

Understanding β -hairpin formation

(CHARMM/implicit solvent/multicanonical Monte Carlo/protein folding)

AARON R. DINNER*[†], THEMIS LAZARIDIS*[‡], AND MARTIN KARPLUS*^{†§¶}

*Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138; [†]Committee on Higher Degrees in Biophysics, Harvard University Cambridge, MA 02138; [‡]Department of Chemistry, City College of New York, Convent Avenue at 138th Street, New York, NY 10031; and [§]Laboratoire de Chimie Biophysique Institut le Bel, Université Louis Pasteur, 4 Rue Blaise Pascal, 67000 Strasbourg, France

Contributed by Martin Karplus, May 28, 1999

ABSTRACT The kinetics of formation of protein structural motifs (e.g., α -helices and β -hairpins) can provide information about the early events in protein folding. A recent study has used fluorescence measurements to monitor the folding thermodynamics and kinetics of a 16-residue β -hairpin. In the present paper, we obtain the free energy surface and conformations involved in the folding of an atomistic model for the β -hairpin from multicanonical Monte Carlo simulations. The results suggest that folding proceeds by a collapse that is downhill in free energy, followed by rearrangement to form a structure with part of the hydrophobic cluster; the hairpin hydrogen bonds propagate outwards in both directions from the partial cluster. Such a folding mechanism differs from the published interpretation of the experimental results, which is based on a helix-coil-type phenomenological model.

The general mechanism by which proteins find their unique native states rapidly (1 s or less) among the astronomically large number of configurations accessible to the polypeptide chain is now thought to be understood (1–4). The essential concept is that only a small fraction of the configurations is visited by any given trajectory because of an overall bias in the (free) energy toward the native state. Although this mechanism is reasonable, there is little direct experimental evidence for its validity. Moreover, even if it is essentially correct, it is important to determine how possible folding scenarios are realized by specific proteins. Such information has been difficult to obtain, though much has been learned recently from a wide range of experiments, among which NMR (5) and protein engineering (6) have played an essential role.

To obtain a detailed understanding of the elementary steps that can contribute to the folding process, it is useful to study peptide fragments that have well defined transitions to stable structures in solution. Their small size relative to proteins makes them more amenable to detailed experimental analyses (7–9) and computational treatments with realistic models (10). The information gained from such studies is expected to be of direct relevance to the folding of proteins, because stable sequences that are fragments of known proteins are likely to correspond to the elements that appear early in the folding reaction (11).

Although the helix-coil transition has been studied for many years (7), it was not until recently that a β -hairpin that is stable in aqueous solution (on average 42% of the native backbone hydrogen bonds are formed at 278 K) was identified (12). It is a 16-residue peptide that corresponds to the second hairpin (residues 41 to 56) of the Ig-binding domain of streptococcal protein G (13). Because there is a single tryptophan that is partly buried in a hydrophobic cluster in the native structure,

fluorescence can be used to probe the folding reaction. On the basis of such measurements, Muñoz *et al.* (14) estimated the (assumed temperature-independent) thermodynamic parameters for the folding transition to be $\Delta H = -11.6$ kcal/mol and $\Delta S = -39$ cal/mol/K, which yield a β -hairpin population close to 80% at 278 K. The time course of the relaxation of the fluorescence after temperature jumps was found to decrease with simple exponential kinetics. A 6- μ s folding time was estimated at the midpoint of the equilibrium thermal denaturation transition (297 K); this figure is about 30 times slower than α -helix formation in systems of similar size (15, 16). To supplement the fluorescence experiments, which yield information only about the solvent exposure of the tryptophan residue and the change in its distance from a C-terminal dansyl group; Muñoz *et al.* (14, 17) developed a helix-coil-type model to provide a structural interpretation for their equilibrium and kinetic data. The model yields a mechanism in which folding initiates at the turn and propagates toward the tails, so that the hydrophobic cluster (from which most of the stabilization derives) forms relatively late during the reaction.

As Muñoz *et al.* (14, 17) point out, alternative mechanisms (such as one in which the hydrophobic cluster forms before the β -turn) are possible. It is important, therefore, to obtain additional information concerning the atomic details of the folding mechanism. In the present study, we use a polar-hydrogen representation for the peptide (18) with an implicit solvent model (19) to examine the equilibrium folding behavior of the β -hairpin. The accessible configuration space is sampled with multicanonical Monte Carlo (MC) simulations (20). At 300 K, native-like structures have a probability of 75%, and on average 51% of the native backbone hydrogen bonds are formed, in reasonable agreement with the experimental estimates (12, 14). The analysis suggests that there is a rapid collapse followed by formation of part of the hydrophobic assembly, from which point the hairpin hydrogen bonds propagate in both directions.

METHODS

The peptide (GEWTYDDATKFTVTE without terminal blocking groups) is modeled with a modified form of the CHARMM 19 polar-hydrogen topology and parameters (18, 21), and the solvent is treated implicitly by the addition of a term to the energy function; thus, the calculated (effective) energy of each configuration corresponds to a free energy (or potential of mean force) for the peptide in the presence of equilibrated solvent. The solvation free energy term is based on a Gaussian solvent exclusion model, the details of which are given in ref. 19. The implicit solvent model has been shown to give good results in a variety of applications (19, 22, 23).

Because the solution structure of the peptide has not been determined (12), we constructed a “native” structure from the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: MC, Monte Carlo; ASA, accessible surface area.
[¶]To whom reprint requests should be addressed. E-mail: marci@tammy.harvard.edu.

corresponding protein coordinates (residues 41 to 56 of Protein Data Bank code 2gb1) (13). Polar hydrogens were added, and the system was minimized for 300 steps with the adopted-basis Newton-Raphson method. The overall heavy atom rms deviation of the resulting structure (Fig. 1*a*) from that in the protein is 0.90 Å; the main chain deviation is 0.54 Å. The effective energy of the minimized structure is -516.24 kcal/mol, of which -217.42 kcal/mol comes from the solvent term, -87.11 kcal/mol comes from the van der Waals term, and -259.32 kcal/mol comes from the electrostatic term (the bond length, bond angle, dihedral, and improper dihedral terms are 3.13, 23.67, 17.70, and 3.11 kcal/mol, respectively; they are similar for folded and unfolded conformations at a given temperature). The solvent term, which is a large part of the nonbonded energy even in the native state, opposes the van der Waals and electrostatic terms; it favors more open structures, which typically have solvent contributions of about -245 kcal/mol.

The simulations sample the distribution of accessible states of the model in the multicanonical ensemble, in which the probability of observing a single conformation with effective energy W is proportional to $1/\omega(W)$, where $\omega(W)$ is the number of states at that effective energy (20). Because one does not know $\omega(W)$ *a priori*, one proceeds iteratively and weights conformations during the MC simulations with probability $P_{MC}(W) \propto 1/\omega_k(W)$, where $\omega_k(W)$ is an estimate for $\omega(W)$ from the k th set of MC trials. Once the weights have converged, the total probability of populating conformations with effective energy W is equal to that of populating conformations at any other W because the product $\omega(W)P_{MC}(W)$ is a constant, so that the system makes an unbiased random walk in energy. Canonical averages of observables can be calculated from the multicanonical data and the known density of states (20).

The calculations were performed with the MC module (24) in the program CHARMM (Version c27a1) (18). The MC move set consisted of single atom displacements, rotations of all freely rotatable dihedral angles, and concerted rotation of seven (or in the case of a chain end, six) sequential main chain

dihedral angles that leave the positions of atoms outside the selected window unchanged (24–26). An initial estimate for the density of states was obtained from a series of Metropolis MC simulations at 300, 400, 500, and 600 K that started in the native structure. The estimate was then refined iteratively in a series of multicanonical simulations, the number and length of which varied with the iteration. In the final set of simulations with converged weights, 30 independent trials were performed. In each such trial, the system began in the native structure, was simulated with standard Metropolis weighting for 200×10^3 MC steps to loosen the structure (this tended to accelerate equilibration), and was then equilibrated with the multicanonical weighting for 3×10^6 MC steps. After the equilibration, the configuration was recorded every 1,000 MC steps for 7×10^6 MC steps. Details of the simulations and the procedure used to refine the density of states will be given elsewhere (A.R.D., T.L., and M.K., unpublished work).

The structures obtained by taking every tenth recorded structure in the simulations were clustered with a neural-network-based method in CHARMM (18, 27). The method partitions the data so as to minimize the deviation of a set of conformational properties for each structure from the average values in the corresponding cluster. The structures were characterized by nine pairwise distances (Fig. 1*b*). The cluster “radius” (the limit of the rms deviation of the nine distances for each member from their average values for the cluster) was 1 Å.

RESULTS AND DISCUSSION

Structures. Application of the cluster analysis to the 21,000 structures yielded 1,005 clusters. Because the nine pairwise distances monitor only the structure of the middle 12 residues (W43 to V54), this result suggests that there are about $1,005^{1/12} = 1.78$ accessible main chain configurations for each amino acid, consistent with extrapolations from simulations of β -amino acid peptides (28). To determine the significant structures at room temperature, the canonical probability of each cluster at 300 K and its average structure were determined. For the latter, each member was aligned such as to minimize its overall rms deviation from the running average structure. Because the averaging process yielded structures with unrealistic bond lengths and angles, they were relaxed with 500 steps of steepest descents minimization (18).

At 300 K, clusters with average structures that deviate from the minimized protein coordinates by less than 2.5 Å (overall rms) account for 75% of the population. The native-like clusters include the nine most probable ones, which account for 63% of the population ($1.61 \leq \text{rms} \leq 2.07$ Å). The most probable cluster, which contains 816 structures, has a probability of 10.7% (Fig. 2*a*). The average structure of this cluster has a closely packed hydrophobic assembly that is made up of Y45, F52, W43, and V54, like the minimized protein coordinates. Most of the remainder of the population (20.1% of the total at 300 K) consists of compact structures (radii of gyration $R_g < 8.5$ Å) that deviate from the native structure by more than 3.0 Å (overall rms). Several representative structures are shown in Fig. 2. Three of them (Fig. 2*b–d*) are hairpins in which the turn is shifted toward the N terminus (Fig. 2*b* and *c*) and/or the strands are reflected (with the turn on the left, W43 and Y45 are closer to the viewer than F52 and V54 in Fig. 2*c* and *d*, in contrast to Fig. 2*a*). In the structures in which the turn is shifted in sequence, the aliphatic part of the side chain of K50 packs against the hydrophobic assembly. It is important to note that, although these structures are β -hairpins, they have only nonnative hydrogen bonds. The structure shown in Fig. 2*e* is more globular, with W43 and Y45 packed perpendicularly to F52 between the strands. The aliphatic part of the K50 side chain packs against the tyrosine ring and its NH_3 group participates in a network of hydrogen bonds involving the

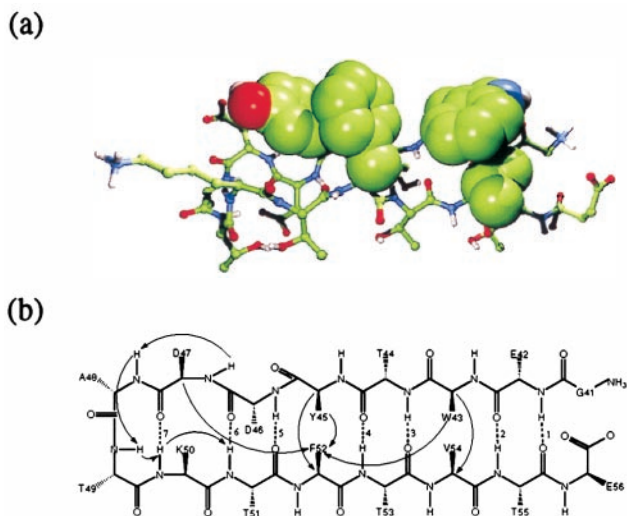


FIG. 1. Native hairpin structure. (a) Space-filling model of the hydrophobic cluster for the minimized coordinates extracted from the NMR structure of the complete Ig-binding domain of protein G (Protein Data Bank code 2gb1) (13). (b) Schematic. Backbone hydrogen bonds contributing to the sheet are indicated by dashed lines and numbered. In general, we count a hydrogen bond if the corresponding heavy atoms are within 3.4 Å of each other and the out-of-line angle ($180^\circ - \chi_{\text{DHA}}$) is less than 70° . Arrows indicate the pairwise distances used in the cluster analysis: D47H–A48H, A48H–T49H, T49H–K50H, K50H–T51H, W43C α –V54C α , Y45C α –F52C α , W43C ϵ_3 –F52C β , Y45C β –F52C γ , and D47C α –F52C γ .

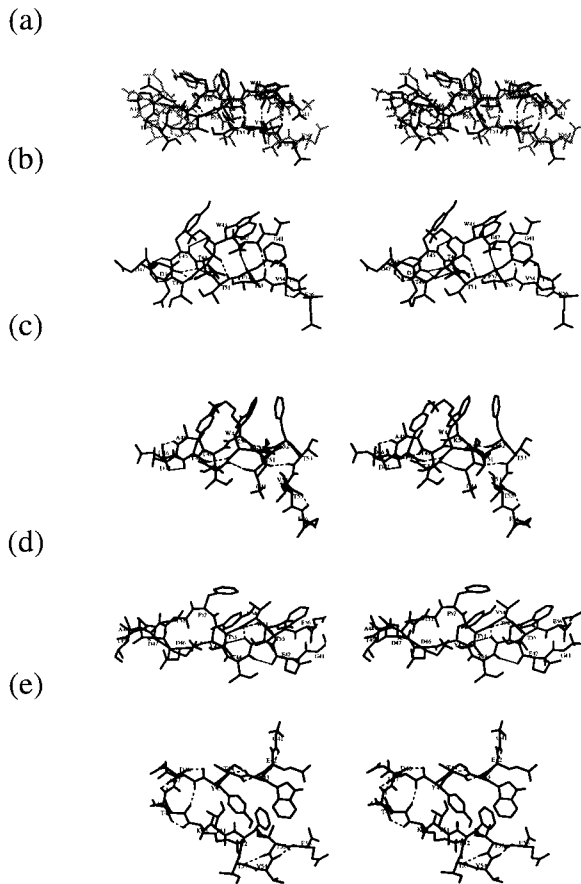


FIG. 2. Stereo views of minimized average structures of high probability clusters at 300 K. (a) The most probable cluster (black) (10.7%) compared with the structure shown in Fig. 1 a (gray). (b–e) Clusters with overall rms deviations from the native structure of more than 3.0 Å.

hydroxyl groups of Y45 and T55 and the carbonyl groups of F52 and T53. These electrostatic interactions make such structures comparable in effective energy to the lowest native-like clusters.

Thermodynamic Parameters. At 300 K, the free energy of folding is estimated to be $\Delta G = -0.40$ kcal/mol, by using a cutoff of 2.5 Å (overall rms) for separating native from nonnative structures (in the complete set of 210,000 structures sampled); the corresponding effective energy of the transition is $\Delta W = -4.32$ kcal/mol, and the contribution to the free energy from the configurational entropy is $-T\Delta S = 3.92$ kcal/mol. The melting temperature of the model is somewhat higher than the estimate of Muñoz *et al.* (297 K) (14). The specific heat curve (not shown) exhibits two peaks: one at 335 K, which corresponds to the folding transition, and one at 395 K, which corresponds to a collapse transition. The latter is significantly above the experimental range (up to 328 K) and thus would not have been observed.

The fluorescence measurements do not provide details of the transition between the folded and unfolded states of the β -hairpin. Consequently, Muñoz and *et al.* (14, 17) introduced a simple model and used it to fit the data and obtain parameters for the interactions that they assumed to be important. In their model, a conformation is defined by the states of the 15 “peptide bonds,” each of which can be either native (denoted “h” for “hairpin”) or nonnative (denoted “c” for “coil”). The thermodynamic quantities that determine the free energy contribution of a conformation (binary sequence of peptide bond states) are: (i) the free energy of native hydrophobic side chain packing interactions (W43–F52, W43–V54, and Y45–

F52); (ii) the enthalpy of native backbone hydrogen bonds, and (iii) the entropy of constraining peptide bonds. The pairwise interactions are counted if, and only if, the two residues involved are connected by a continuous stretch of h peptide bonds; for example, in the conformation hhhchhhhhhhchhhh, the hydrogen bond between residues 47 and 50 would be counted (because it is within the five central h peptide bonds) but not the hydrogen bonds or side chain interactions between residues 41 and 45 and residues 52 to 56. The details of the model are not important because the thermodynamic parameters are determined primarily by the folded state (with all of the native interactions) and the unfolded state (with none).

To evaluate the contribution of the pairwise (hydrogen bond and side chain packing) interactions to the stability of the hairpin, we calculated average effective energies for subsets of atoms. In the case of the side chain interactions, we summed the van der Waals, electrostatic, and solvent terms over all possible pairs of atoms in which one was in the first side chain (including C_β) and the other was in the second side chain for each of the 210,000 configurations and then calculated the canonical average interaction as a function of C_β – C_β distance (binned to a resolution of 0.5 Å). The W43–F52 interaction exhibits two minima: one of -2.24 kcal/mol at 3.75 Å and one of -2.01 kcal/mol at 6.25 Å; however, the barrier dividing these minima is relatively low (0.27 kcal/mol relative to the latter), so that they can be considered part of the same broad basin. Consistent with this idea, all the high probability clusters with C_β – C_β distances corresponding to these two minima are native like (the native distance is 6.95 Å) and vary only with respect to whether F52 is packed against W43 or Y45 (Figs. 1 and 2a). The average effective energy of each of the other side chain pairs exhibits only a single minimum; that for W43–V54 is -1.51 kcal/mol at 5.25 Å (the native distance is 5.19 Å), and that for Y45–F52 is -2.23 kcal/mol at 5.75 Å (the native distance is 5.61 Å). The effective energy includes the solvation entropy, so that the calculated quantities can be compared directly with the value obtained with the simple model (in which all three pairs interact with the same free energy): $\Delta G_{sc} = -2.1$ kcal/mol (14), which is quite close.^{||}

In the case of the hydrogen bonds, we summed the non-bonded effective energy terms over all possible pairs of atoms in which both were in the backbone (HN– C_α –CO) and averaged as a function of the number of native hairpin hydrogen bonds (N_H) (Fig. 1b). Even though the simple model considers only the enthalpy of hydrogen bonds, the calculation includes the solvation free energy, which plays a significant role in determining the strength of these interactions. The average effective energy per hydrogen bond is $[W(7) - W(0)]/7 = (-4.16$ kcal/mol)/7 = -0.60 kcal/mol. This result is roughly half the enthalpy of a native hydrogen bond in the simple model: $\Delta H_{hb} = -1.1$ kcal/mol (14).

To determine the entropy of folding, we calculated the thermodynamics in terms of the number of native peptide bonds (N_ω). Evaluation of N_ω for a specific three-dimensional structure requires that one define what is “native.” We take a peptide bond to be native if its flanking ϕ and ψ dihedral angles are within $\pm 30^\circ$ of their values in the minimized average structure of the most probable cluster (Fig. 2a); this cutoff corresponds approximately to the range sampled by backbone dihedral angles during Metropolis MC simulations in the native basin at 300 K. The choice of cutoff and reference structure does not affect the results significantly. The average decrease in entropy per native peptide bond is $[S(N_\omega=15) -$

^{||}The values in ref. 17 are slightly different: for the “standard” single sequence approximation, $\Delta G_{sc} = -2.19$ kcal/mol, $\Delta H_{hb} = -0.86$ kcal/mol, and $\Delta S = -3.09$ cal/mol/K per peptide bond, and, for the “modified” single sequence approximation, $\Delta G_{sc} = -1.94$ kcal/mol, $\Delta H_{hb} = -0.96$ kcal/mol, and $\Delta S = -2.74$ cal/mol/K per peptide bond.

$S(N_{\omega}=0)]/15 = (-26.2 \text{ cal/mol/K})/15 = -1.75 \text{ cal/mol/K}$. This is about half that obtained with the simple model: $\Delta S = -3.2 \text{ cal/mol/K}$ (14). The simulations thus suggest that the simple model overestimates the strength of the hydrogen bonds and compensates for this error by increasing the entropic cost associated with constraining the backbone.

Folding Mechanism. To obtain insights into the folding mechanism, we project the calculated many-dimensional free energy surface onto one or two structural coordinates at a time. Fig. 3a shows the free energy at 300 K as a function of the overall rms

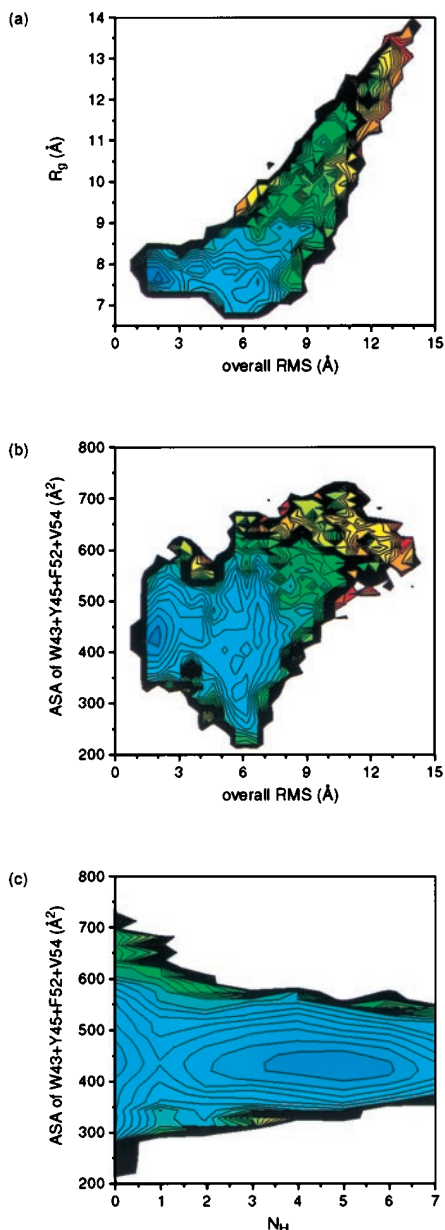


FIG. 3. Two-dimensional projections of the free energy. (a) The free energy as a function of the overall rms deviation from the native structure and the radius of gyration (R_g). (b) The free energy as a function of the overall rms deviation from the native structure and the ASA of the hydrophobic side chains (W43, Y45, F52, and V54); the ASA is measured with a 1.4-Å radius probe. (c) The free energy as a function of the number of native hairpin hydrogen bonds (N_H) and the ASA of the hydrophobic side chains; see Fig. 1 for hydrogen bond definitions. The contours are spaced at intervals of 0.5 kcal/mol and range from blue for low (favorable) values to red for high (unfavorable) values; contours more than 15 kcal/mol above the global free energy minimum are not drawn.

deviation from the native structure and the radius of gyration. Overall, the surface slopes steadily downward from open structures ($R_g \geq 10 \text{ Å}$) to compact ones ($R_g \leq 8 \text{ Å}$). The global minimum is at $(\text{rms}, R_g) = (2.0 \text{ Å}, 7.75 \text{ Å})$, consistent with the large native-like population observed in the cluster analysis. There are significant local minima at $(4.0 \text{ Å}, 7.75 \text{ Å})$, $(6.5 \text{ Å}, 7.50 \text{ Å})$, and $(9.5 \text{ Å}, 10.75 \text{ Å})$; these show that the free energy surface has some “roughness.” The three compact minima at $R_g \leq 7.75 \text{ Å}$ are separated from the open one at $(9.5 \text{ Å}, 10.75 \text{ Å})$ by a barrier of about 2 kcal/mol at $R_g \approx 9.6$ (Fig. 3a). Decomposition of the free energy as a function of R_g (not shown) reveals that this barrier derives from a failure of the configurational entropy to compensate completely for peaks in the bonded energy.

The low free energy region of Fig. 3a ($R_g < 8.5$) can be expanded by plotting on the vertical axis the accessible surface area (ASA) of the hydrophobic assembly (W43, Y45, F52, and V54) (Fig. 3b); we use the ASA as a variable because it is close to what is measured by the fluorescence quenching experiments. The global minimum is at $(\text{rms}, \text{ASA}) = (2.25 \text{ Å}, 430 \text{ Å}^2)$, which is consistent with the ASA of the minimized coordinates (412.83 Å^2). There are significant local minima at $(3.75 \text{ Å}, 450 \text{ Å}^2)$, $(6 \text{ Å}, 330 \text{ Å}^2)$, and $(6.75 \text{ Å}, 420 \text{ Å}^2)$. The minimum at $(3.75 \text{ Å}, 450 \text{ Å}^2)$ contains clusters similar to (and including) the structure shown in Fig. 2b, that at $(6 \text{ Å}, 330 \text{ Å}^2)$ contains clusters similar to (but typically more compact than) the structure shown in Fig. 2e, and that at $(6.75 \text{ Å}, 420 \text{ Å}^2)$ contains clusters similar to the structure shown in Fig. 2e. The structure shown in Fig. 2d maps to the slightly higher local free energy minimum at $(5.5 \text{ Å}, 490 \text{ Å}^2)$.

To determine the factors that stabilize these minima, we examine the components of the free energy in Fig. 4a. The effective energy and entropy contributions span a range of about 80 kcal/mol, but there is significant compensation (as in proteins), so that the free energy difference between the native state and extended conformations is only about 16 kcal/mol. The misfolded conformations at $\text{rms} \approx 6 \text{ Å}$ are separated from the native basin ($\text{rms} \approx 2 \text{ Å}$) by a barrier of about 0.8 kcal/mol at $\text{rms} \approx 5 \text{ Å}$. This barrier derives from incomplete compensation of a peak in the intramolecular nonbonded energy by the configurational entropy and the solvation free energy. The one-dimensional profile as a function of the ASA of the hydrophobic assembly (not shown) is smooth and exhibits a single minimum at native-like values.

In an alternative projection that is closely related to the model of Muñoz *et al.* (14, 17), we plot the free energy at 300 K as a function of the number of native hairpin hydrogen bonds (N_H) and the ASA of the hydrophobic assembly (Fig. 3c). There are two minima, both of which fall at native-like ASA; they are separated by a barrier of about 1.5 kcal/mol at $(1, 430 \text{ Å}^2)$. This barrier derives primarily from the interplay of the intramolecular nonbonded and solvation free energy terms (Fig. 4b Inset), although the bonded terms also exhibit a maximum of a few kcal/mol in this region because of structural distortions.

To determine which of the native hydrogen bonds are most likely to play a role in directing folding in the barrier region (assumed to include the transition states), we calculated the probability of observing each over the temperature range $260 \leq T \leq 430 \text{ K}$ (Fig. 5). The native sheet population decreases steadily because of the small size of the system. Thus, consistent with the gradual decrease in fluorescence quantum yield (Fig. 2 of ref. 14), there is no cooperative transition. The overall population of the sheet hydrogen bonds in our model is 54% at 280 K, which is slightly higher than but comparable to the estimate of 42% from NMR based on chemical shift data (no structure was calculated) (12). In our model, the stability of individual hydrogen bonds appears to depend primarily on their proximity to the hydrophobic assembly, in particular Y45 and F52; the probabilities of hydrogen bonds in the turn region are markedly lower than those of the others (Figs. 1b and 5).

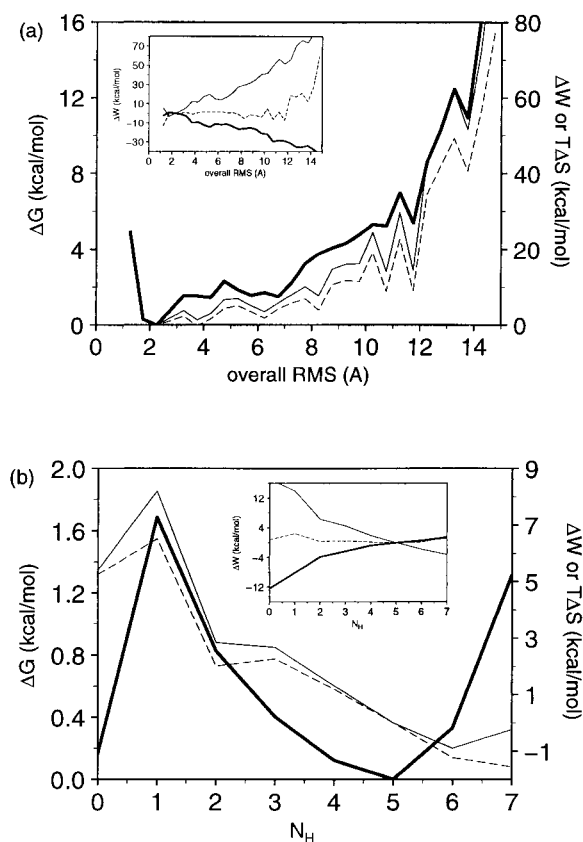


FIG. 4. Free energy (ΔG , bold solid line, left scale) and its effective energetic (ΔW , thin solid line, right scale) and entropic ($T\Delta S$, thin dashed line, right scale) components as functions of (a) the overall rms deviation from the native structure (Fig. 1a) and (b) the number of native sheet hydrogen bonds. (Inset) The average total bonded energy (dashed line), the average total intramolecular nonbonded energy (thin solid line), and the average solvent effective energy (bold solid line).

The folding mechanism that emerges from the projected surfaces in Fig. 3 involves a collapse that is downhill in free energy followed by rearrangement within the minimum at (0, 430 Å²) in Fig. 3c to form the hydrophobic assembly. This nascent structure would bring together the main chain in this region, and the hairpin hydrogen bonds would then propagate outwards from Y45 and F52. This folding scheme is confirmed

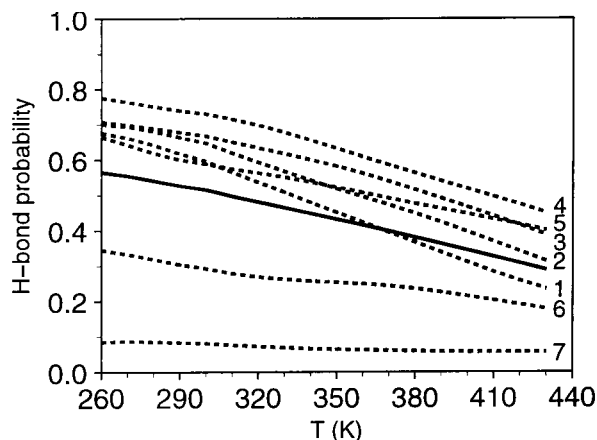


FIG. 5. Temperature dependence of total (solid line) and individual (dashed line) populations of native sheet hydrogen bonds; numbers correspond to those in Fig. 1b. See Fig. 1 legend for hydrogen bond criteria.

by Metropolis MC trajectories at 500 K that monitor unfolding from the native state (not shown) and is consistent with earlier studies of β -turns (29, 30). This mechanism contrasts with the model of Muñoz *et al.* (14, 17) in which folding initiates at the turn and propagates toward the tails, and the stabilizing hydrophobic assembly forms late in the reaction.

In modeling the kinetics as well as the thermodynamics, Muñoz *et al.* (14) used the “standard” single-sequence approximation, which was developed to study the helix-coil transition (31). It assumes that conformations with more than one segment of native peptide bonds can be ignored, which reduces the number of states of the β -hairpin model from $2^{15} = 32,768$ to 121. Since an earlier study found that the “standard” single-sequence approximation underestimates the entropy of the denatured state (32), Muñoz *et al.* (17) “modified” the approximation by enumerating all the binary sequences with more than one segment of native peptide bonds and counting them as part of the coil state (cccccccccccccc). However, the results from the full and reduced treatments are similar because no stabilizing interactions between residues sequentially separated by nonnative peptide bonds are included (see *Thermodynamic Parameters*). The free energy profiles exhibit minima at $N_\omega = 0$ and $N_\omega = 11$ that are separated by a barrier of about 4 kcal/mol (see Fig. 4 of ref. 14). The barrier derives from two aspects of the model: (i) the hydrogen bonds are strongly destabilizing because the configurational entropy loss associated with forming such interactions is much larger than the enthalpy gain; and (ii) the favorable native side chain interactions are not counted until $N_\omega = 5$.

Fig. 6 shows corresponding free energy profiles calculated from the simulation data. The single sequence curves (which differ from each other by definition only at $N_\omega = 0$) are similar to the free energy profile of the model of Muñoz *et al.* (14, 17) in that both have minima at low (coil state) and high (hairpin state) numbers of native peptide bonds that are separated by a large barrier (about 7 kcal/mol). When all the structures are included, whether or not they have more than one stretch of native peptide bonds, the results are strikingly different: the free energy decreases smoothly from a coil state with $N_\omega = 0$ to a hairpin state with $N_\omega = 13$ ($N_\omega = 14$ and $N_\omega = 15$ are higher in free energy because they have very low entropy). The difference results from the fact that there are very few conformations with a single segment of native peptide bonds (only 25,881 of the 210,000 sampled structures, which account for 9.2% of the population at 300 K), particularly with $N_\omega \approx 7$. No barrier is observed when the rest of the structures are included because many (previously excluded) native-like con-

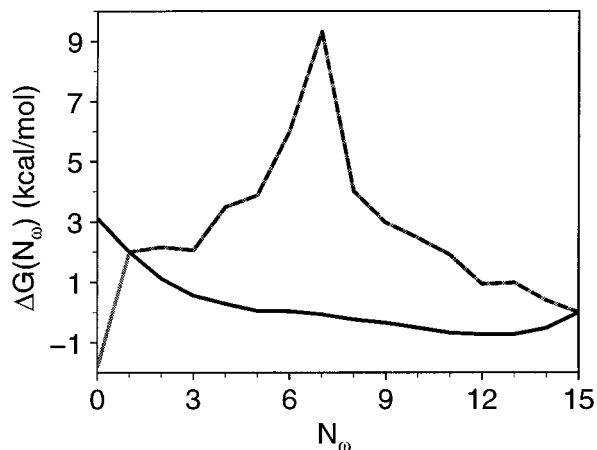


FIG. 6. Free energy as a function of the number of native peptide bonds (N_ω). The black solid line is for the total number of native peptide bonds. The dashed line represents the “standard” single sequence approximation (31) and the gray line represents the “modified” single sequence approximation (ref. 17; see text).

formations map to $N_{\omega} \approx 7$. Such conformations are essentially ignored (by definition) in the simple model. Fig. 6 thus demonstrates that the two-state behavior of the simple model (a free energy profile with two minima separated by a significant barrier) results directly from its one-dimensional (helix-coil) nature.

Muñoz *et al.* stress that their model “produces a funnel-like free energy surface” (p. 197 and Fig. 4a of ref. 14). Although the effective energy (the height of a protein-folding funnel) and the configurational entropy (the width of a protein-folding funnel) both decrease as the native state is approached (for examples of protein-folding funnels, see refs. 2–4), this behavior is not obvious from Fig. 4a of ref. 14, which is a full free energy surface. Rather, the only “funnel-like” aspect of the diagram is its shape, which results directly from the choice of variables: the length and center of the segment of native peptide bonds (only the “standard” single sequence approximation was considered in ref. 14); by construction, only conformations with the turn in its native position are stabilized by the hydrogen bond and side chain interactions included in the simple model. This figure thus highlights the importance of identifying coordinates for the reaction that yield physically meaningful projections of the multidimensional free energy surface. The poor separation of native and nonnative states in the N_{ω} direction (solid black line in Fig. 6) shows that, for realistic models, that parameter is not a good coordinate for describing the reaction; consistent with this idea, for the set of 210,000 structures, a native hydrogen bond is formed (according to the criteria given in Fig. 1) in only about 60% of the occurrences in which the corresponding peptide bonds are native.

CONCLUSIONS

An analysis of the thermodynamics of folding of the 16-residue β -hairpin based on an atomistic model with implicit solvation has yielded elementary interaction parameters similar to those obtained by Muñoz *et al.* (14, 17) by fitting a simple model to their fluorescence data. However, the mechanism of folding suggested by the simulations is different from the one that results from their one-dimensional helix-coil model. They obtain a folding mechanism in which the hairpin forms the turn first and then “zips up” the remaining native hydrogen bonds (14, 17). If the elementary step of forming a hydrogen bond is as fast in the β -hairpin as in α -helices, their scenario suggests that what makes the former fold slowly relative to the latter is that the turn often dissolves before the stabilizing hydrophobic side chain interactions are made. The present analysis suggests that the hydrophobic assembly, in particular the Y45-F52 interaction, forms early. This nascent structure brings together the main chain in this region, from which point the sheet propagates outwards. The folding rate is dominated by the time required for interconversion between compact conformations. As pointed out by Muñoz *et al.* (14, 17), the experiments do not exclude such a mechanism. Moreover, single exponential behavior has been observed for lattice models of proteins, which fold by collapse and subsequent rearrangement (33). Although the implicit solvent model used in the present study may neglect some aspects of the desolvation process in the formation of a fully shielded hydrophobic cluster (34), its early formation calls into question what portion of the folding reaction is measured by the fluorescence measurements. It is hoped that additional experiments [e.g., monitoring the formation of individual hydrogen bonds with infrared measurements of isotopically labeled peptide (35)] and simulations will yield a more complete understanding of this elementary folding reaction.

We thank Luis Serrano and coworkers for providing their NMR data for comparison and David Chandler, Vijay Pande, David Rokhsar, and William Eaton for comments on the manuscript. A.R.D. is a Howard Hughes Medical Institute Predoctoral Fellow. This work was supported in part by a grant from the National Science Foundation.

- Karplus, M. (1997) *Folding Des.* **2**, S69–S75.
- Dill, K. A. & Chan, H. S. (1997) *Nat. Struct. Biol.* **4**, 10–19.
- Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995) *Science* **267**, 1619–1620.
- Dobson, C. M., Karplus, M. & Šali, A. (1998) *Angew. Chem. Int. Ed.* **37**, 868–893.
- Dobson, C. M., Evans, P. A. & Radford, S. E. (1994) *Trends Biochem. Sci.* **19**, 31–37.
- Fersht, A. R. (1995) *Curr. Opin. Struct. Biol.* **5**, 79–84.
- Scholtz, J. M. & Baldwin, R. L. (1992) *Annu. Rev. Biophys. Biomol. Struct.* **21**, 95–118.
- Eaton, W. A., Muñoz, V., Thompson, P. A., Chan, C.-K. & Hofrichter, J. (1997) *Curr. Opin. Struct. Biol.* **7**, 10–14.
- Callender, R. H., Dyer, R. B., Gilmanshin, R. & Woodruff, W. H. (1998) *Annu. Rev. Phys. Chem.* **49**, 173–202.
- Caflish, A. & Karplus, M. (1994) in *The Protein Folding Problem and Tertiary Structure Prediction*, eds. Merz, K. M., Jr. & Le Grand, S. M. (Birkhäuser, Boston), pp. 193–230.
- Karplus, M. & Weaver, D. L. (1994) *Protein Sci.* **3**, 650–668.
- Blanco, F. J., Rivas, G. & Serrano, L. (1994) *Nat. Struct. Biol.* **1**, 584–590.
- Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T. & Clore, G. M. (1991) *Science* **253**, 657–661.
- Muñoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997) *Nature (London)* **390**, 196–199.
- Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H. & Dyer, R. B. (1996) *Biochemistry* **35**, 691–697.
- Thompson, P. A., Eaton, W. A. & Hofrichter, J. (1997) *Biochemistry* **36**, 9200–9210.
- Muñoz, V., Henry, E. R., Hofrichter, J. & Eaton, W. A. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5872–5879.
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comp. Chem.* **4**, 187–217.
- Lazaridis, T. & Karplus, M. (1999) *Proteins* **35**, 133–152.
- Berg, B. A. & Neuhaus, T. (1992) *Phys. Rev. Lett.* **68**, 9–12.
- Neria, E., Fischer, S. & Karplus, M. (1996) *J. Chem. Phys.* **105**, 1902–1921.
- Lazaridis, T. & Karplus, M. (1997) *Science* **278**, 1928–1931.
- Lazaridis, T. & Karplus, M. (1999) *J. Mol. Biol.* **288**, 477–487.
- Dinner, A. R. (1999) Ph.D. thesis (Harvard University, Cambridge, MA).
- Dodd, L. R., Boone, T. D. & Theodorou, D. N. (1993) *Mol. Phys.* **78**, 961–996.
- Gō, N. & Scheraga, H. A. (1970) *Macromolecules* **3**, 178–187.
- Karpen, M. E., Tobias, D. J. & Brooks, C. L., III (1993) *Biochemistry* **32**, 412–420.
- Daura, X., Jaun, B., Seebach, D., van Gunsteren, W. F. & Mark, A. (1998) *J. Mol. Biol.* **280**, 925–932.
- Tobias, D. J., Sneddon, S. F. & Brooks, C. L., III (1990) *J. Mol. Biol.* **216**, 783–796.
- Lazaridis, T., Tobias, D. J., Brooks, C. L., III & Paulaitis, M. E. (1991) *J. Chem. Phys.* **95**, 7612–7625.
- Schellman, J. A. (1958) *J. Phys. Chem.* **62**, 1485–1494.
- Qian, H. & Schellman, J. A. (1992) *J. Phys. Chem.* **96**, 3987–3994.
- Karplus, M., Šali, A., Shakhnovich, E. & Caflish, A. (1995) in *Modelling of Biomolecular Structures and Mechanisms*, eds. Pullman, A. & Jortner, J. (Kluwer, Boston), pp. 69–84.
- Lum, K., Chandler, D. & Weeks, J. D. (1999) *J. Phys. Chem. B* **103**, in press.
- Dyer, R. B., Gai, F., Woodruff, W. H., Gilmanshin, R. & Callender, R. H. (1998) *Acc. Chem. Res.* **31**, 709–716.