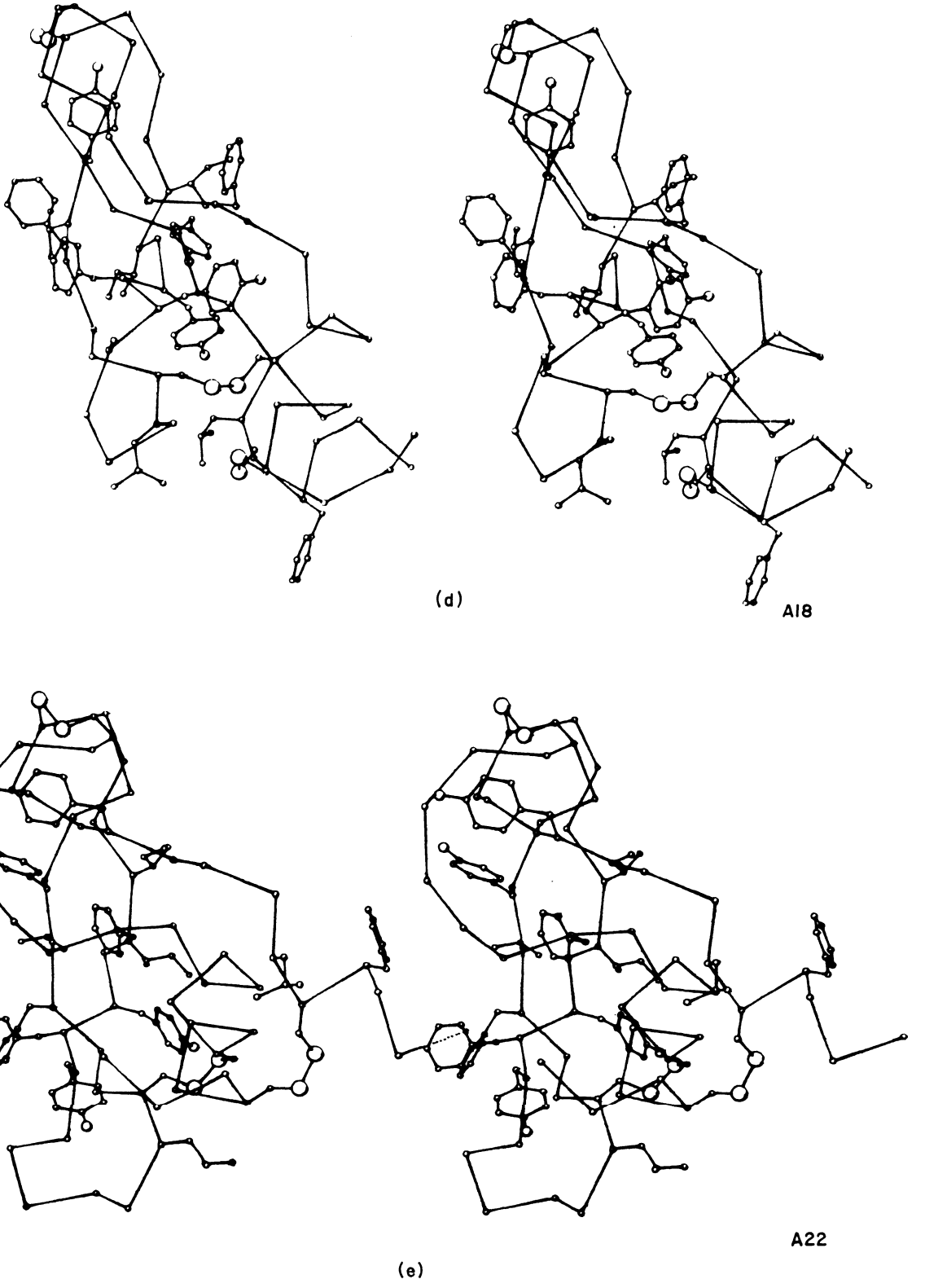


FIG. 11. Stereoscopic drawings showing the main-chain, the disulphide bridges and the large hydrophobic side-chains (Val, Leu, Ile, Met, Phe and Tyr) in the 5 annealed conformations (a) AX, (b) A03, (c) A17, (d) A18 and (e) A22. Interior packing of hydrophobic groups is clearly much better in AX than in the other A conformations.

TABLE 7
Variable *peptide hydrogen bonds*

Property		Conformation AX	A03	Al7	Al8	A22
Number of						
Main/main		22	31	26	23	26
Main/side		7	6	10	9	12
Side/side		2	2	1	2	0
Main/main variable†						
Thrll, 0.. .Gly36, N	(B1)	X				
Gly36, 0. . . Ala16. N	(B1)	X				
Asn24, 0. . . Ala27 N	(B2)	X				
Asn24, 0 . . . Gly28. N	(B2)	X		X		
Arg42, 0. . . Asn44, N		X		X		
Met52 0. . . Gly56, N	(H2)	X				X
Asp3. 0. . . Glu7, N	(H1)		X			
Met52, 0. . . Gly57, N	(H2)		X	X		
Argl, 0.. .Phe4, N	(H1)		X	X	X	x
Leu6, 0. . . Lys46, N			X			
Thrll, 0. . . Asn44, N			X			
Thrll, 0.. .Arg42, N			X			
Pro13, 0. . . Arg39. N	(B1)		X			
Lys15, O . . . Gly37, N	(B1)		X			
Leu29, 0. . . Ala25, N	(B2)		X			
Ala25, 0. . . Leu29, N	(B2)		X			
Ala25, 0. . . Gly28. N	(B2)		X			
Cys30, 0. . . Ala48, N			X			
Ser47, 0. . . Asp50, N	(H2)		X			
Arg53, 0. . . Gly56, N	(H2)		X	X	X	
Arg53, 0. . . Ala58, N	(H2)		X			X
Phe4, 0. . . Cys30, N				X		
Va134, 0. . . Tyr10, N	(B1)			X		
Tyr10, 0. . . Gly36, N	(B1)			X		
Gly36, 0. . . Gly12, N	(B1)			X		
Lys46, 0. . . Asp50, N	(H2)			X		X
Lys41, 0. . . Asn44, N					X	
Pro9, 0. . . Ile19, N					X	
Arg17, 0.. .Thr11, N					X	
Asn43, O . . . Tyr23, N					X	
Arg53, 0. . . Glv57, N	(H2)				X	x
Thrll, 0.. .Lys41, N						X
Cys38, 0 . . . Glv12, N	(B1)					x
Cys14, 0. . . Gly37, N	(B1)					X
Cys14, 0. . . Cys38, N	(B1)					X
Ala25, 0. . . Ala27, N	(B2)					X

† The variable hydrogen bonds listed include all those peptide hydrogen bonds that are formed in *any* of the 5 conformations. The conserved hydrogen bonds are formed in *all* 5 conformations and consist of the 16 that are included as restraints (see Table 1). The variable hydrogen bonds are classified into four classes given in parenthesis after the bond name (see section 4(c)(iii)).

TABLE 8

***Accessible** surface areas† of the X and annealed conformations*

Residue or property	Conformation X	A03	A17	A18	A22
Argl	150·4‡		53	39	83
Pro2	43·6	74	62		67
Asp3	112·2	-48	-28	-51	
Phe4	37·9	67	36	119	74
Cys5	0·0	43		33	
Leu6	99·6	-56	-31		
Glu7	38·3			29	40
Pro8	82·7	-62		-47	
Pro9	52·3	22	24		-29
Tyrl0	75·6	85		109	
Thrll	69·1	-57	50		-66
Gly12	21·5		22		-21
Pro13	86·4	-53			-21
Cysl4	51·2		27	24	
Lysl5	174·9	-61			
Ala16	50·3	31	26		25
Arg17	202·4		-154	-39	-51
Ile18	59·9			-22	
Ile19	113·1		-70	-83	
Arg20	29·6		54		
Tyr2 1	60·1		-52	-54	
Phe22	20·0	61	106	30	37
Tyr23	12·3	65	94		65
Asn24	31·8	49			34
Ala25	50·8			36	
Lys26	180·4				
Ala27	54·0	51			
Gly28	38·8				
Leu29	78·6	29			22
Cys30	23·2		-22		
Gln31	67·9				-41
Thr32	89·9		-85		
Phe33	21·8	62	69	72	72
Val34	81·8	43	-81		
Tyr35	2·9	32	74		
Giy36	0·5	47			
Gly37	39·7				-23
Cys38	49·5	39			-30
Arg39	174·0	-93		33	
Ala40	72·1		-49		-37
Lys41	57·1	55		74	75
Arg42	131·8		62	28	
Asn43	0·8		136	44	85
Asn44	15·2		22		
Phe45	39·0		84	62	
Lys46	145·5	-114	-33	-46	-66
Ser47	23·6			22	
Ala48	29·7				24
Glu49	123·0		-47		-100
Asp50	55-8	-21	-40	-25	-54
Cys51	0·0				38
Met52	62·0	30	47		
Arg53	137·9		28		-92

TABLE 8 *(continued)*

Residue or property	X	Conformation A03	A17	A18	A22
Thr54	48·5	-2 9	-2 5		-2 9
Cys55	0·3	69	43	37	68
Gly56	56·6				
Gly57	63·4		-3 8	-2 5	-2 1
Ala58	89·1			43	
Total area	3779	4098	4171	4206	3862
Non-polar§	975	1528	1423	1252	1368
Polar	1829	1618	1910	1904	1701
Energy‖	0·0	18·2	8·7	4·8	12·4

† The accessible area is in Å^2 and is calculated by the Lee-Richards' method (Lee & Richards, 197 1) using atomic radii given in section 3(b)(iii).

‡ The accessible area of each residue is listed for the X conformation only; for the other conformations the change in area is given if it exceeds 20 Å^2 in magnitude.

§ The non-polar residues are Val, Leu, Ile, Cys, Met, Phe, Tyr and Trp: the polar residues are Asn, Asp, Gln, Glu, Lys and Arg.

‖ The energy is the sum of the change in surface area multiplied by O-024 kcal/mol per Å^2 for non-polar residues and − 0·024 kcal/mol per Å^2 for polar residues (Chothia. 1976).

conformations has more solvent accessible surface than does the X-ray conformation (X). Although this increase is small (between 2% and 11%), there is a much bigger increase in the exposed areas of the non-polar side-chains (between 28% and 57%). Inspection of the changes in surface areas of individual residues shows that some side-chains are more buried or exposed relative to X in *more* than one of the A*i* conformations. Residues that are consistently more exposed include Argl, Pro2, Phe4, Ala16, Phe22, Tyr23, Phe33, Lys41, Asn43 and Cys55. This list includes four of the eight aromatic residues in BPTI. Residues that are more buried include Asp3, Arg17, Lys46, Asp50 and Gly57. This list includes four charged residues. Most of the residues that are more buried than in X occur in the residue ranges 17 to 21 and 46 to 50, whereas those residues that are more exposed occur mostly in the ranges 1 to 4, **22** to **23,** 33 to 35 and **41** to **45.** The much better interior packing of hydrophobic groups in the X-ray structure is shown stereoscopically in Figure 11.

5. Discussion

(a) *Performance of the methods*

(i) *Generation of conformations*

Generation of folded conformations by torsion angle energy minimization with soft-atoms and restraints has worked well. This new method produces a set of atomic Cartesian co-ordinates that has good stereochemistry, few close contacts and obeys the restraints. The energy of these conformations (the Ti conformations) can be used as a selection criterion in that low energy conformations are more likely to be similar to the X-ray conformation. The

method can be used for any restraint that can be expressed as a function of the atomic co-ordinates.

The conformations generated from different randomized sets of initial torsion angles show great diversity although all satisfy the same restraints on S-S bonds, main-chain hydrogen bonds and secondary structure. With such diversity there is a good chance of getting conformations that are like the X-ray structure. This same diversity was found when the method of distance geometry was used to compute a-carbon co-ordinates for BPTI (Have1 et *al.*, 1979). When they used S-S bonds and secondary structure restraints, the r.m.s. deviation (Δd_α^x) from the X-ray conformation was $5{\cdot}6 \pm 0{\cdot}4$ A. Here the corresponding deviation is $4{\cdot}5 \pm 0{\cdot}7$ Å, indicating that the present set of conformations may be more diverse than would have been obtained with the distance geometry method. Other important differences between the two methods include: (1) distance geometry does not provide co-ordinates for all the atoms unless very large matrices are used; (2) distance geometry does not provide an energy value that can be used as a selection criterion; (3) distance geometry can only deal with restraints that are expressed as distances.

(ii) *Refinement of conformations*

The conformations generated by soft-atom restrained energy minimization in torsion angle space can be regarded as starting structures for a more complete refinement aimed at getting conformations of very low potential energy. Here this is done by a combination of energy minimization and molecular dynamics in which all atomic Cartesian co-ordinates are taken as degrees of freedom. This combination provides a powerful tool for annealing conformations, significantly reducing the potential energy and causing changes of conformation of $2{\cdot}1$ Å to $2{\cdot}8$ Å (r.m.s. deviation from the corresponding T conformation). Because the r.m.s. differences between the different T conformations are larger than this change ($4{\cdot}1 \pm 1{\cdot}2$ Å) the refinement does not actually change the overall fold of the starting conformation significantly.

It would have been possible to eliminate torsion angle minimization and start the Cartesian co-ordinate energy minimization from the open random conformations (see Fig. 2). The major advantage of torsion angles is that with only 208 variables, the rapidly convergent variable metric minimizer VA09D can be used. In a previous study it was concluded that the conformational freedom of a protein is greatly restricted with torsion angle co-ordinates (van Gunsteren & Karplus, 1980). We have found that the r.m.s. deviation caused by minimization is comparable with Cartesian and torsion angle co-ordinates provided minimization is continued to convergence (5000 steps of conjugate gradients or 500 steps of variable metric minimization); use of 200 steps of conjugate gradients (van Gunsteren & Karplus, 1980) is simply not sufficient.

(iii) Energy *as a selection criterion*

A most important result obtained here is that the conformations with the lowest energy are also closest to the X-ray conformation (X). Refinement that

lowers the energy also brings the conformation closer to X. The energy of the annealed X-ray conformation, AX, is still lower than that of any of the other annealed conformations, A2, indicating that it is valid to search for the conformation with the lowest possible potential energy value. For any particular value of the r.m.s. deviation there seems to be a lowest value of the minimum potential energy: it is impossible to get a lower value without becoming more similar to the X-ray conformation.

The r.m.s. deviation from X of the lowest energy conformation, A03, is 3 Å, which is significantly better than the best values obtained in other studies (3·8 Å to 6 A). Getting such a low deviation depends critically on the number and position of the restraints used here; without the restraints it would have been much more time-consuming to have found a conformation as close to the X-ray structure. The method can, therefore, only be used on a protein of unknown conformation when either a comparable set of restraint distances are available (perhaps from high-resolution nuclear magnetic resonance; see Wüthrich *et al.*, 1982) or computers become much cheaper and faster. The restraints could in principle be furnished by methods that, for example, predict the regions of secondary structure and pairing of P-strands. Several alternative predictions could be tested as the conformation with the lowest energy value would always be expected to be closest to the X-ray structure. Thus, the method provides the basis for a very general scheme for predicting folded conformations of protein molecules.

A careful analysis of the A03 conformation (including building a Labquip model) showed that it is still easily distinguishable from a real native conformation. The most obvious defects include less regular backbone torsion angles and insufficiently buried aromatic side-chains. Improving the potential energy function to eliminate these defects is relatively straightforward and is expected to make the energy value an even better selector of native conformation.

(iv) *Chain threading and the writhing number*

BPTT has an unusual chain fold in that part of the chain is threaded through a loop formed by another part. Conformations generated by previous studies of BPTI (Levitt & Warshel, **1975;** Levitt, **1976;** Kuntz *et al.*, 1976,1979; Hagler & Honig, **1978;** Goel & Yčas, **1979;** Robson & Osguthorpe, **1979)** do not show this threading, leading to criticism of the methods (Hagler & Honig, **1978).** Six of the **25** Ci conformations generated here do have the correct threading, and three of these (C03, **Cl8** and C22) also have very low energies. The use of soft-atoms during the torsion angle minimization allows the chain to pass through itself and may be responsible for the occurrence of a variety of chain threadings.

In the present calculation, the three S-S bridges act as strong restraints that force the chain into a compact conformation; threading occurs as the soft-atoms allow the main-chain to pass through itself. This is definitely not the way the chain really folds, as in nature the S-S bridge is a short-range bonding interaction and the chain cannot pass through itself. For the real refolding of BPTI the three native S-S bridges are formed in a definite sequence (Creighton, **1978).** Because other non-native S-S bridges also seem to be obligatory during the BPTI

folding pathway, it is not clear whether the results of the calculation could be improved by activating the S-S bridge restraints at different stages of refolding.

The present concern with chain threading lead to the use of the writhing number (W) as a quantitative measure of threading. The novel method used here to calculate W from the a-carbon co-ordinates is simple and shows the writhing of a chain to be an additive property of pairs of line segments. The writhing number distinguishes the different chain threadings directly and provides an automatic way to classify different chain foldings. It is expected that the formula for W presented here will be used to analyse known protein conformation as has already been done with less quantitative methods (Connolly et *al.*, 1980).

(b) *Limitations of the methods*

(i) *Solvent interactions*

The many thousands of water molecules that surround proteins in solution are not treated explicitly, and the effect of solvent interactions is not included in the potential energy function used here. This is done here as a deliberate check on the adequacy of the simpler *in vacuo* potential. Although the simple potential works surprisingly well and the energy value provides a useful indicator of native conformation, the omission of solvent interactions has a noticeable effect on the calculated conformations. More specifically, in the Ai conformations most aromatic residues are insufficiently buried while many charged residues are insufficiently exposed. Both these trends would be opposed by solvent effects in which hydrophobic side-chains are repelled by the solvent, whereas charged side-chains are attracted to it.

Now that this defect of the potential has been identified, it can be remedied by using the Ai conformations themselves. Inclusion of a realistic solvent effect should raise the energy of the Ai conformations relative to that of AX. A simple solvent effect can be obtained by increasing the attractive van der Waals' interaction between the atoms in the large non-polar side-chains (Val, Leu, Ile, Phe, Tyr and Trp) and decreasing this attraction between atoms in polar side-chains (Asp, Asn, Glu, Gln, Lys and Arg) and all other atoms. A more realistic solvent effect may be derived from the potential of mean force calculated from a molecular dynamics simulation of BPTI surrounded by 1850 water molecules (Levitt & Sharon, unpublished work).

(ii) *Computational requirements*

The method used here requires large amounts of computer time. Generation of a single low energy annealed conformation involves soft-atom minimization in torsion angle space (40 min of IBM 370/165 central processing unit time), Cartesian space energy minimization (40 min), molecular dynamics (180 min) and finally more Cartesian energy minimization (40 min). In practice many conformations will have to be generated before the best can be selected on the basis of having the lowest energy value. For BPTI, the present results show that as many as 30,000 starting conformations would be needed to get within 1 Å (r.m.s. deviation) of the native conformation (even when constraints are used).

Certain characteristics of the present scheme can be used to reduce the required computer time. (1) Preselection, in that not all starting conformations need to be refined by Cartesian co-ordinate energy minimization and molecular dynamics. In the present study, it would have sufficed to refine only the five Ti conformations with the lowest energy. Candidates for further refinement could also be selected using the writhing number to classify the starting conformations. (2) Less annealing, in that molecular dynamics need not be continued for 15,000 steps at room temperature. A shorter run at a higher temperature may work equally well at getting to a lower energy minimum. (3) Parallel computation, in that a collection of n independent microprocessors could generate n folded conformations at the same time. This level of parallelism is trivial to implement and only depends on the availability of sufficiently powerful cheap components. It should be noted that the five-hour calculation on the rather old-fashioned IBM 370/165 would take about ten minutes on a Cray 1 or Cyber 205 supercomputer.

(iii) *Is energy minimization a valid approach?*

The present study rests on the working assumption that the native conformation of a globular protein will have a lower value of the potential energy than all other conformations. Proving the validity of the assumption by computation would be very difficult as all energy minima would have to be searched. The present results are encouraging, however, as none of the calculated BPTT conformations has an energy that is lower than that of the annealed X-ray conformation. At the present time we see no workable alternative to the assumption; if globular proteins have metastable conformations that are determined by kinetic factors (Levinthal, 1968), the protein folding problem will be much more difficult. Fortunately, at present, experimental evidence supports the idea that the native conformation is thermodynamically most stable (Baldwin & Creighton, 1980).

(c) *Future applications*

Soft-atom restrained energy minimization introduced here can be regarded as a tool that builds macromolecular conformations having very low energies and obeying arbitrary restraints. The method is extremely robust and has some obvious applications as follows.

(1) Building an unknown protein conformation from a known one. Methods are presently being developed to use the known conformation to detect sequence homology, align the two amino acid sequences and provide starting co-ordinates for atoms in common between the structures. The method is being tested on the serine proteases and on the antibody variable domains. One advantage of the use of the present method compared to other studies (Greer, 1980; Padlan *et al.*, 1976) is that the calculation of a model conformation is completely automatic and can be repeated to give a family of possible conformations.

(2) Extreme perturbation of the native conformation. Small perturbations of the X-ray conformation have helped explain how aromatic side-chains flip over (Gelin & Karplus, 1975), prolines influence rates of refolding (Levitt, 1981*a*) and domains move relative to one another (McCammon *et al.*, 1976). The present

method can easily be used to generate more extensive conformational perturbations. Appropriate restraints are used to give a conformation that is suitably perturbed yet also has a very low energy value. One application of this scheme is a study of the disulphide bond transition states of BPTI in which different pairs of sulphur atoms are restrained to come close together in the native conformation (Levitt, unpublished results).

6. Conclusion

Tt has been shown how soft-atom constrained energy minimization can be used to compute folded conformations of BPTI that are closer to the native structure (3 Å r.m.s. deviation) than obtained before. The value of the in *vacuo* potential energy provides a criterion for selection of the conformations that are more native-like. This selection criterion does *not* depend on knowledge of the X-ray structure; it could be used to predict the unknown conformation of a protein provided suitable restraints were available.

The methods introduced here have many potential applications: soft-atoms allow the polypeptide chain to pass through itself and give a variety of different, chain threadings. The writhing number classifies chain threadings and provides a new measure of chain fold. Annealing dynamics avoids local minima in the potential energy surface and can be applied to other optimization problems.

Improvements to the present method are suggested by the results in that solvent effects are needed to make the calculated structures more native-like, and the starting conformations can be classified by their writhing numbers to reduce the amount of computation. Both these improvements are now being tested on BPTI and another small protein, the C-terminal 72 residues of L7/L12 (Leijonmarck *et al.*, 1980).

Restrained energy minimization can be used to calculate conformations of proteins thought to be homologous to a protein of known conformation and also to study extreme perturbations of the X-ray structure. As such, the method constitutes a completely automatic molecular modeling system- that could be implemented on a small computer and made available to a large community of experimental researchers.

I am grateful to the Weizmann Institute for having provided the ample computer resources needed for this research. I thank F. H. C. Crick for having stimulated the initial phases of the project by critical discussions, and one of the referees for having suggested an improved method of calculating the angles of the spherical quadrilateral (Fig. 3).

REFERENCES

Baldwin, R. L. & Creighton, T. E. (1980). In Protein *Folding* (Jaenicke. R.. ed.), pp. 217-259, Elsevier/North Holland, Amsterdam.
Benham, C. J. (1978). *J. Mol. Biol.* 123, 361-370.
Burgess, A. W. & Scheraga, H. A. (1975). *Proc. Nat. Acad. Sci., U.S.A. 72.* 1221-1225.
Chothia, C. (1976). *J. Mol. Biol.* **105,** 1-14.
Cohen, F. H. & Sternberg, M. J. E. (1980). *J. Mol. Biol.* **138,** 32 l-333.
Connolly, M. L., Kuntz, I. D. & Crippen, G. M. (1980). *Biopolymers*, **19,** 1167-l 182.
Creighton. T. *E.* (1978). *Prog. Biophys. Mol. Biol. 33,* 231-297.
Crippen, G. M. (1977). *J. Comp. Phys. 24,* 96-107.

Davidon, W. C. **(1959).** *A .E.C. Research and Development report,* ANL-5990, Oakridge National Laboratory.
Deisenhofer, J. & Steigemann, W. W. **(1975).** *Acta Crystallogr. sect. B,* **31**, 238–250.
Fletcher, R. **(1970). Computer** *J.* **13, 317-322.**
Fletcher, R. & Powell, M. J. D. **(1963). Computer** *J.* **6, 163-168.**
Fletcher, R. & Reeves, C. M. **(1964). Computer** *J.* **7, 149-154.**
Fuller, F. B. **(1971).** *Proc. Nat. Acad. Sci., U.S.A. 68,* **815-819.**
Gelin, B. R., & Karplus, M. **(1975).** *Proc. Nat. Acad. Sci., U.S. A. 72,* **2002-2006.**
Gibson, **K. D.** & Scheraga, H. A. **(1969).** *Proc. Nat. Acad. Sci., U.S.A. 63,* 9–15.
Goel, N. S. & Yčas, M. **(1979).** *J. Theoret. Biol. 77,* 253–305.
Greer, J. **(1980).** *Proc. Nat. Acad. Sci., U.S.A. 77, 3393-3397.*
Hagler, A. T. & Honig, B. **(1978).** *Proc. Nat. Acad. Sci., U.S. A.* **75**, **554-558.**
Havel, T. F., Crippen, G. M. & Kuntz, I. D. **(1979).** *Biopolymers,* **18, 73-81.**
Hendrickson, J. B. **(1961).** *J. Amer. Chem. Soc. 83, 4537-4547.*
Hestenes, M. R. & Stiefel, E. **(1952).** *J. Res. Nat. Bur. Stand. 49, 409-436.*
Huber, R., Kukla, D., Ruhlmann, A. & Steigemann, W. **(1971).** *Cold Spring Harbor Symp. Quant. Biol. 36,* **141-150.**
Janin, J.. Wodak, S., Levitt, M. & Maigret. B. **(1978).** *J. Mol. Biol.* **125, 357-386.**
Kabsch, W. **(1976).** *Acta Crystallogr. sect. A, 32, 922-923.*
Kendrew, J. C., Dickerson, R. E., Strandberg. B. E., Hart, R. G.. Davies. D. R., Phillips, *D.* C. & Shore, V. *C.* **(1960).** *Nature (London),* **185. 422427.**
Kuntz, I. D., Crippen, G. M., Kollman, P. A. & Kimelman, D. **(1976).** *J. Mol. Biol.* **106, 983-994.**
Kuntz, I. D., Crippen, G. M. & Kollman, P. A. **(1979).** *Biopolymers,* **18, 939-957.**
Le Bret, M. **(1979).** *Biopolymers* **18,** 1709–1725.
Lee, B. & Richards, F. H. **(1971).** *J. Mol. Biol.* **55, 379400.**
Leijonmarck, M., Ericksson, S. & Liljas, A. **(1980).** *Nature (London), 286,* **824-826.**
Levinthal, C. **(1968).** *J. Chim. Phys. 65,* **44-45.**
Levitt, M. **(1976).** *J. Mol. Biol.* **104, 59-107.**
Levitt, M. *(1980).* In *Protein Folding* (Jaenicke, R., ed.), pp. 17-39, Elsevier/North Holland, Amsterdam.
Levitt. M. **(1981a).** *J. Mol. Biol.* **145, 251-263.**
Levitt, M. **(1981** *b). Nature (London), 294, 379-380.*
Levitt, M. **(1982).** *Annu. Rev. Biophys. Bioeng.* **11, 251-271.**
Levitt. M. **(1983a).** *Cold Spring Harbor Symp. Quant. Biol. 47.* **251-262.**
Levitt, M. **(19833).** *J. Mol. Biol.* **168,** 595-620.
Levitt, M. **(1983c).** *J. Mol. Biol.* **168, 621-657.**
Levitt, M. & Warshel, A. **(1975).** *Nature (London),* **253**, *694-698.*
Lifson, S. & Warshel. **A. (1968).** *J. Chem. Phys. 49,* **5116-5129.**
MacKay, A. L. **(1974).** *Acta Crystallogr. sect. A, 30,* **440447.**
McCammon, J. A., Gelin, B. R.. Karplus, M. & Wolynes, P. *G.* **(1976).** *Nature (London),* **262, 325-327.**
McCammon, J. A., Gelin, B. R. & Karplus, M. **(1977).** *Nature (London), 267, 585-589.*
Meirovitch, H. & Scheraga, H. A. **(1981).** *Proc. Nat. Acad. Sci., U.S.A. 78,* **6584-6587.**
Padlan, E. A., Davies, D. R., Pecht, I., Givol, D. & Wright, *C.* E. **(1976).** *Cold Spring Harbor Symp. Quant. Biol.* **41. 627-637.**
Pottle, C., Pottle, M. S., Tuttle, R. W., Kinch, R. J. & Scheraga, H. A. **(1980).** *J. Comp. Chem.* **1, 46-58.**
Rackovsky, S. & Scheraga? H. A. **(1980).** *Macromolecules,* **13,** 1440–1453.
Robson, B. & Osguthorpe. D. J. **(1979).** *J. Mol. Biol.* **132, 19-51.**
Sela, M.. White, F. H. & Anfinsen, C. B. **(1957).** *Science,* **125, 691-692.**
van Gunsteren, W. F. & Karplus, M. **(1980).** *J. Comp. Chem.* **1. 266-274.**
Wüthrich, K., Wider, G., Wagner, G. & Braun, W. **(1982).** *J. Mol. Biol.* **155. 311-319.**

Edited by S. Brenner