

## Understanding $\beta$ -Hairpin Formation: Computational Studies for Three Different Hairpins

Jinhyuk Lee and Seokmin Shin\*

School of Chemistry, Seoul National University, Seoul 151-747, Korea. \*E-mail: sshin@snu.ac.kr

Received December 24, 2007

We have studied the folding mechanism of  $\beta$ -hairpins in the proteins 1GB1, 3AIT and 1A2P by conducting unfolding simulations at moderately high temperatures. The analysis of trajectories obtained from molecular dynamics simulations in explicit aqueous solution suggests that the positions of the hydrophobic core residues lead to subtle differences in the details of folding dynamics. However, the folding of three different hairpins can be explained by a unified mechanism that is a blend of the hydrogen-bond-centric and the hydrophobic-centric models. The initial stage of  $\beta$ -hairpin folding involves various partially folded intermediate structures which are stabilized by both the van der Waals interactions of hydrophobic core residues and the electrostatic interactions of non-native hydrogen bonds. The native structure is obtained by forming native contacts in the final tune-up process. Depending on the relative positions of the hydrophobic residues, the actual mechanism of hairpin folding may or may not exhibit well-defined intermediates.

**Key Words** : Protein folding,  $\beta$ -Hairpin, Unfolding simulations, Hydrophobic core, Free energy

### Introduction

Understanding the mechanism of formation of basic structural elements such as  $\alpha$ -helices and  $\beta$ -sheets can provide useful information for the folding of larger proteins.  $\alpha$ -helix formation has been extensively studied both experimentally and theoretically.<sup>1-3</sup> By contrast, the formation of  $\beta$ -sheet structures has been investigated in detail only recently. It has been proposed that  $\beta$ -turns and  $\beta$ -hairpins act as initiation sites in early protein folding events.<sup>4</sup> A  $\beta$ -hairpin is the simplest form of anti-parallel  $\beta$ -sheet structure and is defined by a loop region flanked by two  $\beta$ -strands. Recent studies suggest that the positioning of the side chain groups in such a way as to promote the formation of a hydrophobic cluster is essential for the folding of hairpin structures.<sup>5</sup>

Concerning the detailed picture of hairpin formation, especially with respect to the relative timing of the formation of the interstrand hydrogen bonds near the turn and the hydrophobic core, two different mechanisms of  $\beta$ -hairpin folding have been proposed. Muñoz *et al.* studied the kinetics of folding a 16-residue  $\beta$ -hairpin from protein GB1 using a nanosecond laser temperature-jump technique.<sup>6</sup> They suggested that the formation of a  $\beta$ -hairpin from protein GB1 is initiated at the  $\beta$ -turn, which then “zips up” the remaining native hydrogen bonds. A turn stabilized by interstrand hydrogen bonds positions the aromatic residues so that they are poised to pack into a hydrocarbon cluster. Bonvin and van Gunsteren studied the stability and folding of a 19-residue  $\beta$ -hairpin fragment of the  $\alpha$ -amylase inhibitor tendamistat.<sup>7</sup> Several unfolding and refolding simulations suggested a model for  $\beta$ -hairpin formation in which the turn is formed first, followed by hydrogen bond formation closing the hairpin, and subsequent stabilization by side-chain hydrophobic interactions. Prevost and Ortman performed refolding simulations of a  $\beta$ -hairpin fragment of

barnase using a simulated annealing method.<sup>8</sup> They found that interstrand side-chain compactness and backbone hydrogen bonding provide concurrent stabilization for  $\beta$ -hairpin formation.

Pande and Rokhsar studied the unfolding and refolding pathway of a  $\beta$ -hairpin fragment of protein GB1 using molecular dynamics simulations.<sup>9</sup> They suggested that during high-temperature unfolding, the  $\beta$ -hairpin undergoes a series of sudden discrete conformational changes. According to their results, the hydrophobic cluster would form without assistance from the interstrand hydrogen bonds, suggesting that the  $\beta$ -hairpin refolds by a “hydrophobic collapse” mechanism. Dinner *et al.* obtained the free energy surface and conformations involved in the folding of the same  $\beta$ -hairpin from multicanonical Monte Carlo simulations.<sup>10</sup> Their results suggested that folding proceeds by a collapse leading to the formation of the hydrophobic assembly; the hairpin hydrogen bonds subsequently propagating outwards in both directions from the hydrophobic core. Ma and Nussinov studied the contributions of three components of a  $\beta$ -hairpin peptide: turn, backbone hydrogen bonding, and side-chain interactions.<sup>11</sup> They examined the structural stability of the  $\beta$ -hairpin under systematic perturbations of the turn region, backbone hydrogen bonds and the hydrophobic core formed by the side-chains. Their results support a side-chain-centric view of the folding of a hairpin structure. Zhou and Linhananta studied the role of hydrophilic and hydrophobic contacts in the folding of a  $\beta$ -hairpin fragment of protein GB1, using molecular dynamics simulations on all-atom based, simplified and reduced models.<sup>12</sup> It was suggested that folding of the hairpin is initiated by hydrophobic collapse before the main-chain hydrogen bonds.

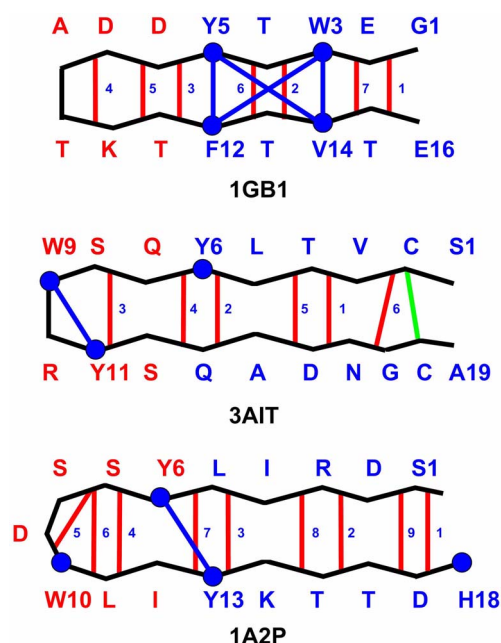
Zagrovic *et al.* used distributed computing techniques and a supercluster of thousands of computer processors to study folding of the C-terminal  $\beta$ -hairpin from protein GB1 in

atomistic detail using an implicit solvent model at 300 K.<sup>13</sup> They simulated a total of nearly 38  $\mu$ s of folding time and obtained eight complete and independent folding trajectories. It was found that partial hydrophobic core formation takes precedence over interstrand hydrogen bonding as the important interaction to initiate folding. They observed that final formation of the complete hydrophobic core occurs cooperatively at the same time that the final hydrogen bonding pattern appears. Garcia and Sanbonmatsu<sup>14</sup> and Zhou *et al.*<sup>15</sup> studied the folding free energy landscape of the same hairpin by using a highly parallel replica exchange method that combines MD trajectories with a temperature exchange MC process. It was suggested that the folding mechanism of the hairpin is a blend of the hydrogen-bond-centric and the hydrophobic-centric models, where the hydrophobic core and the  $\beta$  strand hydrogen bond form at roughly the same time.<sup>15</sup>

It is interesting to note that the three hairpins from proteins GB1 (PDB entry 1GB1: 16 residues), tendamist (PDB entry 3AIT: 19 residues), and barnase (PDB entry 1A2P: 18 residues) seemed to show different folding mechanisms. In a recent study on the thermodynamics and kinetics of off-lattice models for the  $\beta$ -hairpin fragment, Klimov and Thirumalai have suggested that the basic mechanisms of folding depend on the intrinsic rigidity of the hairpin, which is determined by the location of the hydrophobic cluster.<sup>16</sup> Ma and Nussinov investigated the free energy landscape of permutations in the hairpin of protein GB1 by studying six isomers corresponding to moving the hydrophobic cluster along the  $\beta$ -strands.<sup>17</sup> It was found that the energy landscape is dependent on the hydrophobic cluster topology and on the sequence. In a previous study, we have studied the unfolding of the  $\beta$ -hairpins in the above-mentioned three proteins by MD simulations in explicit water solvents at several temperatures.<sup>18</sup> The unfolding trajectories were analyzed by calculating various order parameters, examining the interaction energies and the secondary structure evolution, and using two-dimensional correlation analysis based on essential dynamics (ED) analysis.<sup>19</sup> It was suggested that the positions of the hydrophobic core residues influence the folding dynamics. In the present study, we have performed more extensive simulations on the three hairpin structures. Detailed analyses, including calculating free energy surfaces, are presented in order to provide a better understanding of the folding mechanism of hairpins.

### Model and Simulation Details

The three hairpin structures were obtained from the respective PDB entries (Fig. 1). In the present study, we are concerned with the formation of hydrophobic clusters by favorable interactions among hydrophobic residues. The hydrophobic residues are as follows: 1GB1 (3TRP 5TYR 12PHE 14VAL); 3AIT (6TYR 9TRP 11TYR); 1A2P (6TYR 10TRP 13TYR 18HIS). Except the valine residue of 1GB1, we consider residues having side chains with large rings such as PHE, TYR, TRP, and HIS to be hydrophobic core



**Figure 1.** Schematic descriptions of  $\beta$ -hairpin structures from the proteins GB1 (1GB1: 16 residues), tendamist (3AIT: 19 residues), and barnase (1A2P: 18 residues). The bend turn region (red) and the strand region (blue) are distinguished by different colors of the letters representing residues, while blue circles represent hydrophobic core residues. The red lines represent interstrand native hydrogen bonding with corresponding numbering of the hydrogen bonds. The blue lines represent interactions between proximal hydrophobic residues, while the green line corresponds to disulfide linkage in the 3AIT hairpin.

(See Figure 1 of Reference 18). It is noted that the designations of hydrophobic core residues for 1GB1 is consistent with the previous studies and those for 3AIT and 1A2P follow the similar reasoning. We have also identified the bend turn regions of the hairpins as follows: 1GB1 (6th-11th residues), 1A2P (6th-12th residues) and 3AIT (7th-12th residues). The blue lines in Figure 1 represent the interactions between proximal hydrophobic residues forming the hydrophobic core. Such hydrophobic cores are formed in the strand region for 1GB1, in the bend region for 3AIT and in a region between the bend and the strand for 1A2P. The hairpin of 3AIT has one S-S bond (green line) between two terminal cysteine residues. The unfolding simulations of the 3AIT hairpin were performed without an S-S bond. It can be argued that the formation of the disulfide bond would proceed at the final stage of the folding process. The main focus of the present study is to examine the earlier folding processes of the hairpins starting from the extended configurations. We believe that the exclusion of the S-S bond in our simulations should not influence the main conclusions of the present study. The native structures of the hairpins contain several native hydrogen bonds (red lines).

All simulations were performed with the CHARMM package (v.27).<sup>20</sup> Initial structures obtained from PDB were placed in a cubic box with the dimension of 37.712 Å (1GB1) or 43.4605 Å (1A2P & 3AIT), filled with TIP3P water molecules to a density of  $\sim 1$  g/cm<sup>3</sup>. After removing

the water molecules whose oxygen atoms were within 2.6 Å of any peptide atom, the initial conformation was relaxed by an adapted basis Newton-Raphson method until satisfying tolerance gradient of 0.0001. Bond lengths were constrained through the SHAKE algorithm.<sup>21</sup> A cutoff distance of 8 Å (13 Å for generating neighbor list) was used for the non-bonded interactions and the long-range electrostatic interactions were attenuated between 8 Å and 12 Å using a switching function. The temperature of the minimized system is raised from 0 K to a specified temperature by increasing 5 degrees every 5 steps. All simulations for analysis were carried out at a temperature of 450 K. The temperature of the system was controlled by use of a weak coupling to an external bath, using the method of Berendsen *et al.*<sup>22</sup> with the coupling constant of 5.0 ps. A time step of 1 fs was used and the trajectories were saved every 1 ps. Independent simulations of 2 ns duration were performed 11 times (1GB1), 10 times (3AIT) and 9 times (1A2P). For independent simulations, starting with the minimized native structure, separate heating and production runs were performed.

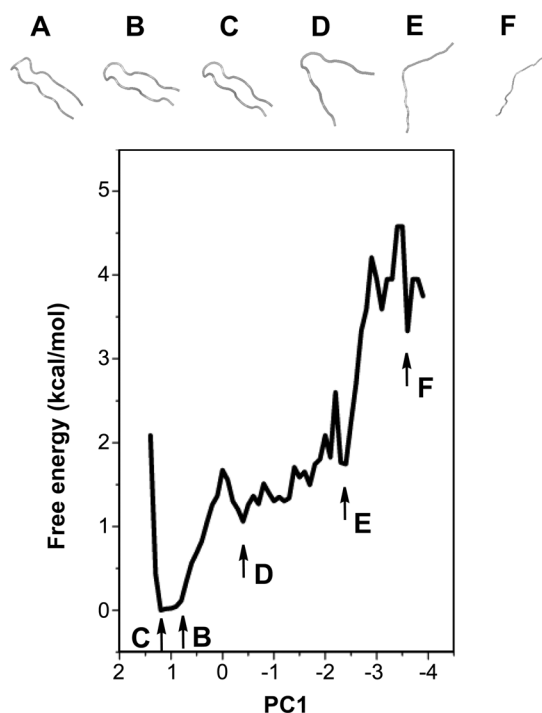
To analyze the trajectories, we calculated several quantities as order parameters: the radius of gyration ( $R_G$ ) of a whole hairpin, the radius of gyration ( $R_{\text{core}}$ ) of the hydrophobic core residues, the number of hydrogen bonds ( $N_{\text{HB}}$ ), the number of native hydrogen bonds ( $N_{\text{HB}}(\text{NAT})$ ), the root mean square displacement (RMSD) compared with the native structure and the averaged interaction energy components (INRE) per residue. Principal component analysis (PCA) was performed using an ED (Essential Dynamics) module in the WHATIF package.<sup>23</sup> The average interaction energies components were calculated for different regions of a hairpin structure: the bend region, the strand region, and the hydrophobic residues. We calculated a "protein-protein" (PP) component of energy by summing the average of the total interaction energies of each residue with the solvent coordinates excluded from the trajectories. The difference between the average energy of the whole system and the protein-protein energy can be defined as "protein-solvent" (PS) energy. The free energy was calculated by  $A_i = -RT \ln(P_i)$  where  $P_i$  is the normalized probability of finding the conformation at the given values of a reaction coordinate such as  $R_G$  or principal component. Some of the details of the analysis tools can be found in our previous paper.<sup>24</sup>

## Results and Discussion

We carried out several independent unfolding simulations of the three hairpins starting from the native structures. At high temperatures ( $> 600$  K), the trajectories of the simulations usually result in a completely unfolded state within very short time, while the folded states with native-like structures are maintained during the simulations at room temperature (300 K). At an intermediate temperature (400–500 K), partial unfolding and refolding of the hairpin are observed in typical trajectories.<sup>18,24</sup> In recent studies, it was shown that *ab initio* fast folding simulations are possible at moderately high temperatures for various peptides and small

proteins.<sup>25,26</sup> Performing folding simulations at low (physiological) temperature usually takes very long time since the system can be trapped in local minima along the rugged energy landscape. At moderately high temperatures, the free energy barriers get smoother and there might exist faster folding pathways leading to the native structures. If the temperature is too high, one may not observe folding events to stable native structures. With the temperatures in the appropriate range, it may be possible to observe reversible folding and search for most of the relevant local minimum structures involved in the final tune-up process of protein folding. In the present unfolding simulations, the simulation temperature was set to 450 K. Such temperature is high enough for initiating unfolding processes while allowing partial refolding processes during the trajectories. Within our simulation time (2 ns), it is not likely to observe complete unfolding/refolding process. However, by performing several ( $\sim 10$ ) independent simulations, one expects to obtain sufficient configurational sampling to generate relevant free energy surfaces for final stages of folding.

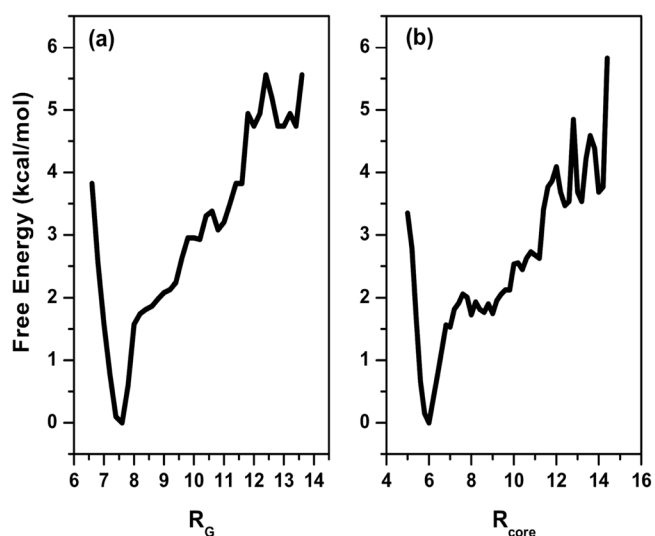
In order to characterize the folding dynamics of  $\beta$ -hairpins, various free energy surfaces were obtained as a function of different structural reaction coordinates. We calculated probability distributions as a function of principal components, describing the essential motions of  $\beta$ -hairpins. These PCs were obtained from 'combined analysis' ED, which was performed using  $C_\alpha$  trajectories of all simulations. From the trajectories of 1GB1, it was found that



**Figure 2.** Free energy of a 1GB1 hairpin as a function of the first principle component (PC1) obtained from essential dynamics (ED) analysis using  $C_\alpha$  trajectories of unfolding simulations at 450 K. Also shown are the representative structures (pictures A through F) of the various local energy minima along the free energy curve. See the text for the details of these structures.

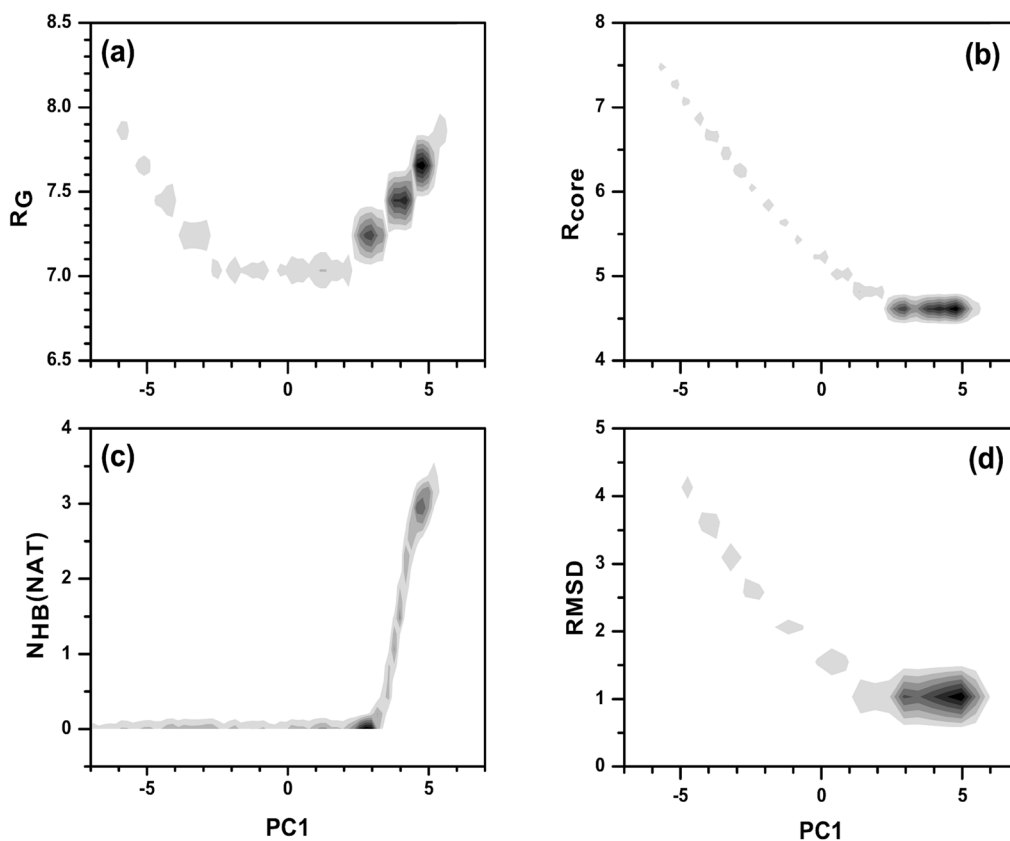
the first (PC1) component shows overall folding motion, while the second component corresponds to hydrogen bond formation in the strand. Figure 2 shows the free energy of 1GB1 as a function of PC1. We found five characteristic structures of the various local energy minima. In Figure 2, we have compared the structures of local minima with the structure of native structure: (A) native  $\beta$ -hairpin extracted from PDB 1GB1, having native hydrogen bonds and a closely packed hydrophobic core; (B) averaged structure on the free energy surface with the same value of PC1 (0.93) with the native structure; (C) averaged folded structure for the global minimum on the free energy surface with the value of 1.2 for PC1; (D) structure for a local minimum at  $-0.4$  of PC1 before crossing a barrier to a folded structure; (E) structure formed after initial collapse of extended structure with a value of  $-2.3$  for PC1; (F) extended structure with the value of  $-3.6$  for PC1. These characteristic structures were taken from averaging all structures at chosen values of PC1. The sequence of structures from (F) to (C) illustrates the folding mechanism of the hairpin. The free energy difference between the folded (structure C) and the unfolded (structure F) is estimated to be 3.3 kcal/mol.

Figure 3 shows the free energy of 1GB1 as a function of  $R_G$  and  $R_{core}$ . We found the folded structure (global minimum) appeared at an  $R_G$  of 7.6 Å and  $R_{core}$  of 6.0 Å, as compared with the values of the native structure ( $R_G \sim 7.79$  Å and  $R_{core} \sim 5.44$  Å). The free energy surface as a function of  $R_{core}$  is more similar to the free energy surface as a



**Figure 3.** Free energy of 1GB1 as a function of (a)  $R_G$  and (b)  $R_{core}$ , calculated from the trajectories of unfolding simulations at 450 K.

function of PC1, compared to the case of  $R_G$ . The free energy difference between the folded and unfolded structures is estimated to be 3.5 kcal/mol in the case of  $R_{core}$ , while the difference is about 4.7 kcal/mol for the surface with  $R_G$ . We also calculated the normalized probability and the free energy as a function of PC's obtained from ED analysis including all the atoms rather than using only the  $C_\alpha$  atoms. The energy difference between the folded and unfolded

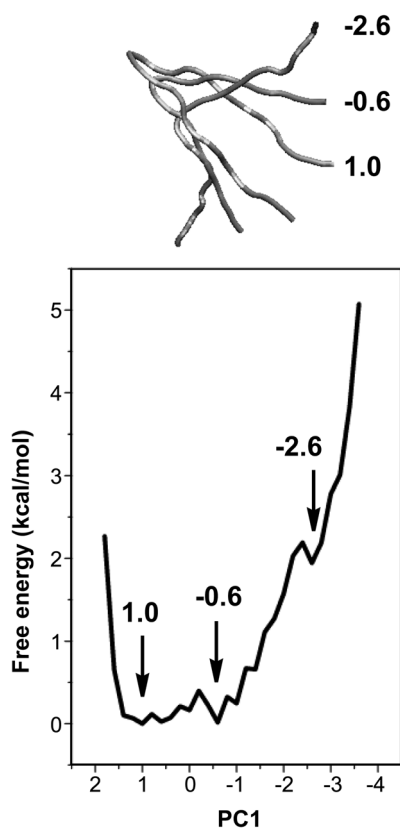


**Figure 4.** Normalized probability surface of 1GB1 as a function of (a) PC1 vs.  $R_G$ , (b) PC1 vs.  $R_{core}$ , (c) PC1 vs.  $N_{HB}(NAT)$ , and (d) PC1 vs. RMSD, calculated from the trajectories of unfolding simulations at 450 K.

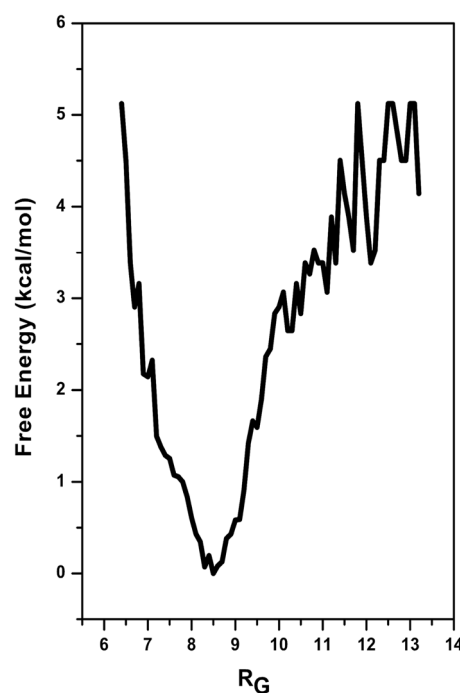
structures on the free energy surface as a function of PC1 is almost the same for the all-atom and  $C_\alpha$ -only cases.

In order to compare different reaction coordinates, we calculated normalized probability distributions as a function of PC1 and one of the other structural quantities (Fig. 4).  $R_{\text{core}}$  and RMSD are found to decrease monotonically as PC1 approaches the global minimum (native structure). It should be noted that  $R_{\text{core}}$  is a better reaction coordinate than  $R_G$ . Figure 4(c) shows that the number of native hydrogen bonds remains virtually zero until PC1 reaches a value near the global minimum. This result is consistent with a folding mechanism for a 1GB1 hairpin, where a collapsed ('molten globule') intermediate structure exists.

The oxidized form of native structure for 3AIT has a disulfide bridge closing the hairpin. Bonvin and van Gunsteren<sup>7</sup> compared the stabilities of the oxidized and reduced (without disulfide bridge) forms of the hairpin. In the unfolding simulations at 400 K, the oxidized native structure with the disulfide bridge was found to be less stable than the reduced form with open ends. The difference in stability was attributed to entropic effects: the reduced form lacking the disulfide bridge is configurationally less restricted than the oxidized form. For our purpose, we used the native structure without the disulfide bridge in order to search for larger configuration space.



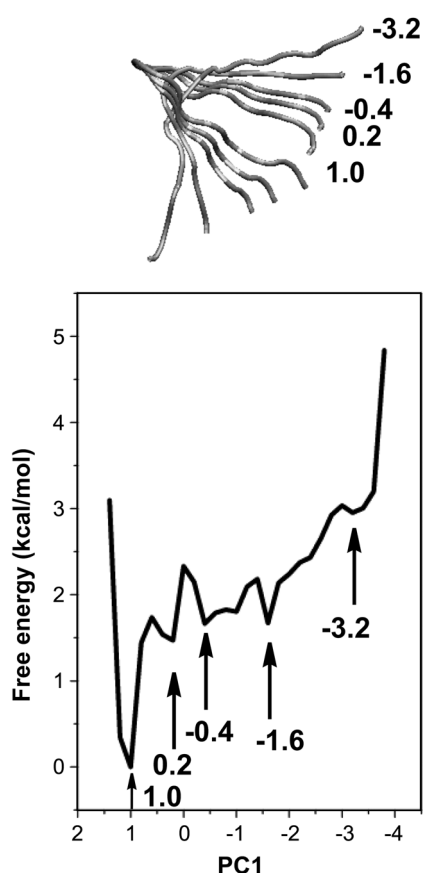
**Figure 5.** Free energy of a 3AIT hairpin as a function of the first principle component (PC1) obtained from essential dynamics (ED) analysis using  $C_\alpha$  trajectories of unfolding simulations at 450 K. Also shown are the representative structures of the various local energy minima along the free energy curve.



**Figure 6.** Free energy of 3AIT as a function of  $R_G$ , calculated from the trajectories of unfolding simulations at 450 K.

Figure 5 shows the free energy of a 3AIT hairpin as a function of PC1, with the averaged structures characteristic of various local energy minima. The structure with the value of 1.0 for PC1 is the native structure in the global minimum. The structure with the value of  $-0.6$  for PC1 is an intermediate, which has similar bend region to the native structure but has no hydrogen bonds between strand regions. The structure with the value of  $-2.6$  for PC1 exhibits an extended strand region with a partially formed bend region. Figure 6 shows the free energy of 3AIT as a function of  $R_G$ . The global minimum has the value of  $8.5 \text{ \AA}$  for  $R_G$ , while the value of  $R_G$  for the native structure from PDB is  $8.86 \text{ \AA}$ . The free energy curve as a function of  $R_G$  has a different shape than the curve of PC1. This suggests that  $R_G$  is not a good reaction coordinate for the folding of 3AIT. Figure 7 shows the free energy of a 1A2P hairpin as a function of PC1, with the averaged structures characteristic of various local energy minima. The structure with the value of 1.0 for PC1 is the native structure in the global minimum.

We calculated the probabilities of native hydrogen bonding for important intermediate structures represented by PC1 from all-atom principle component analysis (Table 1). In the case of 1GB1, the hydrogen bonds #2, 3, 6, and 7 (see Fig. 1) have higher probabilities. For 3AIT, hydrogen bond #3 has the highest probability while 1A2P shows large probabilities for hydrogen bonds #3, 4, 7, and 8. In all three cases, the hydrogen bonds close to the hydrophobic core region have higher probabilities. It can be argued that, for the folding mechanism represented by the reaction coordinate of PC1, the formation of early hydrogen bonding is closely related to the hydrophobic core. The probabilities of hydrogen bonding calculated from the real trajectories are somewhat differ-



**Figure 7.** Free energy of a 1A2P hairpin as a function of the first principle component (PC1) obtained from essential dynamics (ED) analysis using  $C_{\alpha}$  trajectories of unfolding simulations at 450 K. Also shown are the representative structures of the various local energy minima along the free energy curve.

ent from those obtained from PCA analysis. We calculated the relative probabilities of different hydrogen bonding for a fixed number of native hydrogen bonds (Table 2). For 1GB1, H-bonds #3 and 6 have the highest probabilities for  $N_{\text{HB}}(\text{NAT})=2$ , while H-bonds #2, 3, 6, and 7 have comparably large probabilities for  $N_{\text{HB}}(\text{NAT})=4$ . For the initial formation of native hydrogen bonding for 3AIT, H-bond #3 has the highest probability. The probabilities for formation of H-bonds #4, 2, and 5 become larger as  $N_{\text{HB}}(\text{NAT})$  increases. 1A2P shows larger probabilities for H-bonds #3

**Table 1.** Probabilities of native hydrogen bonding for important intermediate structures represented by PC1 from all-atom principle component analysis

N	1GB1	3AIT	1A2P
1	0	21.4	0
2	35.9	17.3	0
3	50.3	46.0	48.1
4	0	26.2	50.0
5	0	7.2	0
6	42.5	0	0
7	18.8		51.0
8			45.3
9			0

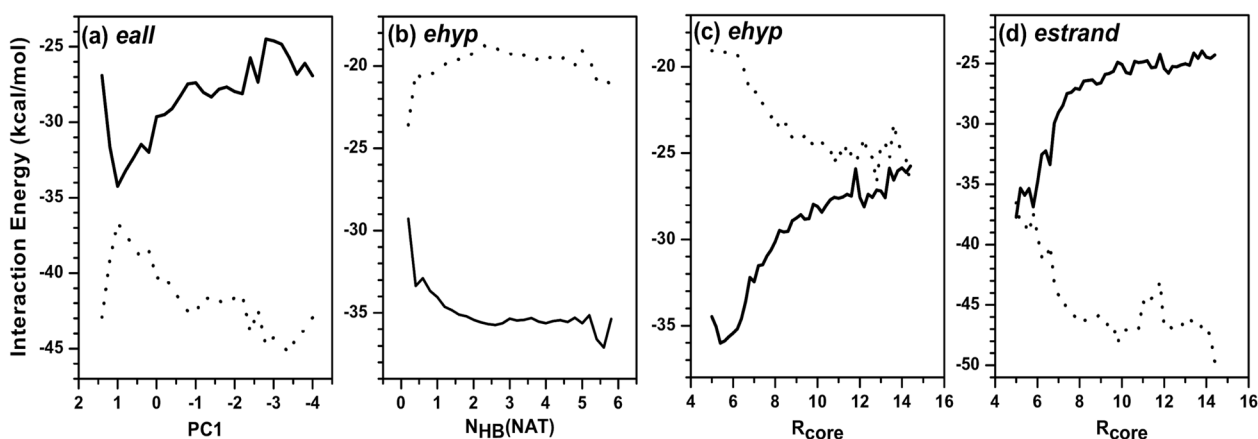
and 7 with  $N_{\text{HB}}(\text{NAT})=2$  and H-bonds #3, 6, and 7 with  $N_{\text{HB}}(\text{NAT})=3$ .

The protein-solvent (PS) energy usually favors the unfolding of a protein, while the native state is favored by the protein-protein (PP) component of the energy.<sup>27</sup> This observation is consistent with the structural features of unfolding: as the protein unfolds, PS interactions occur in preference to the previously favorable PP interactions; the protein becomes more exposed to the solvent, thereby increasing the overall protein-solvent energy. Figure 8 shows the PP and PS components of interaction energies for a 1GB1 hairpin as a function of various reaction coordinates. In Figure 8(a), the total interaction energies (*eall*) for the whole hairpin show that the PP component decreases and the PS component increases as PC1 approaches the folded structure from the unfolded structure. The interaction energies for the hydrophobic region (*ehyp*) show rapid decreases in the PP component (and a corresponding increase in the PS) as the number of natural hydrogen bonds changes from 0 to 2, while subsequent changes as a function of  $N_{\text{HB}}(\text{NAT})$  are relatively small (Fig. 8(b)). As shown in Figure 8(c), the behavior of *ehyp* as a function of  $R_{\text{core}}$  is similar to that of *eall* vs. PC1. The interaction energies for the strand region (*estrand*) show rapid decreases in the PP component as a function of  $R_{\text{core}}$  in the range of  $R_{\text{core}} < 9.5$  Å. There are relatively small changes in *estrand* for  $R_{\text{core}} > 9.5$  Å.

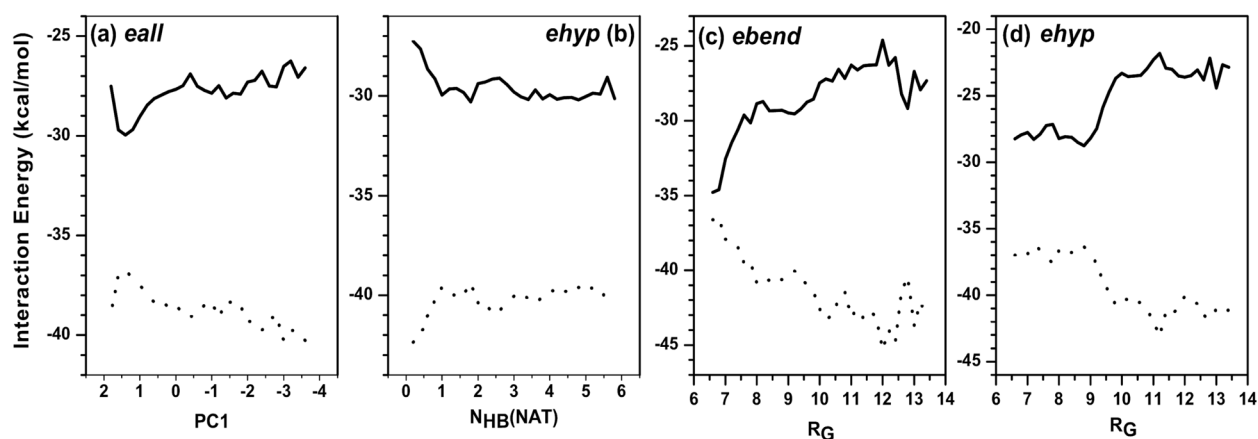
Figure 9 shows the PP and PS components of interaction energies for a 3AIT hairpin as a function of various reaction

**Table 2.** Relative probabilities of different hydrogen bonding for a fixed number of native hydrogen bonds, as given by the number in the parenthesis

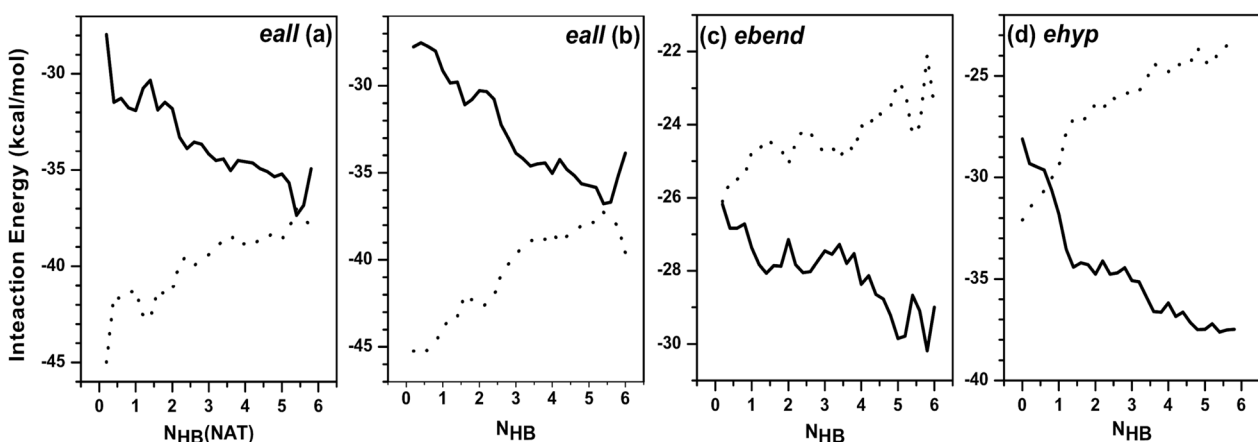
N	1GB1(2)	(3)	(4)	3AIT(1)	(2)	(3)	1A2P(2)	(3)
1	0.00214	0.02863	0.15537	0.01657	0.05921	0.24153	0	0
2	0.28725	0.83206	0.96328	0.186	0.63596	0.8465	0.09573	0.39658
3	0.71383	0.69275	0.93503	0.40884	0.3136	0.50339	0.59145	0.87476
4	0.05573	0.02672	0	0.186	0.79605	0.87359	0.31111	0.32068
5	0.01286	0.08969	0.13277	0.20258	0.19518	0.53047	0.01709	0.03985
6	0.80171	0.91412	0.95198	0	0	0.00451	0	0
7	0.12647	0.41603	0.86158				0.62222	0.71727
8							0.36239	0.64137
9							0	0.00949



**Figure 8.** Comparison of the protein-protein (PP: solid line) and the protein-solvent (PS: dotted line) components of interaction energies for 1GB1 as a function of various reaction coordinates. (a) Total interaction energies (*eall*) for the whole hairpin as a function of PC1; (b) interaction energies for the hydrophobic region (*ehyp*) as a function of  $N_{HB}(NAT)$ ; (c) *ehyp* as a function of  $R_{core}$ ; (d) interaction energies for the strand region (*estrand*) as a function of  $R_{core}$ .



**Figure 9.** Comparison of the protein-protein (PP: solid line) and the protein-solvent (PS: dotted line) components of interaction energies for 3AIT as a function of various reaction coordinates. (a) Total interaction energies (*eall*) for the whole hairpin as a function of PC1; (b) interaction energies for the hydrophobic region (*ehyp*) as a function of  $N_{HB}(NAT)$ ; (c) interaction energies for the bend region (*ebend*) as a function of  $R_G$ ; (d) *ehyp* as a function of  $R_G$ .



**Figure 10.** Comparison of the protein-protein (PP: solid line) and the protein-solvent (PS: dotted line) components of interaction energies for 1A2P as a function of various reaction coordinates. (a) Total interaction energies (*eall*) for the whole hairpin as a function of  $N_{HB}(NAT)$ ; (b) *eall* as a function of  $N_{HB}$ ; (c) interaction energies for the bend region (*ebend*) as a function of  $N_{HB}$ ; (d) interaction energies for the hydrophobic region (*ehyp*) as a function of  $N_{HB}$ .

coordinates. The total interaction energies of the whole hairpin (*eall*), as a function of PC1, and the interaction

energies for the bend region (*ebend*), as a function of  $R_G$ , show monotonic decrease in PP (and a corresponding increase

in PS) as the reaction coordinates change from the unfolded to the folded states (Figs. 9(a) and 9(c)). The interaction energies for the hydrophobic region (*ehyp*), as a function of  $R_G$ , and *ebend*, as a function of  $N_{HB}$ , show relatively large variations in the PP and PS components only for limited ranges of the reaction coordinates:  $N_{HB} = 0 \rightarrow 1$  in Figure 9(c) and  $R_G = 11 \rightarrow 8.5 \text{ \AA}$  in Figure 9(d). Figure 10 shows the PP and PS components of interaction energies for a 1A2P hairpin. The total interaction energies of the whole hairpin (*eall*), as functions of both  $N_{HB}(\text{NAT})$  and  $N_{HB}$ , show a similar monotonic decrease in PP (and a corresponding increase in PS) as the reaction coordinates change from the unfolded to the folded states (Figs. 10(a) and 10(b)). Monotonic changes in the PP and PS components for both the bend (*ebend*) and hydrophobic (*ehyp*) interaction energies as a function of  $N_{HB}$  suggests that the folding of 1A2P does not involve intermediates with partial formation of either bend or strand region only. Instead, the formation of a hydrophobic core between bend and strand regions, stabilized by non-native hydrogen bonds initially, leads to a folding mechanism different from either zip-down or zip-up mechanisms.

### Concluding Remarks

We have studied the folding mechanisms of  $\beta$ -hairpins in the proteins of 1GB1, 3AIT and 1A2P by conducting unfolding simulations at high temperatures. The analysis of trajectories obtained from molecular dynamics simulations in explicit aqueous solution suggests that the positions of the hydrophobic core residues lead to subtle differences in the details of folding dynamics. It was proposed that the positions of the large (hydrophobic) side-chains may be a crucial factor in determining the folding mechanism.<sup>18</sup> The hydrophobic interactions can be also important for designing efficient enzyme catalysis.<sup>28</sup> In the  $\beta$ -hairpin of protein GB1, the hydrophobic residues are located in the middle of the two strands. The stability and the initial folding of the peptide are dominated by the formation of the hydrophobic core. The  $\beta$ -hairpin of tendamistat (3AIT) has its hydrophobic residues clustered near the turn (loop) region, which facilitates the early formation of the turn and subsequent hydrogen bond formation closing the hairpin. For the hairpin of barnase (1A2P), the hydrophobic residues do not seem to form a cluster around one region of the peptide. In this case, the folding cannot be initiated at the particular position of the hairpin by forming a hydrophobic core. Instead, the best way of folding in this case would be the formation of hydrogen bonds assisted by the concomitant side-chain hydrophobic interactions. In other words, the folding process would proceed globally throughout the chain for this hairpin.

The results of the present simulations suggest that the folding mechanisms of the three hairpins may not be totally mutually exclusive. In the process of forming the native structures of hairpins, there exist partially folded intermediate structures. The formation of these intermediate structures is initiated by nonspecific hydrophobic interactions among hydrophobic residues with large side-chains.

Depending on the relative positions of the hydrophobic residues, one may observe distinct intermediate structures, such as a hydrophobic core in a hydrophobic collapse mechanism, or a turn structure in a zip-up mechanism. The partially folded hairpin structures are first stabilized by non-native hydrogen bonds before leading to the native structure by forming native contacts in the final "tune-up" process. It may be more appropriate to propose a new mechanism of  $\beta$ -hairpin folding that is a blend of the hydrogen-bond-centric and the hydrophobic-centric models.<sup>13,15</sup>

**Acknowledgements.** This work was supported by grant R01-2006-000-10418-0 from the Basic Research Program of the Korea Science & Engineering Foundation. This work was also supported by a grant from MarineBio 21 Project funded by Ministry of Maritime Affairs and Fisheries of Korean government.

### References

- Muñoz, V.; Serrano, L. *Curr. Opin. Biotechnol.* **1995**, *6*, 382.
- Williams, S.; Causgrove, T. P.; Gilmanishin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1996**, *35*, 691.
- Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200.
- Blanco, F.; Ramírez-Alvarado, M.; Serrano, L. *Curr. Opin. Struct. Biol.* **1998**, *8*, 107.
- Galzitskaya, O. V.; Higo, J.; Finkelstein, A. V. *Curr. Protein Pept. Sci.* **2002**, *3*, 191.
- Muñoz, V.; Serrano, L. *Curr. Opin. Biotechnol.* **1995**, *6*, 382.
- Bonvin, A. M.; van Gunsteren, W. F. *J. Mol. Biol.* **2000**, *296*, 255.
- Prevost, M.; Ortmans, I. *Proteins Struct. Funct. Genet.* **1997**, *29*, 212.
- Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9062.
- Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9068.
- Ma, B.; Nussinov, R. *J. Mol. Biol.* **2000**, *296*, 1091.
- Zhou, Y.; Linhananta, A. *Proteins* **2002**, *47*, 154.
- Zagrovic, B.; Sorin, E. J.; Pande, V. J. *Mol. Biol.* **2001**, *313*, 151.
- Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345.
- Zhou, Y.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14931.
- Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2544.
- Ma, B.; Nussinov, R. *Protein Sci.* **2003**, *12*, 1882.
- Lee, J.; Jang, S.; Pak, Y.; Shin, S. *Bull. Korean Chem. Soc.* **2003**, *24*, 785.
- Lee, J.; Lee, K.; Shin, S. *J. Phys. Chem. B* **2002**, *106*, 8796.
- Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187.
- Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327.
- Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
- Vriend, G. *J. Mol. Graph.* **1990**, *8*, 52.
- Lee, J.; Shin, S. *Biophys. J.* **2001**, *81*, 2507.
- Jang, S.; Shin, S.; Pak, Y. *J. Am. Chem. Soc.* **2002**, *124*, 4976.
- Jang, S.; Kim, E.; Shin, S.; Pak, Y. *J. Am. Chem. Soc.* **2003**, *125*, 14841.
- Boczko, E. M.; Brooks III, C. L. *Science* **1995**, *269*, 393.
- Kim, M. G.; Kim, M.-S.; Park, H.; Lee, S.; Suh, J. *Bull. Korean Chem. Soc.* **2007**, *28*, 1151.