PROTEINS: Structure, Function, and Genetics 47:305–314 (2002)

Weak Temperature Dependence of the Free Energy Surface and Folding Pathways of Structured Peptides

Andrea Cavalli, Philippe Ferrara, and Amedeo Caflisch*
Department of Biochemistry, University of Zurich, Zurich, Switzerland

ABSTRACT **The thermodynamics and energetics of a 20-residue synthetic peptide with a stable three-stranded antiparallel β-sheet fold are investigated by implicit solvent molecular dynamics (MD) at 330 K (slightly above the melting temperature in the model) and compared with previous simulation results at 360 K. At both temperature values, the peptide folds reversibly to the NMR solution conformation, irrespective of the starting conformation. The sampling of the conformational space (2.3 μs and 25 folding events at 330 K, and 3 μs and 50 folding events at 360 K) is sufficient to obtain a thermodynamic description of minima and transition states on the free energy surface, which is determined near equilibrium by counting populations. The free energy surface, plotted as a function of two-order parameters that monitor formation of either of the β-hairpins, is similar at both temperature values. The statistically predominant folding pathway and its frequency (about two-thirds of the folding events) are the same at 330 K and 360 K. Furthermore, the main unfolding route is the reverse of the predominant folding pathway. The effective energy and its electrostatic and van der Waals contributions show a downhill profile at both temperatures, implying that the free energy barrier is of entropic origin and corresponds to the freezing of about two-thirds of the chain into a β-hairpin conformation. The average folding rate is nearly the same at 330 K and 360 K, while the unfolding rate is about four times slower at 330 K than at 360 K. Taken together with previous MD analysis of α-helices and β-hairpins, the present simulation results indicate that the free energy surface and folding mechanism of structured peptides have a weak temperature dependence. Proteins 2002;47:305–314.**

Key words: folding pathways; free energy; weak temperature dependence

INTRODUCTION

In the postgenomic era, the prediction of protein structure and function is expected to have an enormous practical importance in biomedical and pharmaceutical research. A better understanding of the protein-folding process is important in planning new experiments[1] and will probably help to improve computational methods for structure prediction. Even for a small protein, it is not yet feasible to simulate the complete process of folding with a high-resolution approach, e.g., molecular dynamics (MD) simulations with an all-atom model. The practical difficulties in performing such brute force simulations have led to several types of computational approaches and/or approximative models to study protein folding. One common approach taken in the past is to unfold starting from the native structure.[2,3] Related studies deal with very small protein fragments for which the conformational space is sufficiently small to permit full searches and/or transitions of interest on a manageable time scale.[4] The thermodynamic properties of two peptides (an α-helix and a β-hairpin of 13 and 12 residues, respectively) have been determined using an implicit solvation model and adaptive umbrella sampling.[5] Furthermore, the free energy surface of Betanova, an antiparallel three-stranded β-sheet peptide, has been constructed starting from conformations obtained during unfolding simulations in explicit water at elevated temperatures (350–400 K).[6]

The designed amino acid sequence (Thr_1-Trp_2-Ile_3-Gln_4-Asn_5-Gly_6-Ser_7-Thr_8-Lys_9-Trp_{10}-Tyr_{11}-Gln_{12}-Asn_{13}-Gly_{14}-Ser_{15}-Thr_{16}-Lys_{17}-Ile_{18}-Tyr_{19}-Thr_{20}), called GS peptide henceforth, has been studied in aqueous solution by nuclear magnetic resonance (NMR).[7] Nuclear Overhauser enhancement (NOE) and chemical shift data indicate that at 10°C, the GS peptide populates a single structured form, the expected three-stranded antiparallel β-sheet conformation with turns at Gly6–Ser7 and Gly14–Ser15, in equilibrium with the random coil. Furthermore, the GS peptide was shown to be monomeric in aqueous solution by equilibrium sedimentation and nuclear magnetic resonance (NMR) dilution experiments.[7] Recently, we have performed implicit solvent molecular dynamics simulations of the GS peptide at 360 K and could demonstrate the reversible folding[8] to the NMR conformation. Our results were close to equilibrium conditions because MD simulations at 360 K yielded a large amount of folding and unfolding transitions. This allowed to determine the free energy surface and the identification of an entropic barrier in the folding reaction. Two folding pathways emerged from the simula-

Grant sponsor: NCCR Structural Biology; Grant sponsor: Théodore Ott Foundation.

*Correspondence to: Amedeo Caflisch, Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. E-mail: caflisch@bioc.unizh.ch

Received 19 March 2001; Accepted 14 September 2001

tions, and for both the folding mechanism was found to involve the almost complete formation of one of the two β-hairpins, followed by consolidation of the unstructured strand.[8] The folding through β-hairpin formation is in agreement with recent NMR data for another designed three-stranded antiparallel β-sheet peptide of 24 residues.[9]

The main objective of the present study is to analyze the folding process at different temperature values. For this purpose, the free energy profile and mechanism of folding of the GS peptide are analyzed at 330 K and compared with the previous simulation results at 360 K. The following questions are pertinent to the present analysis: is the overall folding mechanism the same at different temperature values? Are the thermodynamically localized transition states the same? Is the predominance of one of the two pathways higher at 330 K than at 360 K? Is the unfolded state more compact at lower temperature? Which energy contributions are responsible for the more favorable effective energy at the lower temperature? The good agreement between the NMR data and the folded structures obtained in the simulations, irrespective of the starting conformation,[8] justifies the use of the MD results, to try to reply to these questions.

METHODS

Model

The peptide was modeled by explicitly considering all heavy atoms and the hydrogen atoms bound to nitrogen or oxygen atoms.[10] The remaining hydrogen atoms are considered as part of the carbon atoms to which they are covalently bound (extended atom approximation). An implicit model based on the solvent accessible surface[11] is used to describe the main effects of the aqueous solvent on the solute. In this approximation, the solvation free energy is given by

$$G_{\mathrm{solv}}(\mathbf{r}) = \sum_{i=1}^{N} \sigma_i A_i(\mathbf{r}) \qquad (1)$$

for a molecule having N heavy atoms with Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$. $A_i(\mathbf{r})$ is the solvent-accessible surface area of heavy atom i, computed by an approximate analytical expression[12] and using a 1.4-Å probe radius. Furthermore, ionic side-chains were neutralized,[13] and a linear distance-dependent screening function [$\epsilon(r) = 2r$] was used for the electrostatic interactions. The CHARMM PARAM19 default cutoffs for long-range interactions was used; i.e., a shift function[10] was employed with a cutoff at 7.5 Å for both the electrostatic and van der Waals terms. This cutoff length was chosen to be consistent with the parameterization of the force field. The model contains only two σ parameters: one for carbon and sulfur atoms ($\sigma_{C,S} = 0.012$ kcal/mol Å^2), and one for nitrogen and oxygen atoms ($\sigma_{N,O} = -0.060$ kcal/mol Å^2).[14] The σ parameters do not have a temperature dependence since this was shown to be weak in the 330–360 K range by a previous implicit solvent model calibrated on amino acid hydration free energies.[15] The model is not biased toward any particular secondary structure type. In fact, exactly the same force field, implicit solvation model and values of the σ parameters have been used recently to fold reversibly to the correct conformation by standard molecular dynamics six α-helical peptides (ranging in size from 15 to 31 residues),[16,17] a β-hairpin of 12 residues,[16] and another triple-stranded antiparallel β-sheet whose sequence identity with the one of the present study is only 15%.[18] Furthermore, the non-Arrhenius behavior of the temperature dependence of the folding rate of two structured peptides was demonstrated with the same force field and implicit solvation model.[16]

Simulations

The simulations and part of the analysis of the trajectories were performed with the CHARMM program.[10] Twenty runs of ≥100 ns each were performed starting from random conformations. The temperature was kept close to 330 K by weak coupling to an external bath with a coupling constant of 5 ps.[19] The SHAKE algorithm[20] was used to fix the length of the covalent bonds having hydrogen atoms at one end. The leapfrog algorithm and an integration time step of 2 fs were used to integrate the Newton equation of motion. The nonbonded interactions were updated every 10 dynamics steps and coordinate frames were saved every 10 ps. A 100-ns run requires approximately 7 days on a 500-MHz Pentium III processor.

Random Initial Conformations

Five-thousand structures were generated by randomizing the dihedral angles of the rotatable bonds, followed by thousand steps of energy minimization. Structures with one or more native contacts (see below) were discarded. The 40 structures with the most favorable energies were retained as starting conformations. Their average C_α root-mean-square deviation (RMSD) from the mean NMR model is 7.4 Å. The 20 random conformations used as starting point for the 330 K runs are a randomly chosen subset of the 40 random structures used for the 360 K simulations.

Native Contacts and Order Parameters

The conformations sampled previously at 300 K were used to define a list of 26 native contacts, of which 11, 11, 2, and 2 involve residues in strands 1–2, 2–3, 2–2, and 3–3, respectively (see Table 2 of ref. 8). These include 10 backbone hydrogen bonds (five on each β-hairpin) and 16 contacts between side-chains.

Most of the analysis in this paper is based on thermodynamics. For kinetics, one would have to define a reaction coordinate and look for the time sequence of conformations that follow a direct transition to the folded state. For thermodynamics, one needs to define only an order parameter or progress variable, not a reaction coordinate.[21] The folding order parameter Q is defined as the fraction of contacts common to both the current conformation and the native structure. It is usually plotted as a function of simulation time to monitor the change in structure during folding. For a given snapshot along a trajectory, a native

hydrogen bond is considered formed if the O· · ·H distance is <2.6 Å. A native side-chain contact is considered formed if the distance between geometrical centers is <6.7 Å. The following subset of native contacts are particularly appropriate for a clear description of the folding pathways. Q_{1-2} is defined as the fraction of the 11 native contacts (5 hydrogen bonds and 6 side-chain interactions) formed between strands 1 and 2, while Q_{2-3} as the fraction of the 11 native contacts between strands 2 and 3. Q_{distal} and Q_{proximal} are the fraction of the 11 native contacts far away from the turns (4 hydrogen bonds and 7 side-chain interactions) and the 8 native contacts close to the turns (4 hydrogen bonds and 4 side-chain interactions), respectively.

Folding Time

A folding event is considered completed when Q reaches a value larger than 0.85 ($Q > 22/26$), while an unfolding event is considered completed when Q reaches a value smaller than 0.15 ($Q < 4/26$). The rather strict criteria are used to neglect transient events. In the simulations started from random conformations, the time of the first folding event corresponds to the simulation time when Q reaches a value larger than 0.85 for the first time. The unfolding time is defined analogously, i.e., the time difference between the unfolding and previous folding event. In the case of multiple folding events in the same run, the folding time for the events after the first one is defined as the temporal difference between the folding and previous unfolding event. The three simulations that did not reach the folded state within 100 ns were restarted and run for an additional 100 ns, for a better estimation of the folding time. After 200 ns, only one of the simulations did not reach the Q value of 23/26. The mean folding time and unfolding time, were computed using all of the events. For all the remaining analysis, only the first 100 ns of the 20 runs (2 μs) were used.

Effective Energy and Free Energy

The effective energy and free energy surfaces, determined by simulations and experiments, play an important role for the understanding of the protein folding reaction.[22] The effective energy is the sum of the intramolecular energy (CHARMM PARAM19 force field energy) and the solvation free energy. The latter is approximated by the solvent accessible surface term of equation 1 and contains the free energy contribution of the solvent within the approximations of an implicit model of the water molecules. The effective energy does not include the configurational entropy of the peptide which consists of conformational and vibrational entropy contributions.[13] Owing to the complexity of the protein folding process, it is necessary to group states and project the conformational space onto one- or two-order parameters that characterize the system. The value of the effective energy is then averaged within a bin defined by discretizing the reduced space.

For a system in thermodynamic equilibrium, the difference in free energy in going from state A to state B is proportional to the natural logarithm of the quotient of the probability of finding the system in state A divided by the probability of state B. The free energy surface is obtained by counting populations and is usually plotted as a function of the aforementioned two-dimensional space of order parameters by using an arbitrarily chosen reference point (the full unfolded state), as the denominator of the probability quotient.

Cluster Analysis

The method for the cluster analysis is based on structural similarity.[18] The C_α RMSD is evaluated for each pair of structures after optimal superposition. For each conformation, the number of neighbors is then calculated using a C_α RMSD cutoff of 2.0 Å. The conformation with the highest number of neighbors is defined as the center of the first cluster. All the neighbors of this conformation are removed from the ensemble of conformations. The center of the second cluster is then determined in the same way as for the first cluster. This procedure is repeated until each structure is assigned to a cluster.

RESULTS AND DISCUSSION

A detailed analysis of the energy surface and folding mechanism at 360 K and preliminary simulation results at 330 K have been presented elsewhere.[8] In that study, a 200-ns run at 300 K from the folded state was used to derive time averaged inter-proton distance violations as $d_{\text{viol}} = \langle r(t)^{-6} \rangle^{-1/6} - r_{\text{exp}}$, where $r(t)$ is the inter-proton distance at simulation time t, r_{exp} is the NOE upper distance limit,[7] and $\langle\ \rangle$ represents a time average. At 300 K, it was shown that 23 of the 26 NOE restraints are satisfied (d_{viol} <0.0 Å for 16 distances and d_{viol} <1.0 Å for seven distances).[8] The 200-ns run from the folded state at 300 K was also used to define a list of 26 native contacts. The present analysis uses the same list of 26 native contacts and focuses on a set of 20 simulations of the GS peptide at 330 K started from random conformations.

Folding Mechanism

A total of 22 folding ($Q > 22/26$) and 13 unfolding ($Q < 4/26$) events were sampled in the twenty 100-ns trajectories at 330 K. There were 5, 12, and 3 runs with 2, 1 and 0 folding events, respectively. Unfolding takes place twice in 2 runs and once in 9 runs. There are two main folding pathways. Both consist of first almost complete formation of one of the two β-hairpins, which then templates docking of the unstructured strand. About two-thirds (one-third) of the folding events begin with the stabilization of β-hairpin 2–3 (1–2). The statistical weights of the folding pathways are very similar at 330 and 360 K, and the same is true for the unfolding pathways (Table I).

Figure 1 shows time series of the order parameter for trajectory 12 (Tr12) and Tr17. In Tr12, the folding transition starts at about 30 ns with the formation of most of the native contacts between strands 2 and 3 at 40 ns. This is an on-pathway intermediate (characterized by $Q_{2-3} > 0.6$ and $Q_{1-2} < 0.2$) that promotes the complete folding, i.e., the sudden stabilization of strand 1 into the preformed β-hairpin 2–3 at 55 ns. From Figure 1, it is clear that the

TABLE I. Frequency of Folding and Unfolding Pathways

Folding		
Temperature (K)	First formation of β-hairpin 2–3 (%)	First formation of β-hairpin 1–2 (%)
330	64	36
360	66	34

Unfolding		
Temperature (K)	First rupture of β-hairpin 2–3 (%)	First rupture of β-hairpin 1–2 (%)
330	31	69
360	41	59

unfolding event in Tr12 is the reverse of folding; i.e., there is first a sudden rupture of the contacts between strands 1 and 2 after ~80 ns, while the β-hairpin 2–3 remains stable until 90 ns. There are six other runs where the β-hairpin 2–3 intermediate is present for ≥10 ns, the longest period being about 35 ns in Tr13. The β-hairpin 1–2 intermediate is present for ≥10 ns in four runs, the longest period being ~35 ns in Tr9.

There are two folding and two unfolding events in Tr17 (plots on the right of Fig. 1). The time series of Q_{1-2} and Q_{2-3} show that the β-hairpin 1–2 intermediate is present from about 20–40 ns, while the β-hairpin 2–3 intermediate from about 60–75 ns. It is interesting to note that formation of contacts between side-chains precedes hydrogen bond formation (Fig. 1) and drives the folding process in agreement with the explicit water simulations of Betanova.[6]

Energy Surface

The free energy surface at 330 K [Fig. 2(a)] is similar to the one at 360 K.[8] The free energy minimum corresponding to the folded state is located at $Q = 0.81$ and $Q = 0.77$ at 330 K and 360 K, respectively [Fig. 3(a)]. It is not expected to be at $Q = 1.0$ because of entropic reasons and the fact that the native contacts were defined for the folded state at 300 K. Because of the presence of two main pathways, the one-dimensional plot [Fig. 3(a)] cannot describe the folding mechanism, which is much clearer in the two-dimensional plot. This explains the utility of two-order parameters, Q_{1-2} and Q_{2-3}, and the importance of an accurate choice of progress variables and subsets thereof. The folded (51,203 conformations with $Q > 0.7$) and unfolded (87,778 conformations with $Q < 0.3$) states have comparable free energy, which indicates that 330 K is only slightly above the melting temperature in the present model. At 10°C, the β-sheet population was estimated to be 13–31%, based on NOE intensities, and 30–55%, based on the chemical shift data.[7] One of the reasons for the overstability might be the simple model of the solvent, which does not have an explicit temperature dependence.

At both temperature values of 330 and 360 K, two thermodynamically defined transition state regions are located at $Q_{1-2} \approx 0.4$ and $0.5 < Q_{2-3} < 0.8$, and $0.5 < Q_{1-2} < 0.8$ and $Q_{2-3} \approx 0.4$. The barriers arise from the loss of conformational entropy associated with fixing about two-thirds of the chain into a β-hairpin. The free energy surface and folding pathways of the GS peptide at 330 K are consistent with the folding of a 24-residue three-stranded antiparallel β-sheet peptide investigated by NMR at temperatures ranging from 280 to 320 K.[9]

The average effective energy (intramolecular plus solvation) as a function of the Q_{1-2} and Q_{2-3} order parameters has an almost downhill profile at 330 K and 360 K with a single minimum corresponding to the fully folded conformation [Fig. 2(b)]. Both the screened electrostatic [$\epsilon(r) = 2r$] and van der Waals contributions become more favorable along the folding reaction. The plot of the effective energy as a function of Q [Fig. 3(b)] indicates that for $Q < 0.5$ the downhill profile is less pronounced at 330 K than at 360 K in agreement with previous simulation results on shorter structured peptides.[16] This is mainly due to the van der Waals interaction which upon folding improves on average by about −6 kcal/mol at 330 K and −9 kcal/mol at 360 K. The changes in the solvation term upon folding are almost negligible because of the small size of the peptide. Furthermore, the favorable effect originating from the decrease of the nonpolar surface (hydrophobic side-chain contacts) is compensated by the penalty due to the reduction of the polar surface (nitrogen and oxygen atoms involved in backbone hydrogen bonds).

The free energy surface as a function of the distal and proximal contacts at 330 K is shown in Figure 4. It is similar to the corresponding surface at 360 K (not shown). In most trajectories, folding is initiated by the formation of three to four proximal contacts in one of the two native turns and zero or one distal contacts. Although the relative importance of interactions close to, and distal from, the turn might depend on the amino acid sequence, the early formation of the turn contacts is in accord with experimental data on the WW domain,[23] a three-stranded antiparallel β-sheet. It is also consistent with implicit solvent MD simulations of the GS peptide and two other β-sheet forming synthetic peptides,[24] as well as a statistical mechanical model developed to explain laser temperature-jump experiments of the folding of the β-hairpin fragment (41–56) of protein G B1.[25] On the other hand, recent computational studies on the same fragment of protein G B1 have shown different behavior.[26,27]

Denaturated State

To characterize the unfolded state, a cluster analysis was performed on 2,195 structures, i.e., one conformation every 40 of the 87778 with $Q < 0.3$. A total of 1,424 clusters were found using a C_α RMSD cutoff of 2.0 Å (see section 2.7). The largest cluster incorporates 1.4% of the conformations (0.8% at 360 K) and the ten largest clusters 8.6% (3.1% at 360 K). The choice of order parameters and the projection of the free energy surface might hide trapped metastable states with very low values of Q. The very small populations in the largest clusters of the denaturated state indicate that this is not the case. The representatives of clusters 1, 2, 3, and 4 at 330 K have a C_α

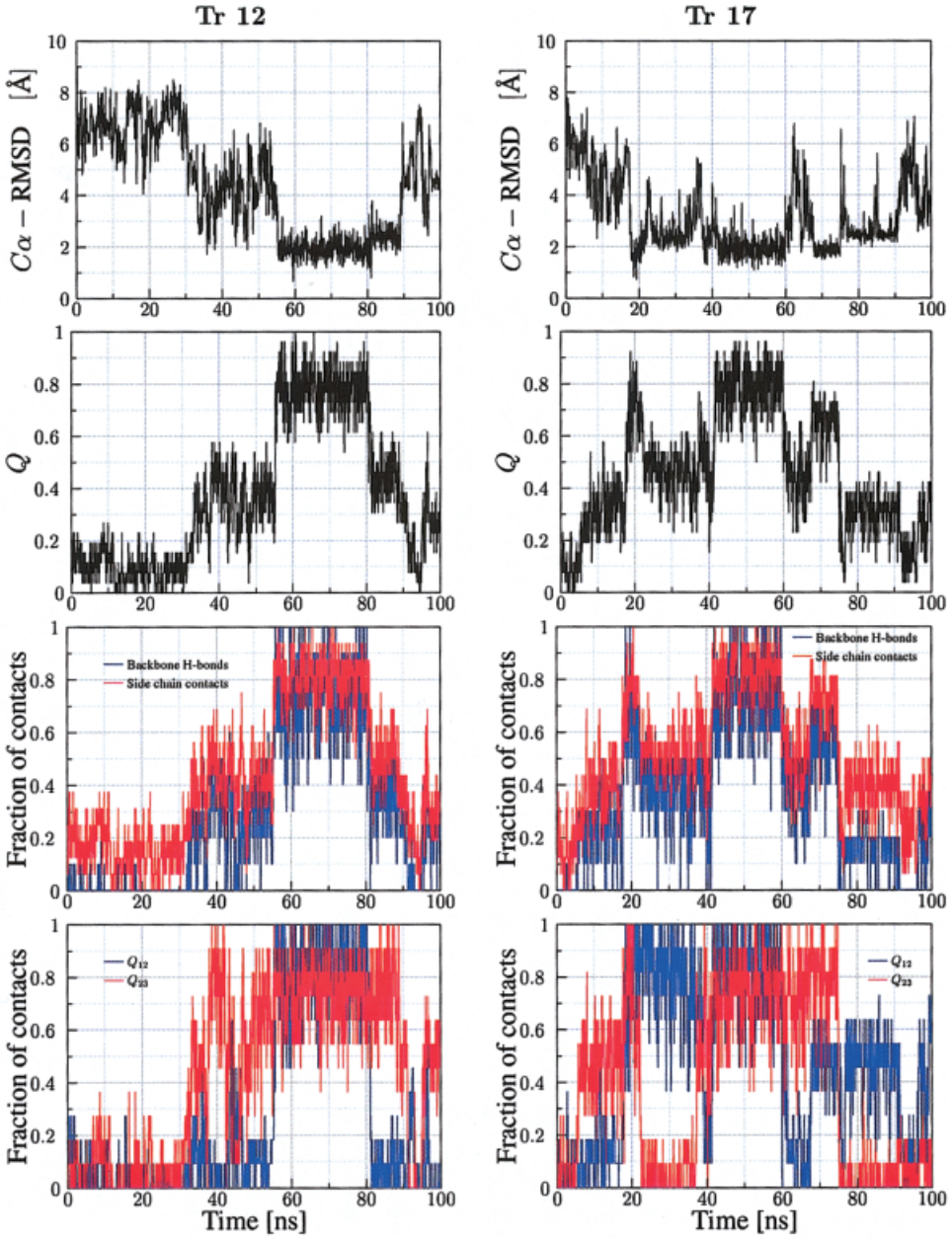

Fig. 1. Time dependence of from top to bottom the C_α RMSD, fraction of native contacts Q, fraction of native backbone hydrogen bonds and side-chain contacts, and fraction of native contacts formed between strands 1 and 2 (Q_{1-2}) and strands 2 and 3 (Q_{2-3}), for trajectory 12 (left column plots) and trajectory 17 (right column plots).

RMSD smaller than 3.5 Å from the 360 K cluster centers 4, 3, 1, and 2, respectively. They represent non-native three-stranded β-sheets.[18] This is consistent with the high occurrence of $i - i + 2$ side-chain contacts (native and non-native) in the conformations with $Q = 3/26$, the minimum of the one-dimensional free energy projection (Fig. 3a). Among the 22014 conformations with $Q = 3/26$, the contacts between side chains of residues 18–20, 15–17, 13–15, 1–3, 16–18, and 5–7 are present in 55%, 55%, 55%, 54%, 53%, and 49% of the structures, respectively.

The average radius of gyration of the conformations in the denaturated state ensemble is 7.79 ± 0.40 Å (8.00 ± 0.50 Å at 360 K). Hence, the denaturated state at 330 K is rather compact and only slightly expanded with respect to the folded state at 300 K, which has a radius of gyration of 7.66 ± 0.13 Å. As a basis of comparison, the 20 NMR conformations have a radius of gyration of 7.71 ± 0.21 Å. The compactness of the unfolded state is in agreement with explicit solvent MD simulations of the β-hairpin fragment (41–56) of protein G at 400 K.[26]

Folding and Unfolding Times

At 330 K, the folding time varies among different trajectories from 2 ns to 140 ns with an average of 39 ns,

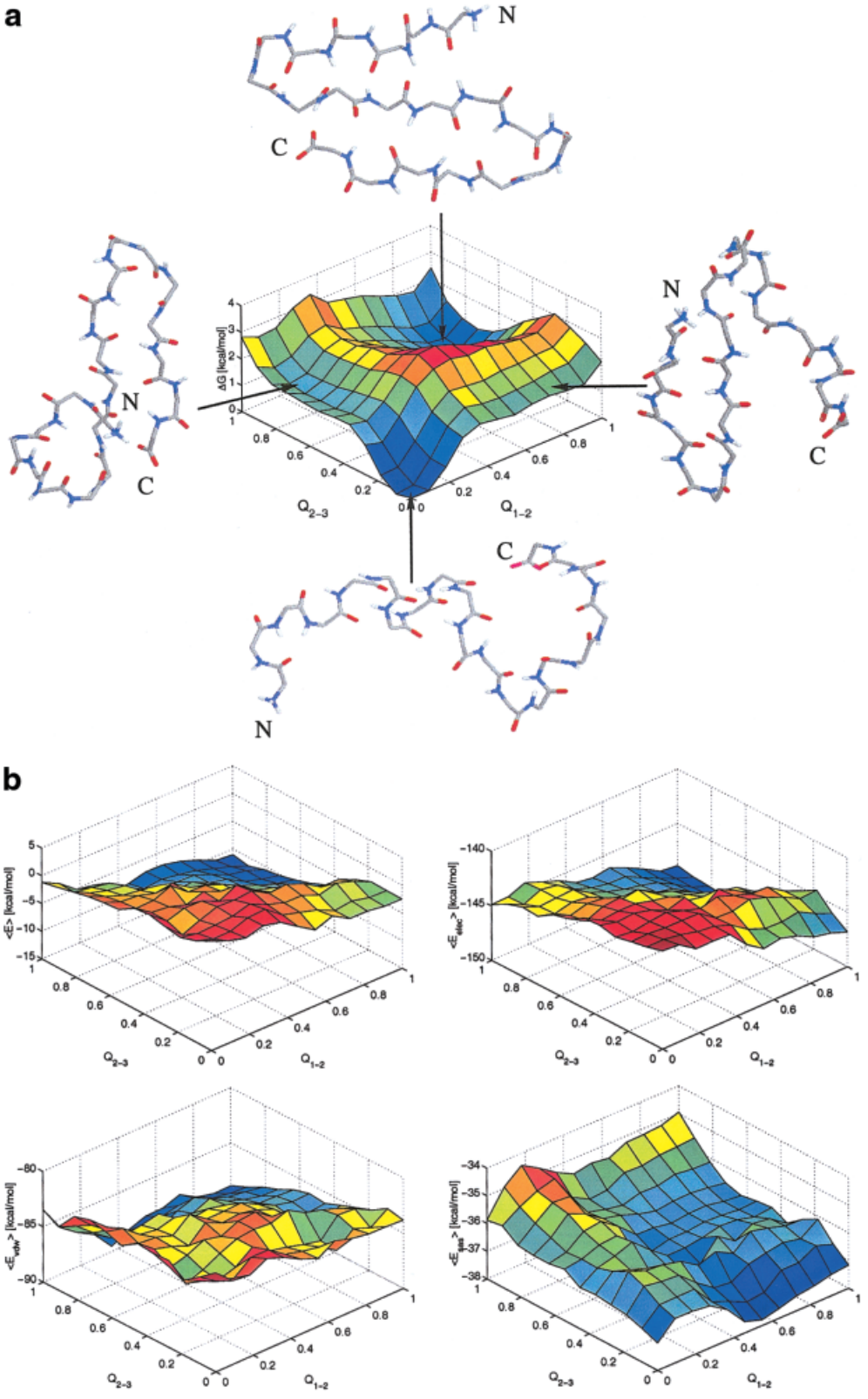
a
N
C
N
C
N
C
C
N
ΔG [kcal/mol]
Q2-3
Q1-2
b
<E> [kcal/mol]
<Eelec> [kcal/mol]
<Evdw> [kcal/mol]
<Egas> [kcal/mol]
Q2-3
Q1-2

Fig. 2.

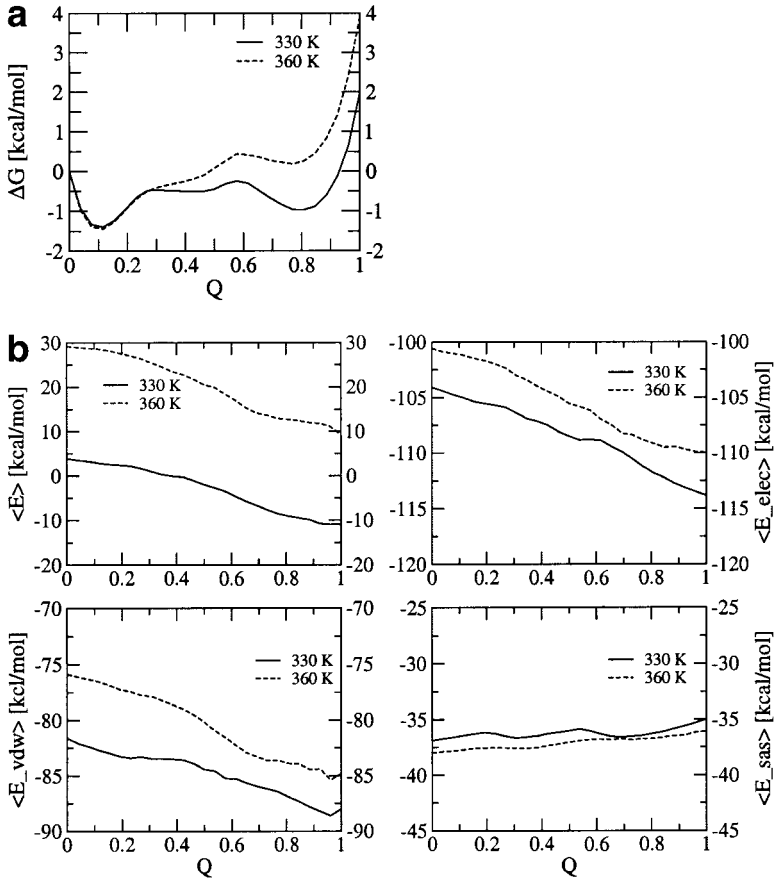

Fig. 3. Energy plots of the 2.0×10^5 and 3.0×10^5 conformations sampled during the 20 runs at 330 K and the 40 runs at 360 K, respectively. **a:** Free energy as a function of the fraction of native contacts between residues. ΔG was computed as

$$-k_b T \ln\left(\frac{N_n}{N_0}\right),$$

where N_n denotes the number of conformations with n native contacts. The statistical error in ΔG was estimated as explained in Figure 2. The average and maximal error of ΔG are 0.6 and 0.7 ($n = 2$) kcal/mol at 330 K and 0.6 and 0.9 ($n = 26$) at 360 K, respectively. **b:** Average effective energy and its contributions plotted as a function of the fraction of native contacts. The statistical error was estimated as explained in Figure 2. The average and maximal errors are: $\langle E \rangle$, 0.2 and 0.7 kcal/mol at 330 K and 0.4 and 0.5 kcal/mol at 360 K; $\langle E_{\text{elec}} \rangle$, 0.1 and 0.4 kcal/mol at 330 K and 0.2 and 0.4 kcal/mol at 360 K; $\langle E_{\text{vdw}} \rangle$ 0.2 and 0.5 kcal/mol at 330 K and 0.2 and 0.4 kcal/mol at 360 K; $\langle E_{\text{sas}} \rangle$, 0.1 and 0.3 kcal/mol at 330 K and 0.1 and 0.2 kcal/mol at 360 K.

Fig. 2. **a:** Free energy at 330 K as a function of the fraction of native contacts between residues in strands 1 and 2 (Q_{1-2}) and between residues in strands 2 and 3 (Q_{2-3}). A total of 2.0×10^5 conformations sampled during the 20 runs at 330 K were used. ΔG was computed as

$$-k_b T \ln\left(\frac{N_{n,m}}{N_{0,0}}\right),$$

where $N_{n,m}$ denotes the number of conformations with n (m) contacts formed between strands 1 and 2 (2 and 3). The statistical error in ΔG was estimated by separating the 20 simulations in two sets of 10. The average and maximal error of ΔG are 0.6 and 1.0 kcal/mol (bin $n = 2, m = 5$), respectively. The backbone of a random structure ($Q_{1-2} = Q_{2-3} = 0$) used as starting conformation for one of the runs is also shown together with three conformations saved along the trajectories and representing the β-hairpins and the folded state. **b:** Average energies at 330 K ($\langle E \rangle$ total energy, intramolecular plus solvation, $\langle E_{\text{elec}} \rangle$ electrostatic term with $\epsilon(r) = 2r$, $\langle E_{\text{vdw}} \rangle$ van der Waals term, $\langle E_{\text{sas}} \rangle$ solvent accessible surface solvation term) as a function of Q_{1-2} and Q_{2-3}. The statistical error was estimated by separating the 20 runs in two sets of 10 simulations each. The average and maximal errors are: $\langle E \rangle$ 0.7 and 3.8 kcal/mol (bin $n = 1, m = 7$), $\langle E_{\text{elec}} \rangle$ 0.4 and 2.0 kcal/mol (bin $n = 2, m = 5$), $\langle E_{\text{vdw}} \rangle$ 0.5 and 3.2 kcal/mol (bin $n = 1, m = 7$), and $\langle E_{\text{sas}} \rangle$ 0.2 and 0.6 kcal/mol (bin $n = 6, m = 1$).

while the unfolding time varies from 2 ns to 65 ns with an average of 28 ns. The mean folding time at 330 K is similar to the value at 360 K (32 ns[8]), whereas the unfolding time is four times longer (28 ns vs 7 ns at 360 K). Although the present model neglects the temperature dependence of the hydrophobic interaction it shows non-Arrhenius temperature dependence of the folding rate and Arrhenius behavior for the unfolding rate. This is in agreement with previous MD simulation results on shorter structured peptides[16] and experimental data obtained using a variety of spectroscopic techniques on peptides[25,28] and proteins.[29–31]

All the folding events were used to analyze the time dependence of folding, i.e., the decay of the unfolded population. In runs with more than one folding event, each event was considered separately, to increase the population size. The fraction of unfolded molecules at time t is the

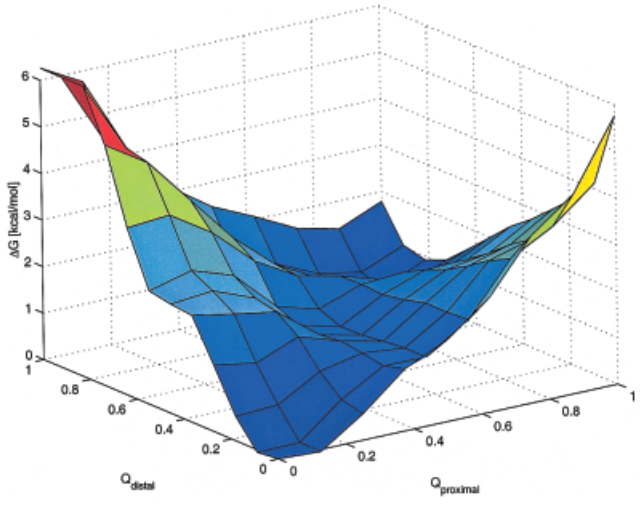

Fig. 4. Free energy surface as a function of the fraction of the 8 native contacts close to ($Q_{proximal}$) and the 11 contacts distant from (Q_{distal}) the turns. A total of 2.0 $\times$ 10^5 conformations sampled during the 20 simulations at 330 K were used.

percentage of the population that has not reached the folded state at time t. An analogous definition was used for the decay of the folded population to analyze the time dependence of unfolding. At 330 K, a monoexponential fit to the fraction of unfolded population yields a characteristic time of the decay of 38 ns, in accord with the average folding time of 39 ns, and a correlation factor of 0.99. For unfolding, the characteristic time of the decay of the folded population is 29 ns and the correlation factor is 0.98. Thus, both folding and unfolding show an exponential time dependence; i.e., the decay of the unfolded and folded population is well represented by a single exponential, in agreement with the presence of a dominant barrier for each of the two main pathways in the free energy surface [Fig. 2(a)]. This is also consistent with previous simulation results at 360 K.[8]

An upper limit for the in vitro folding time of 16–45 μs at 5°C and 4–14 μs at 10°C was estimated for the GS peptide using the one-dimensional ^{1}H-NMR spectra.[7] The shorter folding time in the simulations is due in part to the higher temperature. Furthermore, the implicit solvent model does not take explicitly into account the van der Waals interactions and hydrogen bonds between the atoms of the peptide and the water molecules so that the friction originates solely from the intrapeptide interactions. As a basis of comparison, the conformational transition between the A- and B-DNA forms is about 20 times faster when using the generalized Born implicit solvation model than explicit water molecules.[32]

Following the suggestion of an anonymous referee, the friction effects were investigated by six 330 K Langevin dynamics runs of 400 ns each. At each timestep in Langevin dynamics, the force on each atom is supplemented by an additional stochastic force, approximating the random collisions of water molecules and solute, and a drag force proportional to the atom velocity. A friction coefficient of 6 ps^{-1} was used on all atoms. In only two of the six Langevin dynamics simulations the folded state ($Q > 22/26$) was reached within 400 ns. This happened at 231 ns in one run and at 329 ns in the other run. Hence, the average folding time according to the Langevin simulations at 330 K is $\sim \geq 0.5$ μs. This is closer to the experimentally measured folding time, and more than an order of magnitude slower than the average folding time of 39 ns obtained by Newtonian dynamics at 330 K.

Temperature Dependence of the Free Energy of an α-Helix and a β-Hairpin

In a previous work, the folding thermodynamics and kinetics of two structured peptides were investigated at several temperature values by 862 MD simulations (total of about 4 μs) with the same implicit solvation model as the one used for the GS peptide.[16] For the helical peptide Ace-$(AAQAA)_3$-$NHCH_3$ the plot of the free energy as a function of native contacts (helical hydrogen bonds) has a single minimum whose position depends on the temperature [Fig. 5(a)]. An increase in temperature shifts the position of the minimum toward a smaller amount of helical hydrogen bonds but preserves the overall profile of the free energy. The β-hairpin Ace-$V_5{}^D PGV_5$-NH_2 has similar energy profile at 330 and 360 K; there are two minima separated by a barrier at about 30% of the native contacts [Fig. 5(b)].

Figure 5(c,d) shows the order of contact formation during folding as a function of the C_α RMSD from the folded state. Native contacts that on average form early (late) during folding are close to the top of Figure 5(c,d); i.e., they appear at high (low) values of the C_α RMSD. The sequence of events does not depend either on the temperature or on the direction of the process (folding vs. unfolding). The large standard deviations indicate that multiple folding routes are possible for both peptides and more so for the α-helix. No preferred pathway emerges from the simulation results of the helical peptide if one excludes the late formation of the hydrogen bonds at the termini. This might be related to their intrinsic instability. The β-hairpin initiates folding mainly at the β-turn and then progressively zips up.

These simulation results provide further support to the observation that the free energy profile and folding mechanism (multiple pathways, ubiquitous nucleation of the α-helix and early formation of the turn in the β-hairpin) of structured peptides have a relatively weak temperature dependence.

CONCLUSIONS

The MD simulations of the GS peptide, totaling >5 μs, indicate that the free energy surface, transition state regions, and folding mechanism are similar at 330 K and 360 K. Moreover, the predominant folding pathway is the same at both temperature values and the same is true for its statistical weight. The unfolded state is slightly more compact at the lower temperature. The electrostatic and van der Waals energy contributions show a downhill profile at both temperature values. In a previous MD analysis of two structured peptides, a 15-residue α-helix

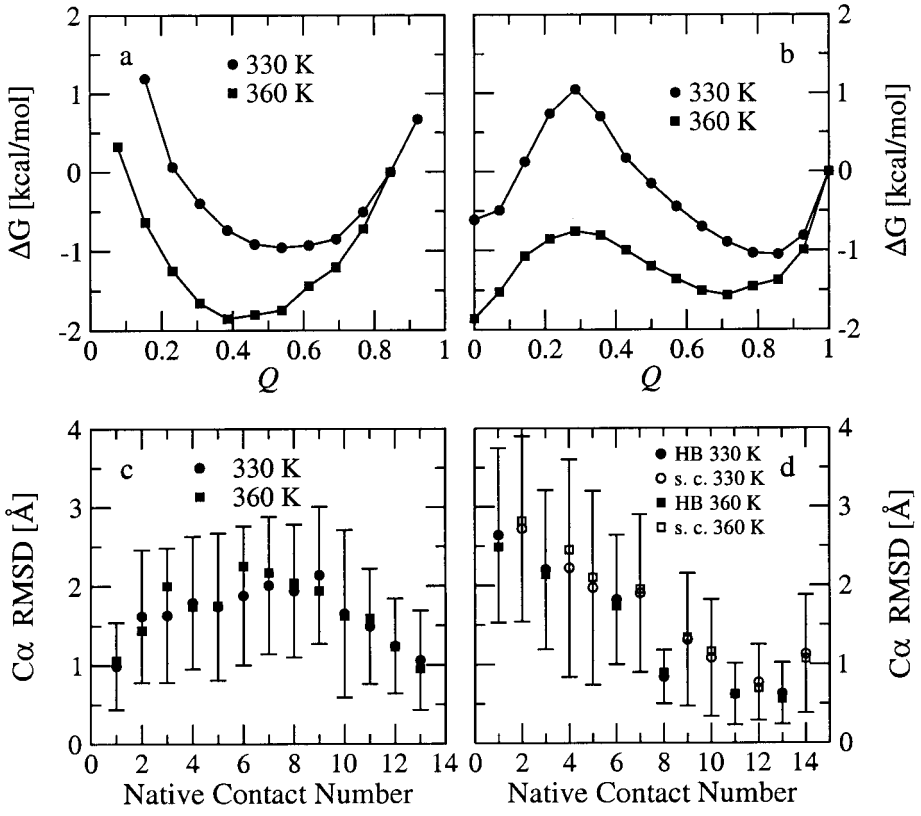

Fig. 5. **Top:** Free energy of the α-helical peptide Ace-$(AAQAA)_3$-$NHCH_3$ (**a**) and the β-hairpin Ace-$V_5{}^DPGV_5$-NH_2 (**b**) as a function of the fraction of native contacts. A few data points for the α-helix are not shown because of insufficient statistics. The free energy was arbitrarily set to zero at $Q = 0.85 = 11/13$ for the α-helix (due to the insufficient statistics at $Q = 1$), and at $Q = 1 = 14/14$ for the β-hairpin. **Bottom:** Average value of the C_α RMSD from the folded state at the last disappearance of the α-helical contacts of Ace-$(AAQAA)_3$-$NHCH_3$ (**c**) and the β-hairpin contacts of Ace-$V_5{}^DPGV_5$-NH_2 (**d**) in the simulations started from random conformations. The native contacts of Ace-$(AAQAA)_3$-$NHCH_3$ are the 13 hydrogen bonds present in the α-helix and are ordered on the *x*-axis from the N-terminus to the C-terminus. For Ace-$V_5{}^DPGV_5$-NH_2, the β-hairpin contacts are ordered on the *x*-axis from the turn to the terminus, and filled and empty square symbols represent backbone hydrogen bonds and side chain contacts, respectively. The last disappearance of a contact is defined as the last interval of ≥50 ps, during which the contact is not present. A contact is said to be present if the distance between the two atoms defining the contact is less than that in the folded state times 1.3. The standard deviations are given only at 330 K for the sake of clarity. They are similar at 360 K. Bars = 2 SD.

and 12-residue β-hairpin, it was shown that the free energy profile and the folding mechanism have a marginal temperature dependence between 300 and 360 K.[16] Taken together, the MD results suggest that the folding mechanism proposed here may be valid at room temperature. The presence of two main pathways, involving first the almost complete formation of either of the β-hairpins, is consistent with recent NMR data of another designed 24-residue three-stranded antiparallel β-sheet.[9]

Although the simulation results provide answers to the questions formulated in the Introduction, it is not clear how much their validity is limited to structured peptides which differ from proteins mainly because of the high symmetry in the native topology and the absence of fully buried hydrophobic cores. Hence, it cannot be excluded that proteins, even small ones, with well-defined hydrophobic cores might show substantial variability in the folding mechanism and free energy landscape at different values of the temperature. That the pathway frequency is the same at 330 K and 360 K might be attributable to the symmetry of the sequence and conformation of the GS peptide and may not hold in general. Monte Carlo simulations of a 125-residue protein model on a lattice have shown that, at very high temperatures, the unfolding trajectories are less diverse than the multiple routes of folding found at temperature values corresponding to those of temperature-jump experiments.[33]

The simulation of proteins with a well-defined hydrophobic core will require a less crude approximation of solvation.[14] We also plan to simulate other miniprotein motifs and mutants whose folding thermodynamics and kinetics have been characterized by circular dichroism and nanosecond laser temperature-jump experiments.[23]

ACKNOWLEDGMENTS

The simulations were performed on a Beowulf cluster running Linux and we thank Dr. N. Budin for his invaluable help in setting up the cluster and computer support. We thank Dr. J. Gsponer and U. Haberthür for interesting discussions. We also thank Dr. M.A. Jiménez for providing

the NMR conformations of the GS peptide. This work was supported in part by the Swiss National Science Foundation (grant 31-53604.98 to A.C.) and the Théodore Ott Foundation.

REFERENCES

1. Dobson CM, Karplus M. The fundamentals of protein folding: bringing together theory and experiment. Curr Opin Struct Biol 1999;9:92–101.
2. Caflisch A, Karplus M. Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. J Mol Biol 1995;252:672–708.
3. Lazaridis T, Karplus M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science 1997;278:1928–1931.
4. Mohanty D, Elber R, Thirumalai D, Beglov D, Roux B. Kinetics of peptide folding: computer simulations of SYPFDV and peptide variants in water. J Mol Biol 1997;272:423–442.
5. Schaefer M, Bartels C, Karplus M. Solution conformation and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. J Mol Biol 1998;284:835–848.
6. Bursulaya BD, Brooks CL III. Folding free energy surface of a three-stranded β-sheet protein. J Am Chem Soc 1999;121:9947–9951.
7. De Alba E, Santoro J, Rico M, Jiménez MA. De novo design of a monomeric three-stranded antiparallel β-sheet. Protein Sci 1999; 8:854–865.
8. Ferrara P, Caflisch A. Folding simulations of a three-stranded antiparallel β-sheet peptide. Proc Natl Acad Sci USA 2000;97: 10780–10785.
9. Griffiths-Jones SR, Searle MS. Structure, folding, and energetics of cooperative interactions between the β-strands of a de novo designed three-stranded antiparallel β-sheet peptide. J Am Chem Soc 2000;122:8350–8356.
10. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4: 187–217.
11. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature 1986;319:199–203.
12. Hasel W, Hendrickson TF, Still WC. A rapid approximation to the solvent accessible surface areas of atoms. Tetrahedron Comput Methodol 1988;1:103–116.
13. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35:133–152.
14. Ferrara P, Apostolakis J, Caflisch A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. Proteins 2002; 46:24–33.
15. Elcock AH, McCammon JA. Continuum solvation model for studying protein hydration thermodynamics at high temperatures. J Phys Chem B 1997;101:9624–9634.
16. Ferrara P, Apostolakis J, Caflisch A. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. J Phys Chem B 2000;104:5000–5010.
17. Hiltpold A, Ferrara P, Gsponer J, Caflisch A. Free energy surface of the helical peptide $Y(MEARA)_6$. J Phys Chem B 2000;104:10080–10086.
18. Ferrara P, Caflisch A. Native topology or specific interactions: what is more important for peptide folding? J Mol Biol 2001;306: 837–850.
19. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
20. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of *n*-alkanes. J Comp Phys 1977;23:327–341.
21. Chen S, Dill KA. RNA folding energy landscapes. Proc Natl Acad Sci USA 2000;97:646–651.
22. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. Understanding protein folding via free-energy surfaces from theory and experiment. TIBS 2000;25:331–339.
23. Jäger M, Nguyen H, Crane JC, Kelly JW, Gruebele M. The folding mechanism of a β-sheet: the WW domain. J Mol Biol 2001;311:373–393.
24. Wang H, Sung S. Molecular dynamics simulations of three-strand β-sheet folding. J Am Chem Soc 2000;122:1999–2009.
25. Muñoz V, Thompson PA, Hofrichter J, Eaton WA. Folding dynamics and mechanism of β-hairpin formation. Nature 1997;390:196–199.
26. Pande VS, Rokhsar DS. Molecular dynamics simulations of unfolding and refolding of a β-hairpin fragment of protein G. Proc Natl Acad Sci USA 1999;96:9062–9067.
27. Dinner AR, Lazaridis T, Karplus M. Understanding β-hairpin formation. Proc Natl Acad Sci USA 1999;96:9068–9073.
28. Lednev IK, Karnoup AS, Sparrow MC, Asher SA. α-Helix peptide folding and unfolding activation barriers: a nanosecond UV resonance raman study. J Am Chem Soc 1999;121:8074–8086.
29. Segawa SI, Sugihara M. Characterization of the transition state of lysozyme unfolding. I. Effect of protein–solvent interactions on the transition state. Biopolymers 1984;23:2473–2488.
30. Oliveberg M, Tan YJ, Fersht AR. Negative activation enthalpies in the kinetics of protein folding. Proc Natl Acad Sci USA 1995;92:8926–8929.
31. Matagne A, Jamin M, Chung EW, Robinson CV, Radford SE, Dobson CM. Thermal unfolding of an intermediate is associated with non-Arrhenius kinetics in the folding of hen lysozyme. J Mol Biol 2000;297:193–210.
32. Tsui V, Case DA. Molecular dynamics simulations of nucleic acids with a Generalized Born solvation model. J Am Chem Soc 2000;122:2489–2498.
33. Dinner AR, Karplus M. Is protein unfolding the reverse of protein folding? A lattice simulation analysis. J Mol Biol 1999;292:403–419.