PROTEINS: Structure, Function, and Genetics 42:345–354 (2001)

Exploring the Energy Landscape of a β Hairpin in Explicit Solvent

Angel E. García[1*] and Kevin Y. Sanbonmatsu[2]
[1]*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico*
[2]*Applied Physics Division, Los Alamos National Laboratory, Los Alamos, New Mexico*

ABSTRACT **We studied the energy landscape of the peptide Ace-GEWTYDDATKTFTVTE-Nme, taken from the C-terminal fragment (41–56) of protein G, in explicit aqueous solution by a highly parallel replica-exchange approach that combines molecular dynamics trajectories with a temperature exchange Monte Carlo process. The combined trajectories in *T* and configurational space allow a replica to overcome a free energy barrier present at one temperature by increasing *T*, changing configurations, and cooling in a self-regulated manner, thus allowing sampling of broad regions of configurational space in short (nanoseconds) time scales. The free energy landscape of this system over a wide range of temperatures shows that the system preferentially adopts a beta hairpin structure. However, the peptide also samples other stable ensembles where the peptide adopts helices and helix-turn-helix states, among others. The helical states become increasingly stable at low temperatures, but are slightly less stable than the beta turn ensemble. The energy landscape is rugged at low *T*, where substates are separated by large energy barriers. These barriers disappear at higher *T* (~330 K), where the system preferentially adopts a "molten globule" state with structures similar to the beta hairpin. Proteins 2001;42:345–354.**

Key words: protein folding; protein folding kinetics; energy landscape

INTRODUCTION

The process by which proteins attain their functional structure has been the subject of numerous experimental and theoretical studies over the last few decades. Proteins, the molecular machines that perform most of the biologically relevant functions in living organisms, attain a unique tertiary structure in short time scales. The fastest folding time is microseconds (μs).[1] However, not all proteins spontaneously adopt their folded structure, because many proteins adopt a functional structure upon binding to target proteins or nucleic acids.[2] Disfunctional states of proteins that fold into the incorrect structures are believed to be the cause of degenerative diseases.[3–5] Recent theoretical advances and modern experimental techniques that probe proteins at different stages during the folding process have shed light on the nature of the physical mechanism and relevant interactions that determine the kinetics of folding, binding, function, and thermodynamic stability. Much of this theoretical understanding has been tested in minimalist lattice and off-lattice models of proteins.[6–8] All atom simulation of protein unfolding using realistic models, although limited by insufficient configurational space sampling, has also revealed much about the complexity of the protein folding.[9–11] Short peptides have been shown to mimic many of the properties of proteins and provide computationally simpler systems to test theoretical models of folding in atomic detail.[10,12–17]

The simulation of biological macromolecular systems by molecular dynamics (MD) and Monte Carlo (MC) techniques is limited by multiple time scale relaxation processes that lead to insufficient sampling of the important configurational space. At room temperature, biomolecular systems get trapped in many local minima. This trapping limits the capacity to effectively sample configurational space. One way to overcome this limitation has been to perform simulations using non-Boltzmann sampling techniques.[18] These algorithms have been applied to proteins in vacuo or in implicit solvent models.[15,19,20] Other algorithms include replica exchange methods[21,22] and have been used to study polymers, spin glass systems, and peptides. The relationship between spin glass systems and proteins has been used to derive the energy landscape theory of protein folding.[23–26] MC implementations of the replica method have been shown to produce ergodic sampling of states, thus avoiding kinetic traps. Sugita and Okamoto[22] developed a formulation for a combined MD/MC implementation of the replica exchange method. Here we apply this method to study the equilibrium thermodynamics of a blocked 16 amino acid fragment of GB1 protein that forms a stable hairpin. The kinetics and stability of this peptide have been studied extensively by time-resolved spectroscopy[16] and theoretical studies.[10,15,27,28] By analyzing the free energy landscape of this system over a wide range of temperatures, we have found that the system

Abbreviations: PMF, potential mean force; *T*, temperature.

Grant sponsor: U.S. DOE; Grant number: W-740-ENG-36; Grant sponsor: LDRD program at Los Alamos; Grant sponsor: IHPC.

*Correspondence to: Angel E. García, Theoretical Division, MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545. E-mail: angel@t10.lanl.gov

Received 13 July 2000; Accepted 22 September 2000

preferentially adopts a β hairpin structure at low temperatures. We also identify other stable ensembles in which the peptide adopts helices, helix-turn-helix states, among other structures. The helical states become increasingly stable at low temperatures. The ensembles containing helical structures are slightly less stable than the β turn ensemble. These states are separated by large energy barriers at low temperature. These barriers disappear at higher T (~330 K).

METHODS AND SYSTEM DESCRIPTION

Description of the Replica MD/MC Method

The replica exchange method has been implemented with a number of M replicas distributed over a number of M processors, where each replica system is a peptide, with explicit solvent. Each system is simulated at different T distributed over a broad range (270–525 K). Replicas are sorted according to T and couple to each other via a temperature exchange MC procedure. At fixed time intervals, neighboring systems, i and j, with temperatures T_i and T_j, respectively, can exchange configurations, such that system i changes to temperature T_j, and system j to temperature T_i. The probability that this exchange occurs satisfies detailed balance, $W(X)w(X, X') = W(X')w(X', X)$, where $W(X)$ is the weighting factor for the state X, and $w(X, X')$ is the transition probability of exchanging system X to system X'. $W(X)$ is given by the product of the Boltzmann factors for each of the M replicas $W(X) = \exp[-\Sigma_i^M \beta_i E(\vec{x}_i, \vec{p}_i]$, with $E = E_{kin} + U$. This gives $w(X, X')/w(X', X) = \exp(-\Delta)$, *where* $\Delta = (\beta_i - \beta_j)(U_i - U_j)$ and U_i is the potential energy of system i, before the exchange. Notice that $(\beta_i - \beta_j) = 1/\kappa_B T_{eff}$, where $T_{eff} = T_i T_j/|T_i - T_j|$, is on the order of 15,000 K for $T_i = 300$ K, and $T_i - T_j \sim 6$ K, and therefore we can have replica exchanges for systems with differences in energy of ~50 $\kappa_B T_i$. These transition probabilities are implemented using the Metropolis criterion. Only replicas having neighboring temperatures can exchange. The direction of exchange of neighboring replicas is chosen at random.

The Hairpin From Protein GB1

The blocked peptide with Ace-G^{41} *EWTYDDATKTFTVTE*^{56}-Nme, taken from the C-terminal fragment (41–56) of protein GB1 forms a β hairpin in solution.[29] Thermodynamics, structural, and time-resolved studies have shown that this peptide captures much of the basic physics of protein folding, including stabilization by hydrogen bonding and hydrophobic interactions, and a funnel-like, partially rugged energy landscape.[16] The folding time for this hairpin is 6 μs, which is about a factor of 100–1000 longer than we can routinely sample in a single MD simulation. The longest MD simulation had reached 1 μs.[30] The system of 16 amino acid residues contains 260 atoms in the peptide and 1,423 water molecules, for a total of 4,529 atoms per replica. We use the force field of Cornell et al.[31] Simulations were done at constant (N, T, V), with a density of ~1.0 g/cm^3. The temperature coupling was set to 0.1 ps.[32] Electrostatic interactions were cut off at 9.0 Å, with interaction pairs evaluated every 10 integration steps of 0.002 ps. We simulated this system with 32 replicas, with T ranging from 270–525 K at steps ranging from 6–10 K, for 3.5 ns after a 60 ps equilibration period. Simulations were started from an energy minimized structure obtained by nuclear magnetic resonance (NMR) (1GB1).[33] We first performed 200-ps simulations at 300 K at constant pressure (1 atm). The final configuration of this system was then subjected to 60 ps of simulation at constant T in the temperature range indicated above. The initial configurations for the 32 replicas were taken from the final configurations of these runs. Average energies and energy fluctuations over this T range were calculated from these runs to estimate the optimal differences in temperature between neighboring replicas needed to obtain a replica-exchange acceptance ratio of ~20%. T-exchanges were attempted every 0.25 ps. Protein configurations were saved at a rate of 4/ps, for a total of 0.45 million configurations. The total integration time of all replicas is 112 ns.

Principal Components (PC) Analysis

To study the system fluctuations, we use a set of directions, $\vec{m}$, in the 3N-dimensional space of the protein, that best represent (in a least-square sense) the ensemble of protein structures observed in the simulations. The construction of these coordinates has been described previously.[34,35] PC coordinates, m, are obtained by solving an eigenvalue equation, $\sigma \cdot \vec{m} = \lambda \vec{m}$, where the matrix $\sigma_{\alpha,\beta}$ is positive semi-definite, and defined by $\sigma_{\alpha,\beta} = \frac{1}{S}\Sigma_{i=1}^{S}(r_i - y_0)_\alpha (r_i - y_0)_\beta$. Here, $y_0 = \frac{1}{S}\Sigma^S \vec{r}_i$ is the average configuration, S is the number of configurations, $\vec{r}_i$, and α and β refer to Cartesian components of the 3N protein atomic coordinates. PC are systematically ranked in the order of decreasing eigenvalues. The mean square fluctuations are given by $(1/N)Tr(\sigma) = (1/N)\Sigma_i \lambda_i$. A generalization of this method to represent two-dimensional and three-dimensional projections of the configurational space, as planes and volumes that better represent the fluctuations of the system, has been presented previously.[35] These coordinates are specific to the molecule and trajectory sampled during an MD simulation, and can be seen as PC axes of the protein configurational space along a given trajectory. The PC are calculated by including the fluctuations of all replicas at all temperatures together. This set of PC is used to describe energy surfaces at each temperature. Only fluctuations of the peptide's nonhydrogen atoms (136 atoms) were included in the PC analysis. The PMF is calculated from the population densities at each temperature, $\beta W(X_1, X_2) = -\log P(X_1, X_2)$, where $P(X)$ is the normalized probability as a function of X, and X is any set of parameters describing the peptide conformations.

RESULTS

We describe free energy surfaces as a function of various structural reaction coordinates, including the number of native hydrogen bonds in the β hairpin (N_{HB}^{β}), the radius of gyration of the hydrophobic cluster containing Phe, Tyr, and Trp (Rg_{hp}),[10] the radius of gyration of all the peptide

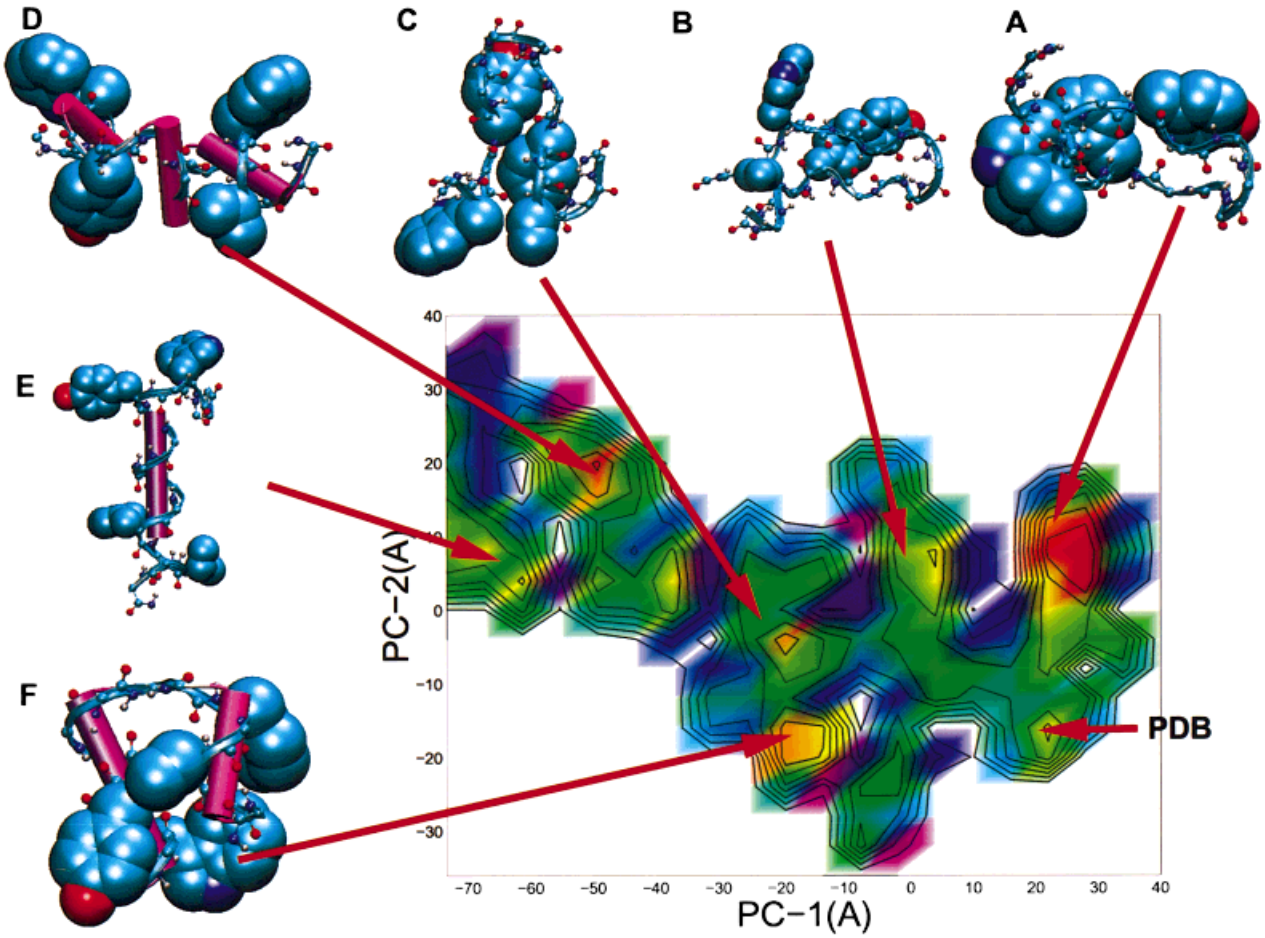

Fig. 1. Free energy surface maps as a function of the two PCs of the system at temperature $T = 270$ K. We identify various local minima: (**A**) at (PC_1, PC_2) ~ (25 Å, 10 Å), corresponding to the β hairpin "folded" ensemble, (**B**) (0,10), a β hairpin with a partially disordered hydrophobic cluster, (**C**) (−20 Å, 0 Å) an end-twisted disordered turn, (**D**) (−50 Å, 20 Å) structures containing various short helical regions, (**E**) (−60 Å, 20 Å), helical and other extended structures ensemble, and (**F**) (−20 Å, −20 Å), a helix-turn-helix motif ensemble. The first PC coordinate, PC_1, distinguishes well between various local energy minima observed, with the exception of structures C and F. The region labeled PDB refers to the peptide structure in protein G. Helical structures are defined using the program Stride[42] as implemented in VMD.[43] The side chains of W43, Y45, F52, and V54 are shown in space filing mode.

nonhydrogen atoms (*Rg*), the C_α atoms root-mean-square distance (*rmsd*) from the most stable ensemble of structures, and PC, as described above. We first obtain free energy surfaces as a function of the PC of the system by collecting histograms of the average occupation of the PC space as a function of temperature. Figure 1 shows the PMF as a function of two PCs (PC_1 and PC_2), at 270 K. At low temperatures ($T < 330\ K$) the PMF is rugged with well-defined local energy minima and high barriers separating these minima. Characteristic structures of the various local energy minima observed are also shown in Figure 1. In this plot, we emphasize helical structures, the formation and disruption of the hydrophobic core, and the overall fold of the chain.

Figure 2 shows the PMF at various temperatures. At high temperatures, the energy landscape is less rugged, and multiple paths connect one local energy minimum basin to another. At very high temperatures, only one basin, covering a wide region of the PC space near the β hairpin ensemble, is populated. This state is similar to the "molten globule" *H* state described by Pande and Rokhsar.[10]

We define the folded structure ensemble as the most probable ensemble at low *T*. A typical folded structure is shown in Figure 1A. This structure maintains on average four native hydrogen bonds. In addition, an i, $i + 3$ backbone hydrogen bond is formed between D46 and T49. Side-chain hydrogen bonds are formed between T49 OG and D46 OD, T49 OG and T51 OG, and T44 OG and T55 OG. A hydrophobic core composed of W43, Y45, F52, and V54, with multiple nonpolar atom pairs ($d < 5$ Å) between W-V, W-F, and F-Y is also formed. The Cα *rmsd* of this structure from the NMR structure is 3.0 Å, and 1.5 Å for amino acids 47–54. This *rmsd* deviation from the peptide in protein GB1 has also been noticed by Roccatano et al.[28] Other stable secondary structural ensembles identified contain α helices (Fig. 1E), helix-turn-helix motifs (Fig. 1F), and other structures containing multiple, short segments of helices (Fig. 1D). These ensembles may serve as free energy traps in the folding process at low tempera-

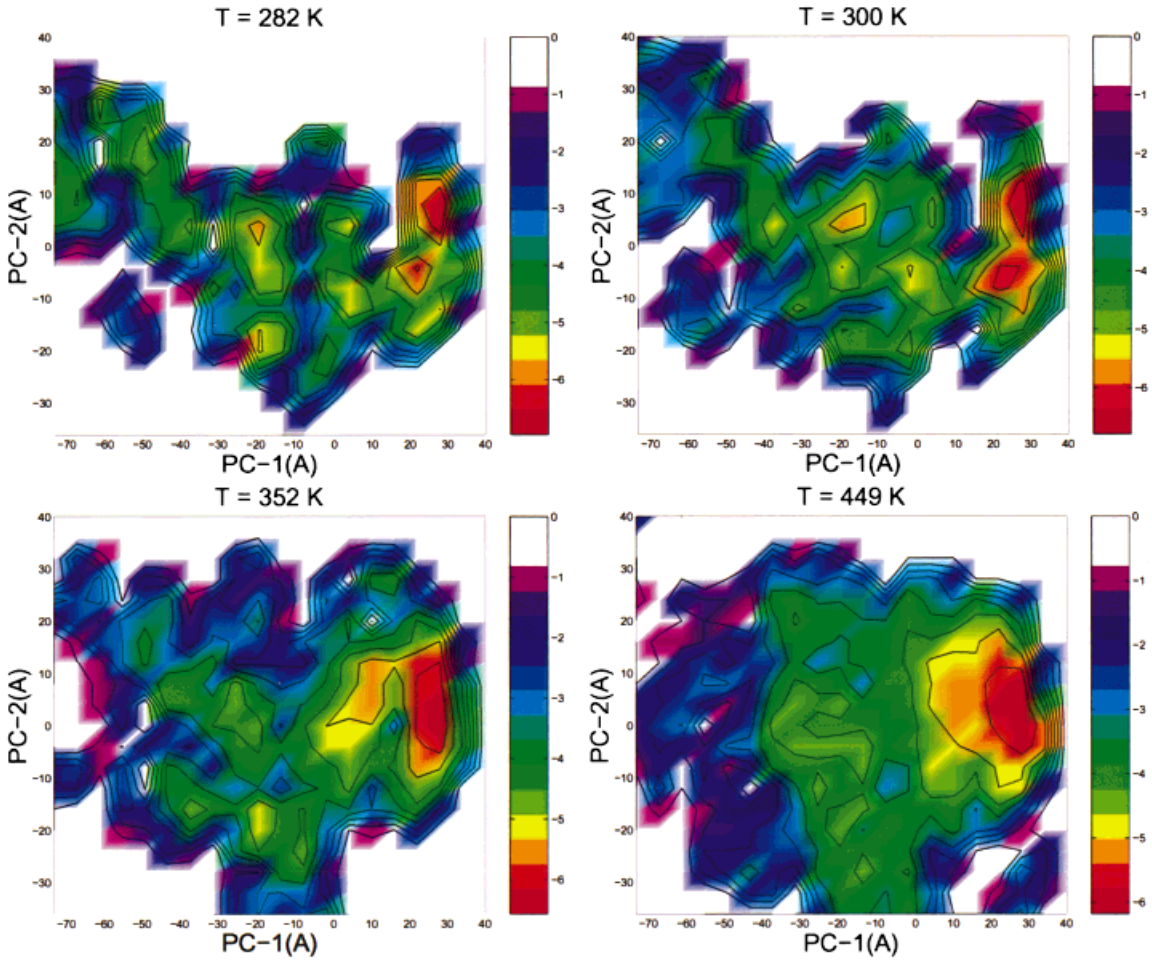

Fig. 2. Free energy surface maps as a function of the two PCs the system at temperatures T = 282, 300, 352, and 449 K.

ture. The presence of other secondary structural motifs is significant because it shows that differences in free energy among structures are small. The secondary structure can change upon changing of the local environment because of solvent conditions, binding to proteins, or by changing the context of the peptide within a larger protein.[2,36,37]

Previous calculations on this peptide have not observed the alternative helical structures we observe. This suggests that the enhanced sampling at low temperatures accomplished by the replica-exchange method enable us to find these states. Other possible reasons are the use of different force fields, explicit versus implicit[15] solvent models, and the simulation of different thermodynamic states, because Pande and Rokhsar[10] used a low solvent density ($\rho = 0.87$ g/cm^3) and high T, versus $\rho \sim 1.0$ g/cm^3 in our simulations.

In the replica-exchange method, interbasin jumps are correlated with large displacements in the temperature coordinate of the replica. This is illustrated in Figure 3, showing the trajectory followed by one of the replicas during the simulation (replica 2) projected on the two PCs of the system. This replica spans the whole T range (270–525 K) during the 3.5-ns simulation, and reversibly samples the basins containing the "folded" β-hairpin state, and the "molten globule" state. At long times, it samples states containing a "helical" state. Transitions between basins occurs fast, with long (0.5 ns) waiting periods between transitions. The combined trajectories in T and configurational space allow a replica to overcome a free energy barrier present at one temperature (e.g., $PC_1 = 30$ Å at $T = 270$ K) by increasing T, moving along the PC at higher T when there is no free energy barrier, and then lowering T after it reaches other local minima (e.g., $PC_1 = -20$ Å). The variations in temperature in the replica-exchange algorithm resembles the simulated annealing method.[38] However, instead of following a predetermined heating and cooling schedule, this process is self-regulated by the replica simulated tempering method used in our calculations. The replica exchange method ensures a Boltzmann sampling at each temperature. The free energy surface, as a function of PC_1 and T, explored by the replicas in the hybrid MD/MC approach is shown in Figure 4.

Figure 5A shows free energy profiles as a function of the first principal component (PC_1). Five dominating free

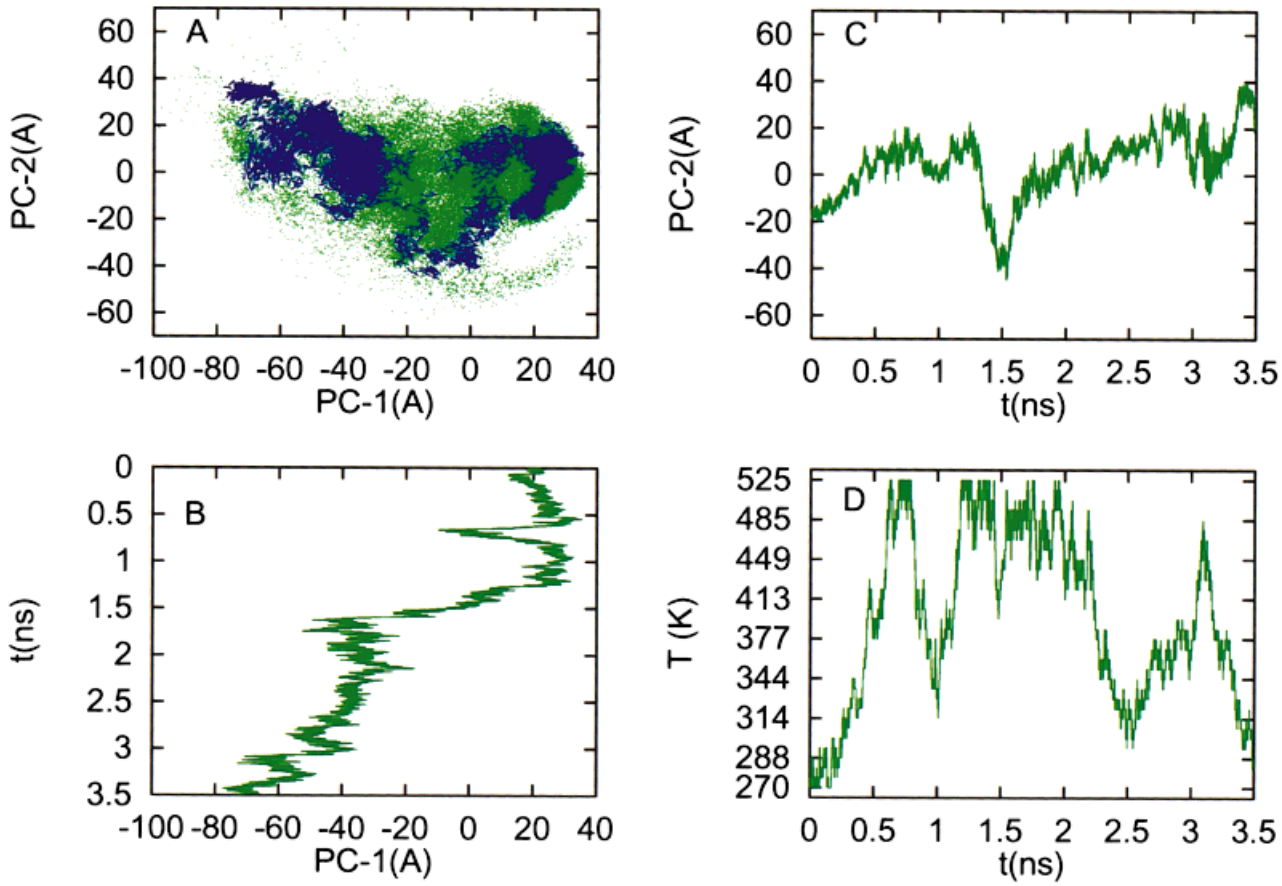

Fig. 3. Characterization of the trajectory followed by one of the replicas during the simulation (replica 2). **A:** Projection of the system trajectory on the PC-space. The green background illustrates the region of space sampled by all replicas. **B:** Projection of replica trajectory on the first PC, and **C:** on the second PC that spans the space shown on (A). **D:** The trajectory of the replica in temperature space.

energy minima can be observed at $T = 270$ K. These energy profiles show features characteristic of a rugged energy landscape at low T ($T < 330$ K), and a downhill collapse into a compact, "molten globule" state with structures close to the β-hairpin structure at higher T. This PMF is similar to the profile obtained by Nymeyer et al.[39] for a minimalist model of a frustrated β sheet protein that showed a glass transition at $T \sim T_{folding}$. For proteins in aqueous solution, the glass transition is strongly coupled to the solvent viscosity.[24] At higher T, the large barrier separating the β turn from the other ensembles is drastically reduced. The drastic change in this barrier height with T and the complete disappearance within 30 K suggest a cooperative change in the system. Given that the peptide conformations are easily accessed at higher T, we believe that this barrier may be caused by water interactions with the β hairpin (e.g., water-mediated hydrogen bonding). This barrier has not been observed with implicit solvent models. Figure 5B shows the percentage of the population on each of the five local minima as a function of T. Ensembles containing helices (basin I) and the helix-turn-helix motif (basin III) are 15–20% populated at 282 K, whereas the basing containing the "folded" β hairpin (basin V) is 30% populated. At 300 K, the population of the β hairpin increases to 40%, whereas the helix is reduced to 5%. The identification and characterization of minor secondary structure populations, as those shown here, by spectroscopy may be difficult.

Applying the diffusive theory of folding[23,40] on the energy profiles shown in Figure 5A, we can get an order of magnitude estimate of the mean folding time from $\tau_f = \int_{x_u}^{x_f} dx \int_0^x dx' \exp[\beta F(x) - \beta F(x')]/D(x)$, where $x = \mathrm{PC}_1$ is chosen as a reaction coordinate, $F(x)$ is the free energy as a function of the reaction coordinate, and $D(x)$ is the diffusion coefficient in configurational space. Following the applications of this theory by Socci et al.,[40] within the quasi-harmonic approximation for the free energy well around the folded state, the diffusion coefficient can be approximated by $D_{\mathrm{PC}_1} = \delta(\mathrm{PC}_1)^2/\tau_{corr}$, where $\delta(\mathrm{PC}_1)^2$ is the variance of the reaction coordinate about the folded state, and τ_{corr} is the reaction coordinate autocorrelation time.[40] $\langle\delta(\mathrm{PC}_1)^2\rangle$ and τ_{corr} are calculated from a simulation at constant (N, T, V), for 15 ns at $T = 300$ K. We obtained $\tau_{corr} \sim 1.5$ ns ($\pm$ 0.5 ns), and $\langle\delta(\mathrm{PC}_1)^2\rangle \sim 28$ Å^2 $\pm$ 1, resulting in $D_{\mathrm{PC}_1} \sim 19$ Å^2/ns (corresponding to a configurational *MSD* of 0.14 Å^2/ns). We obtained τ_f from $F(\mathrm{PC}_1)$ at $T = 300$ K, and from averaging $F(\mathrm{PC}_1)$ over three temperatures ($T = 294$, 300, and 307 K). The resulting

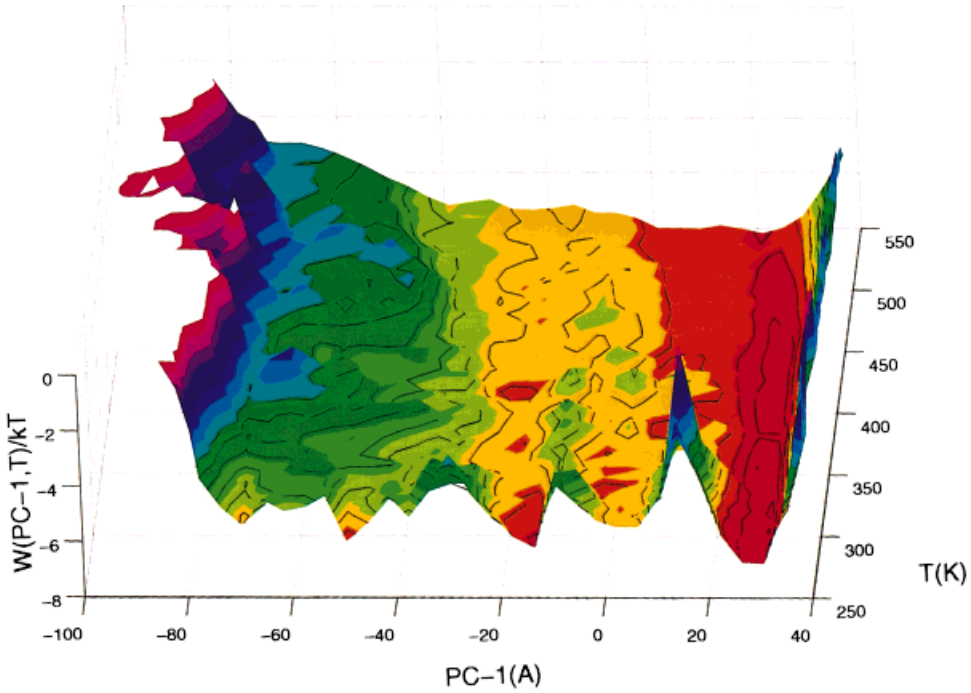

Fig. 4. Illustration of the free energy surface sampled by the hybrid MD/MC replica system as a function of the PC (PC_1) and temperature (T). At a constant low T the energy landscape is rugged, with high energy barriers separating local minima. However, the combined T and conformations space sampling of the replica exchange method allows the peptide to overcome kinetic traps by moving around energy barriers.

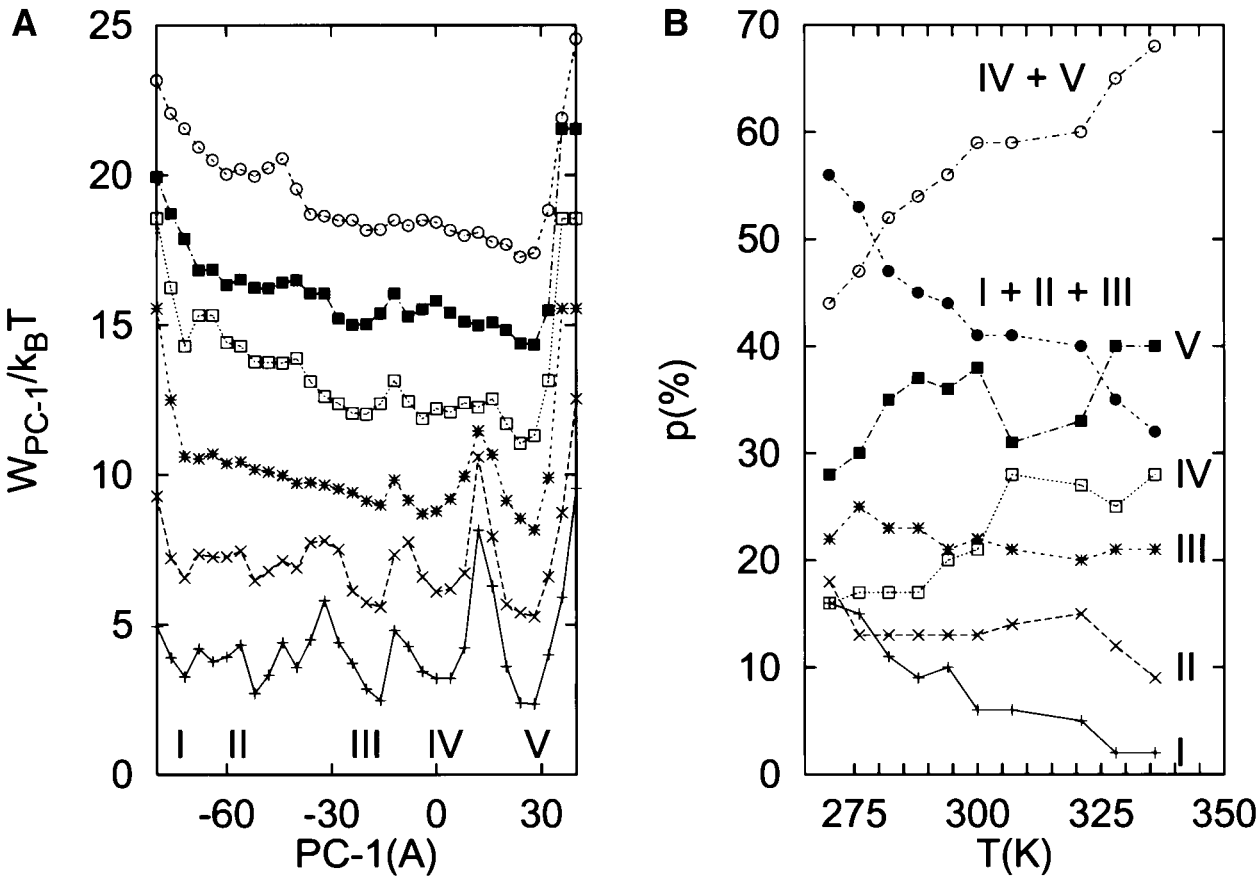

Fig. 5. **A:** Free energy profiles as a function of the PC, PC_1. Free energy profiles are calculated at $T = 270$, 282, 307, 336, 377, and 422 K. Curves are shifted up from each other by $3k_BT$. **B:** Percentage of the population of basins at: I) $PC_1 > -56$ Å, II) -56 Å $< PC_1 \leq -32$ Å, III) -32 Å $< PC_1 \leq -12$ Å, IV) -12 Å $< PC_1 \leq 12$ Å, V) $< PC_1 \leq 40$ Å. The basins I–V are identified from the minima and barriers in the free energy profiles shown in A) for $T = 270$ K.

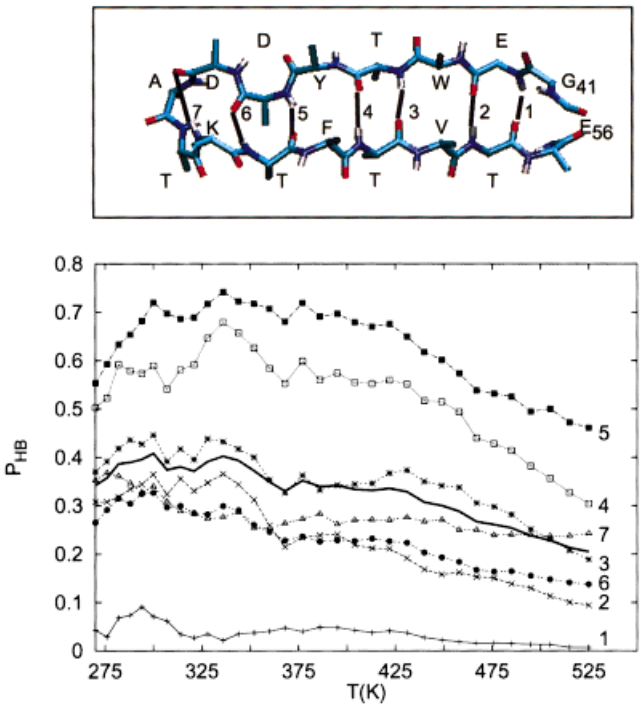

Fig. 6. Probability of forming individual native hydrogen bonds as a function of *T*. The thick solid line shows the average probability over all native hydrogen bonds, shown in the top of the figure. At $T = 282$ K, we obtain an average native hydrogen bond formation probability of 40%.

folding times at 300 K are 1–7 μs (the lower bound is obtained when averaging $F(\mathrm{PC}_1)$ over three *T*). This τ_f is 100–1,000 times longer than our simulation time (3.5 ns/replica). Muñoz et al.[16] estimated the folding time to be 6 μs, which falls within our estimates. However, our calculations of τ_f only provide an order of magnitude estimate because we have not determined the adequacy of the quasi-harmonic approximation or of the PC_1 as a reaction coordinate.

Figure 6 shows the probability of formation of each of the seven native β-hairpin hydrogen bonds, and their overall formation probability average, as a function of temperature in the range $270 \leq T \leq 525$ K. We find that hydrogen bonds 3–6, flanking the hydrophobic cluster (W43, Y45, F52, and V54) are the most probable, with overall probabilities above average for native hydrogen bonds. The lowest probabilities are observed for hydrogen bonds 1 and 2, at the far end from the turn, and hydrogen bond 6 and 7 near the turn, in agreement with previous calculations.[28] The hydrogen bond nearest to the turn (HB = 7) shows a probability of formation that decreases with temperature for $T < 350$ K, and remains constant at higher temperatures. The decrease at lower temperatures may be attributed to participation of this hydrogen bond in $i, i + 3$ hydrogen bonds in helical structure ensembles that are significantly populated at the lowest temperatures. This analysis of the hydrogen bond probabilities suggests that the central hydrogen bonds are stabilized and might be driven to form by the presence of the hydrophobic core. Our results agree with those obtained by Dinner et al.[15] in that the central hydrogen bonds (4–5) are the most stable, but contrast in that they are observed with above average probabilities for hydrogen bonds 1–5, and below average for 6 and 7 (for $T < 360$ K). The average hydrogen bond probability at 282 K is 40%, in agreement with NMR estimates[29] and previous simulations using an implicit solvent model.[15] Muñoz et al.[16] and Klimov and Thirumalai[27] have built models in which the hydrogen bond near the turn are the most stable. Also, their models describe a zippering of hydrogen bonds starting from the turn side and advancing toward the opposite end. Analysis of the replica trajectories do not show evidence of this in our model.

To compare our simulation results with those previously reported for this hairpin, we now calculate the free energy surfaces as a function of a pair of order parameters, the hydrophobic cluster radius of gyration (Rg_{hp}) and the number of native hydrogen bonds (N^{β}_{HB}), $W(Rg_{hp}, N^{\beta}_{HB})$.[10] Figure 7 shows contour plots of $\beta W(Rg_{hp}, N^{\beta}_{HB})$ at various temperatures. At 282 K, we observe three minima in the energy surface, corresponding to the folded state ($Rg_{hp} \sim 6$ Å, $N^{\beta}_{HB} \sim 4 - 6$), a compact state ($Rg_{hp} \sim 5$ Å, $N^{\beta}_{HB} \sim 1 - 2$), and an extended state ($Rg_{hp} \sim 6 - 9$ Å, $N^{\beta}_{HB} \sim 9$), in agreement with the observations of Pande and Rokhsar.[10] At $T \sim 330$ K, only one minimum containing compact structure with $Rg_{hp} \sim 5 - 6$ Å and 2–6 native hydrogen bonds is observed, in addition to the extended structures. At high temperatures, only one minimum is observed.

To simplify the comparison of these energy surfaces with our results using PC as an order parameter, we calculate free energy surfaces at 282 K as a function of PC_1 paired with N^{β}_{HB}, Cα *rmsd*, *Rg*, and Rg_{hp}, shown in Figure 8. By combining these four plots, we can now identify various minima in the free energy with the structures shown in Figure 1. First notice the linear correlation between the Cα *rmsd* and PC_1 for *rmsd* > 2.5 Å. Second, notice the clear separation of configurational space provided by $\beta W(\mathrm{PC}_1, N^{\beta}_{HB})$. We can clearly identify five local minima described next. 1. The minimum at $\mathrm{PC}_1 = 20$ Å corresponds to the folded structure ensemble containing structures that are within 2.5 Å from the folded reference state, $Rg \sim 7$ Å, and $Rg_{hp} \sim 7$ Å, with 4–6 native hydrogen bonds (shown in Fig. 1A). 2. The minimum at $\mathrm{PC}_1 = 0$ Å with an *rmsd* of 2–3 Å, larger overall *Rg* than the "folded" state, but with a more compact hydrophobic cluster and 2–4 native hydrogen bonds (shown in Fig. 1B). This minimum coalesces with the minimum at $\mathrm{PC}_1 = 20$ Å and together form the dominating ensemble at high *T*, which correspond to a "molten globule" state (state *H* in Ref. 10). The minima at $\mathrm{PC}_1 = 30$ Å can be separated into two minima: 3. containing structures with four native hydrogen bonds and a Rg_{hp} similar to the folded state (shown in Fig. 1C), and 4. a minimum containing structures with 0–1 native hydrogen bonds, a compact *Rg* of all heavy atoms (shown in Fig. 1F). 5. The minima containing ensemble of structures with *rmsd* > 6 Å, no hydrophobic core, a large *Rg* and 0–1 native hydrogen bonds. This ensemble contains helical

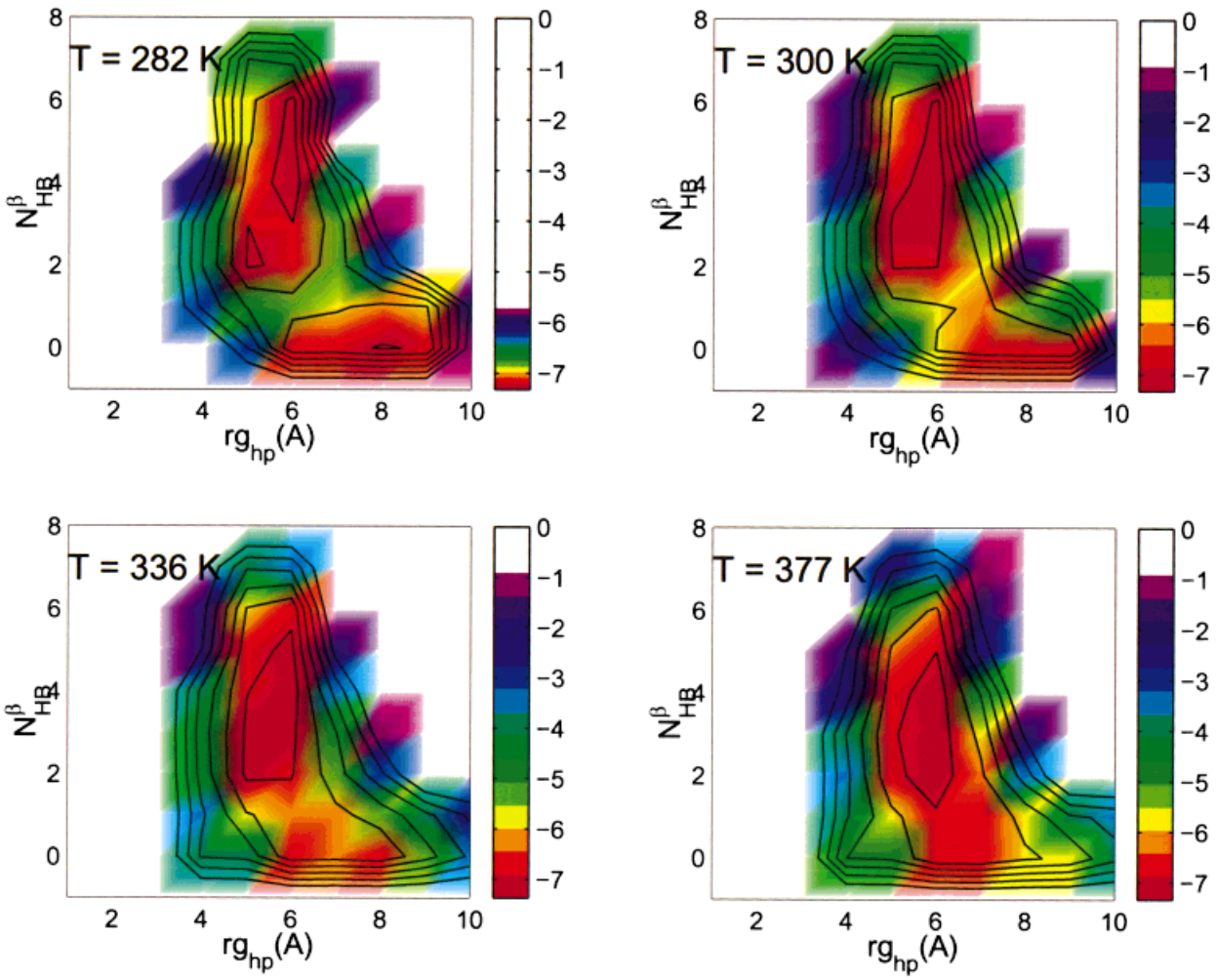

Fig. 7. PMF energy surfaces as a function of the hydrophobic cluster radius of gyration, Rg_{hp}, and the number of hydrogen bonds in common with the β hairpin at various temperatures. In agreement with the F, H, and U states described by Pande and Rokhsar,[10] we observe three minima, one with $R_g \sim$ 6 Å and with 4–6 native hydrogen bonds (state F), one with a smaller R_g but only 2–4 native hydrogen bonds (state H), and another with $R_g \sim$ 8–9 Å and 0–1 hydrogen bonds (state U).

structures (shown in Fig. 1D) that contribute significantly to this ensemble population at low T. The ensemble of structures with a large hydrophobic cluster Rg_{hp} and 0–1 hydrogen bonds present at low T, shown in Figure 7, corresponds to helical structures.

CONCLUSIONS

We have studied the free energy of a peptide that predominantly forms a β hairpin at low T by using a replica exchange method. This method allows for the efficient sampling of configurational space in a time scale that is of the order of 100–1,000 times faster than sampling at constant T. This method can be easily implemented in parallel computers, giving a 95% parallel efficiency. Our results are in agreement with other all-atom MD simulations of protein unfolding,[10] but, in addition, our system samples a much broader region of configurational space. The replica exchange method leads to thermal equilibrium of the set of replicas, where T-jumping and mixing of replicas give proper weighting of configurations. Interbasin jumps are correlated with large displacements in the temperature coordinate, thus enhancing the probability of overcoming energy barriers. Different replicas evolve in heating and cooling cycles that lead to "low T" secondary structure motifs, far from the β hairpin that is the dominating structure at low T. Various ensembles with different characteristic secondary structures coexist in thermal equilibrium. Previous calculations on this peptide have not produced the low energy helical structures that we observed.[10,15,28] Schaefer et al.[14] have found significant α helical populations in another sequence that preferentially forms a β hairpin. It has been argued that the Cornell et al. force field enhances the stability of α helices.[41] However, the calculations used to draw such conclusions were done in the absence of water and cannot be considered to be conclusive. In our calculations, the equilibrium population of the β hairpin at low T agrees with experimental data available[29] and with previous

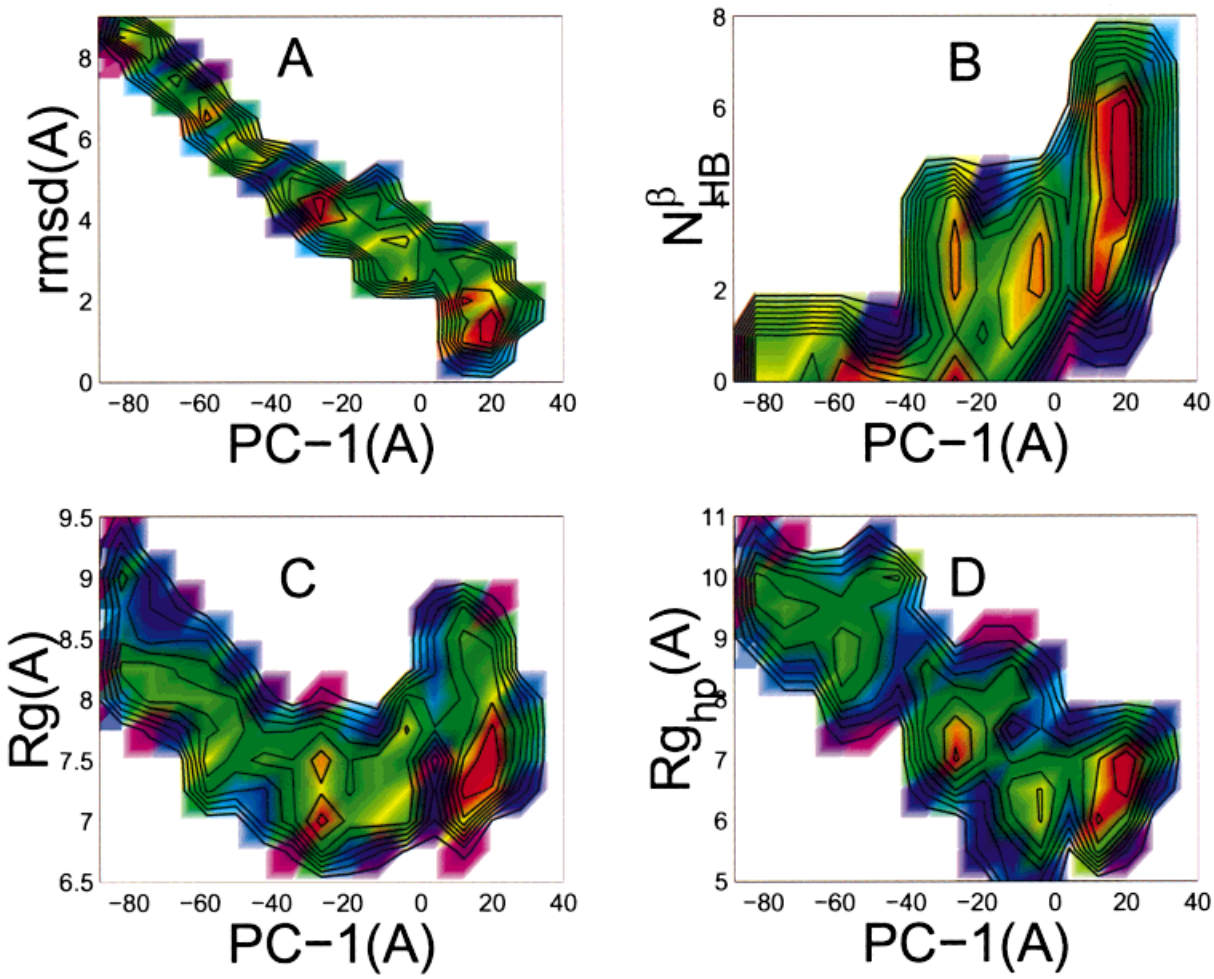

Fig. 8. PMF energy surfaces, at $T = 282$ K, as a function of a pair of reaction coordinates including the first PC of the system (PC_1) combined with the (**a**) the Cα *rmsd* from the folded structure, (**b**) the number of native hydrogen bonds, N^{β}_{HB}, (**c**) the overall radius of gyration, R_g, and (**d**) the hydrophobic cluster radius of gyration, Rg_{hp}.

calculations.[15,28] Therefore, we conclude that the potential energy functions used in this study seem appropriate to describe the equilibrium properties of this hairpin. The efficiency of sampling configurational space and the highly parallel nature of the replica-exchange method makes possible the study of larger proteins in explicit solvent.

This peptide shows a rugged free energy landscape landscape at low T. This energy landscape shows multiple local energy minima, separated by large energy barriers, similar to the energy landscape of a model protein that exhibited a glass transition at $T \sim T_{\text{folding}}$.[39] At higher T, the energy landscape is smooth. Our results suggest that, at $T \sim 300$ K, folding occurs by forming a hydrophobic cluster. Many structures with various degrees of nativeness form the hydrophobic cluster, but not the native hydrogen bonds. At lower T, the number of native hydrogen bonds increases, but other secondary structural elements are also present. We see that by lowering T, we get a lower propensity of forming the "folded" β hairpin structure, as indicated by the native hydrogen bond probabilities as a function of T. The coexistence of multiple local minima with different secondary structures shows that solvent environment and the "context" of peptides within proteins may change secondary structure. This has been seen in Arc repressor,[36] proteins that fold upon binding,[2] and the conversion of a helical protein to β sheet protein by Dalal et al.[37]

ACKNOWLEDGMENTS

We thank B. McMahon, J.N. Onuchic, and Y. Okamoto for helpful discussions. Computer access to the Los Alamos ASCI Nirvana supercomputer through an IHPC grant is gratefully acknowledged.

REFERENCES

1. Burton RE, Huang GS, Daugherty MA, Calderone TL, Oas TG. Nat Struct Biol 1997;4:305.
2. Wright P, Dyson H. J Mol Biol 1999;293:321.
3. Prusiner S. Science 1997;278:245.
4. Dobson C. Trends Biochem Sci 1999;24:329.
5. Rochet J, Lansbury P. Curr Opin Struct Biol 2000;10:60.
6. Dill KA, Chan HS. Nat Struct Biol 1997;4:10.

7. Onuchic JN, Nymeyer H, García A, Chahine J, Socci N. Adv Protein Chem 2000;53:88.
8. Wolynes PG, Onuchic JN, Thirumalai D. Science 1995;267(5204): 1619.
9. Daggett V, Levitt M. Curr Opin Struct Biol 1994;4:291.
10. Pande V, Rokhsar D. Proc Natl Acad Sci USA 1999;96:9062.
11. Brooks CL. Curr Opin Struct Biol 1998;8:222.
12. Dyer RB, Gai F, Woodruuf WH, Gilmanshin R, Callender RH. Acc Chem Res 1998;31:709.
13. Eaton WA, Munoz V, Thompson P, Chan CK, Hofrichter J. Curr Opin Struct Biol 1997;7:10.
14. Schaefer M, Bartels C, Karplus M. J Mol Biol 1998;284:835.
15. Dinner A, Lazaridis T, Karplus M. Proc Natl Acad Sci USA 1999;96:9068.
16. Munoz V, Thompson P, Hofrichter J, Eaton W. Nature 1997;390:196.
17. Serrano L. Adv Protein Chem 2000;53:50.
18. Berg B, Neuhaus T. Phys Lett B 1991;267:249.
19. Hansmann L, Okamoto Y. Curr Opin Struct Biol 1999;9:177.
20. Bartels C, Karplus M. J Phys Chem B 1998;102:865.
21. Hukushima K, Nemoto K. Exchange Monte Carlo method and application to spin glass simulations. J Phys Soc Japan 1996;65: 1604–1608.
22. Sugita Y, Okamoto Y. Chem Phys Lett 1999;314:141.
23. Bryngelson JD, Wolynes PG. J Phys Chem 1989;93(19):6902.
24. Frauenfelder H, Sligar SG, Wolynes PG. Science 1991;254:1598.
25. Leopold PE, Montal M, Onuchic JN. Proc Natl Acad Sci USA 1992;89(18):8721.
26. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Annu Rev Phys Chem 1997;48:545.
27. Klimov D, Thirumalai D. Proc Natl Acad Sci USA 2000;97:2544.
28. Roccatano D, Amadei A, Di Nola A, Berendsen H. Protein Sci 1999;8:2130.
29. Blanco F, Rivas G, Serrano L. Nat Struct Biol 1994;1:584.
30. Duan Y, Kollman PA. Science 1998;282:740.
31. Cornell WD, Cieplak P, Bayley CI, et al. J Am Chem Soc 1995;117:5179.
32. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. J Chem Phys 1984;81(8):3684.
33. Gronenborn A, Filpula D, Essig N, et al. Science 1991;253:657.
34. García AE. Phys Rev Lett 1992;68:2696.
35. García AE, Hummer G, Blumfield R, Krumhansl JA. Physica D 1997;107:225.
36. Cordes M, Walsh N, McKnight C, Sauer R. Science 1999;284:325.
37. Dalal S, Balasubramanian S, Regan L. Nat Struct Biol 1997;4:548.
38. Kirkpatrick S, Gelatt C, Vecchi M. Science 1983;220:671.
39. Nymeyer H, García AE, Onuchic JN. Proc Natl Acad Sci USA 1998;95:5921.
40. Socci ND, Onuchic JN, Wolynes PG. J Chem Phys 1996;104(15): 5860.
41. Beachy M, Chasman D, Murphy R, Halgren T, Friesner R. J Am Chem Soc 1997;119:5908.
42. Frishman D, Argos P. Proteins 1995;23:566.
43. Humphrey W, Dalke A, Schulten K. J Mol Graph 1996;14:33.