Folding Free Energy Surface of a Three-Stranded β-Sheet Protein

Badry D. Bursulaya and Charles L. Brooks III*

Contribution from the Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received May 27, 1999. Revised Manuscript Received August 24, 1999

Abstract: We present a molecular dynamics (MD) simulation study of the folding thermodynamics of the three-stranded β-sheet protein Betanova. The protein and solvent are explicitly described by employing all atom models. An umbrella sampling technique was employed to probe thermodynamically relevant states at different stages of folding. A database for the sampling was generated by conducting four high-temperature simulations. The initial conditions for the umbrella sampling were selected from this database of structures by employing hierarchical clustering. Sampling of conformational space was then carried out at 275 K and the generated data were combined with the weighted histogram method to produce the two-dimensional folding free energy landscape. We found that the folding of the protein Betanova occurs in two collapse stages. The first collapse brings the protein into a basin that contains various structures differing in their size and elements of secondary structure. At the transition state from this basin of collapsed states to the native basin, the protein adopts a native-like fold and size and forms ≈60% of native contacts. Thus the formation of native-like structure is concurrent with the secondary collapse. The overall stability of protein Betanova is found to be about 1 kcal/mol, in agreement with the experimental estimate. We found that the native side chain contacts are the primary factor in driving Betanova folding and stabilizing its native three-stranded β-sheet conformation. By contrast, hydrogen bonding is found to play a minor role in the folding of Betanova. Solvent is observed to be present in the protein core until late in folding.

1. Introduction

Understanding the mechanism of protein folding is extremely important for predicting its three-dimensional conformation under physiological conditions. Much work has been done in both experiment[1,2] and theory[3−10] to elucidate folding mechanisms. According to the modern view, protein folding is regarded as a gradual descent down the folding funnel, whose major features are the local minima which could transiently trap the protein in misfolded states (the landscape roughness) and the overall downhill slope toward the native state. A key notion of this theory is that in all but the final stages of folding there exists an ensemble of structures and, consequently, that protein folding occurs via multiple pathways.[10]

Insights into the mechanism of protein folding and the connection between experiment and theory are provided by numeric simulations. However, due to the size of proteins and the time scale on which folding occurs, its direct study is prohibitive and different approaches have been invented to circumvent this problem. Various minimalist models[5,11−17] which use a reduced (minimal) protein description and effective interaction potentials have been employed in one approach. In these models the protein is envisioned as a series of beads occupying lattice sites[5,11,14,16] or embedded in a continuum[12,13,15−17] (off-lattice models). Although the minimalist models oversimplify the protein structure and the interaction potential, they have been successful in reproducing such essential aspects of protein folding as the existence of the unique native state and cooperative folding.[5] Such models continue to be instrumental in establishing possible scenarios for protein folding.[7,10]

Another approach is all-atom molecular dynamics simulation in explicit solvent. Since the folding takes place on a very large time scale, early research in this area was mainly focused on protein unfolding,[18−24] which can be accelerated by conducting

* Corresponding author. E-mail: brooks@scripps.edu. Phone: 619/784-8035. Fax: 619/784-8688.

(1) Fersht, A. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 3−9.
(2) Eaton, W. A.; Muñoz, V.; Thompson, P. A.; Chan, C.-K.; Hofrichter, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 10−14.
(3) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct. Design* **1995**, *1*, 167−195.
(4) Karplus, M.; Sali, A. *Curr. Opin. Struct. Biol.* **1995**, *5*, 58−73.
(5) Shakhnovich, E. I. *Curr. Opin. Struct. Biol.* **1997**, *7*, 29−40.
(6) Dill, K. A.; Chan, H. S. *Nature Struct. Biol.* **1997**, *4*, 10−19.
(7) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Res. Phys. Chem.* **1997**, *48*, 545−600.
(8) Brooks, C. L., III *Curr. Opin. Struct. Biol.* **1998**, *8*, 222−226.
(9) Brooks, C. L., III; Grubele, M.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11037−11038.
(10) Dobson, C. L.; Karplus, M. *Curr. Opin. Struct. Biol.* **1999**, *9*, 92−101.
(11) Kolinski, A.; Skolnick, J. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 338−352.
(12) Thirumalai, D.; Guo, Z. *Biopolymers* **1995**, *35*, 137−140.
(13) Guo, Z.; Brooks, C. L., III *Biopolymers* **1997**, *42*, 745−757.
(14) Kolinski, A.; Galazka, W.; Skolnick, J. *J. Chem. Phys.* **1998**, *108*, 2608−2617.
(15) Zhou, Y.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14429−14432.
(16) Gutin, A.; Abkevich, V.; Shakhnovich, E. *Fold. Des.* **1998**, *3*, 183−194.
(17) Shea, J.-E.; Nochomovitz, Y. D.; Guo, Z.; Brooks, C. L., III *J. Chem. Phys.* **1998**, *109*, 2895−2903.
(18) Fan, P.; Kominos, D.; Kitchen, D. B.; Levy, R. M. *Chem. Phys.* **1991**, *158*, 295−301.
(19) Mark, A. E.; van Gunsteren, W. F. *Biochemistry* **1992**, *31*, 7745−7748.
(20) Daggett, V.; Levitt, M. *J. Mol. Biol* **1993**, *232*, 600−619.
(21) Caflish, A.; Karplus, M. *J. Mol. Biol.* **1995**, *252*, 672−708.
(22) Li, A.; Daggett, V. *J. Mol. Biol.* **1996**, *257*, 412−429.
(23) Triado-Rives, J.; Orozco, M.; Jorgensen, W. L. *Biochemistry* **1997**, *36*, 7313−7329.
(24) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928−1931.

10.1021/ja991764l CCC: $18.00
Published on Web 10/15/1999

the simulation at high temperature, extreme pH, or in the presence of denaturant. Reviews of papers on protein unfolding and what can be concluded from such studies are given in refs 4, 8, 25, and 26. In addition, several papers have recently been published in which direct protein folding has been explored.[27−31] Both unfolding and direct folding studies require multiple MD runs of very long duration and this requirement sets a limit to their utility. Also, the conformational space sampled in the protein unfolding simulations could differ from that of the protein under physiological conditions.

Alternative to studying the kinetics of folding, all-atom molecular dynamics simulation in explicit solvent can be used to map out the folding free energy landscape along various coordinates (radius of gyration, amount of secondary structure, etc.). For example, this approach was employed to study the folding free energy landscape for the three-helix bundle fragment B of *Staphylococcal* protein A,[32,33] segment B1 of *Streptococcal* protein G (GB1)[34,35] and major cold shock protein A (CspA).[36] The folding coordinates used in these studies included the radius of gyration and the fraction of native contacts. Comparison of these studies reveals that the topology of the protein native state influences the folding mechanism. In particular, in protein A, which is an all-α protein, the formation of tertiary (native) structure occurs concurrently with the formation of the secondary structure and its free energy folding landscape is close to a diagonal shape. In contrast, for the mixed α/β protein GB1, the collapse, accompanied by the formation of ≈35% of the native structure, occurs first and is followed by evolution toward the native state without significant change in size. Similar behavior is seen for CspA. We note that the collapsed states of these proteins have a native-like topology and that their folding surface is close to an L shape.

The aim of the present work is to study, by all-atom MD simulations, the folding free energy landscape for a designed all-β protein and relate it to the previously studied cases of "natural" all-α and mixed α/β proteins. As a model for our study we have chosen the designed 20-amino acid three-stranded β-sheet protein Betanova.[37] The native structure of this protein was determined by NMR spectroscopy and it is depicted in Figure 1. It contains two turns (Asn7-Gly8 and Asn13-Gly14) and three β-strands (β1, β2, and β3).

The thermodynamics and kinetics of Betanova folding were determined via thermal and chemical denaturation monitored by CD and fluorescence spectroscopies, respectively. It was found that Betanova exhibits cooperative two-state folding−unfolding behavior and that its thermodynamic stability is around 0.6 −0.7 kcal/mol at 278 K. Due to its marginal stability this designed protein represents a challenge for MD study.

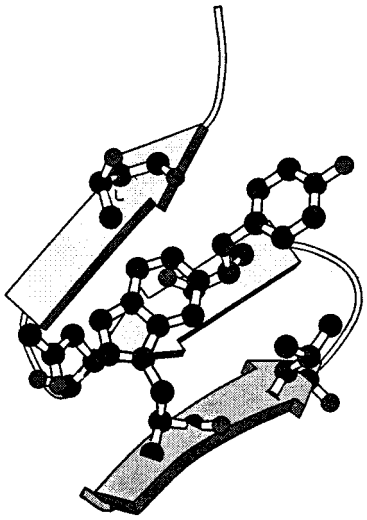

Figure 1. Minimized average NMR structure of protein Betanova from ref 37. Key side chain interactions are shown explicitly.

Table 1. List of Native Contacts for Protein Betanova

no.	1st residue	2nd residue	no.	1st residue	2nd residue
1	Gly2	Ser4	10	Gly8	Tyr10
2	Trp3	Tyr10	11	Lys9	Thr16
3	Trp3	Asn12	12	Lys9	Glu18
4	Trp3	Thr17	13	Tyr10	Thr17
5	Ser4	Gln6	14	Tyr10	Arg20
6	Ser4	Thr11	15	Thr11	Thr16
7	Val5	Gly8	16	Asn12	Gly14
8	Val5	Tyr10	17	Asn12	Thr16
9	Gln6	Thr11	18	Asn12	Thr17

In what follows we describe models and methods employed in this study and present our findings on the thermodynamics and mechanism of folding of the designed protein Betanova.

2. Methods

In this section we briefly outline the methodology used for the sampling of the folding free energy landscape of protein Betanova. We follow the techniques developed in this group from previous studies and provide only a sketch of the methods used here. For a more complete account of these methods, the reader is referred to the refs 32−35.

(i) First we characterized the native state of the protein by conducting two native MD simulations (275 K, see below), using the averaged NMR structure of Betanova[37] as a starting point. By analyzing the generated native trajectories we determined which residues were in contact with probability higher than 56% (native contacts) or hydrogen bonded with probability higher than 66% (native hydrogen bonds). Two residues not adjacent in the sequence were considered to be in contact or hydrogen bonded if the centers of geometry of their side chains were within 6.5 Å (Cα distance for gly) or if the distance between their backbone hydrogen and oxygen atoms was within 2.5 Å. Native contacts and native hydrogen bonds are essential for defining the native fold. We identified 18 native contacts (see Table 1) and 7 native hydrogen bonds.

(ii) Next we prepared the initial conditions for the sampling. This was done by running four high-temperature unfolding trajectories and partitioning the generated database of structures based on the number of native contacts and radius of gyration. The total number of partitions (bins) was 24. For each bin a hierarchical clustering was performed to choose the representative structures that could serve as the initial conditions for the sampling. Following the previous work on protein

(25) Daggett, V.; Levitt, M. *Curr. Opin. Struct. Biol.* **1994**, *4*, 291−295.
(26) Caflish, A.; Karplus, M. In *The protein folding and tertiary structure prediction*; Merz, K., Jr., Le Grand, S., Eds.; Birkhäuser: Boston, 1994; pp 193−230.
(27) Demchuk, E.; Bashford, D.; Case, D. A. *Fold. Des.* **1997**, *2*, 35−46.
(28) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925−932.
(29) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740−744.
(30) Wu, X.; Wang, S. *J. Phys. Chem. B* **1998**, *102*, 7238−7250.
(31) Bartels, C.; Schaefer, M.; Karplus, M. *Theor. Chem. Acc.* **1999**, *101*, 62−66.
(32) Boczko, E. K.; Brooks, C. L., III *Science* **1995**, *269*, 393−396.
(33) Guo, Z.; Brooks, C. L., III; Boczko, E. K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10161−10166.
(34) Sheinerman, F. B.; Brooks, C. L., III *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1562−1567.
(35) Sheinerman, F. B.; Brooks, C. L., III *J. Mol. Biol.* **1998**, *278*, 439−455.
(36) Brooks, C. L., III. Manuscript in preparation.
(37) Kortemme, T.; Ramirez-Alvardo, M.; Serrano, L. *Science* **1998**, *281*, 253−256. The protein sequence is as follows: arg gly trp ser val gln asn gly lys tyr thr asn asn gly lys thr thr glu gly arg.

A[32,33] and GB1,[35] we employed the number of native contacts, the number of native hydrogen bonds, and protein solvation energy as descriptors for the protein structure and parameters for the dissimilarity function. Thus, overall 26 protein descriptors (18 native contacts, 7 native hydrogen bonds, plus the solvation energy) were defined. Upon the completion of clustering we obtained 74 cluster centers, which were used as initial conditions for the sampling.

(iii) The initial conditions were resolvated and reequilibrated as described at the end of this section. Molecular dynamics simulations with a biased (umbrella) potential were performed for a duration of 300 ps to explore the configurational space in the vicinity of each initial condition. The biasing potential was identical in form with that employed by Sheinerman and Brooks.[35] A force constant of 500 kcal/mol was employed.

(iv) Finally, the sampling data were combined using the weighted histogram analysis method (WHAM).[38−40] This method provides a robust estimate of the density of states projected onto specific reaction coordinates (radius of gyration, number of native contacts, etc.), as has been demonstrated in a number of biophysical applications.[13,32−35,40]

The molecular dynamics simulations were performed with the CHARMM package using a parameter set TOPH19/PARAM19.[41] The protein was solvated with TIP3P water molecules[42] in a truncated octahedron constructed from a cubic box with an edge length of 56.57 Å. Corresponding periodic boundary conditions were employed. The simulations were performed with a constant volume and constant solvent density. The trajectories were integrated with a 2 fs time step using the Verlet leap-frog algorithm.[43] The covalent bonds between hydrogen and heavy atoms were fixed with the SHAKE algorithm.[44] The nonbonded interactions were truncated at 11 Å with an electrostatic shifting function and van der Waals switching, with the switch beginning at 9 Å. The nonbonded interactions list was extended to 12 Å and updated every 20 steps. A constant temperature was maintained by reassigning atom velocities from a Gaussian distribution if the average temperature drifted outside of a window ±5 K. The temperature was maintained at 275 K for the native and sampling simulations. The length of the two native trajectories was 2.4 and 2.8 ns, respectively. The length of each of the 74 sampling runs was 100 ps equilibration plus 300 ps for data collection. Four unfolding simulations were conducted at temperatures 350, 375, 385, and 400 K. Their respective lengths were 2, 1, 1.5, and 2 ns. Thus, a total of ≈45 ns of dynamics was used to construct the free energy landscape. In all MD simulations the generated data were saved every 100 steps for subsequent analysis. Calculations were done on the CRAY T3E supercomputer at the San Diego Supercomputing Center. Sixty picoseconds of molecular dynamics required ≈1 CPU hour on 64 processors of the T3E.

The following procedure was employed to start each of the simulations. First, the protein in its initial conformation was inserted into a box containing 2942 water molecules. The molecules overlapping with protein atoms were removed. A constant number of water molecules, namely 2831, was maintained for each of the initial conditions. This was done by varying the cutoff distance for the removal of water between 2.3 and 2.7 Å. Next the protein/solvent system was minimized for 200 steps with the steepest descent algorithm and then equilibrated by running a 100 ps MD trajectory. Harmonic restraints with a gradually decreasing force constant were applied to the protein heavy atoms during the first 40 ps of the equilibration.

(38) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.
(39) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(40) Boczko, E. K.; Brooks, C. L., III *J. Phys. Chem.* **1993**, *97*, 4509−4513.
(41) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(42) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(43) Allen, M. P.; Tldesley, D. J. *Computer Simulation of Liquids*; Clarendon: Oxford, 1989.
(44) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

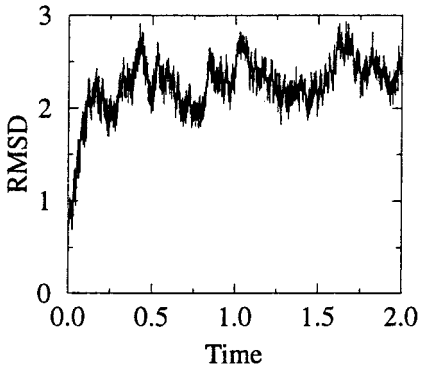

Figure 2. The time evolution of the backbone root-mean-square deviation relative to the initial configuration for the native simulation at 275 K.

3. Results and Discussion

We start by analyzing the properties of the protein Betanova derived from the native trajectories. In Figure 2 the backbone root-mean-square deviation (rmsd) between the initial and instantaneous protein conformations as a function of evolution time along the trajectory is shown. We see that the backbone rmsd reaches a value of 2−2.5 Å very quickly and stays at that level throughout the simulations. This suggests that the native state of Betanova is both stable and very mobile. The radius of gyration is found to be ≈6.8 Å. The average structure obtained form the native trajectories is very similar to the initial (experimental) one. This is based on the fact that the following six contacts, Trp3-Tyr10, Trp3-Asn12, Trp3-Thr17, Val5-Tyr10, Tyr10-Thr17, and Asn12-Thr17, are present in both initial and average structures (see Table 1). These contacts constitute the hydrophobic cluster, as was observed in the experimental studies on this system.[37]

Next we consider the folding free energy landscape of Betanova. The potential of mean force (PMF) as a function of the fraction of native contacts, ρ, and radius of gyration, R_g, is given in Figure 3. First we note that the basin of unfolded states, which is located in the upper left corner of the PMF plot, is relatively narrow along the ρ coordinate and extends from 16 to 11 Å along the R_g coordinate. It contains a local minimum, centered at $\rho \approx 0.15$ and $R_g \approx 12$ Å. A typical conformation adopted by the protein in this local minimum is depicted in Figure 4. One can see that it is curved at two locations, which roughly correspond to the turns in the native form, and lacks any discernible secondary structure. Thus, the presence of these metastable conformations suggests some intrinsic turn propensities exist in this sequence. To escape from this local minimum to a basin of more compact (collapsed) states, the protein has to overcome a small, ≈1 kcal/mol, barrier.

The basin of collapsed states is wide in terms of both fraction of native contacts and radius of gyration. The former varies between 0.1 and 0.6 and the latter varies between 7 and 11 Å. By contrast to the basin of unfolded states, the basin of collapsed states is populated by protein conformations having various elements of secondary structure. Initial secondary structure, attained by the protein as it enters the basin of collapsed states, involves formation of a native-like turn, Asn13-Gly14, and orientation of the β2 and β3 strands in such a way that some of their residues are in contact. No native hydrogen bonds are formed between β-strands 2 and 3 at this point. Further progress along the folding coordinate ρ without significant compaction brings the protein into a local free energy minimum at $\rho \approx 0.6$ and $R_g \approx 9.5$ Å. This evolution yields a well-formed β2−β3

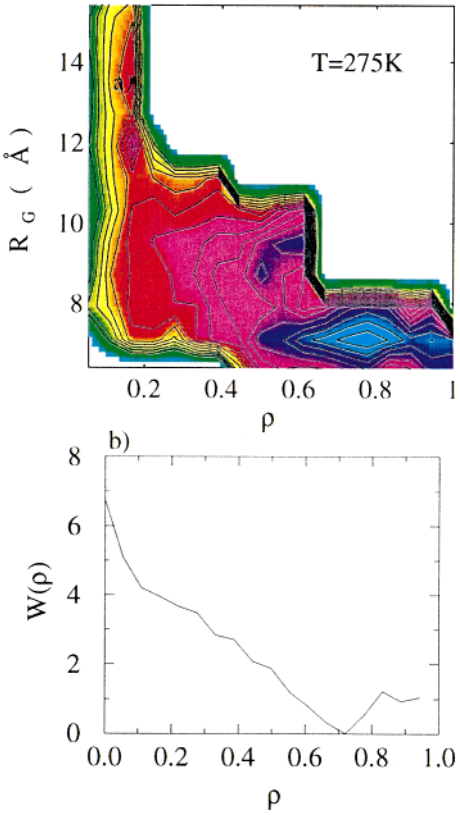

Figure 3. (a) Potential of mean force as a function of radius of gyration, R_g, and fraction of native contacts, ρ. The contours are drawn every 0.5 kcal/mol. (b) Potential of mean force as a function of fraction of native contacts, ρ.

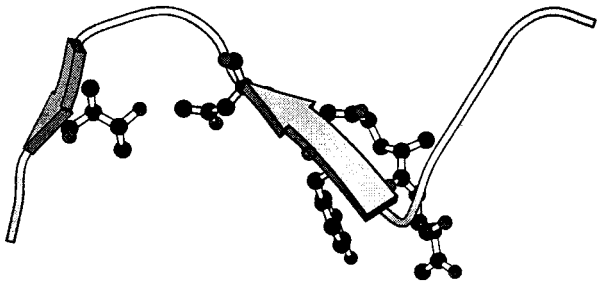

Figure 4. A typical conformation, adopted by the protein at the local minimum with $\rho \approx 0.15$ and $R_g \approx 12$ Å. Some of the specific side chain interactions occurring for these conformations are shown in atomic detail.

hairpin having interstrand hydrogen bonds. We note in passing that the Betanova sequence had been built on a stable hairpin, containing most of the residues which constitute the β2 and β3 strands of protein Betanova. These are the very same strands that participate in the early collapse of the protein. The analysis of more compact structures populating the basin of collapsed states reveals the following. Typically at smaller values of ρ, each structure has either one of the native turns and the native contacts between corresponding β-strands. At larger values of ρ all structures contain both native turns, although only two β-strands are aligned to form native contacts and hydrogen bonds between each other, while the third one is frayed away. In particular for small values of ρ and R_g, structures containing the Asn7-Gly8 turn and corresponding native-like orientation of β1 and β2 strands dominate. At even larger values of folding coordinate, ρ approaching 0.6, structures containing all three β-strands in correct native-like orientation start to appear. This

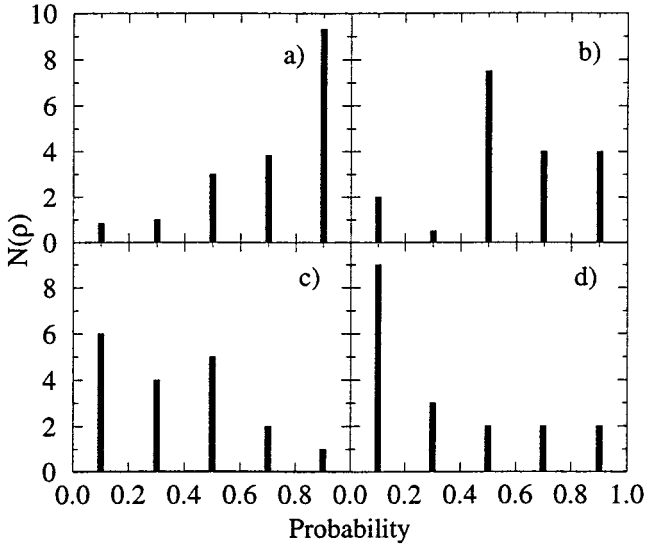

Figure 5. Number of native contacts having a given probability of formation for various values of folding coordinate ρ: (a) $\rho = 0.7$; (b) $\rho = 0.6$; (c) $\rho = 0.4$; and (d) $\rho = 0.2$.

signifies that the protein approaches a basin of native states. Thus 60% of native contacts is necessary to achieve a stable three-stranded β-sheet conformation.

The global minimum on the folding free energy surface, corresponding to the basin of native states of the protein, is located in the lower right corner. The basin of native structures is relatively narrow along the R_g coordinate and begins at $\rho = 0.6$. There is not a significant barrier between the native basin and the basin of collapsed structures, suggesting that the folding of Betanova is likely to be diffusion controlled. It should be emphasized here that at the transition point between the basins of collapsed and native states the protein adopts native-like fold and compactness.

From a large variety of structures populating the basin of collapsed states in the PMF surface of Betanova, we conclude that its folding occurs via multiple pathways, at least in the thermodynamical sense. To be more quantitative, we plot in Figure 5 the number of native contacts with given probability of formation for different values of folding coordinate ρ. At values of ρ corresponding to the basin of collapsed states, we observe a broad distribution of contacts with probability ranging from 0.1 to 0.7. This suggests that the contacts are distributed between various structures. By contrast, the distribution becomes more localized at higher values of probability for values of ρ near the native state ($\rho = 0.7$, see Figure 5), which indicates the formation of a stable core.

In Figure 6 we present the cumulative probability, $P(\rho^*)$, for the protein to occupy states with $1 \leq \rho \leq \rho^*$:

$$P(\rho^*) = \frac{\int_1^{\rho^*} d\,\rho \exp(-\beta W(\rho))}{\int_1^0 d\,\rho \exp(-\beta W(\rho))} \qquad (1)$$

where $\beta = 1/k_B T$ is the Boltzmann factor and $W(\rho)$ is the potential of mean force (see Figure 3b). We see that the accumulated population of protein Betanova in the native basin, $\rho > 0.6$, is $\approx$90%. This translates into $\approx$1 kcal/mol stability of the native state. As was mentioned in the Introduction, the corresponding experimental estimate is 0.6−0.7 kcal/mol at 278 K. Thus we observe good agreement between our MD result and the experimental estimates for Betanova stability.

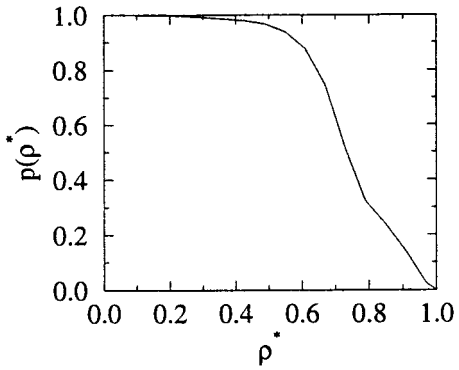

Figure 6. Cumulative probability to observe a protein conformation along the folding coordinate ρ.

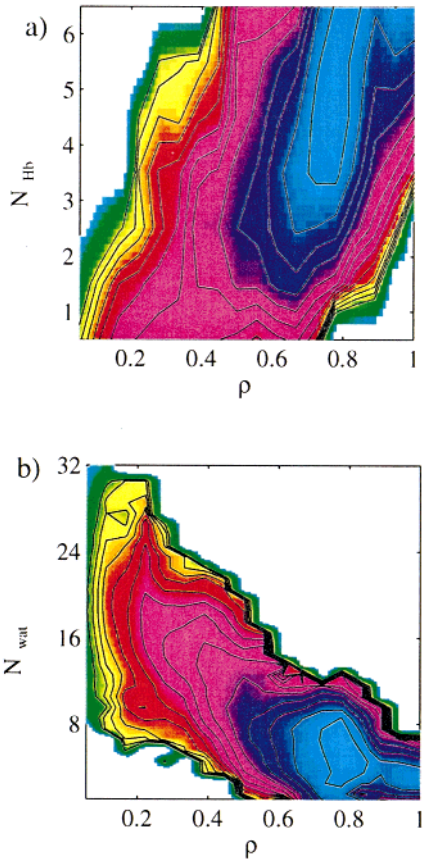

Figure 7. (a) Potential of mean force as a function of total number of native hydrogen bonds, N_{hb}, and fraction of native contacts, ρ. (b) Potential of mean force as a function of number of core water molecules, N_{wat}, and fraction of native contacts, ρ. The contours are drawn every 0.5 kcal/mol.

Turning to the issue of the role played by interstrand hydrogen bonds in the folding of β-sheet proteins, we investigate the folding free energy as a function of the total number of native hydrogen bonds, N_{hb}, and fraction of native contacts, ρ (see Figure 7a). First we note that the free energy surface has only one basin centered around $\rho = 0.7$. This basin corresponds to the native state of the protein. In the native basin N_{hb} varies between 3 and 7, suggesting that hydrogen bonding may play a role in stabilizing the protein native topology. We also see that early in the folding the free energy is independent of N_{hb} up to a value of $\rho = 0.6$. This suggests that hydrogen bonding is not a driving force in β-sheet protein folding. Our finding is somewhat similar to that of Honig et al.,[45] who showed that the primary factor determining the stability of β-sheet proteins is the van der Waals and hydrophobic interactions, while electrostatic interactions including hydrogen bonding were found to be destabilizing.

To better understand what is driving the folding of this designed β-sheet protein, and to establish a connection between our results and those of Honig et al., we investigate the role of solvent. We characterize the solvation of the "protein core" by the number of water molecules, N_{wat}, inside of a sphere of radius 6 Å drawn around the center of mass of the protein. While such a small protein really does not possess a well-defined buried core in the native state, our geometric definition of the core serves to deliniate the role of interactions with solvent during folding. In the native state $N_{wat} = 5 \pm 2$ and it increases to 30 as the protein unfolds. The potential of mean force as a function of N_{wat} and ρ is shown in Figure 7b. By contrast to the PMF in Figure 3a, it has only one free energy minimum, which corresponds to the native state of the protein, and no free energy barriers. The lack of the minima corresponding to the initial collapsed ($\rho \approx 0.15$ and $R_g \approx 12$ Å) and partially folded ($\rho \approx 0.6$ and $R_g \approx 9.5$ Å) states suggests that there is no correlation between the stabilization of those states and solvation. From this and from our previous conclusion regarding the role of hydrogen bonding, we suggest that the contacts between residue side chains are responsible for the stability of the structures in the local minima in Figure 7a and are driving the folding in this protein.

4. Conclusions

We have examined the PMF surface for the folding of a designed protein, Betanova, which has three-stranded β-sheet topology in the native state. By contrast with all-α and mixed α/β protein systems, the folding of Betanova involves two-stage collapse. The first collapse results in a variety of structures which have between 10 and 60% of the native contacts. To enter the native basin the protein needs to undergo a second collapse. It also adopts a native-like fold and has ≈60% of native contacts at this transition point. The PMF surface does not exhibit a free energy barrier between the basins of native structures and collapsed states. The overall stability of Betanova is found to be about 1 kcal/mol, which agrees well with the experimental estimate. We also found that interstrand hydrogen bonds do not assist the folding of Betanova, rather to a certain degree they stabilize a native fold. Water is present in the protein core until late in the folding and its expulsion from the protein interior coincides with the formation of the tertiary structure.

Acknowledgment. We are very thankful to Drs. L. Serrano and T. Kortemme for providing detailed data for initial structural models of Betanova. We are grateful to the National Institute of Health (GM48807) for financial support. Computational support from a NRAC resources allocation at the San Diego Supercomputer Center is acknowledged.

JA991764L

(45) Yang, A−S.; Honig, B. *J. Mol. Biol.* **1995**, *252*, 366−376.