predictions on the temperature and molecular weight dependence of the force law, which are currently being tested experimentally.

Acknowledgment. P.P. is pleased to express his thanks to Professors P.-G. de Gennes and T. Witten for helpful conversations. He has also benefited from several discussions with R. Cantor on polymer adsorption. J. Klein is grateful to Professor P. J. Flory for an illuminating remark and to Professor A. Silberberg for a useful discussion. We acknowledge partial support from the National Science Foundation.

References and Notes

(1) For a good review on steric stabilization see: Vincent, B. *Adv. Colloid Interface Sci.* **1974,** *4,* 193.

(2) Ninham, B. W.; Mahanty, J. "Dispersion Forces"; Academic Press: London, 1976.

(3) Lyklema, H.; Van Vliet, T. *Faraday Discuss. Chem. Soc.* **1978,** *No. 65,* 25.

(4) Cain, F. W.; Ottewill, R. H.; Smitham, J. B. *Faraday Discuss. Chem. Soc.* **1978,** *No. 65,* 33.

(5) Israelachvili, J. N.; Tandon, J. N.; White, L. R. *Nature (London)* **1979,** *277,* 120.

(6) (a) de Gennes, P.-G. *Macromolecules* **1981,** *14,* 1637. (b) *Ibid.* **1982,** *15,* 492.

(7) Jones, I. S.; Richmond, P. *J. Chem. Soc., Faraday Trans. 2* **1977,** *73,* 1062.

(8) de Gennes, P.-G. "Scaling Concepts in Polymer Physics"; Cornell University Press: Ithaca, N.Y., 1979.

(9) Moore, M. *J. Phys. A: Math., Nucl. Gen.* **1977,** *10,* 305.

(10) Klein, J. *Nature (London)* **1980,** *288,* 248.

(11) We use units with the Boltzmann constant taken to be unity.

(12) Helfand, E.; Tagami, Y. *J. Chem. Phys.* **1971,** *56,* 3592. *Ibid.* **1972,** *57,* 1812.

(13) Cahn, J. *J. Chem. Phys.* **1977,** *66,* 3667.

(14) Edwards, S. F. *J. Phys. A: Math., Gen. Nucl.* **1975,** *8,* 1670.

(15) If the wall is repulsive, the monomer density near the wall may enter the unstable region of the bulk phase diagram. This may be investigated by the method used here for a dilute solution and an attractive interface.

(16) Napper, D. H. *J. Colloid Interface Sci.* **1977,** *58,* 390.

(17) Schultz, A. R.; Flory, P. J. *J. Am. Chem. Soc.* **1952,** *74,* 4760.

(18) Recently, direct experiments of adsorption of polystyrene on mica at the Θ temperature using a microbalance technique have yielded similar values of the adsorbance.

(19) Derjaguin, B. V. *Koilloidn. Zh.* **1934,** *69,* 155.

(20) Scheutjens, J. M. H. M.; Fleer, G. J. *Adv. Colloid Interface Sci.*, in press.

(21) Klein, J., unpublished data.

(22) Klein, J. *Adv. Colloid Interface Sci.*, in press.

Monte Carlo Simulation of Protein Folding Using a Lattice Model

William R. Krigbaum* and Stephen F. Lin[†]

Gross Chemical Laboratory, Duke University, Durham, North Carolina 27706.

Received December 1, 1981

ABSTRACT: Monte Carlo folding simulations are performed with a bcc lattice model of pancreatic trypsin inhibitor. We compare the results obtained with centrosymmetric and local interaction potentials in five folding runs with different sets of random numbers. The four initial structures investigated are the best lattice representation of the native molecule, a random coil, a randomized structure having the disulfide bonds intact, and one having the helix intact. Both potentials result in smooth folding to globular conformations having root-mean-square deviations equivalent to, or smaller than, those previously obtained by similar methods using multifactor potentials. This observation is ascribed to restriction of the available conformational space by the lattice model. The indices used to compare the generated and idealized native structures indicate no preference between the two types of folding potentials. Retention of the correct disulfide bonds in the starting structure strongly directs folding toward the native conformation.

Introduction

Attempts to predict the native conformation of a protein molecule by minimization of an empirical energy function, using a multiatom representation of each residue, have foundered due to the existence of multiple minima in the energy surface.[1,2] This failure led to the exploration of highly simplified models of the peptide chain for folding simulations based upon Monte Carlo techniques. Levitt and Warshel[3] pioneered the use of simplified models, using as a test structure the small, single subunit protein pancreatic trypsin inhibitor (PTI). In the first of these papers, each of the $N = 58$ residues that is not glycine is replaced by two spheres taken to represent the peptide backbone and side chain, respectively. The bonds of this model chain are the virtual bonds connecting α-carbon atoms, and the torsion angle, α_i, is determined by the coordinates of the four contiguous α-carbon atoms $i - 1$ to $i + 2$. A conformation corresponding to minimum energy is sought by molecular dynamics techniques, taking the $N - 4$ values of the torsional angles as independent variables. During their folding simulations, "thermalizations" were performed periodically to allow the conformation to escape local minima, and the course of the process could be guided by application of "holding" and "pushing" potentials.[4] The basic argument advanced by these authors is that the computer requirements for simulation can be reduced to manageable bounds, and the energy surface will contain fewer subsidiary minima, if the number of independent variables is restricted by simplifying the model. A somewhat similar procedure was subsequently used by Kuntz, Crippen, Kollman, and Kimelman,[5] although they adopted $3N$ Cartesian coordinates as independent variables, where N is the number of residues in the molecule. Robson and Osguthorpe[6] performed folding simulations using an angular variable, γ, which couples the variation of ϕ and ψ of the same residue. They applied the equivalent of Levitt and Warshel's "holding" and "pushing" potentials, but at regular intervals during the simulation.

The use of simplified models in folding simulations is based upon the assumption that the empirical energy function corresponds to an energy surface having a global minimum and that the conformation corresponding to this global minimum will closely resemble the "idealized" native state. Hence, these structures have been compared with the idealized native conformation using a root-mean-square deviation of distances. Claims of success for this type of

[†] Permanent address: Department of Chemistry, North Carolina Central University, Durham, N.C.

0024-9297/82/2215-1135$01.25/0 © 1982 American Chemical Society

simulation appear to have diminished steadily with time. Levitt and Warshel[3] and Levitt[4] stated that their procedure simulated the kinetics of the actual folding pathway of proteins and led to conformations closely resembling that of the native protein. Kuntz et al.[5] did not attempt to simulate the folding pathway, and their claims of correspondence to the native structure were more circumspect. A subsequent paper by Hagler and Honig[7] concluded that the criteria of success used in the previous papers was overly permissive, since all of the generated structures showed significant differences in topology from the native structure. Further, they showed that the superficial similarity between the generated and "idealized" native conformations reported by Levitt and Warshel arose from the introduction by these authors of artificial torsional potentials having single minima that guaranteed the introduction of bends in appropriate locations and otherwise favored the extended chain conformation present in the starting structure. In fact, Hagler and Honig were able to achieve folded structures of similar quality using only a two amino acid representation (Ala and Gly) of PTI. These authors followed Levitt and Warshel[3] in replacing Gly, Asp, and Asn residues by Gly. They also observed that bends could be introduced in the appropriate positions without resorting to the use of an artificial torsional potential having a single minimum. Robson and Osguthorpe[6] emphasized the importance of "hinge points", which are coil-type residues having sufficient flexibility to permit regions of the partially folded protein to swing together to create a globular structure.

A quite different approach was taken by Gō and coworkers in a series of papers dealing with a self-avoiding random walk model on a two-dimensional square lattice[8–10] and a three-dimensional cubic lattice.[11] Their attention was directed toward study of the folding process per se, and they designated a completely arbitrary geometric structure as the "native" conformation. The energy of their system can be reduced by the formation of near-neighbor contacts between selected units of the chain (long-range interactions) and by adoption of the "correct" rotational angles determined by three (or four) chain units (short-range interactions). Also, the entropy of their system can be reduced by the occurrence of vacant lattice sites adjacent to certain chain units arbitrarily designated as "hydrophobic". They simulated the folding by the Monte Carlo process of Metropolis et al.[12] and allowed translation and rotation of both single units and sequence of units. The results of these studies, as summarized by Gō et al.,[13] are as follows:

(a) The folding process is accelerated by short-range interactions and "correct" long-range interactions, while inclusion of additional "incorrect" long-range interactions decelerates folding.

(b) Cooperativity of the transition is increased by the inclusion of "correct" long-range interactions but reduced by short-range and "incorrect" long-range interactions.

Renaturation of the two-dimensional model chain was observed to occur over a range of reduced temperatures. For a judicious choice of reduced temperature, several renaturation and denaturation events could be observed in a single computer run. A similar study[11] using a three-dimensional lattice model gave renaturation from a partially unfolded state but not from the fully denatured state. This failure was attributed in part to the lack of flexibility of the lattice model, which led to trapping of partially folded conformations, and in part to the formation of mirror image partial structures (nuclei), which are isoenergetic in the model system but cannot be combined to produce the native structure.

Recently, Dashevskii[14] reported a study of conformations generated on a tetrahedral lattice for trypsin inhibitor and ribonuclease S. The model adopted requires one site per residue. In placing α-carbon i, a target function is evaluated for the 81 possible locations of α-carbons i through $i + 3$, and α-carbon i is assigned the site that occurs in the sequence of four "moves" producing the maximum value of the target function. This procedure produces a unique lattice chain conformation for a given target function and a fixed number of steps scanned (four in this case). The target function included a contribution (not further defined) favoring β strands and a term representing hydrophobic interactions. For this purpose the residues were divided into nonpolar, indifferent, and polar categories, and the interaction was calculated using a 3×3 matrix incorporating four adjustable parameters. The nonpolar–polar interaction was assigned the value $-\infty$, and it was found that all the other parameters had to make positive contributions to the target function in order to obtain a compact structure. Eighteen arbitrary parameter sets were examined for trypsin inhibitor and three for ribonuclease S. Some of the generated structures contained features which resembled the native structures, but the parameter set yielding the best results for PTI gave poor results for RNAse S and vice versa. The authors conclude that the predictive power of their model is low due to the limited number of conformational states available using the tetrahedral lattice model and to the neglect of many of the specific interactions stabilizing tertiary structure. They do not mention what would appear to be the most serious shortcoming of this method, namely, the neglect of interactions more remote than three residues away.

In summary, while the use of simplified models has enjoyed only partial success in tracing the folding pathway and in predicting native-like conformations, it has furnished some useful insights into how the various terms in the empirical energy expression may influence the process of folding. It is probably a valid, and useful, conclusion from the foregoing work that the folding process can be performed more quickly and smoothly if the region of conformational space that must be sampled is limited by simplification of the model.

The latter conclusion has led us to further examine the lattice model for folding simulations. This choice has both advantages and disadvantages, which we will enumerate. Concerning the former, the use of a lattice model is a very effective way to limit the volume of the available conformational space, since the chain units are constrained to lie at discrete lattice sites. Secondly, the requirement that the chain be self-avoiding can be met expeditiously by moving only one randomly chosen chain segment in any cycle, as suggested by King[15] and tested by Verdier and Stockmayer.[16] This property is not so advantageous in the study of random flight chains, as pointed out by the latter authors, but it becomes more powerful in the present application in which the subset of conformations corresponding to a collapsed globule is of particular interest. This procedure permitted us to perform folding simulations for PTI using a computer having only 32K words of memory, which is much smaller than the computers used in previous molecular dynamics folding simulations. Third, comparison of the folded and "idealized" native conformations was restricted in earlier work to root-mean-square deviations of distances because this property is independent of the relative orientation of the two structures. Since only 24 rotations are required to cover the full ranges of the Euler angles, a more sensitive root-mean-square criterion based upon vectors can be used with the lattice model.

The disadvantages of our procedure accrue primarily from the artificialities of the lattice model. If the step distance is fixed to correspond to the virtual bond length, 3.8 Å, then the average density of quasi-globular conformations will depend upon the type of lattice used. Preliminary investigation revealed that, for a model occupying one lattice site per residue, globular conformations on the tetrahedral lattice have too large a radius of gyration and are too compact in the case of simple and body-centered cubic (bcc) lattices. If each residue except glycine occupies two lattice sites, globular structures on the bcc lattice are more nearly of the correct size, although still somewhat too compact. Secondly, attempts to represent types of secondary structure by geometrically regular conformations inevitably result in an incorrect projected length per residue. This is not a serious problem in the present application, since we will not be particularly concerned with secondary structure. A final disadvantage arises from the procedure involving, in each cycle, moving only the two points representing one residue. Two multiunit secondary structures (e.g., helix or β strand) cannot be displaced relative to one another by this procedure without disrupting at least one of the secondary structure units.

In earlier work from this laboratory,[17-19] protein crystal structure data were examined to test the effectiveness of van der Waals interactions between side chains as a structure determinant. One outcome of that work was a parameter set[17,18] furnishing an estimate of the magnitude of these interactions for all possible side chain–side chain pairs. It was also found that the radially averaged interaction parameter (or polarity) of the side chains increased monotonically with distance from the center of gravity of the molecule and that the gradient was steeper for smaller proteins. Some of these same conclusions were subsequently rediscovered, using essentially identical procedures, by Scheraga and co-workers.[20,21] Krigbaum and Komoriya[18] also found that if the side chains were divided into polar (P), indifferent (I), and nonpolar (N) categories, each category exhibited a unique radial dependence. Again, the form of these functions varied according to protein size. It should be possible to use these radial dependences to construct a centrosymmetric potential for use in protein folding simulations. Indeed, such a term was included (along with others) in the error function of Kuntz et al.[5] However, it was our belief that these radial profiles for the N, I, and P categories have no fundamental significance but arise as a result of local interactions involving side chains. Evidence for this view includes a fairly accurate prediction[18] of the observed radial profiles and their molecular weight dependences from consideration of local interactions in a simple, three-shell model of the protein molecule, and the success achieved in using local interactions to predict the relative position and orientation of the ribonuclease S–peptide and S–protein as they form the native complex.[19] This led us to expect that more successful folding simulations would be obtained with a potential based upon local interactions. Hence, the objective of this work is to compare the speed of folding and the quality of the structures generated with centrosymmetric and local interaction potentials. We have selected PTI as a model protein since this choice permits comparison with earlier simulation work.

Methods

Two empirical centrosymmetric folding potentials and one based upon local interactions were constructed. Monte Carlo folding simulations,[22] starting from the native structure and a random coil conformation, were performed for each potential using five sets of pseudorandom numbers. An idealized lattice model counterpart of native PTI was constructed by computer. The lattice model structure consists of 58 backbone units occupying contiguous nearest-neighbor sites, with all nonglycine residues having a side chain also occupying a nearest-neighbor site. We obtain the translation from a site to its nearest-neighbor by incrementing or decrementing each of the Cartesian coordinates of the site by one unit. Hence the nearest-neighboring distance, 3.8 Å, corresponds to $3^{1/2}$ lattice units (abbreviated lu below). Several criteria of quality were calculated by comparison with this idealized structure. Finally, in order to provide closer comparison with the results of previous workers, additional folding simulations were performed, starting from two lattice chain conformations that retain some features of the native structure.

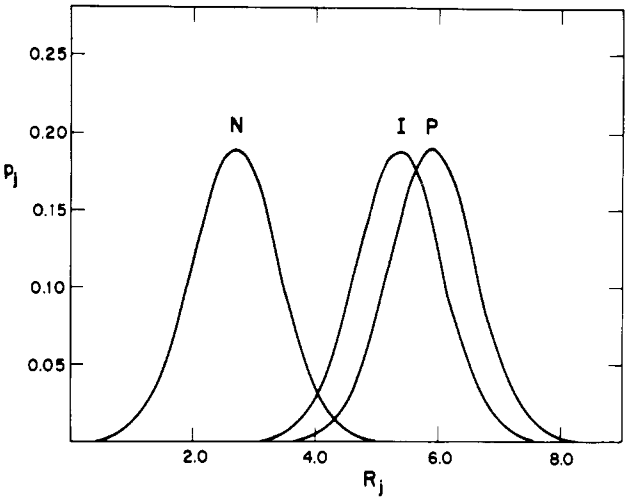

Figure 1. Radial dependence (in lu) of the probability profiles for nonpolar (N), indifferent (I), and polar (P) side chains used in the simpler centrosymmetric potential.

Centrosymmetric Potential. We follow Krigbaum and Komoriya[18] in dividing the 20 side chains into three categories:

N: Ile, Cys, Met, Phe, Leu, Val, Trp, and Tyr

I: Ala, Thr, His, Ser, Gln, Asp, Gly, and Glu

P: Asn, Arg, Pro, and Lys

The unnormalized probability assigned to a side chain of category j, when located a distance r from the center of gravity of the molecule, is

$$p_j \sim \exp(-V_j) \qquad (1)$$

where

$$V_j = (r - R_j)^2 \qquad (2)$$

Here R_j, the target radius for category j, is assigned the values (in lattice units) of $1.414S_R$, $2.828S_R$, and $3.111S_R$ for N, I, and P side chains, respectively, where S_R is an arbitrary radius scaling factor adjusted to match the radius of gyration of the generated structures to that of the idealized structure. These three probability functions, which are shifted Gaussians, are illustrated in Figure 1 for S_R = 1.9, the value adopted in the folding simulations.

A second, and more complicated, set of centrosymmetric probability functions was constructed in an attempt to more closely represent the profiles for the three categories

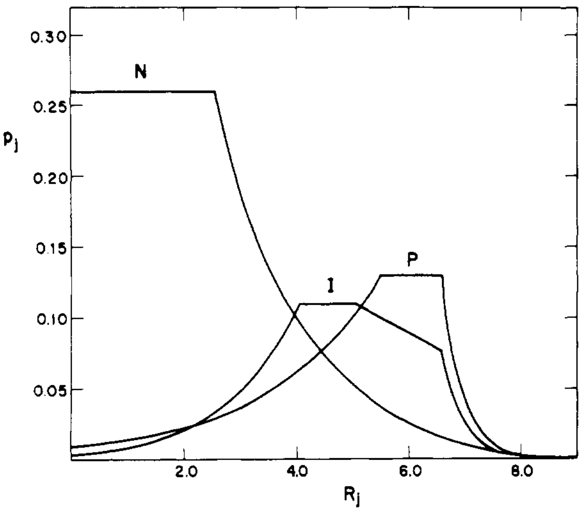

Figure 2. Second centrosymmetric probability functions for the three classes of side chain.

reported by Krigbaum and Komoriya[18] for a protein of this molecular weight:

Nonpolar (N)

$$p = 0.26 \qquad 0 < r < 2.3S_R$$

$$p = 0.26 \exp[-(r - 2.3S_R)/1.4S_R] \qquad 2.3S_R < r < 6.0S_R$$

$$p = 0.26e^{-2.64} \exp[-(r - 6.0S_R)/S_R] \qquad r > 6.0S_R$$

Indifferent (I)

$$p = 0.11 \exp[-(0.88/S_R)(3.17S_R - r)] \qquad 0 < r < 3.7S_R$$

$$p = 0.11 \qquad 3.7S_R < r < 4.6S_R$$

$$p = 0.11[1 - (0.22/S_R)(r - 4.6S_R)] \qquad 4.6S_R < r < 6.0S_R$$

$$p = 0.11(0.69) \exp[(3/S_R)(r - 6.0S_R)] \qquad r > 6.0S_R$$

Polar (P)

$$p = 0.13 \exp[-(5.0S_R - r)/1.8S_R] \qquad 0 < r < 5.0S_R$$

$$p = 0.13 \qquad 5.0S_R < r < 6.0S_R$$

$$p = 0.13 \exp[-(3/S_R)(r - 6.0S_R)] \qquad r > 6.0S_R$$

The probability profiles for $S_R = 1.12$, the value used in the folding simulations, are illustrated in Figure 2.

Local Interactions. The present model differs from the two-point representation we used previously mainly in having the distance between a backbone unit and its side chain fixed at 3.8 Å, as required by the lattice model. We believe this change should be inconsequential, and hence the contribution to the free energy from and i–j contact pair was calculated in terms of ΔG_{ij} given by eq 2 of Krigbaum and Komoriya,[18] using values of the necessary parameters taken from that paper. Since ΔG_{ij} is defined in terms of a process in which 1 mol each of side chain types i and j is removed from water to form i–j contacts, the contribution from one i–j pair is taken as $\Delta G_{ij}/\gamma NkT$, where γ is an effective coordination number, N is Avogadro's number, k is the Boltzmann constant, and T is the absolute temperature (taken as 300 K). Since the contact of a side chain with its own backbone is discounted, as well as those with the side chain and backbone units of the neighboring residues, we have assigned $\gamma = 6$. For use in the computer folding program, all 210 values of $\Delta G_{ij}/\gamma NkT$ were stored as elements of a triangular array. The probability assigned to side chain location t is

$$p_t \sim \exp(-u_t) \qquad (3)$$

where the potential u_t is the product of a folding factor, ff, multiplied by the sum of the local interaction free energy terms arising from side chain–side chain contacts formed at site t. The method of assigning a magnitude to ff will be described below. In previous work any nonadjacent pair of side chains having a separation less than 5.6 Å was considered to form a contact. For the present purpose the interaction distance is taken as that of second-neighbor sites (separation of 2 lu or 4.4 Å). This distance requirement may have been somewhat overly restrictive; hence an additional set of folding simulations was performed with a larger interaction distance for the first 20K cycles.

Monte Carlo Folding Procedure. The first set of folding runs had as its initial structure an idealized lattice model. One cycle of the folding process consists of moving a randomly selected backbone unit and its side chain. If the residue i (selected by random number) is not a chain end, the number of possible locations for the backbone unit is governed by the distance between backbone units $i - 1$ and $i + 1$. If this distance in lattice units is $12^{1/2}$, there is only one possible backbone location, with six locations for the side chain unit. If the distance is $8^{1/2}$ lu, there are two possible backbone site locations, each having six possible sites for the side chain. One of these was eliminated because it did not occur in idealized lattice model representations of native structures. Finally, if the distance is 2 lu there are four possible sites for the backbone. Of the six possible side chain sites about each of these, two were eliminated for the reason stated above. If the selected residue is at one of the chain ends, there are seven sites available to the backbone unit and, for each of these, seven side chain sites. The appropriate combinations of possible backbone and side chain sites is examined and the number, m, of *allowed* combinations is ascertained by eliminating any which involve sites occupied by the backbones or side chains of other residues. The normalized probability of choice t from among the m allowed moves is given by

$$p_t = \exp(-u_t)/\sum_{t=1}^{m}\exp(-u_t) \qquad (4)$$

One of the m allowed combinations is selected by a random number in the range 0 to 1, and the new conformation is characterized by calculating certain properties of interest.

Idealized Lattice Structure. An idealized lattice structure conforming to the constraints mentioned above and furnishing a best representation of the native structure[23] was created to test the quality of the folded structures. This idealized lattice structure was generated by computer with a two-step process. As a preliminary, a set of modified crystallographic coordinates was obtained by transforming the crystallographic coordinates of the α carbon and a selected[17] side chain atom to give a common separation. A coordinate framework was adopted with the first α-carbon atom at the origin, and the lattice chain was constructed, one residue at a time, on the same coordinate frame. The lattice point representing α-carbon i was chosen to give the best representation of the vector joining backbone chain unit $i - 1$ and the modified crystallographic coordinates of α-carbon i. If the preferred lattice site was occupied, the next best site was tested until a vacant site was located. A similar procedure was used to select a site for side chain i. The quality of the fitted lattice structure will vary as the modified crystallographic coordinates are rotated relative to the lattice. Hence, a root-mean-square deviation was calculated for each fitted structure as the three Euler angles were incremented to cover their full ranges. The best fitted lattice structure, obtained with $\theta = 80.8°$, $\phi = 10.4°$, and $\psi = 55.2°$, had $R_g = 11.28$ Å and a vector root-mean-square deviation of 2.46 Å from the rotated, modified crystallographic coordinates and 2.52 Å from the rotated crystallographic coordinates.

An inherent postulate of the local interaction model is that a small number of side chain–side chain contacts involving pairs of nonpolar side chains makes the predominant contribution to the stability of the native structure. We have arbitrarily defined NP–NP contacts as the 31 pairs for which $-\Delta G_{ij} > 1140$ cal/mol. These are Ile, Cys, Met, or Phe with Ile, Cys, Met, Phe, Val, Trp, or Tyr, and Trp with Val, Trp, or Tyr. Counting all side chain contacts within $2^{3/2}$ lattice units (6.2-Å separation), the best-fitting structure had 17 NP–NP contacts, of which 10 matched those having a separation not exceeding 5.6 Å as calculated from the crystallographic coordinates.

One might suspect that such a serial production method may not locate the best-fitting lattice representation of a dense, globular native conformation. We therefore used a second Monte Carlo procedure to seek a better fit, taking for the initial conformation the best-fitting representation described above. The backbone and side chain of a randomly selected residue could be moved in any cycle, using the allowed moves described above. The unnormalized probability of a particular move t of side chain and backbone units is taken as

$$p_t \sim \exp(-U_t) \qquad (5)$$

where the potential U_t assigned to choice t is the product of an arbitrary folding parameter, A, and the root-mean-square deviation of the backbone and side chain sites of that choice from the corresponding modified, rotated coordinates. The conformation resulting from 35 cycles was nearly identical with the starting structure. It had $R_g = 11.34$ Å, a vector root-mean-square deviation from the rotated crystallographic coordinates of 2.45 Å (as compared to 2.52 Å for the starting structure), and the same numbers of NP–NP and matching contacts. This structure, which will be referred to below as the idealized lattice model, is compared with the crystallographic structure in Figure 3. Only the course of the backbone, as defined by the α-carbon coordinates, is shown for clarity. Our experience in creating this idealized lattice structure indicates that the most important step is finding the Euler angles that transform the coordinate set to give the best fit to the lattice. The root-mean-square deviations based upon vector differences calculated for various folded conformations turn out to be nearly the same, whether compared to the idealized lattice model or the modified, rotated crystallographic coordinates.

Criteria of Quality. Reference has already been made to the discussion in the literature regarding criteria to assess the quality of folded conformations. Since the criteria to be used in different situations have different constraints, no single one can be universally useful. For example, any criterion used to monitor the course of the folding process must be rapidly calculable. The criteria used to assess a smaller number of selected conformations can be allotted more computer time, and it is these that have been the major subject of discussion. Most workers compared their generated and idealized structures by means of the root-mean-square deviation based upon distances. Cohen and Sternberg[24] suggest that the root-mean-square deviation based upon vectors would provide a more significant comparison. However, this has not been much utilized because it requires rotation of one of the structures through the full range of Euler angles to locate the minimum root-mean-square deviation. Robson and Osguthorpe[6] propose the use of a ratio of root-mean-square deviations based upon distance, D, as a measure of "native-like" character:

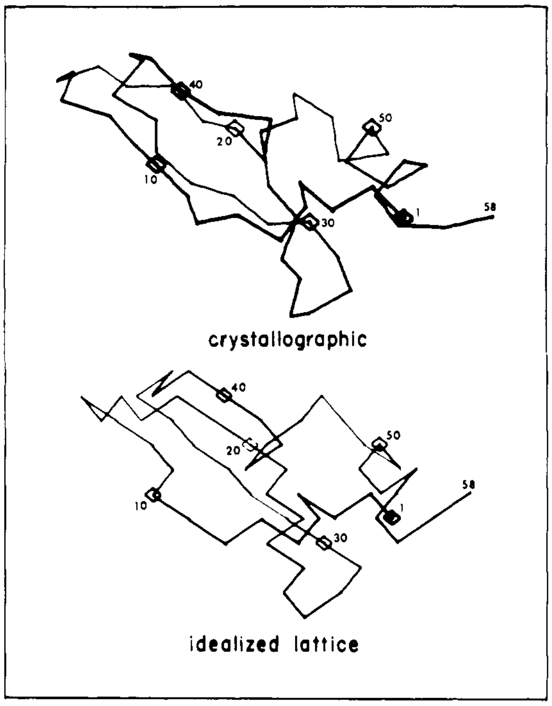

Figure 3. Backbone tracing of the idealized lattice representation of PTI (below) compared with that of the crystallographic structure (above).

$$N = \frac{D_{\max} - D_{\text{calcd}}}{D_{\max} - D_{\min}} \times 100$$

Here, $D_{\min}$ is the root-mean-square deviation of the idealized structure while $D_{\max}$ is that obtained for a Monte Carlo simulation using only dipeptide interactions. They cite for PTI $D_{\min} = 1.1$ and $D_{\max} = 23.6$. The latter is said to represent a random flight conformation, although it appears to be unusually large. Cohen and Sternberg[24] recommend the use of a quality index Q defined by

$$Q = 1 - \log N_p / \log N_R$$

where N_p is the number of conformations having a smaller root-mean-square vector deviation, V_{rms}, than that of the generated structure, while N_R is the number of conformations having a smaller root-mean-square vector deviation than the class of randomized compact structures. This suggestion has much to recommend it, but we have not used it because we do not know the statistical distribution of compact structures as a function of V_{rms} for the lattice model used here.

We have regularly used five criteria to evaluate folded conformations, and our experience with a sixth criteria will be mentioned later in the paper. These five criteria and the columns in which they appear in Tables I–IV are as follows:

1. Column 3 lists R_g, the radius of gyration, defined as[25]

$$R_g = (3.8 \text{ Å}/3^{1/2})[(1/2N)\sum_{i=1}^{N}\{(x_i + x_i' - 2x_0)^2 + (y_i + y_i' - 2y_0)^2 + (z_i + z_i' - 2z_0)^2\}]^{1/2}$$

where the coordinates of the α-carbon and side chain are unprimed and primed, respectively, and subscript zero

Table I
Best Structures Folded by the Simple Centrosymmetric Potential from the Idealized Lattice Structure

run	cycle of structure $\times 10^3$	R_g, Å	D_{rms}, Å	V_{rms}, Å	V'_{rms}, Å	NP-NP contacts (matching)
1	16	11.31	5.88	10.33	10.47	24 (3)
2	23.5	11.04	5.64	10.07	10.08	22 (5)
3	23.5	11.43	5.64	10.17	9.94	29 (4)
4	26	11.31	5.23	9.27	9.25	26 (4)
5	21	11.24	5.34	8.30	8.29	29 (5)
av	22	11.27	5.54	9.63	9.61	26.0 (4.2)

Table II
Best Structures Folded from an Initial Random Coil Conformation

run	cycle of collapse $\times 10^3$	R_g^0, Å	D_{rms}, Å	V_{rms}, Å	V'_{rms}, Å	NP-NP contacts (matching)
A. Simple Centrosymmetric Potential						
1	9.5	11.38	6.08	12.55	12.61	26 (4)
2	3	11.43	6.48	13.47	13.61	23 (2)
3	8.5	11.34	6.13	12.80	12.73	22 (5)
4	6	11.38	6.16	12.94	13.05	25 (4)
5	2.5	11.34	5.86	12.73	12.97	25 (5)
av	5.9	11.37	6.15	12.90	12.99	24.2 (4.0)
B. Second Centrosymmetric Potential						
1	1.5	11.17	6.31	11.70	11.86	14 (3)
2	1.5	11.19	5.75	12.52	12.80	18 (3)
3	2	11.07	5.54	12.11	12.24	15 (2)
4	2	11.42	6.01	13.30	13.44	19 (3)
5	2.5	11.39	6.19	12.61	12.65	27 (5)
av	1.9	11.25	5.96	12.45	12.60	18.6 (3.2)
C. Local Interaction Potential (Interaction Distance 2 lu)						
1	37	10.60	5.71	12.02	11.80	52 (5)
2	37	10.41	5.92	11.72	11.58	56 (6)
3	58	11.44	6.55	13.91	13.95	38 (5)
4	44	11.01	6.28	11.31	11.37	41 (2)
5	36	11.62	6.77	13.85	13.80	33 (6)
av	42.4	11.02	6.23	12.56	12.50	44.0 (4.8)
D. Local Interaction Potential (Interaction Distance $2^{3/2}$ lu)						
1	7	10.94	6.21	11.96	12.24	43 (5)
2	62	10.72	6.59	10.95	11.09	45 (4)
3	6	10.60	6.13	12.31	12.08	38 (5)
4	15	10.59	5.71	11.30	11.24	54 (4)
5	50	10.57	6.38	11.35	11.47	42 (2)
av	28	10.68	6.21	11.57	11.62	44.4 (4.0)

designates coordinates of the center of gravity of the model structure, given by

$$x_0 = (1/2N)\sum_{i=1}^{N}(x_i + x_i')$$

etc.

2. Column 4 lists D_{rms}, the root-mean-square deviation by distance of the generated and idealized lattice structures, given by

$$D_{rms} = \frac{3.8\ \text{Å}}{3^{1/2}}\left[\left(\frac{1}{N-1}\right)\sum_{i=1}^{N-1}\frac{1}{2(N-i)}\sum_{t=1}^{N-i}\{(d_t - d_t^0)^2 + (d_t' - d_t'^0)^2\}\right]^{1/2}$$

Here, $N = 58$ is the number of residues in the chain, d_t and d_t^0 are the separations (in lattice units) of backbone units i and $i + t$, and the superscript zero refers to the idealized lattice structure. The separations d_t and d_t^0 between backbone units are given by relations of the form

$$d_t = [(x_i - x_{i+t})^2 + (y_i - y_{i+t})^2 + (z_i - z_{i+t})^2]^{1/2}$$

$$d_t^0 = [(x_i^0 - x_{i+t}^0)^2 + (y_i^0 - y_{i+t}^0)^2 + (z_i^0 - z_{i+t}^0)^2]^{1/2}$$

Relations for the separation between side chains (indicated above by a prime) take a similar form.

3. Column 5 gives V_{rms}, the root-mean-square deviation of vectors of the generated and idealized lattice structures, defined as

$$V_{rms} = \left[\frac{3.8\ \text{Å}}{2(3N)^{1/2}}\right][\{\sum_{i=1}^{N}(x_i - x_i^0)^2 + (y_i - y_i^0)^2 + (z_i - z_i^0)^2\}^{1/2} + \{\sum_{i=1}^{N}(x_i' - x_i'^0)^2 + (y_i' - y_i'^0)^2 + (z_i' - z_i'^0)^2\}^{1/2}]$$

Unprimed and primed parameters indicate coordinates of the backbone and side chain, respectively, and coordinates of the idealized lattice structure are designated by a superscript zero. V_{rms} is calculated as the generated lattice structure is rotated through 24 sets of Euler angles, and the lowest value is accepted.

4. Column 6 lists V'_{rms}, which is the same as V_{rms} in (3) except that comparison is made with the rotated crystallographic coordinates.

5. Column 7 gives the number of NP–NP contact pairs (as defined above) and the number matching those found

Table III
Best Structures Folded from an Initial Randomized Conformation with Disulfides Intact

run	cycle of collapse ($\times 10^3$)	R_g, Å	D_{rms}, Å	V_{rms}, Å	V'_{rms}, Å	NP-NP contacts (matching)
A. Simple Centrosymmetric Potential						
1	0.5	11.48	4.52	7.85	7.79	20 (7)
2	1.0	11.63	4.64	8.00	7.99	21 (7)
3	1.5	11.59	4.60	8.19	8.11	28 (7)
4	0.5	11.55	4.77	8.66	8.41	19 (7)
5	0.5	11.22	4.60	8.28	8.12	25 (7)
av	0.8	11.49	4.63	8.20	8.08	22.6 (7)
B. Second Centrosymmetric Potential						
1	0.5	11.62	4.15	6.79	6.60	18 (8)
2	0.5	11.42	4.52	8.25	8.11	14 (5)
3	0.5	11.50	4.23	7.00	6.86	19 (7)
4	0.75	11.18	4.47	7.38	7.31	28 (7)
5	0.5	11.58	4.30	7.26	7.17	18 (7)
av	0.55	11.46	4.33	7.34	7.21	19.4 (6.8)
C. Local Interaction Potential (Interaction Distance 2 lu)						
1	3	11.77	4.03	6.02	5.95	26 (10)
2	3	11.02	4.20	6.44	6.24	29 (9)
3	3	11.51	4.57	7.62	7.61	42 (7)
4	4	11.93	4.23	6.20	6.20	27 (8)
5	5	11.17	4.37	7.16	7.14	28 (6)
av	3.6	11.48	4.29	6.69	6.63	30.4 (8.0)

Table IV
Best Structures Folded from an Initial Randomized Conformation with the Helix Intact

run	cycle of collapse ($\times 10^3$)	R_g, Å	D_{rms}, Å	V_{rms}, Å	V'_{rms}, Å	NP-NP contacts (matching)
A. Simple Centrosymmetric Potential						
1	43	11.63	5.34	10.43	10.40	25 (5)
2	24	11.49	5.40	10.53	10.67	25 (4)
3	54	11.62	5.84	10.73	10.77	20 (3)
4	49.5	11.85	5.94	11.95	11.75	21 (4)
5	27	11.41	5.33	10.70	10.78	18 (3)
av	39.5	11.60	5.57	10.87	10.87	21.8 (3.8)
B. Second Centrosymmetric Potential						
1	14	11.19	5.72	11.64	11.77	16 (2)
2	5.5	11.50	5.88	11.73	11.72	10 (3)
3	5	11.69	5.77	11.37	11.09	11 (3)
4	15	11.63	5.81	11.33	11.43	15 (2)
5	12.5	11.61	5.57	11.43	11.47	13 (3)
av	10.4	11.52	5.75	11.50	11.50	13.0 (2.6)
C. Local Interaction Potential (Interaction Distance 2 lu)						
1	169	11.21	6.12	12.60	12.56	40 (3)
2	148	11.91	6.12	12.11	12.02	26 (3)
3	138	10.88	6.13	10.66	10.68	42 (4)
4	81	10.68	6.21	11.44	11.25	38 (5)
5	101	10.89	6.12	12.85	12.74	38 (5)
av	127.4	11.11	6.14	11.93	11.85	36.8 (4.0)

in the crystallographic structure.

Folding Simulations

Results with Four Starting Structures. To serve as a bench mark for subsequent results, five folding simulations were performed, starting with the idealized lattice structure and using different random number sets. The simpler centrosymmetric potential given by eq 1 and 2 was used. Preliminary runs gave somewhat too compact conformations, but R_g could be increased into the correct range by increasing S_R from unity to 1.9. Characterization of the best structure obtained in each run appears in Table I. As shown in column 2, an average of 2.2 × 10^4 cycles was required to achieve the best structure. The average radius of gyration of these best structures is R_g = 11.27 Å, as compared with 11.34 Å for the idealized lattice structure. The average values of D_{rms} and V_{rms} are 5.54 and 9.63 Å, respectively.

The next folding simulations were performed with an initial random flight conformation using the three potentials described above. The randomized structure had R_g = 14.3 Å, as compared with 11.4 Å for the idealized lattice model, and D_{rms} = 8.36 Å. As shown in Table IIA, folding with the simpler centrosymmetric potential proceeded rapidly and smoothly to a compact structure. An average of only 5.9 × 10^3 cycles was required for collapse to a globular structure. During the course of any run, small-scale oscillations were superimposed on the overall decrease of R_g, and it seems likely that brief expansions provide an escape from situations in which part of the chain is conformationally trapped. We also explored the possibility of assigning different weights to contributions from the three categories of side chain; however, unit weights proved to be better.

A second set of folding simulations was performed with the more elaborate centrosymmetric potential. Again, the possibility of using different weights was tested, but unit weights provided better results. The best structures obtained with this potential and a radius scaling factor $S_R = 1.12$ are characterized in Table IIB. Comparison of the entries in column 2 of sections A and B of Table II indicates that collapse to a globular structure began more quickly with the second potential. The criteria appearing in columns 4–6 show that the best structures obtained with these two centrosymmetric potentials are quite comparable, with the second potential giving slightly smaller root-mean-square deviations if the results for the five runs are averaged. On the other hand, the second potential results in significantly fewer NP–NP contacts, probably due to the fact that the radial dependence for N-type side chains is spread over a wider range of values for this potential (compare Figures 1 and 2). This may also explain the faster collapse observed with the latter potential.

A third set of folding simulations was performed with the local interaction potential. Preliminary runs showed that collapse to a globular conformation only occurred for a limited range of values of the folding factor, *ff*. The initial portion of the folding process evidently involves the formation and dissolution of small clusters of nonpolar side chains. These clusters do not grow if the attractive force is too small, and the conformation becomes frozen in a local minimum if the attractions are too strong. Best results were obtained with $ff = 5.0$. The program was then modified to separate short-range side chain–side chain contacts (those within four residues) and the remaining long-range contacts. The "obligatory" contacts described by Krigbaum and Komoriya[18] for certain types of secondary structure were added as short-range contacts. These involve side chain i and side chain $i + 4$ if both are within a geometrically regular structure arbitrarily defined as helical and side chain i and side chain $i + 2$ for internal residues in a regular structure arbitrarily defined as a β strand. However, our potential does not particularly favor the formation of these regular structures, so the contribution from this source was probably negligible. Trials with different weights for short-range and long-range interactions indicated that unit weights gave best results. Since folding with this potential is postulated to occur by the formation of clusters of nonpolar side chains, one would expect that the process could be accelerated by biasing the selection of the residue to be moved to favor the nonpolar category (as defined above). This was the case, and we have increased the probability of selecting an N-type residue by a factor of 2.

Table IIC presents a characterization of the best structures obtained in five trials using the local interaction potential. Column 2 indicates that this potential requires many more cycles to form a collapsed, globular conformation. We expect that little progress toward collapse will occur until a cluster of nonpolar side chains of critical size is formed. By contrast, with the centrosymmetric potential each cycle tends to drive the side chain toward a target radius that changes little during the collapse, since to a first approximation the center of gravity of the molecule is stationary during this period of time. Column 3 of Table IIC indicates that the best globular conformations generated with the local interaction potential are somewhat too compact. In this case we have no parameter analogous to the radius scaling factor of the centrosymmetric potential. According to the criteria in columns 4–6, the best structures produced by the local interaction potential are comparable in quality to those obtained with the two centrosymmetric potentials. The local interaction potential produces many more NP–NP contacts (column 7), as expected, but this potential does not appear to lead to better selectivity in terms of those contacts found in the native protein.

We surmised that the longer time required to achieve collapse with the local interaction potential might be due, at least in part, to the rather short interaction distance (2 lu or 4.4 Å) adopted above. Hence, five additional runs were performed with the same local interaction potential, but increasing the interaction distance to $2^{3/2}$ lu (6.2 Å) for the first 2.0×10^4 cycles. Characterization of the best structures from the five runs appears in Table IID. As indicated in column 2, the number of cycles required to achieve a collapsed structure is quite variable, ranging from 6×10^3 to 6.2×10^4. However, the average for the five runs, 2.6×10^4, is clearly less than the 4.24×10^4 average using the shorter interaction distance throughout. Since the same interaction distance was used in both cases during the final "equilibration", no difference in the quality of the best generated structures was expected, and none is found on comparing sections C and D in Table II.

Comparison with Other Folding Simulations. Since our interest has been to compare the effectiveness of centrosymmetric and local interaction potentials, we have omitted all other factors (some of which may play a very important role). Thus, one should not expect our results to be of comparable quality to the previous nonlattice simulations, which incorporated many additional factors. For example, the empirical energy function of Levitt[4] included a repulsive potential, van der Waals interactions of side chains, side chain–water interactions, rotational potentials, a peptide hydrogen-bonding term, and a harmonic potential to force the formation of disulfide bonds. The results he obtained appear to depend upon the particular set of random numbers, and in 30% of the trials no globular conformation was formed. Of the successful runs, folding from an initial extended chain conformation gave $D_{\rm rms} = 8.5$ Å. If a helical region at residues 48–58 was preformed, $D_{\rm rms}$ was reduced to the range 6.2–7.1 Å, and addition of the potential to form the correct disulfide bonds reduced $D_{\rm rms}$ to 5.8–5.9 Å. Hagler and Honig,[7] using only a two-peptide representation of PTI, obtained $D_{\rm rms}$ values in the range 6.2–6.8 Å, starting from a fully extended conformation. The error function of Kuntz et al.[5] included a repulsive potential, a potential relating to the virtual bond length, hydrophobic interactions, a centrosymmetric potential, and a term forcing the formation of the correct disulfide bonds. They started from an extended conformation and presented, for PTI and rubredoxin, only two results: the best conformation obtained using common parameters for both proteins and the best results obtained when all parameters are independently adjusted. These were $D_{\rm rms} = 5.0$ and 4.7 Å for PTI and 4.3 and 4.0 Å for rubredoxin. These lower root-mean-square deviations are somewhat surprising, since a casual inspection of the stereoviews would lead to the conclusion that the folded structure displayed by Levitt and Warshel[3] has a closer superficial resemblance to the native structure than that shown by Kuntz et al.[5] Robson and Osguthorpe[6] included parameters relating to rotational potentials, van der Waals, electrostatic, and solvent-dependent interactions, as well as hydrogen bonding. Their first folding simulation of PTI began from a model having predicted secondary structure (including the C-terminal helix) and included a disulfide closing potential. This gave $D_{\rm rms} = 6.0$ Å. A second simulation starting from a nearly extended chain produced D_{rms} = 7.0 Å.

If the five runs shown in each of the first three parts of Table II are averaged, our folding simulations give D_{rms} values in the range 6.0–6.2 Å. By this one common criterion, either single potential produced structures of somewhat poorer quality than those of Kuntz et al.;[5] however, our results are equivalent to, or better than, those obtained by the other workers.[3,4,6,7] This observation is surprising on two counts. First, as mentioned above, the other workers accounted for many more factors which should be structure directing. Secondly, they performed their folding simulations in such a way as to ensure the formation of certain aspects of the native structure, which would be expected to lower the D_{rms} value. As mentioned above, the work of Levitt[4] indicates that the introduction of the preformed helix reduced D_{rms} from 8.5 to 6.2–7.1 Å. It seems safe to conclude from these observations that the use of the lattice model, which reduces the total conformational space available to the molecule, would produce significantly better folded structures than the virtual bond model if the same potential functions and starting structures were used for both. Further confirmation of this view is provided by the results of the tetrahedral lattice model simulation by Dashevskii.[14] He used a relatively simple target function including hydrophobic interactions and a factor favoring the formation of β strands. His eight results for PTI gave V_{rms} values in the range 10.0–12.1 Å. Based upon our results, this would be equivalent to D_{rms} of 5.4–6.1 Å, which again compares very favorably with the nonlattice results simulations using multifactor potentials.

In an attempt to provide a more direct comparison of our results with previous work, folding simulations were performed using the same values of the parameters S_R and ff for two additional starting structures. The first of these was obtained by maintaining the native coordinates of cystines 5, 14, 30, 38, 51, and 55 while randomizing the coordinates of the remaining residues in 3.4 × 10^4 cycles of single-residue moves. This starting structure had R_g = 12.5 Å and D_{rms} of only 5.10 Å. This small D_{rms} indicates that fixing the disulfide bonds produces a number of correct interunit vectors. The coordinates of these six cystine residues also remained fixed during folding simulations performed with the three potentials, which gave the results shown in Table III. Comparison of the entries in column 2 of Tables II and III indicates that folding to a globular structure occurred much more rapidly from the latter starting structure. The root-mean-square deviations shown in columns 4–6 are significantly smaller with the correct disulfide bonds preformed, the average values of D_{rms} decreasing from 6.0–6.2 to 4.3–4.6 Å and V_{rms} decreasing from 12.4–12.9 to 6.7–8.2 Å. These deviations are also smaller than those appearing in Table I for the best structures obtained on folding from the idealized lattice structure. Evidently, fixing the disulfide bonds provides a substantial bias toward the native conformation. It should be pointed out that the bias in this case is larger than that imposed by the constraints of Kuntz et al.,[5] since these authors required the correct disulfide pairs to form but did not fix the separation or orientation of the three pairs. Shortcomings of the single-residue move per cycle, as described earlier, precluded us from performing the simulation in a way that would more closely resemble their constraints. Nevertheless, comparison of our averaged D_{rms}, 4.3–4.6 Å, with their values, 4.7–5.0 Å, suggests that we would have obtained structures of at least equivalent quality if we had been able to duplicate their treatment of the disulfide bonding. Moreover, our results using a single potential are better than those reported by Levitt and Warshel,[3] Robson and Osguthorpe,[6] and Hagler and Honig,[7] all of whom used multifactor potentials without the lattice restriction.

As judged from the average values for the five runs in Table III, the local interaction potential produces the best results from this partially native starting structure. This potential happens to yield about the correct R_g with this starting structure and is more selective in forming the NP–NP pairs found in the native structure. It might be mentioned that if randomization of the native structure with fixed disulfides is performed for fewer than 1.0 × 10^4 cycles, all three potentials give very rapid collapse of this partially randomized structure to a globular conformation and more spectacular root-mean-square deviations.

A second randomized structure was produced in a similar fashion, but keeping the helical residues 47–56 fixed. This starting structure, obtained after 3.0 × 10^4 cycles of randomization of the remaining residues, had R_g = 16.1 Å and D_{rms} = 9.43 Å. The best structures obtained by folding from this initial conformation are characterized in Table IV. Comparison with Table II shows that folding with a preformed helix required significantly *longer* to reach a globular conformation for all three potentials. This may be an artifact of our procedure in which a single residue is moved in any cycle, since this treats the helix as a stationary object. More surprising, the best folded structures from all potentials are only slightly better than those derived from the fully randomized starting structure, D_{rms} decreasing from 6.0–6.2 and 5.6–6.1 Å and V_{rms} from 12.4–12.9 to 10.9–11.9 Å. At least this level of improvement would be anticipated from the assignment of the native coordinates to the ten helical residues. This appears to confirm the conclusion of Havel, Crippen, and Kuntz[26] that information concerning the secondary structure does not significantly restrict the range of possible tertiary structures.

Comparison with the Native Structure. Figure 4 illustrates backbone tracings of several of the best structures. The idealized lattice structure is labeled c. Structures a, d, and f were obtained with the simple centrosymmetric potential, and structures b, e, and g with the local interaction potential. Structures a and b correspond to entries A3 and C2 of Table II folded from a random flight chain, d and e illustrate entries A1 and C2 of Table III folded from a randomized structure with the correct disulfide geometry, and f and g represent entries A5 and C3 of Table IV folded from the randomized PTI model with helical residues 47–56 fixed. In comparing structures f and g with c, one must bear in mind that the final two residues (57 and 58) are not part of the fixed helical structure. Inspection of Figure 4 reveals that the best structures folded from the random coil and from the starting structure having the helix preformed tend to be too spherical, whereas folding from the starting structure with the correct disulfide geometry more nearly resembles the ellipsoidal geometry of the native PTI molecule. Folded structures d and e also have longer runs of extended chain resembling the native β strands, and the chain tracing more closely resembles that of the idealized lattice structure. Comparison of averaged values from Tables II and IV with that from Table III gives for D_{rms} = 6.1, 5.8, and 4.4 Å, while the corresponding V_{rms} values are 12.6, 11.4, and 7.4 Å. As expected, the root-mean-square vector difference provides a more valid indication of similarity to the native structure. Moreover, it appears to reflect, at least in part, some of the same criteria one uses in making a visual comparison. On the other hand, V_{rms} does

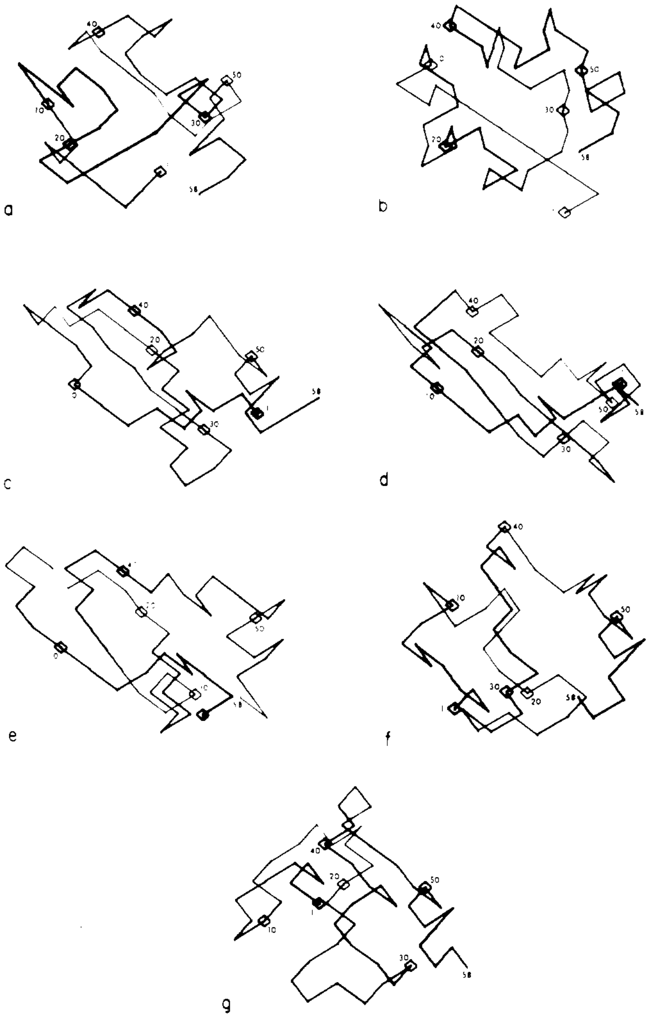

Figure 4. Comparison of the best folded structures with the idealized lattice model of PTI (symbol c). Structures a and b were folded from a random coil, d and e from a randomized structure retaining the correct disulfide geometry, and f and g from a randomized structure retaining the native helix at residues 47–56.

not test the detailed topology, which is much more difficult to evaluate quantitatively. Unfortunately, there are no V_{rms} values from previous work for comparison because this parameter is rotation dependent.

Recently, Rackovsky and Scheraga[27] suggested that two protein conformations can be compared in an objective manner by use of parameters derived from the concepts of differential geometry. In this procedure the coordinates of four successive α-carbon atoms, $i - 1$ to $i + 2$, are employed to describe the main-chain conformation in the vicinity of residue i in terms of the curvature, κ_i, and torsion, τ_i. Subsequently, they introduced a composite parameter, ρ_d, to facilitate this comparison.[28] While they proposed to use these parameters for a detailed comparison of two conformations, residue by residue, it would appear that values averaged over the chain should provide an overall index of the quality of match.

Table V compares the values of these three parameters for the initial structures and for some randomly selected examples of the best folded structures taken from Tables II–IV. The parameters based upon differential geometry exhibit no correlation with any other index of quality we have used. An obvious example occurs in the third group, where the starting structure having a preformed helix has lower rms deviations of both κ and ρ_d than any of the best folded structures. We cannot offer an explanation for this observation. It may be that the parameters advocated by Rackovsky and Scheraga are too short range to detect variations in the overall conformation, or perhaps the procedure is simply not applicable to a lattice model chain.

Table V
Quality Assessment Based upon Differential-Geometric Chain Representation

structure	κ(rms)	τ(rms)	ρ_d
A. Initial Random Coil Conformation (Table II)			
initial	0.3229	0.2943	0.3952
A1	0.2852	0.2526	0.3359
A5	0.2432	0.2906	0.3298
B2	0.3098	0.2747	0.3658
B3	0.3099	0.2619	0.3541
C1	0.3036	0.2325	0.3202
C2	0.2664	0.2464	0.3156
B. Initial Structure with Disulfide Bonds Intact (Table III)			
initial	0.2941	0.2569	0.3439
A1	0.2677	0.2986	0.3460
A3	0.3298	0.2676	0.3489
B1	0.2732	0.2588	0.3038
B3	0.3839	0.2715	0.4062
C1	0.1653	0.2500	0.2388
C4	0.2326	0.2304	0.2550
C. Initial Structure with Preformed Helix (Table IV)			
initial	0.1181	0.2510	0.2181
A1	0.2284	0.2482	0.2784
A5	0.3038	0.2378	0.3171
B3	0.1934	0.2255	0.2452
B5	0.3571	0.2375	0.2986
C3	0.2710	0.2871	0.3665
C4	0.2583	0.2531	0.2957

Conclusions

We proposed to use a lattice model in the belief that it would permit folding simulations to be performed quickly and smoothly by restricting the region of conformational space available to the molecule. Our experience appears to bear out this postulate. Either centrosymmetric potential or the local interaction potential, with an appropriate choice of the folding factor, gives smooth folding to compact, globular structures. Further, the quality of the best structures obtained with different sets of random numbers, when compared using the only available common quality index, D_{rms}, is as good or better than those obtained by other workers who have used the virtual bond model. This difference is particularly striking in view of our use of a potential consisting of a single factor, as compared with the numerous factors included in the nonlattice simulations. We have no a priori knowledge that any of three potentials used in this work has, in fact, a global minimum and, if one does exist, that it lies close to the native state. This aspect of our procedure differs from that of Gō and co-workers,[8–10] who could define their energy function so that it possessed a minimum at the arbitrarily defined native state. We can only hope that our potentials, which are deduced from the examination of native proteins, will direct the folding toward "native-like" conformations. Both potentials do give globular structures having reasonably low root-mean-square deviations based upon distances or vectors; however, what is lacking is a more sensitive criterion to assess the quality of these globular structures. In this regard, we can make no addition to the comments already given by Kuntz and co-workers,[5] Hagler and Honig,[7] and Cohen and Sternberg.[24]

The Monte Carlo method of Metropolis et al.,[12] in conjunction with the procedure first suggested by King,[15] allows the use of a small computer for folding simulations, whereas the molecular dynamics method involves much larger memory requirements. The inability to effect a relative displacement of two regions of secondary structure was a known disadvantage of the use of a one residue move per cycle, but the impediment to folding caused by a preformed helix was not anticipated. It is clearly shown that fixing the disulfide bonds does direct the folding process, but a preformed helix has much less effect in directing tertiary structure.

Our primary objective was to compare the folding process and the resulting structures using centrosymmetric and local interaction potentials. Faster folding with the centrosymmetric potential can be anticipated and was observed. We believed that the local interaction potential is a more fundamental quantity, and thus we anticipated that there would be a significant difference in the quality of the structures obtained with the two potentials. This expectation was not borne out by our results. The local interaction potential leads to more NP–NP contacts, as expected, but this is not accompanied by a noticeable improvement in any of the other indices we have used to assess quality. Only in the case of the starting structure containing the native disulfide bonding did the local interaction potential show any advantage, and even in this case the difference was small.

We initially planned that the next step might involve the use of a centrosymmetric potential to achieve a fast collapse of the conformation, followed by a sufficient number of cycles with the local interaction potential to form optimum side chain–side chain contacts. However, in view of these results, it would appear to be more profitable to combine some type of torsional potential with the centrosymmetric potential in an attempt to improve the quality of the present results.

Acknowledgment. S.L. thanks the Department of Health, Education, and Welfare for a National Research Service award, which permitted his contribution to this study.

References and Notes

(1) Burgess, A. W.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 1221.

(2) Nēmethy, G.; Scheraga, H. A. *Q. Rev. Biophys.* **1977**, *10*, 239.

(3) Levitt, M.; Warshel, A. *Nature (London)* **1975**, *253*, 694.

(4) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59.

(5) Kuntz, I. D.; Crippen, G. M.; Kollman, P. A.; Kimelman, D. *J. Mol. Biol.* **1976**, *106*, 983.

(6) Robson, B.; Osguthorpe, D. J. *J. Mol. Biol.* **1979**, *132*, 19.

(7) Hagler, A. T.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 554.

(8) Taketomi, H.; Ueda, Y.; Gō, N. *Int. J. Pept. Protein Res.* **1975**, *7*, 445.

(9) Gō, N.; Taketomi, H. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 559.

(10) Gō, N.; Taketomi, H. *Int. J. Pept. Protein Res.* **1979**, *13*, 235, 447.

(11) Ueda, Y.; Taketomi, H.; Gō, N. *Biopolymers* **1978**, *17*, 1531.

(12) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.

(13) Gō, N.; Abe, H.; Mizuno, H.; Taketomi, H. In "Protein Folding"; Jaenicke, R., Ed.; Elsevier/North-Holland Biomedical Press: Amsterdam, 1980; p 167.

(14) Dashevskii, V. G. *Mol. Biol. (Engl. Transl.)* **1980**, *14*, 105.

(15) King, G. W. "Research in the Physical and Structural Analysis of High Polymers by Punched Card Methods", Final Report, Office of Naval Research.

(16) Verdier, P. H.; Stockmayer, W. H. *J. Chem. Phys.* **1962**, *36*, 227.

(17) Krigbaum, W. R.; Rubin, B. H. *Biochim. Biophys. Acta* **1971**, *229*, 368.

(18) Krigbaum, W. R.; Komoriya, A. *Biochim. Biophys. Acta* **1979**, *576*, 204.

(19) Krigbaum, W. R.; Komoriya, A *Biochim. Biophys. Acta* **1979**, *576*, 226.

(20) Meirovitch, H.; Rackovsky, S.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 1398.

(21) Meirovitch, H.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 1406.

(22) We thank Steven Meador, Gary Hovis, and Jean-Pierre Auffret, who worked on the development of earlier versions of these programs as their undergraduate Independent Study projects.

(23) Huber, R.; Kukla, D.; Ruhlman, A.; Steigemann, W. *Cold Spring Harbor Symp. Quant. Biol.* **1971**, *36*, 141.

(24) Cohen, F. E.; Sternberg, M. J. E. *J. Mol. Biol.* **1980**, *138*, 321.

(25) It should be noted that Levitt and Warshel[2] and Levitt[3] used a different definition of the radius of gyration, so their values for this quantity cannot be compared with ours.

(26) Havel, T. F.; Crippen, G. M.; Kuntz, I. D. *Biopolymers* **1979**, *18*, 73.

(27) Rackovsky, S.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 1168.

(28) Rackovsky, S.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 1440.

Sorption of Water by Epoxide Prepolymers

Jose L. Garcia-Fierro and Jose V. Aleman*

Instituto de Plasticos y Caucho, Juan de la Cierva 3, Madrid 6, Spain.

Received October 13, 1981

ABSTRACT: Water equilibrium in epoxide prepolymers below the glass transition temperature (T_g = 343 K) is described. The heat of mixing and entropy of mixing show that, at low concentrations, the water may be hydrogen bonded, while at moderate or high concentrations, clustering of the water takes place, with an increase in free volume and a decrease in T_g to 323 K.

Introduction

All polymers may contain water in varying amounts according to the polarity of their macromolecular chains. Water content in solid epoxide prepolymers has not yet been described. Information available refers to "cured" epoxy systems, and results are at times contradictory.[1-4] The subject is of significance since the water adsorbed may interfere with the process of curing and, therefore, with the properties of the final products.

Experimental Section

1. Materials. Liquid epoxide prepolymers are usually prepared by reacting epichlorhydrin, bisphenol A, and sodium hydroxide according to the reaction mechanism shown in Scheme I.[5]

Solid epoxide prepolymers are made by two different processes: (i) by direct addition of epichlorohydrin to bisphenol A and (ii) by reaction of the low molecular weight epoxide resin, the main constituent of which is the diglycidyl ether of bisphenol A with bisphenol A. This last reaction is also known as the fusion process.

One solid epoxide prepolymer prepared = the fusion process (Araldite 6097) was supplied by Ciba-Geigy AG (Switzerland).

The glass transition temperature (T_g = 343 K) was measured as the first change in slope of the specific heat vs. temperature curves obtained in a DuPont 900 differential scanning calorimeter, as described elsewhere.[5] During heating of the polymer some volatile material evolved, which was mainly water and possibly

0024-9297/82/2215-1145$01.25/0 © 1982 American Chemical Society