Computer Simulations of De Novo Designed Helical Proteins

Andrzej Sikorski,* Andrzej Kolinski,*# and Jeffrey Skolnick#
*Department of Chemistry, University of Warsaw, Pasteura 1, 02–093 Warsaw, Poland; and #Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037 USA

ABSTRACT In the context of reduced protein models, Monte Carlo simulations of three de novo designed helical proteins (four-member helical bundle) were performed. At low temperatures, for all proteins under consideration, protein-like folds having different topologies were obtained from random starting conformations. These simulations are consistent with experimental evidence indicating that these de novo designed proteins have the features of a molten globule state. The results of Monte Carlo simulations suggest that these molecules adopt four-helix bundle topologies. They also give insight into the possible mechanism of folding and association, which occurs in these simulations by on-site assembly of the helices. The low-temperature conformations of all three sequences have the features of a molten globule state.

INTRODUCTION

The relationship between amino acid sequence and the native structure of globular proteins is still far from being understood. Although progress has been made toward understanding the interactions responsible for the stability of the tertiary structure, in general, we still are not able to predict the native state from amino acid sequence (Creighton, 1993). However, it has become possible to design and synthesize some predominantly α-helical polypeptides (Betz et al., 1995). Only very recently has some progress been made in designing proteins containing different and more complicated structural motifs (Quinn et al., 1994; Yan and Erickson, 1994; Tanaka et al., 1994; Smith et al., 1995) with prescribed functionality (Choma et al., 1994; Robertson et al., 1994). De novo designed proteins provide an excellent test of and assistance in the further development of our knowledge of the factors responsible for the structural and thermodynamic properties of proteins.

DeGrado and co-workers have developed a hierarchical approach to design and synthesize polypeptides that form four-member, α-helical bundles. First, they constructed short (12–16 residues) polypeptides (called α_1), which were believed to self-assemble to an α-helical tetramer (Ho and DeGrado, 1987). Next, a loop connecting two of these short chains was introduced and optimized. These two model chains (each called α_2), each capable of folding into an α-helical hairpin, should dimerize and form a four-member helical bundle (Ho and DeGrado, 1987). The final step was to connect these two α_2 chains by the same short loop to obtain the α_4 protein that should fold to a monomeric, four-helix bundle (Regan and DeGrado, 1988). It was found experimentally by size-exclusion chromatography and circular dichroism that this designed protein is monomeric, adopts a globular compact structure, and exhibits high helicity. α_4 is also more stable than both α_1 and α_2 (Regan and DeGrado, 1988). The crystal structure of α_1 was also determined at a resolution of 2.7 Å (Hill et al., 1990). Surprisingly, the crystal structure is not the expected tetramer, but is more complicated. There are antiparallel helical dimers forming a hexamer around a threefold axis of symmetry. On the other hand, a pair of dimers from neighboring hexamers forms a tetramer. Energetic considerations led the authors to conclude that this hexamer is more stable than the tetramer. However, it should be pointed out that the experiment concerned the shortened version of α_1 (12 residues) and was performed at low pH. For this modified α_1, the hexamer/tetramer equilibrium also exists in solution (Ciesla et al., 1991).

The design of proteins forming α-helical bundles is based on the hydrophobic pattern found in proteins that adopt α-helical bundle and coiled coil geometries. Leucine residues are located on one face of the helix and stabilize hydrophobic helix-helix contacts (Hodges et al., 1981; Cohen and Parry, 1990). On the solvent-exposed faces, there are glutamic acid and lysine residues that stabilize helices via electrostatic interactions (Cohen and Parry, 1990; Betz et al., 1996). These proteins, as well as most de novo designed proteins, exhibit some features of a molten globule state (Betz et al., 1993), i.e., a state where the system has a protein-like topology (topologies), but the packing of side groups is not specific but liquid-like, and the folding transition is not cooperative (Ptitsyn, 1992). This is explained by the fact that these designed polypeptide chains lack specific interactions in the hydrophobic core, which is built entirely from leucine residues (Betz et al., 1993).

Subsequently, these proteins were redesigned to be more native-like. Additional amino acid diversity in the hydrophobic core of α_2 was introduced. The obtained sequence is called α_2C (Raleigh and DeGrado, 1992). Next, some hydrophilic residues were added in the interfacial positions, and the hydrophobicity pattern was modified to destabilize potential alternative folds. The resulting chain is called α_2D

Received for publication 3 September 1997 and in final form 31 March 1998.

Address reprint requests to Dr. Andrzej Sikorski, Department of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland. Tel.: 48-22-822-22-02-11 ext. 388; Fax: 48-22-822-22-59-96; E-mail: sikorski@chem.uw.edu.pl.

0006-3495/98/07/92/14 $2.00

(Raleigh et al., 1995). According to some experimental evidence, the main difference between natural proteins and these designed proteins is that there are very many energetically close states in the designed proteins, instead of the structurally well defined native state of natural proteins (Betz et al., 1996).

These simple proteins can serve as a useful subject for studying the principles of protein folding both by experiment and by computer simulation. Using a low coordination number lattice to represent the protein, the α_4 designed protein and another improved version of α_4 were simulated by Kolinski et al. (1993). Both simulated polypeptides fold to a stable four-member, α-helical bundle. The existence of the molten globule state in the former case was confirmed, and the transition from molten globule state to native-like state was found for the second sequence; both observations are in agreement with known experimental facts. Thus, the force field of these lattice models appears to capture essential aspects of the physics of protein systems. More recently, a new high coordination lattice model was designed for the simulation of small globular proteins such as protein A, monomeric ROP, and crambin (Kolinski and Skolnick, 1994b), as well as for helical coiled coils such as the GCN4 leucine zipper (Vieth et al., 1994, 1995). The folding pathways were relatively well defined with early intermediate states, and the transition state had several features of a molten globule. In all of these cases, the transition from molten globule to native-like state with fixed positions of the side chains has been observed (Kolinski and Skolnick, 1996).

The success of these simulations encouraged us to refine the model and the simulation algorithm as well as to undertake Monte Carlo simulations of proteins designed by DeGrado and co-workers. Here, we studied the original DeGrado polypeptide chains, the α_4 tetramer, the α_2 dimer, and the α_1 monomer to check whether all these designed proteins assembled to a four-member, α-helical bundle. Then, the stability of different topologies of obtained folds was estimated. In subsequent work, we will undertake the simulation of other designed proteins, especially those that undergo a transition from a molten globule state to a native-like state.

To do our simulations, we chose the 310 hybrid lattice (Kolinski and Skolnick, 1994a) to maintain good representation of the protein chain with realistic side chain representation and with the proper force field. The all-atom reconstruction from this model is rather straightforward (Kolinski and Skolnick, 1996). The new lattice model has a better force field and provides a better representation of the chain geometry. The number of allowed orientations of secondary structure elements was significantly enlarged here. This refinement is especially important for helical proteins; they can now change their orientation almost continuously and undergo deformations as in real chains. In the new model, there is no bias toward the folded state (in the former model, there was a small bias toward the proper local geometry of a chain). This may be an important factor that prevents a possible bias of the model toward the molten globule state. The representation of side groups was also refined (the resolution is now $\sim$1 Å root mean square deviation (RMSD) for the side groups' center-of-masses), and the number of allowed rotamers was enlarged. Thus, the new model has many features of an off-lattice model. Eliminating most of the possible lattice artifacts should provide a more realistic picture of the protein folding pathway and the structure of the folded states. Computer simulations also allow observation of very early folding events that could be difficult to investigate in a real physical experiment, thereby providing some insight into the folding problem.

Description of the model

In our model, a protein chain is represented by the α-carbons of the protein backbone. The α-carbons are connected by 1 of 90 vectors chosen from the following set: {($\pm$3, $\pm$1, $\pm$1), ($\pm$3, $\pm$1, 0), ($\pm$3, 0, 0), ($\pm$2, $\pm$2, $\pm$1), ($\pm$2, $\pm$2, 0)} (with all permutations). The lattice unit corresponds to 1.22 Å. Side groups, which are not confined to the lattice, are single spheres having various rotamer isomeric states. A detailed description of the lattice properties representation was presented elsewhere (Kolinski and Skolnick, 1994a), and only a short outline is given here. Due to geometrical restrictions, some pairs of consecutive vectors are excluded. The geometric resolution of this lattice is very high, and real protein structures can be fitted to the lattice with accuracy close to the experimental resolution of these structures. The RMSD of the α-carbon positions is usually between 0.6 and 0.7 Å (Kolinski and Skolnick, 1994a; Godzik et al., 1993).

The detailed description of the Monte Carlo simulation algorithm was given previously (Vieth et al., 1995; Kolinski and Skolnick, 1994a). The Monte Carlo algorithm employs two-bond motions, three-bond motions, and chain end modifications as the set of local changes of chain conformation. Additionally, we occasionally attempted small distance shifts of larger chain pieces to avoid trapping the model system in local energy minima. For every modification, a random reorientation of the involved side groups was made. Rotamer equilibration was also performed for the entire chain(s) to allow for the adjustment of side groups. This is especially important for folded and partially folded conformations. Thus, the model dynamics do not introduce any bias toward any specific folding mechanism (Kolinski and Skolnick, 1994a). A single Monte Carlo run usually consists of 250,000–500,000 cycles.

Description of force field

Most of the terms in the model potential were derived from a statistical analysis of high-resolution crystal structures from the Brookhaven Protein Data Bank (Bernstein et al., 1977; Brookhaven Protein Data Bank, 1995). The procedure for the extraction of these data is described elsewhere (Kolinski and Skolnick, 1994a). The interactions introduced

into the model can be divided into short- and long-range terms. Short-range terms describe local conformational preferences along the sequence, and long-range terms refer to interactions between residues separated by at least three other residues along the chain contour but that are close in space. The force field used in this work was already described in detail (Vieth et al., 1994); thus, we give only a short outline here.

Long-range interactions

Pairwise potential

The pairwise energy of two interacting residues i and j can be written as follows:

$$E_{\text{pair}} = \sum_{i=1}^{N} \sum_{j=1+4}^{N} E_2^{ij}(A_i, A_j) \tag{1}$$

It is assumed that this potential is nonzero for $|i - j| > 2$. Additional details are found in the paper of Vieth et al. (1995). The scaling factor for this term is 0.5.

Cooperative pairwise interactions

This potential was introduced to permit the cooperative transition for a molten globule state to a native state with protein-like side-group packing (Kolinski and Skolnick, 1995). This potential can magnify protein-like interactions between secondary structural elements but does not prefer any particular type of secondary structure. It is given by

$$E_{\text{tem}} = (\epsilon_{\text{Ai,Aj}} + \epsilon_{\text{Ai+k,Aj+n}})\delta_{ij}\delta_{i+k,j+n}, \quad n = \pm 3, \pm 4, \quad k = \pm 3, \pm 4, \tag{2}$$

where δ_{ij} and $\delta_{i+k,j+n} = 1(0)$ when two specific pairs of residues i, j, and $i + k$, $j + n$ are (not) simultaneously in contact. $\epsilon_{\text{Ai,Aj}}$ is the pair potential (Vieth et al., 1995). It is important to note that the cooperative potential enables but does not enforce side chain fixation (as was demonstrated elsewhere (Kolinski et al., 1993)). The scaling factor for this contribution to the total potential is 4.25.

Burial energy

A residue is treated as buried when the number of side group contacts with other side groups exceeds a residue-specific threshold $n_{\text{con}}(A_i)$. A buried residue has zero energy, whereas a nonburied residue has an energy of E_{unbur}. Values of E_{unbur} and n_{con} for every amino acid are listed in Table 3 of Vieth et al. (1995). The total burial energy of a given chain is given by

$$E_{\text{one}} = \sum_{i=1}^{N} E_{\text{unbur}}(A_i, n_{\text{con}}(A_i)) \tag{3}$$

The scaling factor for the contribution of this potential to the total energy is 0.5.

Harmonic potential

This potential was introduced to maintain a constant high concentration of polypeptide chains. In its absence, due to translational entropy, we would not be able to obtain assembled structures in the available amount of computer time. The harmonic potential is given by the following formula:

$$E_{\text{harm}} = \begin{cases} 0 & \text{if} \quad r_{ij} \le R_{\text{cut}} \\ e_{\text{harm}}(r_{ij} - R_{\text{cut}})^2 & \text{if} \quad r_{ij} > R_{\text{cut}}, \end{cases} \tag{4}$$

where e_{harm} has a small positive value, r_{ij} is the distance between centers of mass of chains i and j, and R_{cut} is the distance below which the potential is turned off. The value of R_{cut} varied from 15 to 20 Å (distances somewhat larger than the size of the bundle). The scaling factor for the contribution of this interaction to the total energy is 1.0.

This kind of harmonic potential can be also used in the case of a single chain system (α_4) as an additional burial potential to speed up the assembly. It was introduced as a harmonic term between the center of mass of the entire chain and the center of mass of those parts of the chain that are expected to form helices (residues number 1–17, 21–36, 40–55, and 59–74, respectively, for the first, second, third, and fourth helix). The cutoff radius of this potential was slightly lower than that used in multichain systems (12–15 Å); therefore, this potential does not enforce any structure but only speeds up simulations. By applying this potential, we can create a system of dimers (for α_2) and tetramers (for α_1). In α_1, monomers, dimers, and trimers are sometimes observed, but more complicated aggregates such as hexamers, etc., were a priori excluded.

Short-range interactions

Rotamer energy

The total rotameric energy (Kolinski and Skolnick, 1994a) is given by

$$E_{\text{rot}} = \sum_{i=1}^{N} E_r(A_i, \theta_i), \tag{5}$$

where E_r is an energy of a given rotameric state for a given residue A_i, and θ_i is the angle between backbone bonds u_{i-1} and u_i. The scaling factor for the contribution of this interaction to the total energy is 0.5.

Local side chain orientational coupling

This potential introduces a preference for some orientations of the neighboring (down the chain) side groups (Kolinski and Skolnick, 1994a). The value of this potential

depends on pairs of amino acids and on the local backbone conformation:

$$E_\beta = \sum_{i=2}^{N-1}\left(\sum_{k=1}^{4} E_\beta^k(\cos(\alpha_{i,i+k}, A_i, A_{i+k})\right), \quad (6)$$

where $\cos(\alpha_{i,i+k})$ is the cosine of the angle between C_α and C_β (side group) vectors of residues i and $i + k$. E_β is the residue-specific orientational coupling energy. The scaling factor for the contribution of this potential to the total energy is 1.0.

Effective Ramachandran potential

This potential was introduced to mimic the protein-like geometry of the backbone because the lack of atomic detail leads to a distribution of intrachain distances, characteristic of a polymeric random coil (Kolinski and Skolnick, 1994a). Thus, we introduced a distance bias between α-carbons i and $i + 3$ and a bias of the chirality of these three bonds toward that of real proteins. The potential is defined as follows:

$$E_{R14} = \sum_{i=2}^{N-2}(4.5E_{14}(R_{14}^i, A_i, A_{i+1}) + E_{14|14}(R_{14}^{i-1}|R_{14}^i)), \quad (7)$$

where A_i is the type of amino acid in the ith position, R_{14}^i is defined as

$$R_{14}^i = |\mathbf{u}_{i-1} + \mathbf{u}_i + \mathbf{u}_{i+1}|^2 X, \quad (8)$$

where $X = \text{sign}\{(\mathbf{u}_{i-1} \times \mathbf{u}_i)\mathbf{u}_{i+1})\}$. The values of the parameter E_{14} and the residue-independent coupling energy $E_{14\,14}$ are listed in Tables 1 and 2, respectively, of Vieth et al. (1995). The scaling factor for these contributions to the total energy is 0.25.

Hydrogen bonds

This potential is residue independent and was introduced to simulate the hydrogen bond network found in proteins. In our model, there is no donor-acceptor differentiation, and every α-carbon can participate in two hydrogen bonds (with the exception of proline, which can form only one bond). We note that the positions of all backbone atoms could be easily determined (Milik et al., 1997). Here, however, we opt for a somewhat simpler definition: two residues, i and j, form a hydrogen bond if

$$\begin{aligned} &|i - j| \geq 3 \\ &4.6\text{Å} \leq |\mathbf{r}_{ij}| \leq 7.3\ \text{Å} \\ &|\mathbf{u}_{i-1} - \mathbf{u}_i) \cdot \mathbf{r}_{ij}| \leq 13.4\ \text{Å}^2 \\ &|\mathbf{u}_{i-1} - \mathbf{u}_i) \cdot \mathbf{r}_{ij}| \leq 13.4\ \text{Å}^2, \end{aligned} \quad (10)$$

where $\mathbf{u}_i$ is the backbone vector and $\mathbf{r}_{ij}$ is the vector between two α-carbons forming hydrogen bonds. This model potential reproduces most of the hydrogen bonds assigned by the Kabsch-Sander method (Kabsch and Sander, 1983) and is very close to the Levitt-Greer method (Levitt and Greer, 1977) of secondary structure assignment.

The energy of the hydrogen bond network can be expressed as follows:

$$E_{HB} = \sum_{i=1}^{N-3}\sum_{j=1,i+3}^{N}(E^H\delta_{ij} + E^{HH}\delta_{ij}\delta_{i+1,j+1}), \quad (11)$$

where $E^H = -0.4\ kT$ is the hydrogen bond energy, $E^{HH} = -0.5\ kT$ is the cooperativity energy, $\delta_{ij} = 0(1)$ if residues i and j are (not) hydrogen bonded. The scaling factor for the contribution of this term to the total energy is 1.0.

Total energy

The total energy of the model chain E_{tot} is given by the following formula:

$$\begin{aligned} E_{tot} = {} & 0.5E_{pair} + 4.25E_{tem} + 0.5E_{one} + E_{harm} + 0.5E_{rot} \\ & + E_\beta + 0.25E_{R14} + E_{HB} \end{aligned} \quad (12)$$

The scaling factors appearing in Eq. 12 were introduced because our force field is a combination of potentials statistically derived from a structural database and semi-empirical terms (e.g., hydrogen bonds). These scale factors are necessary to maintain the proper balance between the short- and long-range energy contributions, to obtain the proper level of secondary structure at high temperatures, and to introduce cooperativity into the folding transition. They were worked out by Vieth et al. (1995). It is important to note that they were derived for different systems (coiled coils), and their application to the systems studied here demonstrates the robustness of the model. We did not change the overall relative balance between secondary and tertiary interactions; however, to achieve this balance, some parameters were adjusted. First, we lowered the strength of the hydrogen bonds and their cooperative terms (from -0.6 to -0.4 and from -0.75 to -0.5, respectively). Second, we enforced the generic (amino-acid-independent) term of Ramachandran potential (we applied a 4.5 factor instead of 3). Third, the values of cutoff radii R_{ij} were lowered by a factor of 20%. These changes were done to maintain similar contributions from long- and short-range interactions and to lower the helix content (which is apparently too high) in the unfolded state. These changes make this potential close to that originally developed for globular proteins by Kolinski and Skolnick (1994a). The importance of particular components of the force field was studied previously for simple polypeptides (Kolinski and Skolnick, 1992). In our initial simulations with the force field exactly the same as used previously, we found that the fraction of helical states in the unfolded chain was too high. To make our model more physical, we introduced the changes described above. It must be noted that the former force field led to the same

folds as the new one. The changes introduced amplify the specific tertiary interactions and thus avoid a bias toward the molten globule state. The above force field was not self-consistently derived. When applied to different kinds of proteins (e.g., globular proteins versus coiled coils), the scaling factors slightly differ. Future work is clearly required to produce a universal force field for all protein types.

RESULTS AND DISCUSSION

General considerations

In this paper, we describe our results concerning the simplest polypeptide chains, i.e., those with a small diversity of amino acid residues. We studied three designed proteins introduced by Regan and DeGrado (1988): 1) α_1, Ac-helix-$CONH_2$; 2) α_2, Ac-helix-loop-helix-$CONH_2$; and 3) α_4, MET-helix-loop-helix-loop-helix-loop-helix-COOH, where helix is GLY-GLU-LEU-GLU-GLU-LEU-LEU-LYS-LYS-LEU-LYS-GLU-LEU-LEU-LYS-GLY, loop is -PRO-ARG-ARG, Ac is acetyl, and $CONH_2$ is carboxamide.

Based on the heptad amino acid reptat (Hodges et al., 1981), it is believed that all of these polypeptide chains should form four-member α-helical bundles. In α_4, a monomeric conformation should be adopted. α_2 should be a dimer of two helical hairpins and α_1 should be a tetramer of four helices.

Monte Carlo simulations were performed for many independent runs. The temperature range was chosen based on results from an initial series of simulations. The temperature was high enough to allow the model polypeptide chain to probe the entire configurational space and not be trapped in deep local energy minima. On the other hand, the temperature was low enough to allow for the formation and dissolution of the elements of folded structures.

During each simulation run, the temperature of the system was gradually lowered. The starting high-temperature, extended configuration was chosen at random. Every time the temperature was lowered, the equilibration of the model system was performed. Subsequently, the actual production Monte Carlo simulation run was executed. The simulation time (the number of Monte Carlo steps) was of the order of 10^5 Monte Carlo cycles at every temperature. A single equilibration run consisted of 10^4–10^5 Monte Carlo cycles.

For α_1, α_2, and α_4 proteins, in more than 50% of the Monte Carlo simulation runs, we obtained structures that folded to a four-member, α-helical bundle (see Table 1, below). We chose these structures because their energy was the lowest and they were the only ones that were adopted repeatedly. All of the remaining low-temperature structures were collapsed, but random. All folds presumed to be correct (on the basis of the average energy) subsequently underwent long isothermal simulations (correspondingly long isothermal runs for other randomly collapsed structures did not change their size, shape, contacts, or other properties). The temperature of these isothermal simulations was below the folding transition. During these isothermal simulations, the harmonic potential (Eq. 4) was turned off to allow the system to relax without additional constraints. Constant-low-temperature simulations were performed to obtain information about the stability of a given fold and to obtain the following statistics for parameters describing the folded state: the mean-square radius of gyration, the total energy of a system, and the helix content. Additionally, we obtained some dynamic (time-dependent) data concerning the behavior of the obtained folds.

TABLE 1 Statistics of α_1, α_2, and α_4 folds obtained in Monte Carlo simulations

Protein	Topology*	% population	Average RMSD (Å)	Radius of gyration (kT units)	Average total energy
α_4	I	84	3.7	12.4	−127.3
	II	16		12.7	−124.7
α_2	I	13	3.8	12.0	−125.1
	II	37	3.7	11.9	−132.1
	III	25	4.1	12.1	−124.2
	IV	19	4.0	11.8	−120.3
	V	6		12.6	−122.8
α_1	I	20	4.2	11.7	−125.1
	II	45	3.8	11.7	−131.2
	III	15	3.6	12.0	−125.3
	IV	20	3.9	11.9	−128.6

*All possible (and obtained) topologies are shown in Fig. 3. The numbering of topologies used here refers to that provided in Fig. 3.

α_4 protein

In the case of a single polypeptide chain containing 74 residues, α_4, in 7 of 15 independent attempts, we obtained four-helical bundle folds. The folding transition occurred over a very narrow temperature range between 2.55 and 2.5. The assembly process was rather fast and cooperative. All of the obtained folds were α-helical bundles. In Fig. 1, we present a typical pathway of the folding of the α_4 protein. In Fig. 1 *a*, the starting conformation can be described as a random coil. At a temperature of 2.6, these structures exhibited a mean-square radius of gyration $\langle S^2 \rangle$ roughly equal to 230, a total energy, $E_{tot} \approx -32$, and a helix content on the level of 10%. Next (Fig. 1 *b*), a helical hairpin is formed by on-site construction (Kolinski and Skolnick, 1996). In most cases, it is a central hairpin composed of helices 2 and 3 (residues 21–54). For this hairpin, the pair of helices folded almost simultaneously and very quickly, but this structure is rather unstable. It forms and dissolves many times. Then (Fig. 1 *c*), the third helix is assembled. This three-helical structure is much more stable and can be described as a well defined, topological intermediate. In some cases, it is not fully in register and, thus, underwent subsequent rearrangements or dissolved. If an out-of-register hairpin survives for longer periods of time, we usually obtain misfolded structures at lower temperatures; these are described below. In the last stage (Fig. 1 *d*), during which the folded state forms,

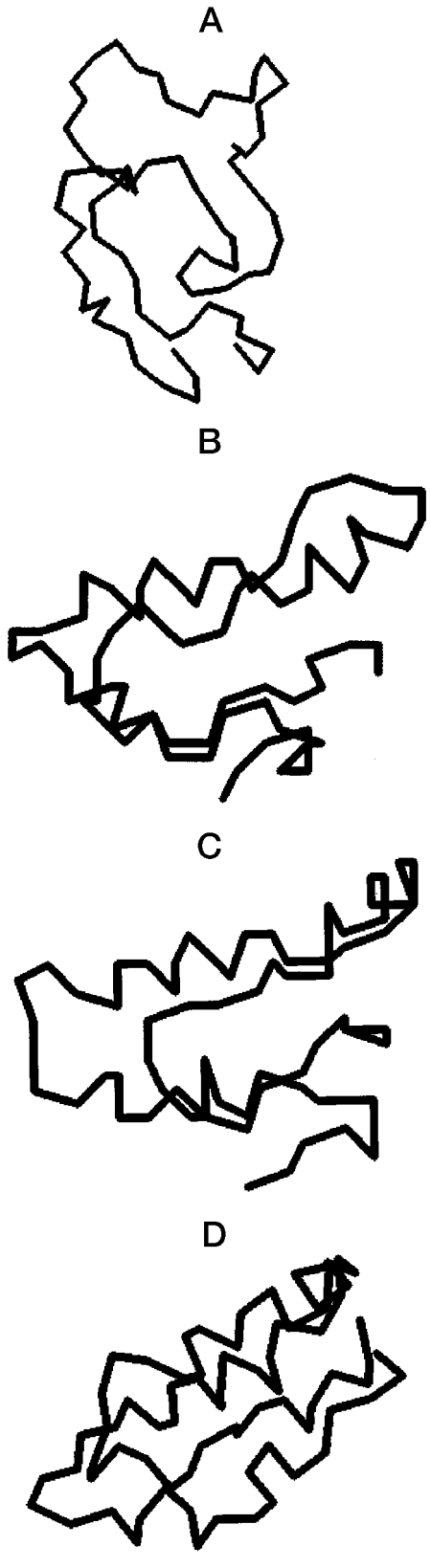


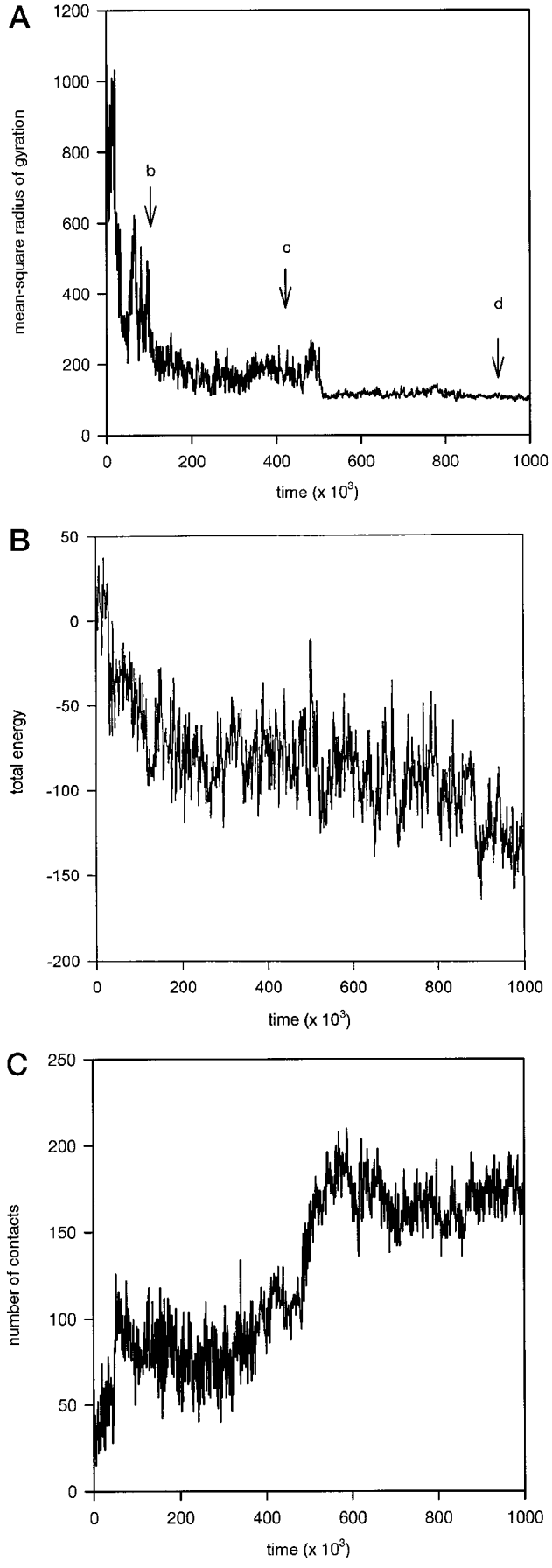
FIGURE 1 Typical folding pathway for α_4. The starting conformation is a random coil. As the final result of simulation, the right-handed topology fold of α_4 is obtained.

the fourth helix is assembled on-site. This pathway is exactly the same as in our previous simulations of simpler models on a tetrahedral lattice (Sikorski and Skolnick, 1989a,b, 1990a,b), on the (201) lattice (Kolinski et al., 1993) and in off-lattice Monte Carlo simulations (Rey and Skolnick, 1993). This is another confirmation of the robustness of lattice models for studying protein properties.

Analysis of the behavior of the mean-square radius of gyration and total energy during the simulation run yields a similar picture of the assembly process. In Fig. 2, *A–C*, we present flow charts of S^2, E_{tot}, and the number of side group contacts. One can see that the mean-square radius of gyration changes from high-temperature values (on the level of a few hundred) to some lower value (on the level of 150). This corresponds to a relatively stable three-member helical bundle. Then, S^2 becomes close to 100, and its fluctuations are less. This means that the system has approached the folded state. The behavior of the total energy of the system is similar in the sense that one can distinguish three regimes: high-temperature, intermediate (three helices assembled), and folded. The main difference between the behavior of S^2 and E_{tot} is that the latter exhibits much greater fluctuations, which do not disappear in the folded state. In the flow chart for the number of contacts versus time, one can distinguish the same regimes and one can also see large fluctuations. In the native state, the number of contacts per residue is close to 3, which corresponds to the maximal value for helical chains in this model.

Apart from out-of-register states, another type of misfolded intermediate contains three helices. This occurs when the centers of the helices are located at the vertices of a triangle instead of at three vertices of a square (see Fig. 3). This three-helical bundle is relatively stable because the hydrophobic faces of these helices are almost completely buried, and the remaining hydrophilic residues are exposed to the solution. The occurrence of these misfolds can be reduced by the proper selection of force field (Rey and Skolnick, 1993). A structure containing three helices was also found experimentally by DeGrado, Eisenberg, and coworkers (Betz et al., 1996) for a de novo designed protein. The other popular misfold, the Z-bundle (Fig. 3 *b*), never appeared in our simulations of the α_4 polypeptide. This Z-bundle structure was obtained by Troyer et al. (1994) in a Brownian dynamics simulation of a model of α_4 based on an α-carbon representation. This structure is very unlikely, and in our hands, it is not very stable. We made some Monte Carlo simulation runs starting from this configuration at the folding temperature. The Z-bundle structure was preserved during the simulation run, but its energy was significantly higher ($-108\ kT$) in comparison with other folds (see below).

In some simulations ($\sim$30% of the simulation runs), we obtained collapsed structures with high helicity and a large number of hydrophobic contacts (but these contacts were not reproduced in the next set of simulation runs). The energies of such structures were approximately $30-40kT$ greater than those of a properly folded state. The existence of these structures was probably caused by system quenching.



It should be pointed out that a four-member, α-helical bundle can exist in two different topological states as a left-handed bundle and as a right-handed bundle. Both topologies are schematically presented in Fig. 3. In previous Monte Carlo simulations of the α_4 sequence (Handel and DeGrado, 1992) on a coarser lattice (201), both topologies were found (Kolinski et al., 1993) and were very close in energy. The prediction was subsequently confirmed by experiment.

Because we obtained only one left-handed bundle in these Monte Carlo simulations, we checked the stability of this alternate fold as follows. For some folds (right-handed bundles), the configuration, not the sequence, of the chain was read backward (Olszewski et al., 1996). As a result, left-handed bundles were generated, and then this system underwent a long isothermal Monte Carlo run just below the folding temperature. In all cases, the folded chain was stable and exhibited structural parameters (the mean-square radius of gyration, the total energy, and side group contacts) that were almost the same as for the right-handed bundles obtained in Monte Carlo annealing experiments. The energy of these left-handed folds was $-124.7kT$, as compared with the average energy of the original right-handed fold of $-127.3kT$. It is impossible to state whether any of these two topologies is correct because we do not have a large number of independently generated structures. Moreover, the energy difference between these two folds is very small ($2.3kT$), and it seems that both forms would appear in real experiments.

In all folded bundles, their secondary structure (helices) was well defined. The differences between a pair of two structures can be expressed in terms of the RMSD between α-carbons. Because we do not know the crystal structure of α_4 or the other proteins under consideration, we confine our discussion to a comparison of folds belonging to the same topology. The results are collected in Table 1. One can see that the average RMSD for the right-handed α_4 folds is 3.7 Å and is 4.1 Å for the left-handed fold. Similar deviations were found between the known crystal structures and folds obtained in simulations for small globular proteins (Kolinski and Skolnick, 1994b).

In Fig. 4, we present a representative contact map for the folded structure of α_4. This map is very typical of an α-helical bundle. But, as mentioned above, we suspected that this structure has the features of a molten globule state because of fluctuations in the radius of gyration and especially because the value of total energy during long isothermal runs is rather high. A molten globule compact state is usually characterized as having substantial secondary structure, but the packing of the side groups in the hydrophobic

FIGURE 2 For α_4, a flow chart of (*A*) the mean-square radius of gyration, (*B*) the total energy, and (*C*) the number of contacts versus time (letters b, c, and d mark times corresponding to the structures b, c, and d in Fig. 1). The left-hand portion shows the behavior of the fold during annealing and the right-hand portion during the long isothermal run.

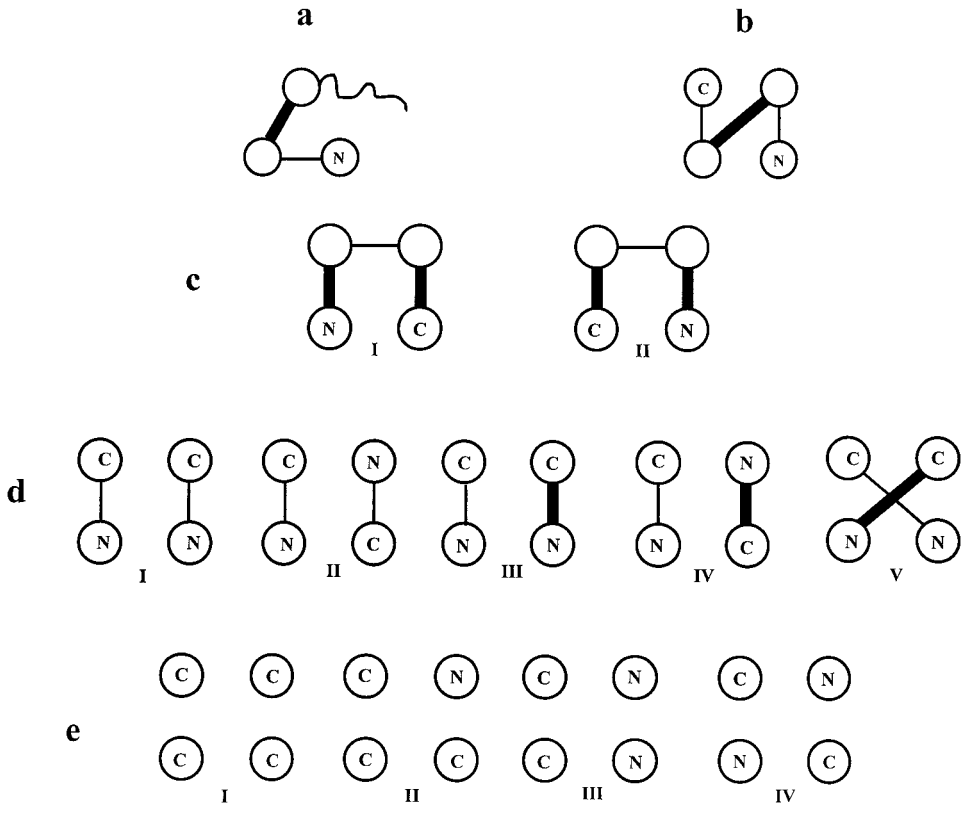


FIGURE 3 All possible topologies of folded four-helix bundles. The schematic representation of the most populated misfolds of α_4 are (*a*) triangle and (*b*) Z-bundle. (*c*) α_4 right-handed bundle (I) and left-handed bundle (II). (*d*) Five topologies of α_2 (all mutual orientations of two helical hairpins). (*e*) Four topologies of α_1 (all combinations of parallel and antiparallel four helices).

core is not unique (Ptitsyn, 1992; Kuwajima, 1989). Here, the liquid-like structure of the hydrophobic core is confirmed by the analyses of the lifetime of side group-side group contacts (Ptitsyn 1992). In Fig. 4, we also present a typical lifetime contact map with contacts marked by squares, with the intensity of gray proportional to the lifetime of a given contact. One can observe that with the sole exception of a few contacts, almost none survive the entire simulation (usually 100,000 Monte Carlo steps). Similar to the work of Kolinski and Skolnick (1994b), the relaxation of rotamers is very fast.

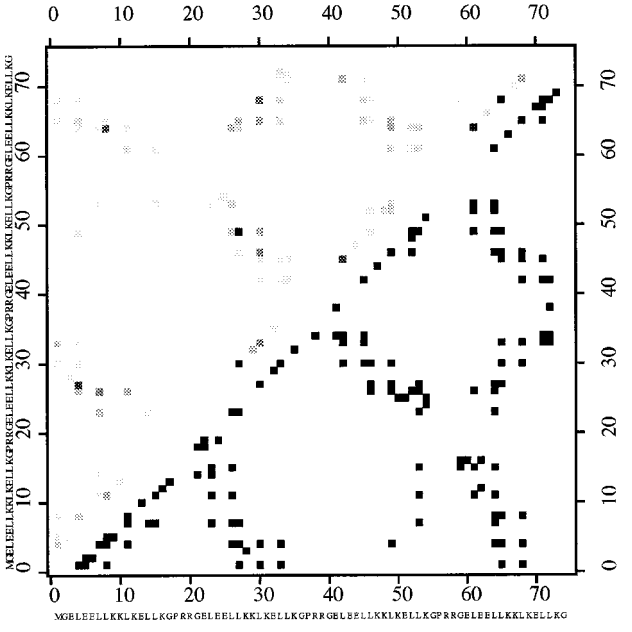


FIGURE 4 For the case of right-handed, four-helix bundle topology 1 of α_4, the average side-chain contact map (lower right corner) and the lifetime contact map obtained from a long isothermal run is presented, with contacts marked by squares, for which the intensity of gray is proportional to the contact lifetime (upper left corner).

It was previously shown (Kolinski and Skolnick, 1994b) that this model applied for natural proteins reproduces the features of a folded native state as well as a molten globule state. Here, we do not know the structure of the folded (native-like) state. To check whether we can also reproduce the properties of a native-like state, we performed some isothermal Monte Carlo simulation runs using our α_2 folds as a starting conformation, but with the sequences of $\alpha_2 C$ and $\alpha_2 D$ (Raleigh and DeGrado, 1992; Raleigh et al., 1995), which are believed to have some features of native proteins. It appeared that the resulting helical bundles are stable, with smaller fluctuations in the radius of gyration and energy. The number of contacts that survived the entire simulation ($\sim 10^6$ Monte Carlo steps) was significantly greater than for our simple sequences.

α_2 protein

The folding of two α_2 protein chains (two identical 35-residue sequences) occurs under almost the same temperature conditions as that of α_4, although the folding temper-

ature is probably slightly lower. Annealing runs were done for temperatures between 2.55 and 2.45. Of 35 independent attempts, 18 resulted in a four-member α-helical bundle built from two hairpins. The remaining structures have much higher energy. A typical folding pathway is presented in Fig. 5. There is no essential difference between the folding of pathway α_2 and that of α_4. Folding consists of the following steps. a) The initial conformation of two chains, each consisting of 35 residues, can be described as a random coil. At a temperature of 2.6, the mean-square radius of gyration $\langle S^2 \rangle$ is roughly equal to 203, the total energy $E_{tot} \approx -59$, and its helix content is 10%. b) Then, one helical hairpin, which is a very unstable and short-lived structure, forms. The first hairpin that appears is usually constructed from one chain (16 cases) but sometimes is built from two helices, each from a different chain (two cases). c) Next, the third helix from the second chain assembles. The three-helical bundle built from α_2 chains has a lifetime similar to that of α_4, and similar triangular misfolds appear during the simulations. d) After a considerably longer time, the last helix assembles on-site.

The flow charts of the mean-square radius of gyration and total energy are presented in Fig. 6, *A–C*. (compare with Fig. 2, *A–C*). The existence of the three-member, helical bundle intermediate state is confirmed. The fluctuations of the size, energy, and the number of contacts of the folded state are slightly higher than in α_4. This reflects the absence of a turn between a pair of helices when α_2 is compared with α_4.

Two α_2 chains can form the five distinguishable four-helix bundle topologies shown in Fig. 3 *d*. This difference arises from the mutual orientations of helical hairpins. A loop (turn) can be located near the other loop or near the COOH and NH_2 termini. Moreover, quite different inter-helical packing is possible (e.g., the helices may be crossed). In our Monte Carlo simulations, we obtained all topologies. The results are presented in Table 1. The populations of all topologies are equally distributed (with one exception, see below), and they are not distinguishable on the basis of their energies or their sizes (mean-square radius of gyration). All exhibit parameters that are very close to those of α_4. These native-like folds are stable, and they do not dissolve during long isothermal simulation runs.

The case of topology 5 is very interesting because this structure, and only this one, was recently found experimentally for a different sequence (W.F. DeGrado, private communication). In our simulation, this structure appeared only once, and its behavior does not differ significantly from the other obtained topologies. As it was pointed out in the previous section, the number of particular folds and their energy differences suggest that all forms could appear in real experiments.

In Fig. 7, we present the contact map of a predicted fold of the $\alpha 2$ dimer that has topology 2. The contact map with information about the contact lifetime is also included. One can see that there is no qualitative difference between these contact maps and those of α_4. This is reasonable given that α_2 and α_4 form very similar helical bundles. Thus, as seen for the α_4, the folded α_2 dimers also have features of molten globules.

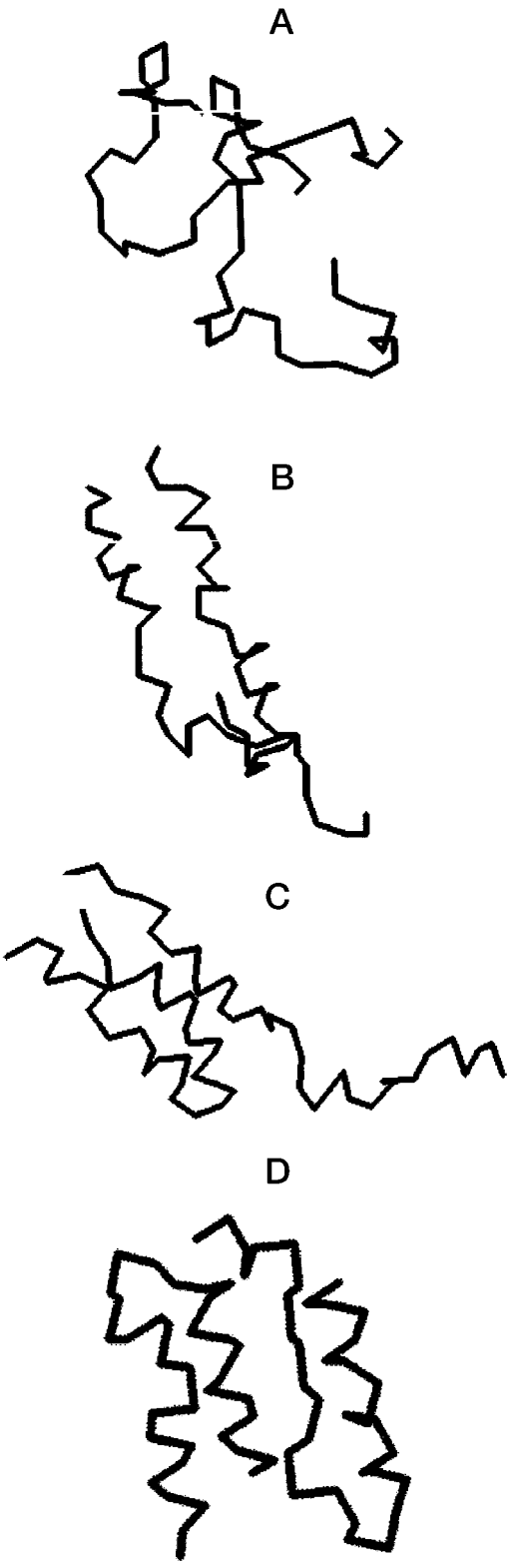


FIGURE 5 Typical pathway of folding of α_2 dimer. The final fold resulting from this simulation has topology 2.

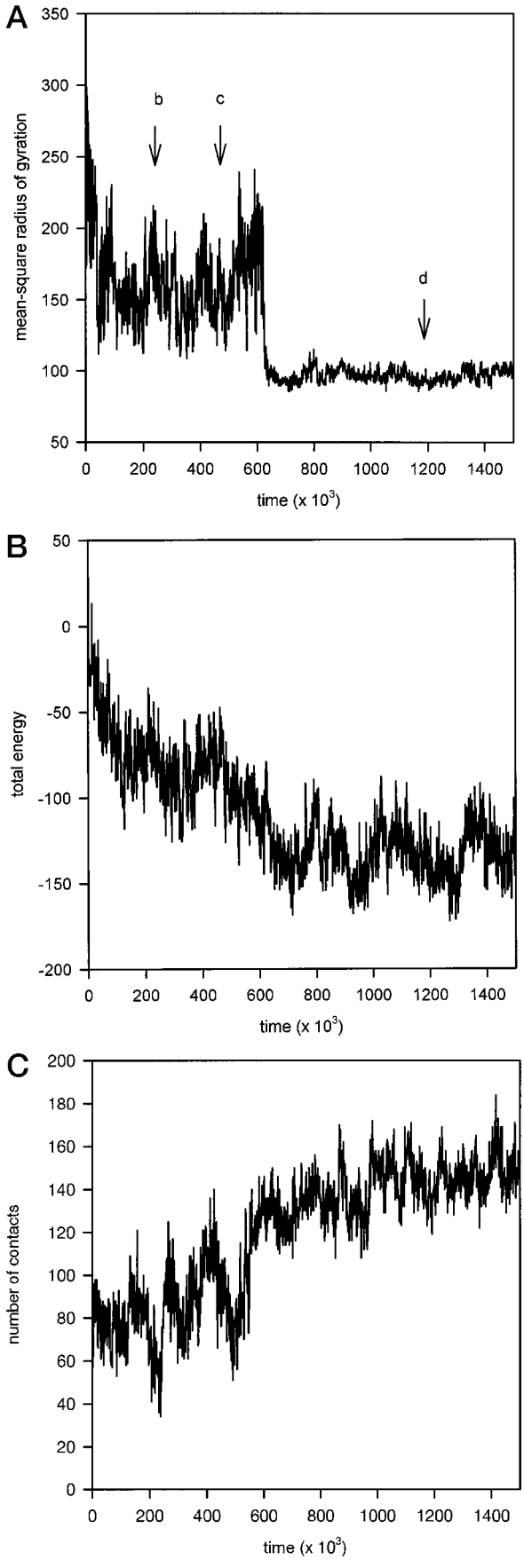


FIGURE 6 For a dimer with topology 2, a flow chart of (*A*) the mean-square radius of gyration, (*B*) the total energy, and (*C*) the number of contacts versus time (letters b, c, and d mark times corresponding to the structures b, c, and d in Fig. 5). The left-hand portion shows the behavior of the fold during annealing and the right-hand portion during the long isothermal run.

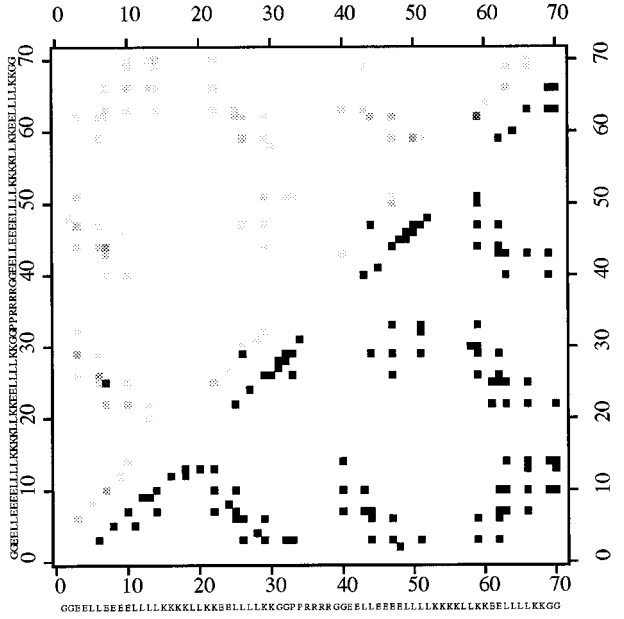


FIGURE 7 The average contact map of the folded structure of α_2 dimer for topology 2 (lower right corner) and the lifetime contact map with contacts marked by squares for which the intensity of gray is proportional to the contact lifetime (upper left corner).

The precision of the prediction of a particular folded state can be measured in terms of the RMSD within a given topology. Similar to an α_4 chain, we do not have crystal structure data. The results are presented in Table 1. One can see that for a given topology, the RMSDs are very close and lie between 3.5 and 4.2 Å. The helices are well defined, and the average distance between them is equal to 11.9 Å.

α_1 protein

The α_1 protein behaves similarly to the preceding two cases. The model system was composed of four identical chains, each with sixteen residues. The range of temperatures in which folding occurs is almost the same as for α_4 and α_2 (between 2.45 and 2.35). At a temperature of 2.6, the system is described by a mean-square radius of gyration $\langle S^2 \rangle$ roughly equal to 120, a total energy $E_{\text{tot}} \approx -50$, and a helix content on the level of 10%, at a temperature of 2.6. The folding pathway shown in Fig. 8 is as follows. a) Folding begins from a high-temperature state. b) In the beginning, a dimer is built from two chains. This process is usually cooperative, and the helices zip up very fast. c) Then, the third helix assembles to form a three-helix bundle, which is

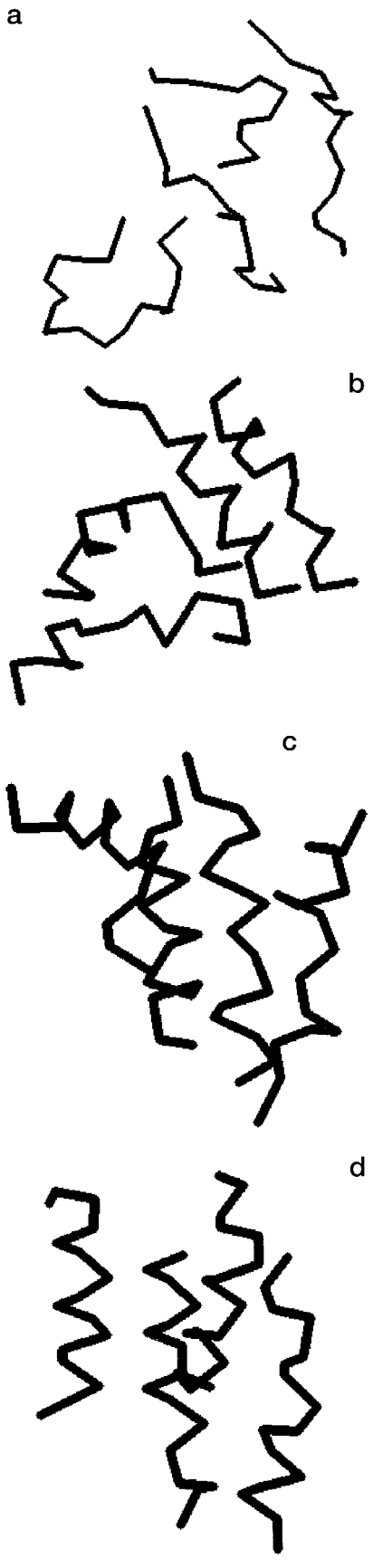


FIGURE 8 Typical folding pathway of an α_1 tetramer having topology 2.

a long-lived intermediate. d) Finally, assembly of the last helix occurs. This process is much faster than in α_4 and α_2. This could be explained by the fact that this fourth chain is not connected to the already folded structure as in α_4 and α_2. And, this fourth chain is kept in the general vicinity of the folded structure by the harmonic potential. The process of building a given helix is very fast; thus, it is difficult to characterize the actual assembly mechanism. But in this case, there is a considerably larger number of misfolds where the helices are out of register, but the average lifetime of these misfolds is shorter. The cutoff radius in the harmonic potential has no influence on the structure of the final fold, but a smaller value does speed up the assembly process. The flow charts of the mean-square radius of gyration and of the total energy presented in Fig. 9, *A–C*, is similar to those of α_4 and α_2. Of course, compared with α_4 and α_2, the fluctuations in the folded state are greater because α_1 has four chain ends and there are no turns connecting helices. The dissolution of the folded structure at the transition temperature was observed as in the α_4 and α_2 systems.

Four α_1 chains can exist in the four different topological states, schematically shown in Fig. 3 *e*. The different topologies arise from the fact that parts of chains (helices) in a tetramer can be assembled in parallel and antiparallel orientations. In our Monte Carlo simulations, we obtained 16 folds (out of 25 attempts), and all the topologies were represented. The properties of all of these folds are presented in Table 1. As one can see from Table 1, there is almost no energy difference among these species. We also present data concerning the dimensions of the folded states and their population. The number of observed different folds is almost equal with the exception of topology 2. The antiparallel bundle might be expected (Cohen and Parry, 1990; Weber and Salemme, 1980). The secondary structure is well defined in all folded topologies, and the average distance between helices does not depend on the topology and is equal to 11.8 Å. Experimental data are available for a shortened version of α_1 only, and this distance is equal to 12.9 Å (which is very large compared with naturally occurring helical bundles) (Hill et al., 1990).

In Fig. 10, the contact map plotted for α_1 (with folded topology 2) is representative of the tetramer and is very characteristic of a helical bundle and very similar to those for the α_4 and α_2 model systems. The lifetime contact map is also plotted for the same topology on this figure. This figure indicates that in the case of the α_1 tetramer, only very few contacts survive the entire simulation run. Hence, this structure also has features of a molten globule, as do α_2 and α_4 described above.

For comparison, we also simulated simple polypeptides containing only glycine (polyG), valine (polyV), and alanine (polyA). These peptides had the same length and number of chains as α_1, α_2, and α_4. The simulations were carried out with the same force field but in different temperature ranges because the temperature of the folding transition was different for these polypeptides. In the case of a single chain version of polyG, the low temperature leads to

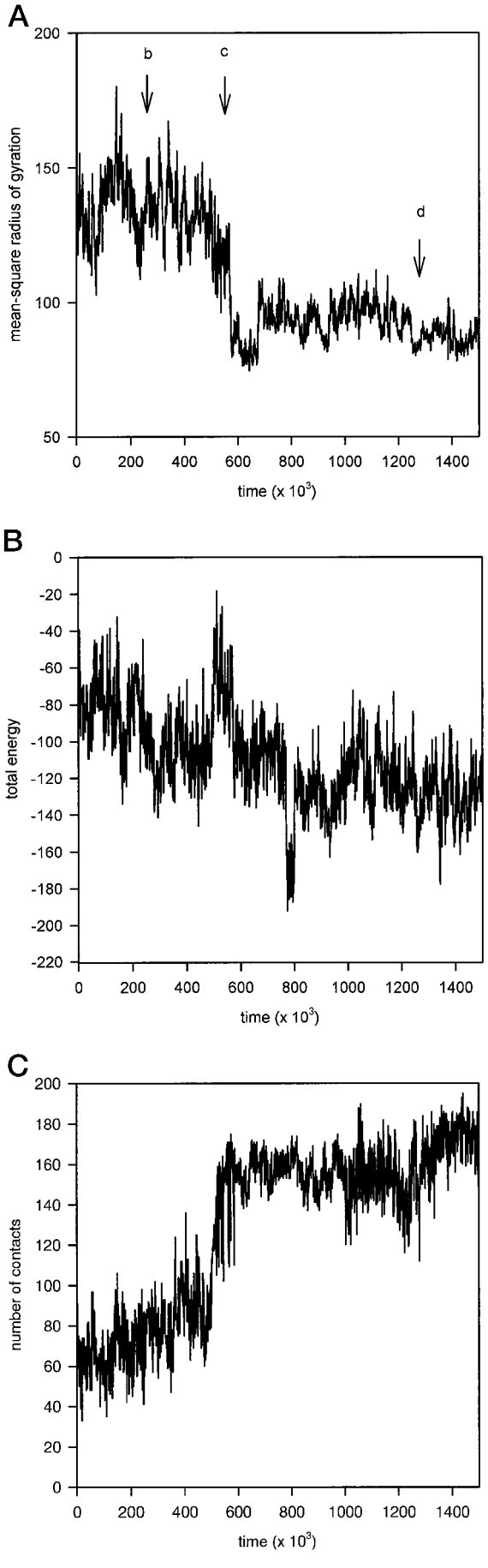


FIGURE 9 For an α_1 tetramer having topology 2, a flow chart of (*A*) the mean-square radius of gyration, (*B*) the total energy, and (*C*) the number of contacts versus time (letters b, c, and d mark times corresponding to the structures b, c, and d in Fig. 10). The left-hand portion shows the behavior of the fold during annealing and the right-hand portion during the long isothermal run.

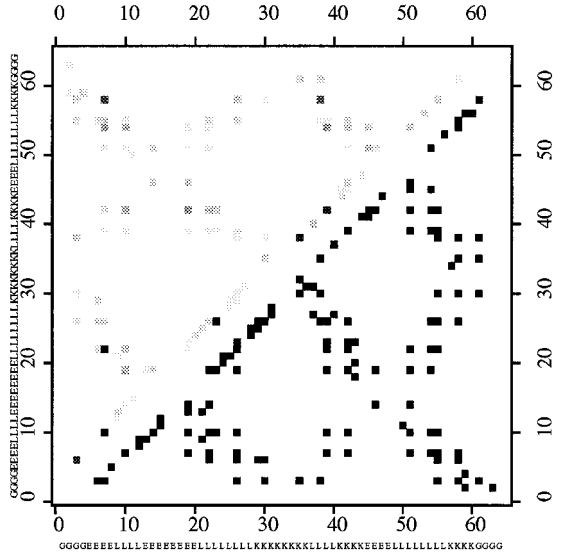


FIGURE 10 The average contact map of the folded structure of an α_1 tetramer having topology 2 (lower right corner) and the lifetime contact map with contacts marked by squares for which the intensity of gray is proportional to the contact lifetime (upper left corner).

the collapse of the chain into a dense globule with no detectable secondary structure. The same result was observed for systems consisting of two and four chains. The topology of the obtained folded structures was in every case quite different. At low temperatures, the polyV chains form compact structures with a high degree of β-like secondary structure. Thus, the simulations for polyG and polyV show that the model is not directly biased toward a four-member helical bundle.

In the case of polyA, all folded chains have a high fraction of helicity. Compared with the De Grado proteins, the helices in the folded polyA are more mobile and their fluctuations are larger. The folded structures are not unique; the RMSD between the particular folds is on the level of 10 Å.

Very similar results were obtained for slightly different lattice models of polyA and polyV (Kolinski and Skolnick, 1992).

CONCLUSION

The α_1, α_2, and α_4 proteins designed by DeGrado and co-workers were expected to form a monomer, dimer, and tetramer four-member α-helical bundle, respectively. According to experiment, their folded structures are compact

with a high degree of helicity, but their hydrophobic core is rather loosely defined; thus, they exhibit many features of the molten globule state. We applied the model of a protein chain based on a (310) lattice α-carbon representation (Kolinski and Skolnick, 1994a) and employed a force field similar to that used recently in simulations of helical coiled coils (Vieth et al., 1995). We performed a series of independent simulated annealing runs for all three chains under consideration: α_1, α_2, and α_4. The obtained folded structures then underwent a long isothermal run. We obtained several native-like folds for all proteins studied. It was found that all possible topologies of a folded four-helix bundle were obtained in the simulation process. With respect to a given topology, we were able to reproducibly fold with an RMSD of 3.5–4.2 Å for these three designed proteins. All are of very similar size and total energy, and thus, their stability should be comparable. The folds have well defined secondary structures (helices). Additional simulations of simple polypeptides (polyglycine, polyvaline, and polyalanine) showed that the simulations were not biased toward helical folds or toward a unique four-member, helical bundle.

All of these folds appear to adopt molten globule states rather than the well defined native states characteristic of natural globular proteins. Within a given topology, side chain contacts are very unstable and short-lived. This could be caused by the very simple design of all of these proteins. Diversification of residues should assist in the selection of one unique, native-like structure. Such studies are currently underway.

It should be pointed out that in simulations of the same models of other helical proteins (protein A, ROP) the final state had native-like packing with a fixed network of tertiary contacts. Thus, it seems that the molten globule-like feature of a designed helical protein is not built into the model but is a result of the lack of specificity of tertiary interactions because of the uniform composition of the hydrophobic core.

The pathways of early-stage fold assembly were found to be very similar to those obtained for other helical proteins studied within the framework of similar models (protein A, ROP). The mechanism of folding can be described as the on-site formation of a short helical hairpin and then the fast assembly of a third helix. This three-helical bundle is a relatively long-lived, topological intermediate.

Comparing the results of present studies with the results obtained in simulations of the coarse (210) lattice model, one should note that apart from a much better lattice representation, the RMSD between two folds belonging to the same class of topology is exactly the same (3.6–4.5 Å). Here, use of a more sophisticated side group model led to the appearance of a few longer-lived contacts, but most contacts still dissolved during the simulation run. These observations are consistent with experiment in that the folds of these simple helical designed proteins have features of molten globules.

The helpful assistance of Dr. Michal Vieth is gratefully acknowledged.
This work was partially supported by National Institutes of Health FIRCA grant TW-00418 and grant GM-38794 and by the University of Warsaw grant BST-532/34/96.

REFERENCES

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Simanouchi, and M. Tasumi. 1977. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Betz, S. F., J. W. Bryson, and W. F. DeGrado. 1995. Native-like and structurally characterized designed α-helical bundles. *Curr. Biol.* 5:457–463.

Betz, S. F., D. P. Raleigh, and W. F. DeGrado. 1993. De novo protein design: from molten globules to native-like states. *Curr. Opin. Struct. Biol.* 3:601–610.

Betz, S. F., D. P. Raleigh, W. F. DeGrado, B. Lovejoy, D. Anderson, N. Ogihara, and D. Eisenberg. 1996. Crystallization of a designed peptide from a molten globule ensemble. *Folding Design*. 1:57–64.

Choma, C. T., J. D. Lear, M. J. Nelson, P. L. Dutton, D. E. Robertson, and W. F. DeGrado. 1994. Design of a heme-binding four-helix bundle. *J. Am. Chem. Soc.* 116:856–865.

Ciesla, D. S., D. E. Gilbert, and J. J. Feigon. 1991. Secondary structure of the designed peptide alpha-1 determined by nuclear magnetic resonance. *J. Am. Chem. Soc.* 113:3957–3961.

Cohen, C., and D. A. D. Parry. 1990. α-Helical coiled coils and bundles: how to design an α-helical protein. *Proteins Struct. Funct. Genet.* 7:1–15.

Creighton, T. E. 1993. Proteins: Structures and Molecular Properties. Freeman, New York.

Godzik, A., A. Kolinski, and J. Skolnick. 1993. Lattice representation of globular proteins: how good are they? *J. Comput. Chem.* 14:1194–1202.

Handel, T., and W. F. DeGrado. 1992. A designed 4-helical bundle shows characteristics of both molten globule and native state. *Biophys. J.* 61:A265.

Hill, C. P., D. H. Anderson, L. Wesson, and W. F. DeGrado. 1990. Crystal structure of a_1: implications for protein design. *Science*. 249:543–546.

Ho, S. P., and W. F. DeGrado. 1987. Design of a 4-helix bundle protein: synthesis of peptides which self-associate into a helical protein. *J. Am. Chem. Soc.* 109:6751–6758.

Hodges, R. S., A. S. Saund, P. C. S. Chong, S. A. St.-Pierre, and R. E. Reid. 1981. Synthetic model for two-stranded α-helical coiled coils. *J. Biol. Chem.* 256:1214–1224.

Kabsch, J., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 22:2577–2637.

Kolinski, A., Godzik, A., and J. Skolnick. 1993. A general method for the prediction of the three-dimensional structure and folding pathways of globular proteins: application to designed helical proteins. *J. Chem. Phys.* 98:7420–7433.

Kolinski, A., and J. Skolnick. 1992. Discretized model of proteins. I. Monte Carlo study of cooperativity in homopolypeptides. *J. Chem. Phys.* 97:9412–9426.

Kolinski, A., and J. Skolnick. 1994a. Monte Carlo simulation of protein folding. I. Lattice model and interaction scheme. *Proteins Struct. Funct. Genet.* 18:338–352.

Kolinski, A., and J. Skolnick. 1994b. Monte Carlo simulation of protein folding. II. Application to Protein A, ROP, and Crambin. *Proteins Struct. Funct. Genet.* 18:353–366.

Kolinski, A., and J. Skolnick. 1996. Lattice Models of Protein Folding, Dynamics and Thermodynamics. R. G. Landes, Austin, TX.

Kuwajima, K. 1989. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins*. 6:87–103.

Levitt, M., and J. Greer. 1977. Automatic identification of secondary structure in globular proteins. *J. Mol. Biol.* 114:181–293.

Milik, M., A. Kolinski, and J. Skolnick. 1997. An algorithm for rapid reconstruction of protein backbone from alpha carbon coordinates. *J. Comput. Chem.* 18:80–85.

Olszewski, K. A., A. Kolinski, and J. Skolnick. 1996. Does a backwardly read protein sequence have a unique native state? *Protein Eng.* 9:5–14.

Brookhaven Protein Data Bank. 1995. *Q. Newslett.* 17, January 1995.

Ptitsyn, O. B. 1992. The molten globule state. *In* Protein Folding. Creighton, T. E., editor. Freeman, New York. 243–299.

Quinn, T. P., N. B. Tweedy, R. W. Williams, J. S. Richardson, and D. C. Richardson. 1994. Betadoublet: de novo design, synthesis, and characterization of a β-sandwich protein. *Proc. Natl. Acad. Sci. U.S.A.* 91: 8747–8751.

Raleigh, D. P., Betz, S. F., and W. F. DeGrado. 1995. A de novo designed protein mimics the native state of natural proteins. *J. Am. Chem. Soc.* 117:7558–7559.

Raleigh, D. P., and W. F. DeGrado. 1992. A de novo designed protein shows a thermally induced transition from a native to a molten globule-like state. *J. Am. Chem. Soc.* 114:10079–10081.

Regan, L., and W. F. DeGrado. 1988. Characterization of a helical protein designed from first principles. *Science*. 241:976–978.

Rey, A., and J. Skolnick. 1993. Computer modelling and folding of four-helix bundles. *Proteins Struct. Funct. Genet.* 16:8–28.

Robertson, D. E., R. S. Fraid, C. C. Moser, J. L. Urbauer, S. E. Mulholland, R. Pidikiti, J. D. Lear, A. J. Wand, W. F. DeGrado, and P. L. Dutton. 1994. Design and synthesis of multi-haem proteins. *Nature*. 368: 425–432.

Sikorski, A., and J. Skolnick. 1989a. Monte Carlo simulation of equilibrium globular protein folding: α-helical bundles with long loops. *Proc. Natl. Acad. Sci. U.S.A.* 86:2668–2672.

Sikorski, A., and J. Skolnick. 1989b. Monte Carlo study on equilibrium globular protein folding. III. The four helix bundle. *Biopolymers*. 28: 1097–1113.

Sikorski, A., and J. Skolnick. 1990a. Dynamic Monte Carlo simulations of globular protein folding/unfolding pathways. II. α-helical motifs. *J. Mol. Biol.* 212:819–836.

Sikorski, A., and J. Skolnick. 1990b. Dynamic Monte Carlo simulations of globular protein folding: model studies of in vivo assembly of four helix bundle and four member β-barrels. *J. Mol. Biol.* 215:183–198.

Smith, D. D. S., K. A. Pratt, I. G. Sumner, and C. M. Henneke. 1995. Greek key jelly roll protein motif design: expression and characterization of a first generation molecule. *Protein Eng.* 8:13–20.

Tanaka, T., Y. Kuroda, H. Kimura, S. Kidokoro, and H. Nakamura. 1994. Cooperative deformation of a de novo designed protein. *Protein Eng.* 4:969–976.

Troyer, J. M., F. E. Cohen, and D. M. Ferguson. 1994. Langevin dynamics of simplified protein models. First Electronic Computational Chemistry Conference, available at http://www.cmpharm.ucsf.edu/~troyer.

Vieth, M., A. Kolinski, C. L. Brooks III, and J. Skolnick. 1994. Prediction of the folding pathways and structure of the GCN4 leucine zipper. *J. Mol. Biol.* 237:361–367.

Vieth, M., A. Kolinski, C. L. Brooks III, and J. Skolnick. 1995. Prediction of the quarternary structure of coiled coils: application to mutants of the GCN4 leucine zipper. *J. Mol. Biol.* 251:448–467.

Weber, P. C., and F. R. Salemme. 1980. Structural and functional diversity in 4 α-helical proteins. *Nature* 287:82–84.

Yan, Y., and B. W. Erickson. 1994. Engineering of betabellin 14D: disulfide-induced folding of a β-sheet protein. *Protein Sci.* 3:1069–1073.