Computer Simulations of the Properties of the α_2, α_2C, and α_2D De Novo Designed Helical Proteins

Andrzej Sikorski,[1]* Andrzej Kolinski,[1,2] and Jeffrey Skolnick[2]
[1]*Department of Chemistry, University of Warsaw, Warsaw, Poland*
[2]*Department of Molecular Biology, The Scripps Research Institute, La Jolla, California*

ABSTRACT **Reduced lattice models of the three de novo designed helical proteins α_2, α_2C, and α_2D were studied. Low temperature stable folds were obtained for all three proteins. In all cases, the lowest energy folds were four-helix bundles. The folding pathway is qualitatively the same for all proteins studied. The energies of various topologies are similar, especially for the α_2 polypeptide. The simulated crossover from molten globule to native-like behavior is very similar to that seen in experimental studies. Simulations on a reduced protein model reproduce most of the experimental properties of the α_2, α_2C, and α_2D proteins. Stable four-helix bundle structures were obtained, with increasing native-like behavior on-going from α_2 to α_2D that mimics experiment. Proteins 2000;38:17–28.**

Key words: lattice models; Monte Carlo method; protein folding; 4-α-helix bundles

INTRODUCTION

A hierarchical approach to the design of -helical proteins was proposed by DeGrado[1] and coworkers. Here, short (*12–16* residues) polypeptides called α_1 were designed to adopt an α-helical tetramer, and thus, on association, they should form a four-helix bundle. Then, a longer polypeptide, called α_2 was built from two α_1 chains connected by a short loop. This polypeptide was expected to fold into an α-helical hairpin and form a similar four-helix bundle upon dimerization. [1] Finally, two α_2 chains were connected by the same short loop. The resulting chain, α_4, should form a tetrameric four-helix bundle.[2] Experiments have shown that α_4 is indeed monomeric and adopts a globular compact structure with high helical content. The folded conformation(s) of α_4 is more stable[2] than both α_1 and α_2. Other, similar de novo designs were recently published.[3,4] Their simplicity, low molecular weight, various native-like features, and the high symmetry of the putative native fold make them excellent subjects for simulation studies.

None of the designed proteins mentioned above exhibits all the features of native helical proteins. For example, α_1 clearly behaves as a molten globule. In the molten globule state even though the molecule adopts a native-like topology, the packing of side groups is not specific; rather, it is liquid-like.[5,6] This behavior results from the fact that the hydrophobic cores of these proteins are loosely packed with mobile contacts between side groups. This effect apparently arises from the presence of all leucines in the hydrophobic core. Leucine side groups can adopt a number of geometrically very similar rotameric states, thereby making possible significant mobility of the core.[3,7] In addition, there is no packing heterogeneity associated with different possible side chain arrangements because the identity of all the hydrophobic residues is the same.

In the GTD-43 polypeptide, Baltzer et al.[8,9] demonstrated experimentally that the introduction of some aromatic amino acid residues restricts the internal mobility of α_2, the polypeptide chain designed to be a helical hairpin. However, the modified polypeptide still binds 8-anilino-l-naphtalenesulfonate (ANS) dye, which suggests that its hydrophobic core is not as well packed as that in naturally occurring globular proteins.

Subsequently, DeGrado and coworkers[4] introduced additional sequence diversity into the hydrophobic core. Some leucine residues were replaced by Val, Ile, Phe, and Trp, thereby considerably improving the design of the helix-helix interfaces in the putative helical bundle built from two hairpins. The resulting polypeptide was called α_2C. It should be pointed out that the folded structure of α_2C is a heterodimer because the two chains that form the bundle have different sequences. The nuclear magnetic resonance (NMR) study of folded α_2C clearly demonstrated that this molecule exhibits molten globule behavior at its folding temperature. At lower temperatures, it undergoes a transition to a more native-like state. Nevertheless, not all the features of the molten globule disappear at low temperatures. The folded state still binds hydrophobic dyes (ANS), thereby suggesting that some hydrophobic residues are exposed to the solvent.[4]

The next step of the DeGrado group was the synthesis of the α_2D protein.[10] The new design was based on the assumption that alternative folds have to be destabilized. A metal binding site was constructed by introducing two His residues at positions 11 and 26 in each chain and placing Glu in the seventh position. It was shown that this polypeptide could exist as a monomer or a dimer, with the monomeric form being a random coil chain and the dimers being highly helical. The α_2D protein exhibits a cooperative thermal denaturation, similar to that seen in small natural globular proteins.

Grant sponsor: Polish State Committee for Scientific Research (KBN); Grant number: 6 PO4A 014 13; Grant sponsor: National Institutes of Health; Grant number: 1 P41 RR12255.

*Correspondence to: Andrzej Sikorski, Department of Chemistry, University of Warsaw, Pasteura 1, 02–093 Warsaw, Poland. E-mail: sikorski@chem.uw.edu.pl

Received 8 April 1999; Accepted 11 August 1999

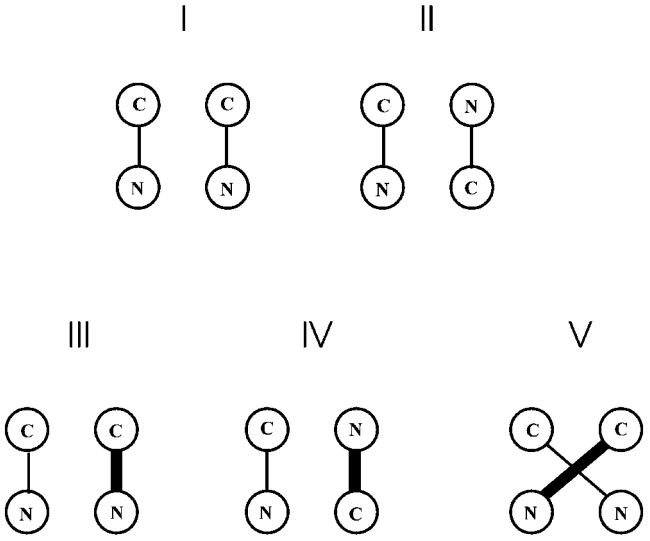

Fig. 1. All five possible topologies of folded four-helix bundles consisting of two polypeptide chains are shown.

Turning to previous theoretical studies, Thirumalai et al.[11–13]proposed a very simple model of a protein chain and applied it to a very simplified α_4 chain (HP model). They found that even in this simple model, there are two low temperature states. the first is a compact randomly arranged globule and the second (at lower temperatures) is an ordered globule that can be treated as a native-like state.

In our previous studies,[14,15] we simulated most of the original DeGrado polypeptide sequences in the context of two quite different models. The first early model[14] employed a relatively coarse (210) lattice (with a coordination number of 56) for the α-carbon representation of the main chain and a coarse library of single-center side chain conformations. Simulations of the α_4 chain showed that it adopted isoenergetic left and right turning, four-helix bundle topologies. The compact states resembled a molten globule. The introduction of amino acid diversity led to the transition from molten globule state to a native-like state. Folds of the improved α_4 formed a unique structure.

The second model used for simulations of designed helical proteins was much more realistic. It was based on the (310) hybrid lattice (coordination number 90) developed by Kolinski and Skolnick.[16,17] Here, the geometric representation of proteins was more accurate. The force field was refined to avoid a possible bias towards the molten globule state. We studied the α_4 tetramer, the α_2 dimer, and the α_1 monomer model systems. It was found that all these simple polypeptides adopt (upon association, when appropriate) a variety of four-helix bundle topologies that were essentially isoenergetic. Features of the molten globule state were found in all proteins studied.

Experimental studies[3,18] indicated that all the folded conformations of α_1, α_2, and α_4 have quite similar properties. This was probably the reason why the DeGrado group focused on studying the α_2 type of polypeptide. In this work, we focus on the α_2 protein and its two improved versions,

TABLE I. Statistics and Some Parameters Describing α_2, α_2C and α_2D Folds Obtained in Monte Carlo Simulations

Topology[a]	I			II			III			IV			V		
Protein	α_2	α_2C	α_2D	α_2	α_2C	α_2D	α_2	α_2C	α_2D	α_2	α_2C	α_2D	α_2	α_2C	α_2D
% Population	14	27	30	36	42	40	23	16	13	18	15	17	9	0	0
Average rms (Å)	3.8 (0.7)[c]	3.9 (0.5)	3.6 (0.7)	3.8 (0.7)	3.7 (0.6)	3.5 (0.4)	4.1 (0.6)	4.3 (0.7)	4.1 (0.7)	4.0 (0.8)	4.1 (0.6)	4.2 (0.6)	4.3 (1.1)	—	—
Radius of gyration (Å)	12.0 (1.5)	12.5 (1.3)	11.7 (1.6)	11.8 (1.2)	12.0 (1.4)	11.7 (1.7)	12.1 (1.5)	12.1 (1.1)	12.3 (1.6)	11.9 (1.8)	11.7 (1.4)	12.0 (1.6)	12.5 (2.1)	—	—
Average energy (kT)[b]	−125 (11)	−200 (17)	−194 (14)	−132 (13)	−198 (18)	−194 (21)	−124 (17)	−192 (12)	−178 (19)	−121 (16)	−202 (15)	−175 (19)	−122 (23)	—	—

[a]All possible topologies of α_2-type dimer are shown in Figure 1. The topology numbers used here refer to those in Figure 1.
[b]Energies for all chains were normalized to the temperature 2.4.
[c]The number in parentheses is the standard deviation.

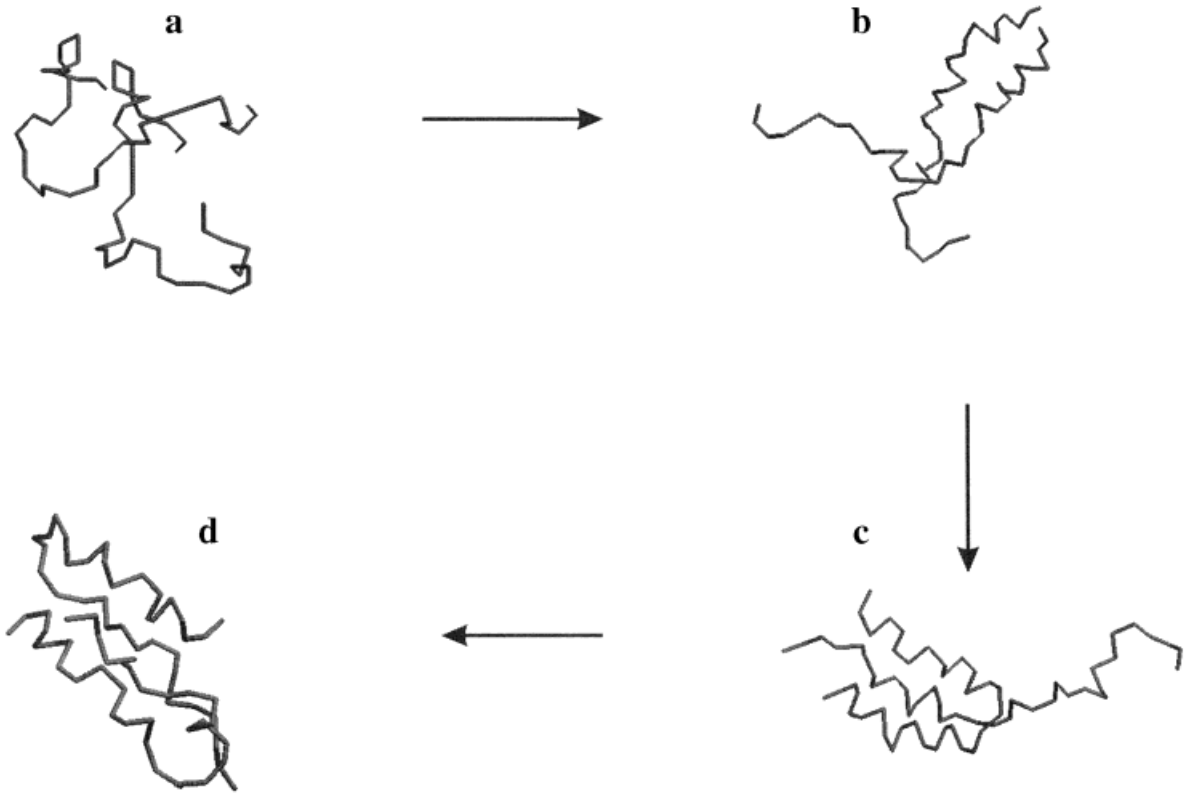

Fig. 2. Typical folding pathway of two α_2C chains. **a:** The initial conformation is a random coil at a reduced temperature of 2.7. **b:** One chain forms a hairpin by an on-site assembly mechanism. **c:** The second chain starts to assemble. **d:** The right-handed fold (topology I in Fig.1) is obtained at a temperature of 2.5.

α_2C and α_2D. The purpose of this study is to obtain insight into the mechanism of folding these polypeptide chains. Another interesting question addressed concerns the stability of the different compact topologies adopted by these sequences. Does the diversification of the sequence really lead to a unique fold? Do all the features of the molten globule state disappear? Since our previous work[15] demonstrated that the (310) lattice model is of sufficient resolution for such studies, we employ this model here.

MODEL AND METHOD

Because the model was described previously in detail,[15–17,19,20] only a short outline is provided here for the reader's convenience.

The protein backbone chain is represented by the α-carbon atoms. The chain is confined to the 310 lattice that comprises 90 basis vectors. If the lattice unit is equal to 1.22 Å, the best fit of protein crystallographic structures to the lattice model is obtained with the average rmsd of the α-carbon trace of the lattice model from the native structure between 0.6 and 0.7 Å.[16] The side groups are represented by single spheres having different sizes and appearing in various rotamer isomeric states. The positions of side groups were not confined to the lattice. The rmsd of the model side groups compared with the corresponding experimental structure is on the level of 1 Å.

The Monte Carlo sampling algorithm is based on the Metropolis scheme described elsewhere.[16,19,20] It uses a set of local micromodifications of protein chain conformation. A randomly chosen high-temperature chain conformation was selected and then slowly cooled. At each temperature, the model chain was carefully equilibrated before the results were collected. The temperature change was chosen to be small enough to avoid quenching of the system. For a given temperature, we performed about 10^5 to 10^6 Monte Carlo steps (each Monte Carlo step contains one attempt of every kind of micromodification per residue in the chain). Folded chains underwent further simulation at the constant temperature below the folding transition.

The force field used in our model consisted of long- and short-range terms. The following long-range interactions are included in the force field:[15]

1. The pairwise specific energy of two interacting residues (which are at least three residues apart along the chain contour) E_{pair}.
2. Cooperative (four-body) side chain interactions E_{tem}, which magnify the interactions of elements of secondary structure. These interactions are important in inducing the cooperative transition from a molten globule state to a native-like structure and act by inducing well defined side chain packing patterns.

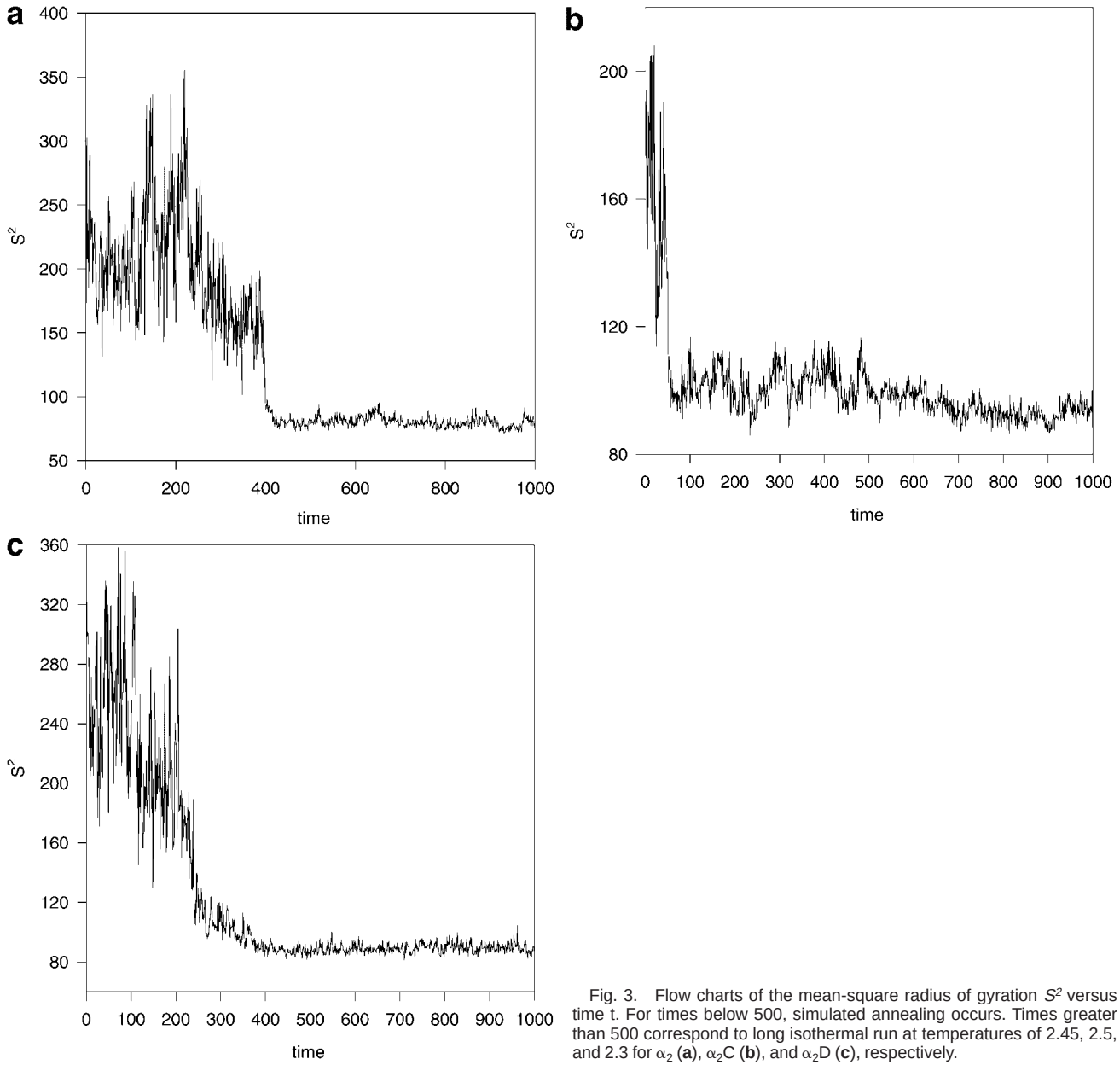

Fig. 3. Flow charts of the mean-square radius of gyration S^2 versus time t. For times below 500, simulated annealing occurs. Times greater than 500 correspond to long isothermal run at temperatures of 2.45, 2.5, and 2.3 for α_2 (**a**), α_2C (**b**), and α_2D (**c**), respectively.

3. A burial potential, E_{one}. It is assumed that a residue is buried when the number of side groups in contact exceeds a certain threshold value.[19]
4. A weak harmonic potential was introduced to keep both chains together. It is turned off at distances below a critical distance that is somewhat larger than the size of the helical bundle (usually between 15 and 20 Å). Thereby it has no effect on final energy of folded conformations. The introduction of this term provides for a constant local chain density in the Monte Carlo box.[15]

Short-range interactions were built from the following contributions:

1. The rotamer energy, E_{rot}, depends on the residue type and on the local conformation of the backbone chain.
2. A local side chain orientational coupling term, E_m, which introduces a preference for some orientations of the neighboring (down the chain) side groups.[16]
3. An effective Ramachandran potential, E_{R14}, introduces pairwise specific secondary conformational propensities for main chain fragments involving α-carbons i and $i+3$.
4. Hydrogen bonds, E_{HB}, simulate the main chain hydrogen bond network found in real proteins. Every α-carbon can participate at most in two hydrogen bonds (with the exception of proline, which can form only one bond).

The weight factors for the potentials were introduced to maintain the proper balance between the short- and long-range energy contributions. The detailed discussion concerning the potentials and weights was given elsewhere.[15] This weighting ensures a proper secondary struc-

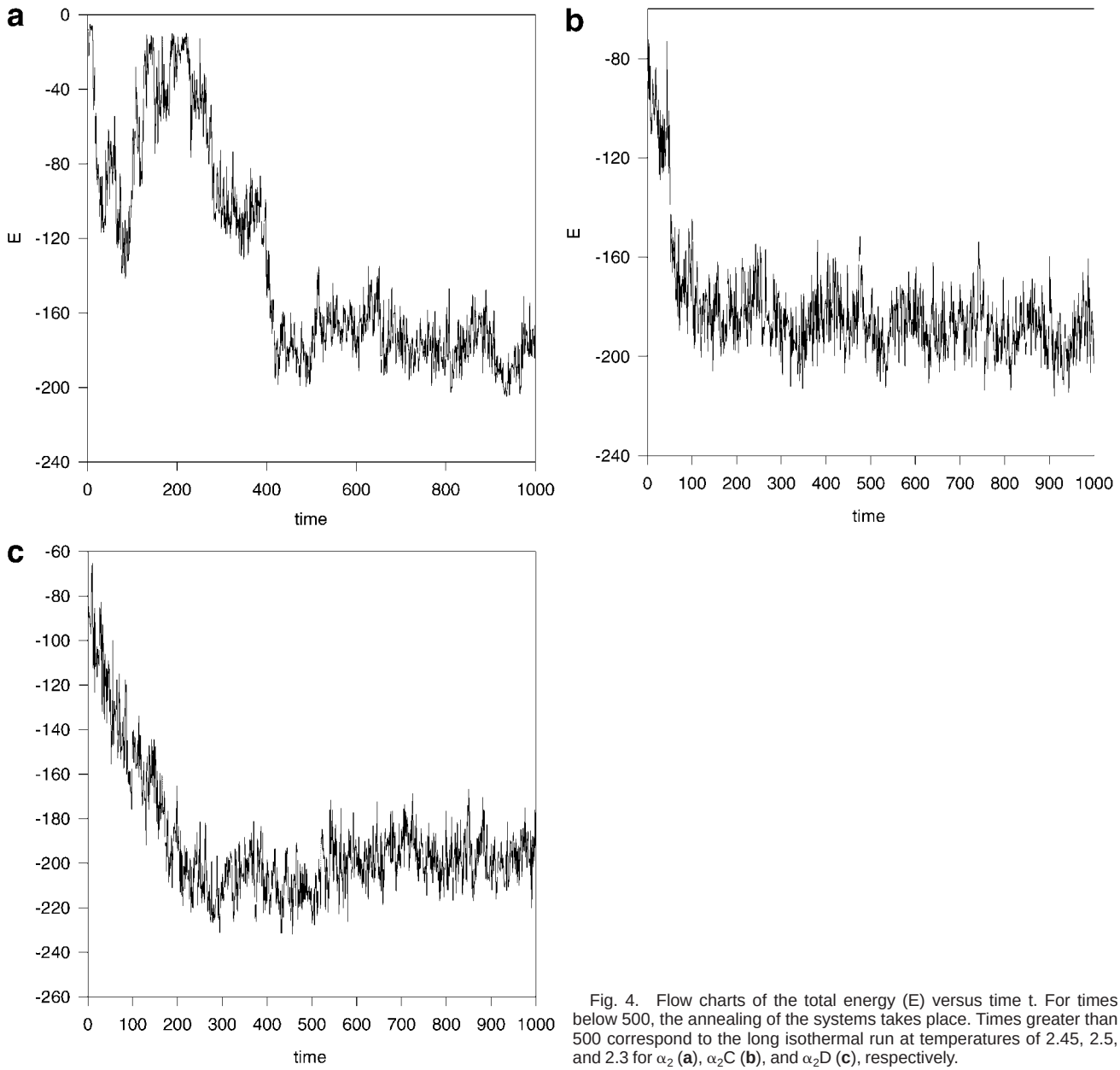

Fig. 4. Flow charts of the total energy (E) versus time t. For times below 500, the annealing of the systems takes place. Times greater than 500 correspond to the long isothermal run at temperatures of 2.45, 2.5, and 2.3 for α_2 (**a**), α_2C (**b**), and α_2D (**c**), respectively.

ture content at high temperatures. It also avoids an artificial bias of the model toward a molten globule state. This might be important in studies of de novo designed proteins that in reality have many of the features of molten globules. Most of the potentials listed above were derived from a statistical analysis of the protein structures from the Protein Data Bank (PDB).[21,22]

RESULTS AND DISCUSSION

The simulation experiments were performed for the following designed helical proteins:

1. α_2 is supposed to form the following secondary structure pattern:

$$\text{Ac-}\textit{Helix}\text{-}\textit{Loop}\text{-}\textit{Helix}\text{-CONH}_2 \qquad (1)$$

where the *Helix* has the following sequence:
GLY-GLU-LEU-GLU-GLU-LEU-LEU-LYS-LYS-LEU-LYS-GLU-LEU-LEU-LYS-GLY
and *Loop* consists of PRO-ARG-ARG, Ac is an acetyl group, and $CONH_2$ is a carboxyamide group.

2. α_2C is supposed to form the following secondary structure:

$$\text{Ac-}\textit{Helix}_1\text{-}\textit{Loop}\text{-}\textit{Helix}_2\text{-CONH}_2 \qquad (2)$$

where $\textit{Helix}_1$ has the following sequence:
GLY-GLU-VAL-GLU-GLU-LEU-LEU-LYS-LYS-PHE-LYS-GLU-LEU-TRP-LYS-GLY
$\textit{Helix}_2$ consists of
GLY-GLU-ILE-GLU-GLU-LEU-PHE-LYS-LYS-PHE-

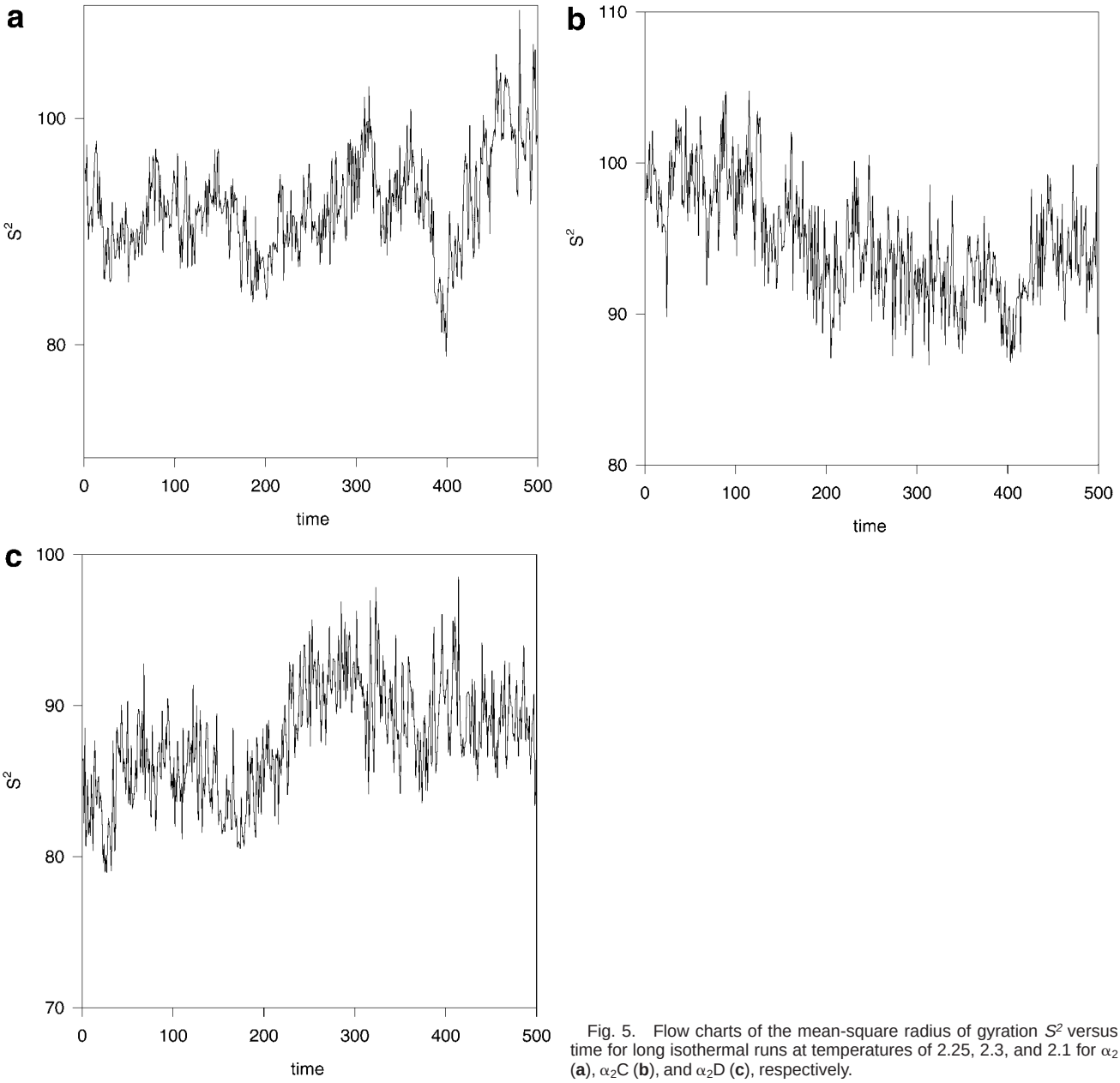

Fig. 5. Flow charts of the mean-square radius of gyration S^2 versus time for long isothermal runs at temperatures of 2.25, 2.3, and 2.1 for α_2 (**a**), α_2C (**b**), and α_2D (**c**), respectively.

LYS-GLU-LEU-ILE-LYS-GLY and *Loop* has exactly the same sequence as in the original α_2 chain.

3. α_2D is supposed to form the following structure:

$$\text{Ac-}Helix_1\text{-}Loop\text{-}Helix_2\text{-CONH}_2 \qquad (3)$$

where $Helix_1$ comprises
GLY-GLU-VAL-GLU-GLU-LEU-GLU-LYS-LYS-PHE-HIS-GLU-LEU-TRP-LYS-GLY
and $Helix_2$ comprises:
GLY-GLU-ILE-GLU-GLU-LEU-HIS-LYS-LYS-PHE-LYS-GLU-LEU-ILE-LYS-GLY

Finally, *Loop* was designed exactly the same as in α_2 and α_2C.

All possible four-helix bundle topologies that a system consisting of two polypeptide chains can adopt are shown in Figure 1. The possible existence of these different folded states arises from the different mutual orientations of the helical hairpins. The short loop connecting both helical fragments can be located near the loop formed by the second chain or near the *C*- and *N*-termini of the second chain. There is also the fifth topology where the loops cross (see Fig.1). In our Monte Carlo simulations, we obtained all the possible topologies.

The results of our simulations are summarized in Table 1. Although the number of simulations is too small to establish anything more than a qualitative trend, for α_2, the populations of all topologies are rather uniformly distributed, with the exception of topology V with crossed

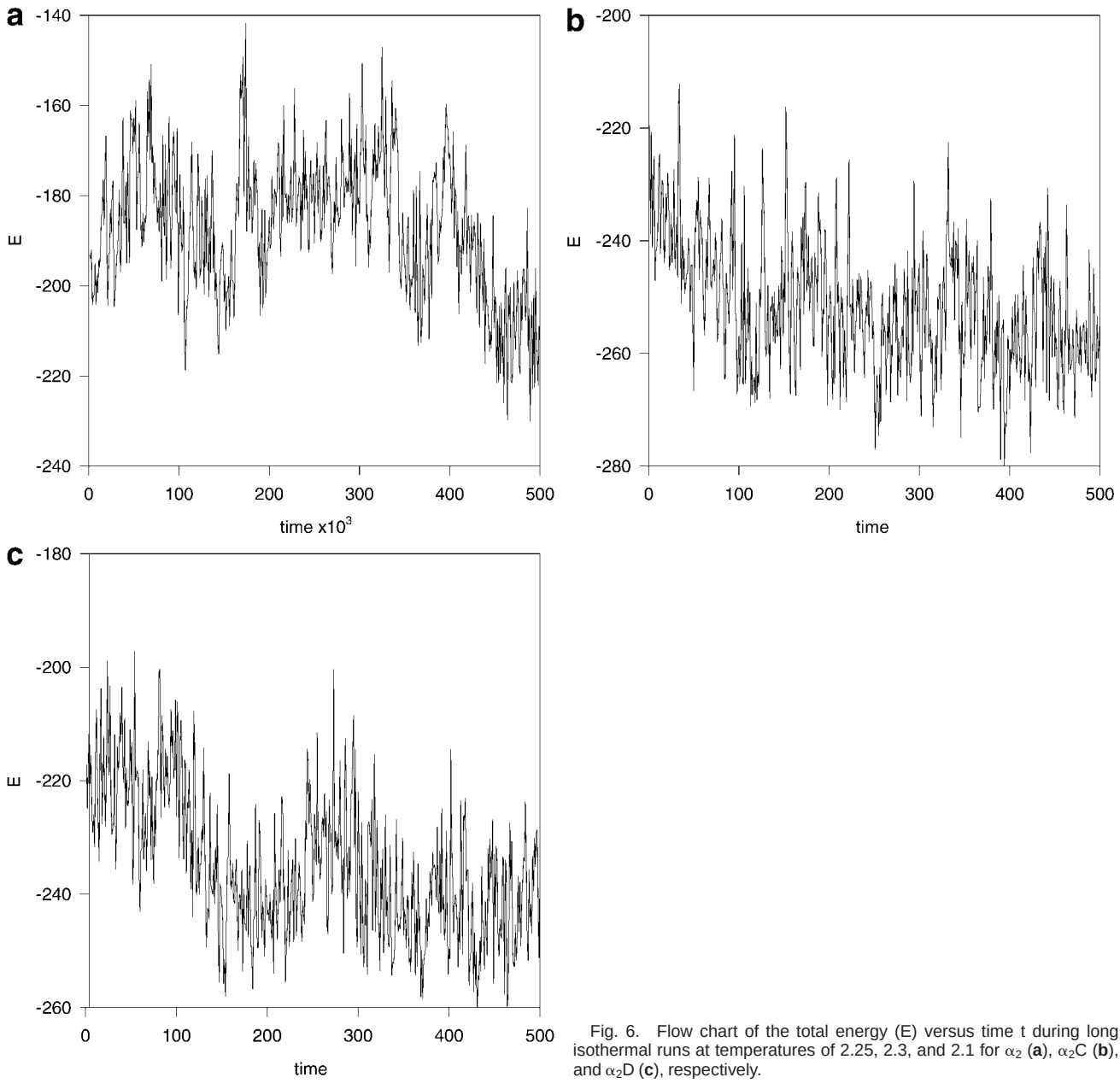

Fig. 6. Flow chart of the total energy (E) versus time t during long isothermal runs at temperatures of 2.25, 2.3, and 2.1 for α_2 (**a**), α_2C (**b**), and α_2D (**c**), respectively.

helices, which was obtained only once. Going from α_2 to α_2C to α_2D, the following comments can be made on the observed population:

1. The topology with crossed loops (topology V on Fig. 1) does not appear in α_2C and α_2D chains. In α_2, it was only found once; therefore, one cannot exclude the possibility that it is actually a misfolded state; however, its energy is very close to that observed in the alternative topologies.
2. Folds in which both loops are on the same end of the bundle (topologies I and II) are the most frequent. For the α_2D protein, topology II (with antiparallel helices) is most frequently obtained, but for α_2C it is slightly less frequent than topology I. An inspection of real protein structures indicates that in natural all four-member helical bundles, only topologies with antiparallel helices are observed.[23,24]
3. Topologies with loops located on the opposite ends of the bundle (topologies II and IV) are significantly less populated in both α_2C and α_2D systems.

The folding of the system consisting of two α_2 protein chains was previously described in detail.[15] Forty annealing simulation runs between temperatures *2.6* and *2.45* were performed for this system. In these simulations, we obtained 22 folded structures consisting of dimeric four-member, α-helical bundles. The remaining misfolded struc-

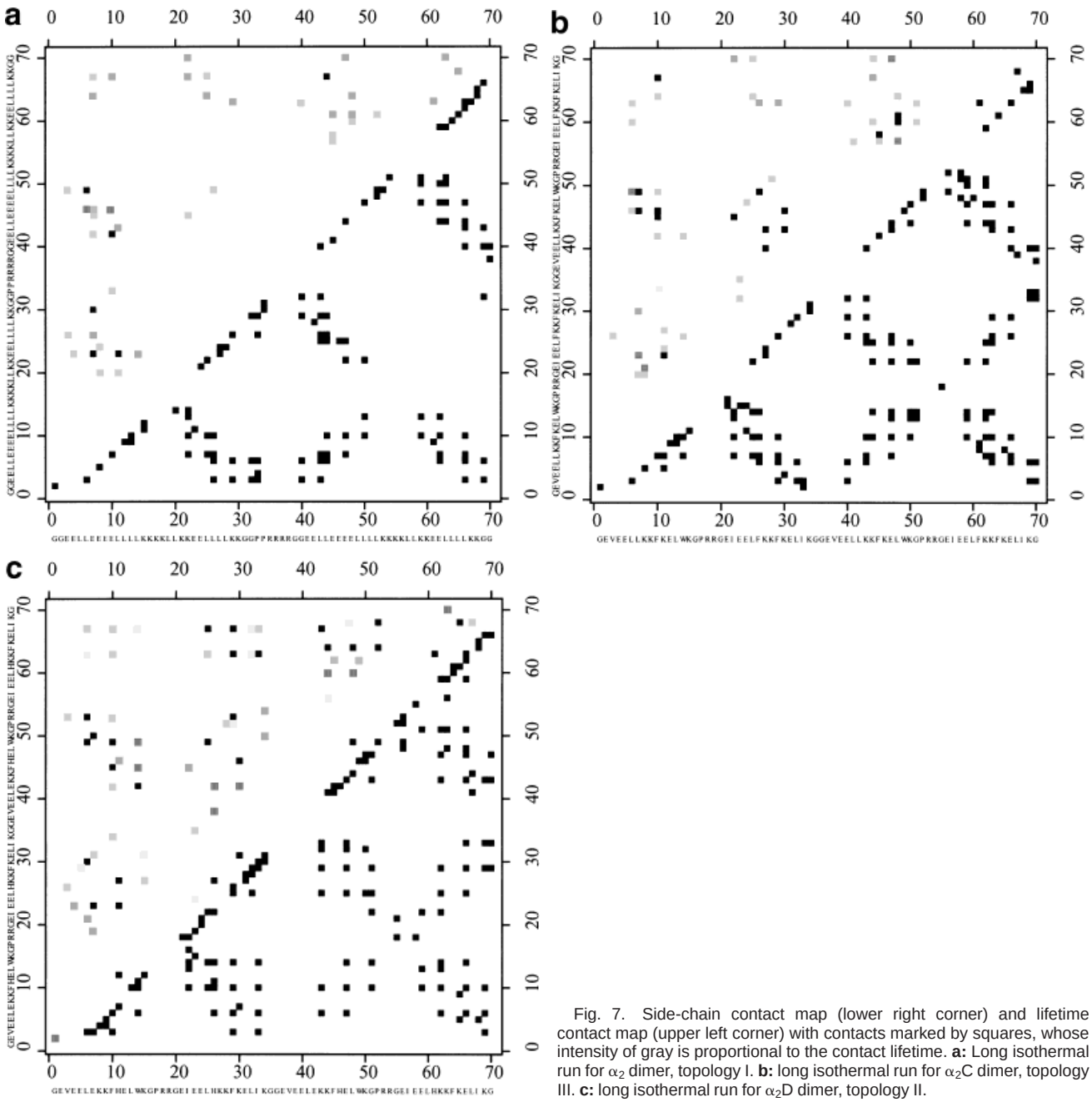

Fig. 7. Side-chain contact map (lower right corner) and lifetime contact map (upper left corner) with contacts marked by squares, whose intensity of gray is proportional to the contact lifetime. **a:** Long isothermal run for α_2 dimer, topology I. **b:** long isothermal run for α_2C dimer, topology III. **c:** long isothermal run for α_2D dimer, topology II.

tures consisted of randomly arranged out-of-register helical systems, frequently with only three helices fully assembled. These could be dismissed because of lack of reproducibility and, more importantly, because of their conformational energy. Given the observed set of four-helix bundle structures, no specific topology could be selected because the differences in energy were not statistically significant. The energy of various folds of different topologies varied from −121 to −132 kT. The radius of gyration varied from 11.8 to 12.5 Å.

For the α_2C system, we performed 28 simulation runs for temperatures between *2.7* and *2.5*. Nineteen of these yielded four-member α-helical bundles. The other runs yielded higher energy structures that can be safely ignored. The total energy of the different classes of folds changed from −188 to −198 kT. The radius of gyration was in the range of 11.7 to 12.5 Å.

The folding of the α_2D protein was simulated in 30 Monte Carlo annealing runs at temperatures between 2.6 and 2.3. Twenty-three simulations were successful in giving a four-member α-helical bundle dimer with the remainder being higher in energy; therefore, they are dismissed. The energies of different topologies varied from −187 to −207 kT. The radius of gyration varies from 11.7

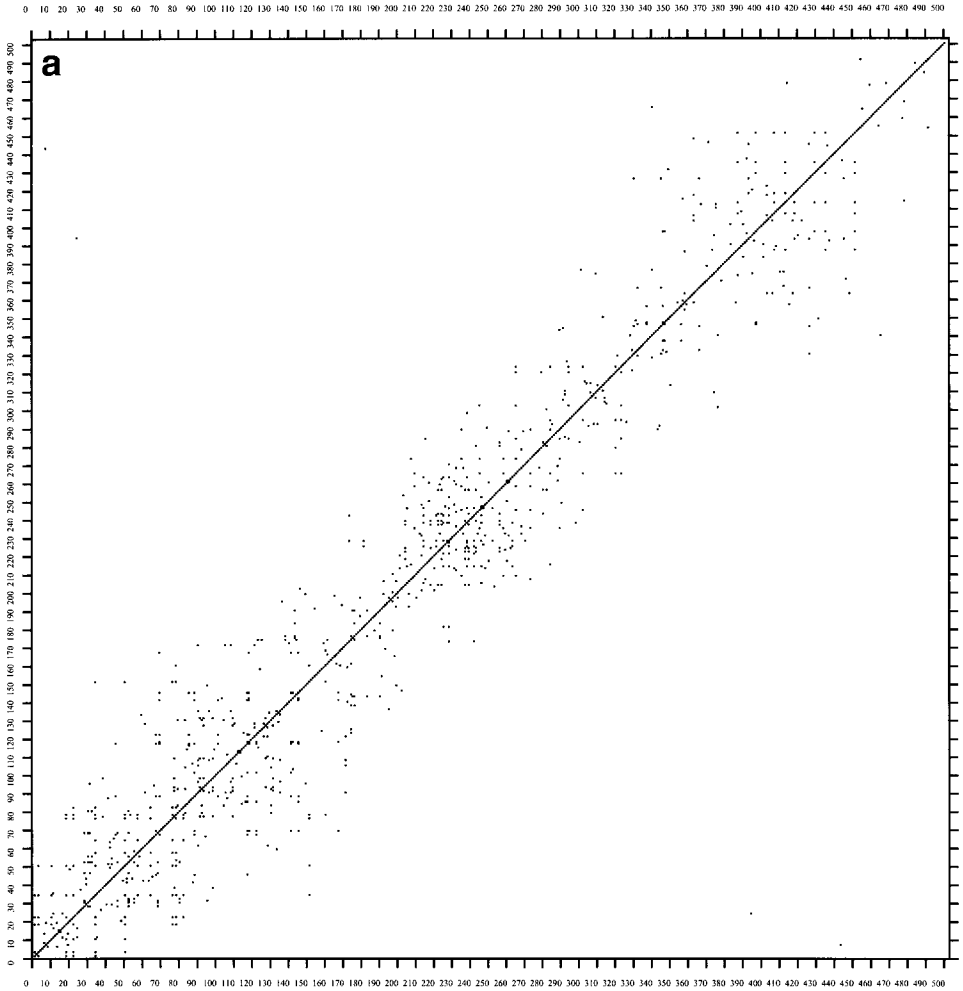

Fig. 8. Clustering of the conformations obtained in a long isothermal run. The results correspond to the same simulation runs as in Figure 7 for α_2 (**a**), α_2C (**b**), and α_2D (**c**), respectively.

to 12.0 Å. In general, the energy changes for different folds are on the same level in all models but the following general remarks can be made:

1. On average, the volume of the folds decreases with increasing sequence diversity, but these changes are very small.
2. The energies of the α_2C and α_2D folds are much lower (by about 20 %) than those of the α_2 folds. The energies of topologies I and II are apparently the lowest, although the differences among them are rather small.
3. The next important difference between all protein models studied concerns the conditions at which chains start to fold and unfold, i.e., the transition temperature. The α_2 chain folds at a temperature of 2.45, α_2C at 2.5, and α_2D at 2.3.

The folding mechanism, common for all three proteins, is the following: all model chains at high temperatures (T = 2.6 to 2.7) are random coils ($<S^2>/<R^2>$ = 0.155), with the radius of gyration $<S^2>^{1/2}$ oscillating around 20 Å. The total energy of the system at that temperature is $E_{tot} \approx -60$, -76, and -88 kT for α_2, α_2C, and α_2D, respectively. The helix content under these conditions does not exceed 10%. There is no quantitative difference in the high-temperature values of all these parameters for all sequences under consideration. An example of a high temperature configuration is shown on Figure 2. Annealing of the system changes the properties of both chains. The average size of the chains and their energies starts to decrease, but their instantaneous values fluctuate rapidly (see discussion below). These changes are accompanied by the fast formation (and dissolution) of some helical elements. At lower temperatures, helices of up to 16 residues in length (the entire putative helix in the polypeptide chain) are frequently formed. Entire helical hairpins (once the chains are fully folded) were also formed as shown on Figure 2b. These helices and hairpins are unstable. The mechanism of folding of a hairpin can be described as on-site assembly. The next step is the assembly of the third helix, which comes from the second chain (Fig. 2c). This three-helix bundle is the first intermediate of the topology

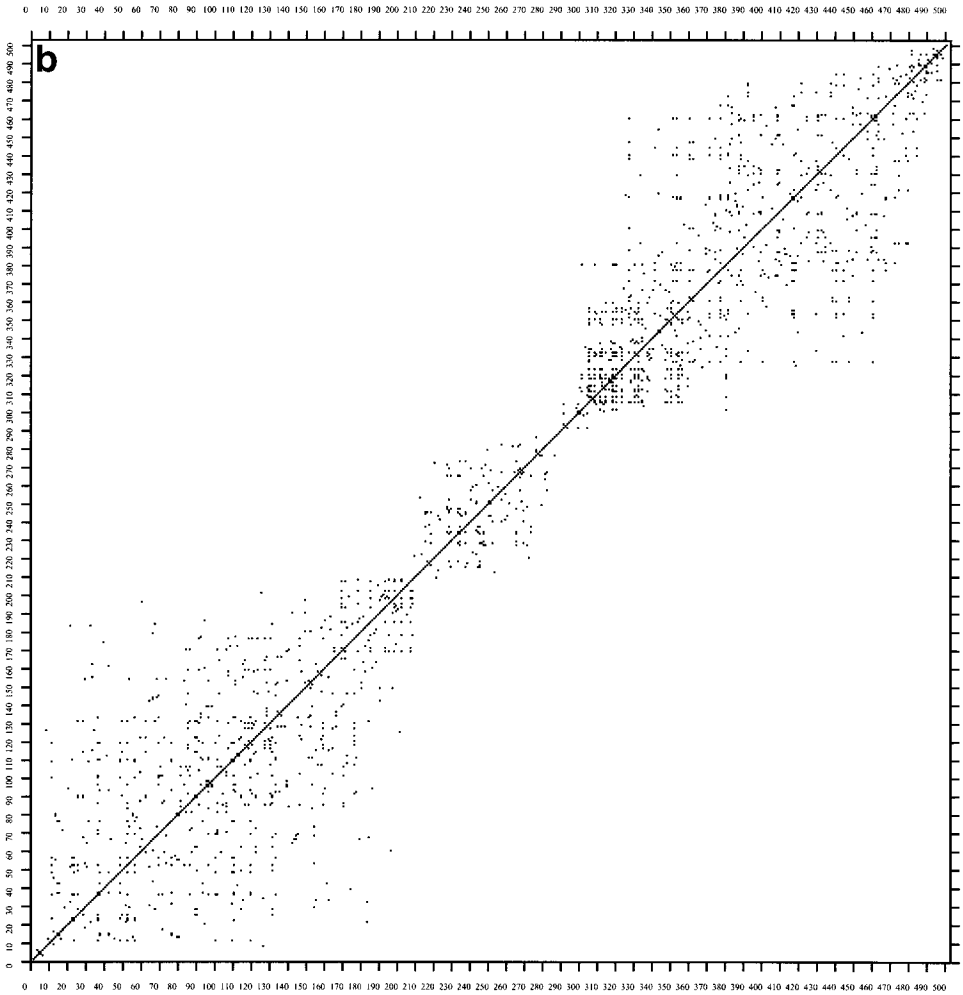

Figure 8. (Continued.)

assembly process. Then, the last helix assembles on site (Fig. 2d). This is the slowest stage of the topology assembly. The mechanism of the opposite process of the thermal denaturation of the folded α_2 (as well as folded α_2C and α_2D) dimer is the same: the dissolution of helices occurs in exactly the opposite sequence of events. The pathways obtained for our current model are qualitatively the same as were observed previously for much simpler models on a tetrahedral lattice.[25–27]

The precision of the prediction of a given folded state can be measured in terms of the rmsd within a given topology. We cannot make a comparison with real experiment because no crystal structures for α_2 are publicly available. The rmsds between α_2 folds (generated in independent simulations) belonging to a given topology are in the range of 3.5 to 4.3 Å. The best fit (3.5 Å) was found for a pair of conformations belonging to topology number I (Fig. 1a) in which the helices are antiparallel. The helices in all folds are well defined with the average distance between their axes equal to 11.9 Å. For α_2C chains, the pairwise rmsd was on the same level as α_2 (between 3.7 and 4.3 Å). For α_2D chains, the rmsd values are smaller and range between 3.5 and 4.2 Å. The main conclusion that can be drawn here is that all obtained folds exhibit a similar level of overall geometrical accuracy.

Some insight into the folding pathway of these systems could be obtained from an analysis of the changes of the radius of gyration and the total energy of the system during the folding experiments. Figure 3 shows the flow charts of the radius of gyration $<S^2>^{1/2}$ for α_2 and α_2C, and α_2D systems, respectively. There are no essential qualitative differences among these flow charts. At high temperatures, the size of all model systems fluctuates within the range of 250 to 350 lattice units. At lower temperatures where intermediate states are formed, the mean value of S^2 is lower and the range of fluctuations is considerably less. Further simulation led to a stable helical fold whose radius of gyrations is close to 90 (in lattice units). There are no correlations between the type of chain (α_2, α_2C, or α_2D) and the range of these fluctuations or the frequency of appearance of structural elements during the folding process. The range of fluctuations of the low-temperature

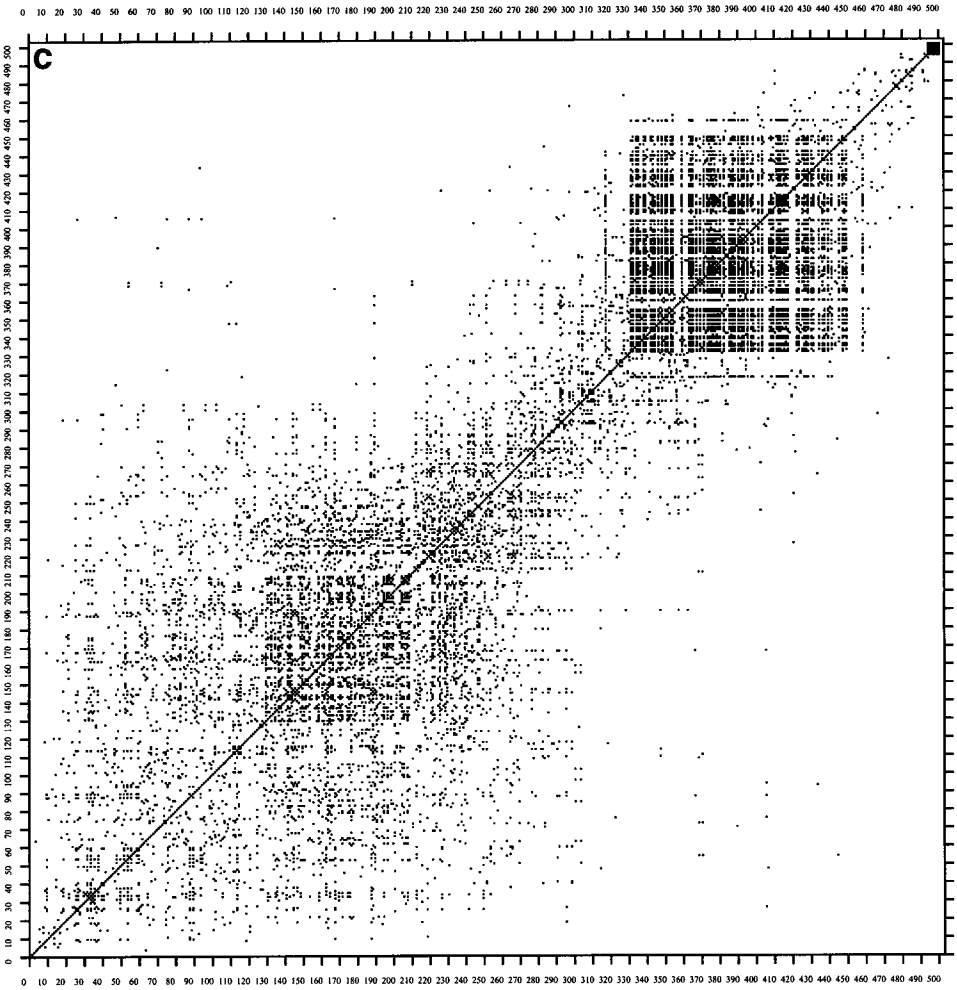

Figure 8. (Continued.)

folds is also very similar. In all cases, the transition is quite abrupt and occurs in a narrow temperature window.

The profiles of the total energy of the model chain systems as a function of simulation time (with linearly decreasing temperature during the initial 500 steps) are presented in Figure 4. The fluctuations in the folded state (energy on the level of -100 or -200 kT) are still large. It is surprising that there is no qualitative difference between all models under consideration. This implies that in all models (α_2, α_2C, and α_2D), the structural fluctuations in the folded state are significant. This kind of behavior is usually identified with the molten globule state.

Further annealing of the system and a subsequent long isothermal run at temperatures well below the folding transition did not significantly change the properties of the model systems. In Figures. 5 and 6, a low-temperature representative example ($T = 2.1$) of flow charts of S^2 and E versus time for the α_2D system are presented. The range of fluctuations of both parameters is the same as the long isothermal runs just below the folding temperature ($T = 2.3$); see Figure 4c.

To further analyze the properties of the folded state, we focus on the dynamic behavior of the side chain packing patterns. Figure 7 shows the side chain contact maps of obtained folds. The data were collected from long isothermal runs at temperatures just below the folding transition. Without a doubt, all contact maps exhibit characteristic α-helical patterns. Figure 7 also contains information about the lifetime of side chain contacts. One can observe that for α_2, only very few contacts survive the entire simulation. The remaining contacts were short lived. For the folded α_2C dimer, one can find more long-lived contacts, but their number is still small. In the case of the α_2D dimer, the number of contacts that remained fixed during the entire simulation is much larger. The lifetime of the remaining contacts in α_2C and α_2D folds is considerably longer than that in the α_2 dimers. Furthermore, the long-lived contacts are distributed over the entire map. Thus, the entire folded structure of α_2C and α_2D has a fixed network of pairwise contacts. This is a strong signal of a transition from the liquid-like structure of α_2 to a more

native-like conformation, with fixed positions of some side chains in these two proteins.

To study the correlation between the folded structures along the simulation run we carried out a structural clustering procedure identical to that used previously.[14] The mutual rms of every pair of conformations (at a certain time interval) is stored and arranged into a "history matrix" containing 200 × 200 elements. This matrix was rearranged so as to minimize its moment of inertia: lower elements were displaced toward the diagonal of the matrix. As a result, we obtained clusters of similar conformations near the diagonal. In Figure 8, we display examples of the rearranged history matrix for the three studied proteins. For clarity, we present only those pairs of conformations for which the rmsd is lower than 4 Å. The main difference between the three systems is the number of pairs of conformations whose rmsd is less than 4 Å: the number of such cases is rather low for α_2, higher for α_2C, and even higher for α_2D. For the α_2 fold, almost all similar conformations live for only a very short time. In contrast, for α_2C and α_2D, one can find quite large clusters of conformations with the same low rms. In all cases, there are several clusters of conformations, but the level of clustering of a large number of states increases from α_2 to α_2D. In the last case, there is one large dominating cluster, indicating significant structural uniqueness (and lower level of contact fluctuations) of the folded state (as measured by the side chain fixation process).

CONCLUSIONS

Our previous simulations of α_1, α_2, and α_4 indicated (in agreement with experimental results) a lack of structural uniqueness of the folded structure of these designed proteins. In the present studies of α_2, α_2C, and α_2D, there is substantial evidence of increasing structural uniqueness within the series of molecules. Although the average conformational energy of the folded states is not conclusive, there is a statistical trend indicating higher stability in the α_2C and α_2D proteins. The same conclusion could be drawn from the fold volume and the distribution of folds: the α_2D chain topology with antiparallel helices is most populated.

A cross-over from molten globule-like character to native-like behavior of the folded state in this series of three proteins is also demonstrated. The number of side group contacts that survive the entire simulation time below the folding temperature becomes higher for α_2C and much higher for α_2D. In spite of this well defined trend, the fraction of long-lived contacts is rather small. The clustering of the rmsd of a pair of conformations confirms its higher stability and the existence of a better defined native-like free energy basin in the α_2C and α_2D molecules relative to the generic α_2 protein.

REFERENCES

1. Ho SP, DeGrado WF. Design of a 4-helix bundle protein: synthesis of peptides which self-associate into a helical protein. J Am Chem Soc 1987;109:6751–6758.
2. Regan L, DeGrado WF. Characterization of a helical protein designed from first principles. Science 1988;241:976–978.
3. Betz SF, Raleigh DP, DeGrado WF. De novo protein design: from molten globules to native-like states. Curr Opin Struct Biol 1993;3:601–610.
4. Betz SF, Raleigh DP, DeGrado WF, et al. Crystallization of a designed peptide from a molten globule ensemble. Folding Design 1996;1:57–64.
5. Kuwajima K. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. Proteins 1989;6:87–103.
6. Ptitsyn OB. The molten globule state. In: Creighton TE, editor. Protein Folding. New York: Freeman; 1990. p .243–299.
7. McGregor MJ, Islam SA, Sternberg MJE. Analysis of the relationship between side chain conformation and secondary structure in globular proteins. J Mol Biol 1987;198:775–791.
8. Dolphin GT, Brive L, Johansson G, Baltzer L. Use of aromatic amino acid residues to restrict the dynamics in the hydrophobic core of a designed helix-loop-helix dimer. J Am Chem Soc 1996;118: 11297–11298.
9. Brive L, Dolphin GT, Baltzer L. Structure and function of an aromatic ensemble that restricts the dynamics of the hydrophobic core of a designed helix-loop-helix dimer. J Am Chem Soc 1997;119: 8598–8607.
10. Raleigh DP, Betz SF, DeGrado WF. A de novo designed protein mimics the native state of natural proteins. J Am Chem Soc 1995;117:7558–7559.
11. Honeycutt JD, Thirumalai D. Metastability of folded states of globular proteins. Proc Natl Acad Sci USA 1990;87:3526–3529.
12. Guo Z, Thirumalai D, Honeycutt JD. Folding kinetics of proteins: a model study. J Chem Phys 1992;97:525–535.
13. Guo Z, Thirumalai D. Kinetics and thermodynamics of folding of a de novo designed four-helix bundle protein. J Mol Biol 1996;263; 323–343.
14. Kolinski A, Godzik A, Skolnick J. 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 1993;98:7420–7433.
15. Sikorski A, Kolinski A, Skolnick J. Computer simulations of de novo designed helical proteins. Biophys J 1998;75;92–105.
16. Kolinski A, Skolnick J. Monte Carlo simulation of protein folding. I. Lattice model and interaction scheme. Proteins1994;18:338–352.
17. Kolinski A, Skolnick J. Lattice models of protein folding, dynamics and thermodynamics. Austin, TX: R.G. Landes Publishing; 1996.
18. Betz SF, Bryson JW, DeGrado WF. Native-like and structurally characterized designed -helical bundles. Curr Biol 1995;5:457–463.
19. Vieth M, Kolinski A, Brooks CL III, Skolnick J. Prediction of the folding pathways and structure of the GCN4 leucine zipper. J Mol Biol 1994;237:361–367.
20. Vieth M, Kolinski A, Brooks CL III, Skolnick J. Prediction of the quaternary structure of coiled coils. Application to mutants of the GCN4 leucine zipper. J Mol Biol 1995;251:448–467.
21. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR. The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
22. PDB. Quart Newslett No 17, January 1995.
23. Presnell SR, Cohen FE. Topological distribution of four helix-bundles. Proc Natl Acad Sci USA 1989;86:6592–6596.
24. Cohen C, Parry DAD. α-helical coiled coils and bundles: how to design an α-helical protein. Proteins 1990;7:1–15.
25. Sikorski A, Skolnick J. Monte Carlo simulation of equilibrium globular protein folding: α-helical bundles with long loops. Proc Natl Acad Sci USA 1989;86:2668–2672.
26. Sikorski A, Skolnick J. Monte Carlo study on equilibrium globular protein folding. III. The four helix bundle. Biopolymers 1989;28: 1097–1113.
27. Sikorski A, Skolnick J. Dynamic Monte Carlo simulations of globular protein folding/unfolding pathways. II. α-helical motifs. J Mol Biol 1990;212:819–836.