J|A|C|S ARTICLES

Published on Web 11/18/2009

Common Folding Mechanism of a β-Hairpin Peptide via Non-native Turn Formation Revealed by Unbiased Molecular Dynamics Simulations

Lipi Thukral,[†] Jeremy C. Smith,*[,†,‡] and Isabella Daidone*[,§]

Interdisciplinary Center for Scientific Computing, University of Heidelberg, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany, Center for Molecular Biophysics, University of Tennesse/Oak Ridge National Laboratory, One Bethel Valley Road, Oak Ridge, Tennessee 37831, and Dipartimento di Chimica Ingegneria Chimica e Materiali, University of L'Aquila, Via Vetoio (Coppito 1), 67010 Coppito (AQ), Italy

Received July 30, 2009; E-mail: daidone@caspur.it; smithjc@ornl.gov

Abstract: The folding of a 15-residue β-hairpin peptide (Peptide 1) is characterized using multiple unbiased, atomistic molecular dynamics (MD) simulations. Fifteen independent MD trajectories, each 2.5 μs-long for a total of 37.5 μs, are performed of the peptide in explicit solvent, at room temperature, and without the use of enhanced sampling techniques. The computed folding time of 1−1.5 μs obtained from the simulations is in good agreement with experiment [Xu, Y.; et al. *J. Am. Chem. Soc.* **2003**, *125*, 15388−15394]. A common folding mechanism is observed, in which the turn is always found to be the major determinant in initiating the folding process, followed by cooperative formation of the interstrand hydrogen bonds and the side-chain packing. Furthermore, direct transition to the folded state from fully unstructured conformations does not take place. Instead, the peptide is always observed to form partially structured conformations involving a non-native (ESYI) turn from which the native (NPDG) turn forms, triggering the folding to the β-hairpin.

Introduction

Peptides forming α-helices and β-sheets provide useful model systems for determining the principles governing initial events in protein folding. β-hairpins are the smallest β-sheet units, typically 12−16 residues in length, and are composed of two short antiparallel β-strands connected by a turn. Although they are relatively small, hairpin peptides may exhibit properties that are representative of globular proteins, e.g. they may contain a hydrophobic core and/or exhibit a cooperative thermal folding−unfolding transition.[1] Furthermore, there is evidence suggesting that β-hairpins could act as nucleation sites in early stages of protein folding.[2] Thus, understanding how β-hairpins fold should shed light on critical initial processes in protein folding.

The structure of the first peptide found to be a β-hairpin in aqueous solution, the C-terminal hairpin (residues 41−56) of protein GB1, was solved using nuclear magnetic resonance in the early 1990s.[3] The folding kinetics of this peptide, determined using a fluorescence T-jump technique, was shown to follow a two-state behavior with a folding time constant of 6 μs at 297 K.[1] These results triggered a number of experimental and computational studies on the GB1 hairpin.[4−10] In contrast, examples of detailed examination of other β-hairpins are comparatively few,[11−16] and β-hairpin folding mechanisms hitherto proposed are based essentially on the studies conducted on GB1.

Two major β-hairpin folding mechanisms have been proposed. In the "hydrophobic collapse" mechanism the first step corresponds to the formation of native hydrophobic clusters, followed by the interstrand hydrogen-bond formation.[4] In contrast, in the second, "zipper" mechanism, the folding initiates from the turn, followed by the formation of interstrand hydrogen bonds and, finally, by the native side-chain packing.[1] In addition, a reptation mechanism, in which the formation of the native

[†] University of Heidelberg.
[‡] University of Tennesse/Oak Ridge National Laboratory.
[§] University of L'Aquila.

(1) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196–199.
(2) Du, D.; Zhu, Y.; Huang, C.-Y.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15915–15920.
(3) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584–590.
(4) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062–9067.
(5) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068–9073.
(6) Ma, B.; Nussinov, R. *J. Mol. Biol.* **2000**, *296*, 1091–1104.
(7) Lee, J.; Shin, S. *Biophys. J.* **2001**, *81*, 2507–2516.
(8) Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12129–12134.
(9) Wei, G.; Mousseau, N.; Derreumaux, P. *Proteins* **2004**, *56*, 464–474.
(10) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *121*, 1080–1090.
(11) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5578–5583.
(12) Santiveri, C. M.; Santoro, J.; Rico, M.; Jiménez, M. A. *J. Am. Chem. Soc.* **2002**, *124*, 14903–14909.
(13) Xu, Y.; Oyola, R.; Gai, F. *J. Am. Chem. Soc.* **2003**, *125*, 15388–15394.
(14) Dyer, R. B.; Maness, S. J.; Peterson, E. S.; Franzen, S.; Fesinmeyer, R. M.; Andersen, N. H. *Biochemistry* **2004**, *43*, 11560–11566.
(15) Olsen, K. A.; Fesinmeyer, R. M.; Stewart, J. M.; Andersen, N. H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15483–15487.
(16) Muñoz, V.; Ghirlando, R.; Blanco, F. J.; Jas, G. S.; Hofrichter, J.; Eaton, W. A. *Biochemistry* **2006**, *45*, 7023–7035.

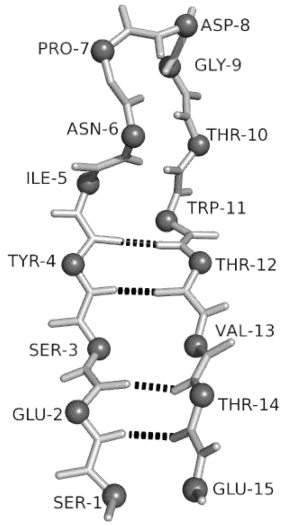

Figure 1. Backbone snapshot of Peptide 1 extracted at the end of the simulation started from the NMR structure. The C_α atoms are shown as spheres for clarity. Four native interstrand hydrogen bonds are shown with dotted lines.

β-hairpin occurs through significant populations of misregistered hairpins, has been identified as a specific folding pathway.[9]

Recently, a 15-residue designed β-hairpin-forming peptide, termed Peptide 1 (SESYINPDGTWTVTE), has been receiving increasing attention (Figure 1). The structure of Peptide 1, solved in aqueous solution also using NMR,[17] is a 3:5 hairpin[18] with a type I + G1 β-bulge turn[19] possessing a loosely packed hydrophobic core. Although Peptide 1 is a designed peptide, the turn sequence, NPDG, is statistically the most abundant type I turn in proteins,[20] accentuating the relevance of the study of its folding mechanism for natural proteins. The folding time of Peptide 1, determined by means of a T-jump technique in combination with IR spectroscopy,[13] has been found to be $\sim$0.8 μs at $T = 300$ K, making it the fastest folding β-hairpin found to date.

On the basis of the above kinetic experiment,[13] a folding mechanism for Peptide 1 was proposed in which the hydrophobic native cluster forms first, followed by the formation of the interstrand hydrogen bonds.[13] However, an alternative mechanism has been proposed from a thermodynamic analysis of the thermal dependence of H1 NMR chemical shifts,[17] in which the first step is the formation of the turn, followed by the simultaneous formation of hydrogen bonding between the strands and the side-chain−side-chain interactions, although it was not possible to rule out that these form sequentially.[17]

The present work reports on a kinetic study of the folding of Peptide 1 by means of all-atom molecular dynamics (MD) simulation in explicit solvent at room temperature. Fifteen MD trajectories, each 2.5 μs long, were generated starting from different unfolded structures, resulting in a combined simulation time of 37.5 μs. Multiple spontaneous folding events to the β-hairpin NMR structure are observed, leading to a mean folding time in very good agreement with the experiments and allowing for a folding scheme to be determined with statistical reliability.

Particular attention is given here to the role of the turn,[21] addressing the question as to whether it actively promotes hairpin folding or, instead, is passively driven by other factors such as the interstrand hydrogen bonding and side-chain packing.[4,5] The results resolve the existing ambiguity regarding the β-hairpin folding mechanism in favor of the turn-driven process. To our knowledge, this is the first time a robust folding kinetic experiment has been conducted *in silico*, i.e. by performing multiple independent MD simulations starting from the unfolded state, each covering a time (2.5 μs) which is several times the experimentally derived folding time (0.8 μs), with a full atomistic description of the peptide and the solvent.

Methods

Molecular Dynamics Simulations. A series of fifteen 2.5 μs-long atomistic MD simulations of Peptide 1 in explicit solvent was performed. Fifteen starting structures representing the unfolded state were extracted randomly from a simulation of 50 ns that was started from a fully extended conformation of the peptide. The MD simulations were performed with the program GROMACS,[22] and the OPLS-AA all-atom force field[23] was used for the peptide. The water was modeled using the TIP4P representation.[24] Each of the 15 starting conformations was placed in a dodecahedral water box (volume = 48 nm^3) large enough to contain the peptide and at least 1.0 nm of solvent on all sides. Each simulation box contained 6647 atoms. Periodic boundary conditions were used, and the long-range electrostatic interactions were treated with the particle mesh Ewald method[25] using a grid spacing of 0.12 nm combined with a fourth-order B-spline interpolation to compute the potential and forces in-between grid points. The real space cutoff distance was set to 0.9 nm and the van der Waals cutoff to 1.4 nm. The bond lengths were fixed,[26] and a time step of 2 fs for numerical integration of the equations of motion was used. Coordinates were saved every 0.2 ps. Simulations were performed in the NVT ensemble with isokinetic temperature coupling[27] keeping the temperature constant at 300 K. Three positive counterions (Na^+) were added, replacing three water molecules so as to produce a neutral simulation box.

All the starting structures were subjected to a two-step minimization protocol using the steepest descent method. The first minimization was performed with the coordinates of the peptide held fixed, allowing only the water and the ions to move, and the second was performed on the atoms of both the peptide and the solvent molecules. The temperature of the system was then increased from 50 to 300 K in 500 ps MD before the 2.5 μs production simulation was started.

To investigate the effect of the force field on the stability of the native NMR structure, we also performed two 100 ns simulations

(17) Santiveri, C. M.; Rico, M.; Jiménez, M. A. *J Biomol NMR* **2001**, *19*, 331–345.
(18) A hairpin is defined here using the classification by Thornton and colleagues (refs 58, 59). The ratio is assigned, depending on whether the end residues of the β-turn are linked by a single or double hydrogen bond. If both hydrogen bonds are formed, then $X = Y$ forming 2:2, 3:3, etc. loops, where X is the number of residues in the loop. If just one hydrogen bond is formed, then $Y = X + 2$. In the case of Peptide 1, $X = 3$ and $Y = 3 + 2 = 5$, i.e., a single hydrogen bond is formed.
(19) A β-bulge is defined as a region between two consecutive β-type hydrogen bonds which includes two residues on one strand and a single residue on the other strand. The G1 Type β-bulge is a common type of β-bulge (ref 60).
(20) Wilmot, C. M.; Thornton, J. M. *J. Mol. Biol.* **1988**, *203*, 221–232.
(21) Marcelino, A. M. C.; Gierasch, L. M. *Biopolymers* **2008**, *89*, 380–391.
(22) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
(23) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926–935.
(25) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(26) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
(27) Brown, D.; Clarke, J. H. R. *Mol. Phys.* **1984**, *51*, 1243–1252.

with the GROMOS96 53a6[28] and OPLS-AA/A[23] force fields. Given that with the OPLS force field the NMR structure was found to be more stable (i.e. it was maintained for a higher proportion of time in the 100 ns simulation) than with GROMOS96, the OPLS force field was used in all further simulations.

Analysis of Trajectories. The trajectories were analyzed using a number of order parameters that capture principal aspects of the folding process of Peptide 1. A robust parameter for identifying conformational transitions is the '*R*' parameter,[29] calculated as follows:

$$R = \sum_{i=1}^{5} \frac{R_i^N}{R_i}$$

where R_i^N is the *i*th interstrand $C_\alpha - C_\alpha$ distance in the native NMR structure and R_i is the same distance in the MD. The five interstrand $C_\alpha - C_\alpha$ pairs in the Peptide 1 hairpin are the following: N6-T10, I5-W11, Y4-T12, S3-V13, and E2-T14. A value of $R \approx 5$ indicates formation of the native β-hairpin. In particular, a cutoff value of $R \geq 4.8$ was used to define folded conformations (see below for this choice).

The N6-P7-D8-G9 and E2-S3-Y4-I5 turns are considered here to be formed when the distance between the C_α atoms of the two end residues is ≤0.7 nm, as usually assumed in previous work.[30] A hydrogen bond is considered to exist when both the donor−acceptor hydrogen angle is <30° and the donor−acceptor distance is <0.35 nm, again as usually assumed in previous work.[31] The middle and end interstrand hydrogen bonds in the hairpin form between residues Y4 and T12 and residues E2 and T14, respectively.

Side chain packing is quantified via the fraction of native side chain contacts, ρ (a contact between two side chains is considered to be formed when the minimum distance between the atoms belonging to the side chains is ≤0.55 nm). A cut off value of $\rho \geq 0.7$ was used to define conformations with a native side chain arrangement (see below for a discussion of the choice of the cutoff). For all the analyses conformations saved every 10 ps were used.

Given a system in thermodynamic equilibrium, the change in free energy on going from a reference state, *ref*, of the system to a generic state, *i*, (e.g. from unfolded to folded) at constant temperature and constant volume was evaluated as

$$\Delta A_{ref \to i} = -RT \ln \frac{p_i}{p_{ref}} \quad (1)$$

where *R* is the ideal gas constant, *T* is the temperature, and p_i and p_{ref} are the probabilities of finding the system in state *i* and state *ref*, respectively. We describe the free energy surface as a function of two order parameters, namely the fraction of native contacts, ρ, and the *R* parameter. All the structures were projected onto the ρ−*R* plane. A grid 20 × 20 was used to divide this plane into 400 cells, and for every cell the number of points was counted and the relative probability calculated, allowing $\Delta A_{ref \to i}$ to be evaluated. The reference state was chosen to be the grid cell with the highest probability, which corresponds to the folded β-hairpin ensemble.

The cutoff values for defining folded conformations were chosen based on the ρ-R free energy landscape: structures populating the folded basin (with free energy values within 5 kJ/mol from the global minimum) have $\rho \geq 0.7$ and $R \geq 4.8$.

To evaluate the probability density of a given structural feature (e.g. the ESYI or NPDG turns, see Results) in the unfolded state as a function of the bidimensional *R*−ρ plane, the following procedure was used. For every position (grid cell) of the plane, the number of structures possessing the given structural feature was taken and divided by the total number of unfolded structures (unfolded structures are those for which $R < 4.8$ and $\rho < 0.7$). By integrating over such distributions, the total fraction of ESYI or NPDG turn structures in the unfolded state was determined.

The statistical error on different properties evaluated from the simulations, such as the mean folding time or the fraction of ESYI or NPDG turn structures in the unfolded state, was estimated through the standard error of their mean, σ_{mean}, calculated over *n* subsets of the trajectories:

$$\sigma_{mean} = \frac{\sigma}{\sqrt{n}} \quad (2)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (\bar{a}_i - \bar{a})^2}{n-1}} \quad (3)$$

$$\bar{a} = \frac{\sum_{i=1}^{n} \bar{a}_i}{n} \quad (4)$$

where $\bar{a}_i$ is the mean of the given parameter evaluated in the *i*th subset, $\bar{a}$ and σ are the mean value over the *n* samples and the sample standard deviation, respectively. In the present case we used three independent subsets of the trajectories ($n = 3$, i.e. three groups each made of six trajectories) which was found to be a good compromise between the statistics within each subset and the sample size, *n*. Note that assuming, as usual, a normal distribution of the mean value, $\bar{a}$, the expected value of *a* is with 95% confidence inside the $\bar{a} \pm 2\sigma_{mean}$ interval.

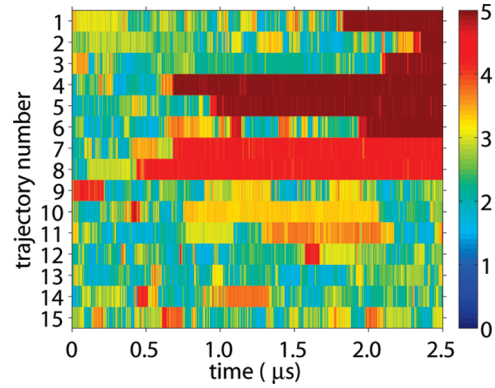

Figure 2. Time evolution of the *R* parameter (see Methods) along the 15 trajectories. The peptide folds into the native NMR structure (dark red, $R \approx 5$) in trajectories 1−6, 10, and 12 (the hairpin is mostly stable in trajectories 1−6). In trajectories 7−9 the peptide folds into a β-hairpin with a shift in the interstrand registry (light red, $R \approx 4$). In the remaining trajectories (11,13−15) transitions into fully folded hairpins are not observed.

Results and Discussion

To determine the kinetics and the folding mechanism of Peptide 1, we performed 15 independent MD simulations in explicit aqueous solvent, each 2.5 μs long, starting from different unfolded structures. The time evolution of the conformational transitions is reported in Figure 2 using the *R* parameter (see Methods) as the conformational order parameter. The peptide folds into the native NMR structure (dark red, $R \approx 5$) in trajectories 1−6, 10, and 12 (the hairpin is mostly stable in

(28) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
(29) Yang, S.; Onuchic, J. N.; García, A. E.; Levine, H. *J. Mol. Biol.* **2007**, *372*, 756–763.
(30) Venkatachalam, C. M. *Biopolymers* **1968**, *6*, 1425–1436.
(31) Daidone, I.; Ulmschneider, M. B.; Nola, A. D.; Amadei, A.; Smith, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15230–15235.

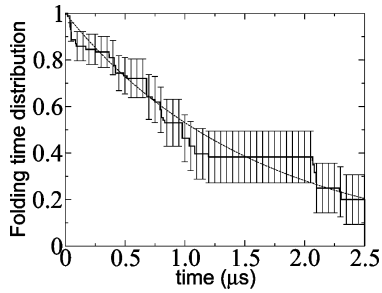

Figure 3. Distribution of the residence times in the unfolded state for the Peptide 1 (solid line). The R parameter is used to define folded ($R \geq 4.8$) and unfolded structures ($R < 4.8$). All 15 trajectories are used, and the error bars shown represent one standard deviation over the 15 trajectories. Given the time resolution of 20 ns of the experimental setup, residence times shorter than 20 ns were not included in the distribution. The dotted line shows a monoexponential fit to the data. The time constant of the exponential fit is taken as the average folding time.

trajectories 1−6). In trajectories 7−9 the peptide folds into a β-hairpin with a shift in the interstrand registry (light red, $R \approx 4$). In the remaining trajectories (11,13−15) transitions into fully folded hairpins are not observed.

In what follows, first the kinetics of the folding process are quantified, and the folding time is compared with experimental results, and then the β-hairpin folding mechanism is presented.

Folding Kinetics. The mean folding time is defined here as the mean residence time in the unfolded state (for folded structures $R \geq 4.8$, for unfolded structures $R < 4.8$ - see Methods). The distribution of the residence times was calculated using all 15 trajectories (Figure 3), and fitted by a monoexponential function (dotted line), assuming a two-state model as was also assumed in the corresponding experiments.[13] The mean folding time, i.e., the mean residence time in the unfolded state, obtained from the fitting function, is 1.57 ± 0.15 μs (for the estimate of the error see Methods). The sensitivity of the folding time to changing the cutoff value used to define folded and unfolded structures was examined by recalculating the time for $R \geq 4.7$, $R \geq 4.6$ and $R \geq 4.5$, resulting in values of 1.45 ± 0.14 μs, 1.31 ± 0.15 μs and 1.22 ± 0.14 μs, respectively. The calculated folding time of $\sim$1−1.5 μs is in good agreement with the experimental value of $\sim$0.8 μs.

Previous computational simulation studies on folding kinetics of other hairpins employed biasing techniques such as transition path sampling,[10,32] enhanced sampling methods such as replica exchange MD simulations,[29,33] coarse-grained models[34,35] or made use of implicit solvent models.[36] To our knowledge, the present work is the first time an unbiased folding kinetic experiment has been conducted *in silico*, i.e. by performing multiple independent fully atomistic MD simulations starting from the unfolded state, each covering a time (of 2.5 μs) which is several times the experimentally derived folding time.

Mechanism of β-Hairpin Folding. The main interactions driving β-hairpin folding identified in previous work are formation of the turn,[1,37−39] formation of the interstrand hydrogen bonds[40] and hydrophobic collapse of the side-chains.[4,5,41] Most of the studies so far documented have been conducted on the GB1 hairpin and some of its mutants, such as the family of the Trp zippers.[1,4−7,11,38,42−44] However, a consensus on the mechanism and the forces initiating the folding is still lacking. In some studies, the role of the turn sequence has been suggested to be crucial.[37] From a structural point of view, mutational studies using NMR demonstrated that the turn sequence is more important in determining the type of β-hairpin than the nature of the side chains.[37] However, the kinetic role of the turn is still unclear.[21] Although it has been suggested that the turn actively promotes the β-hairpin folding,[37] it has also been suggested that the turn passively forms, driven by other forces such as interstrand hydrogen bonds and side-chain interactions.[4,5]

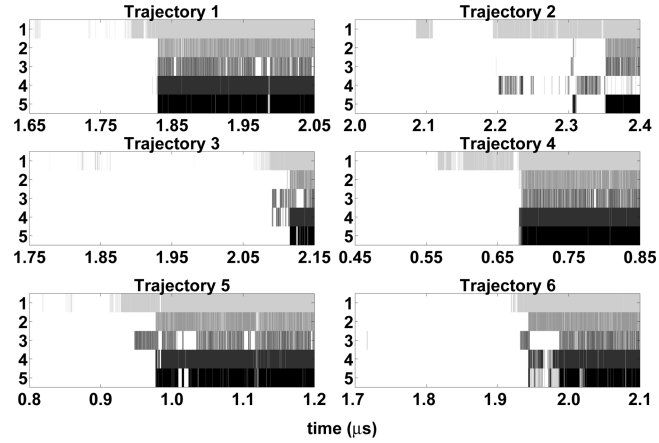

Figure 4. Order parameters (rows) during the folding windows of trajectories 1−6. The rows represent the existence of the following: (1) NPDG turn, (2) middle hydrogen bonds (Y4-T12), (3) end hydrogen bonds (E2-T14), (4) native contacts within the peptide, (5) fully folded hairpin. See Methods for the details on the definition of the different parameters.

To analyze the interplay between the formation of the turn (N6-P7-D8-G9), hydrogen bonds and side-chain packing in Peptide 1, the folding transitions of six trajectories (1−6) in which folding to stable β-hairpin conformations occurred were analyzed in detail. The time evolution of the formation of the turn, of the interstrand hydrogen bonds and of the native side-chain interactions were calculated and are shown in Figure 4. The time evolution of the formation of the full β-hairpin is also reported as a reference for identifying the folding transitions.

Figure 4 shows that the turn always forms first. The turn formation is followed by the almost simultaneous, cooperative formation of the interstrand hydrogen bonds and the side-chain packing. This occurs in all six trajectories and demonstrates that the formation of the turn drives the folding process.

A detailed examination revealed that the hydrogen bonds and side-chain packing form within 100−1000 ps of each other and

(32) Bolhuis, P. G. *Biophys. J.* **2005**, *88*, 50–61.
(33) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6801–6806.
(34) Zhou, Y.; Karplus, M. *Nature* **1999**, *401*, 400–403.
(35) Allen, L. R.; Krivov, S. V.; Paci, E. *PLoS Comput. Biol.* **2009**, *5*, e1000428.
(36) Zagrovic, B.; Sorin, E. J.; Pande, V. *J. Mol. Biol.* **2001**, *313*, 151–169.
(37) de Alba, E.; Rico, M.; Jiménez, M. A. *Protein Sci.* **1999**, *8*, 2234–2244.
(38) Muñoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5872–5879.
(39) Ramírez-Alvarado, M.; Blanco, F. J.; Niemann, H.; Serrano, L. *J. Mol. Biol.* **1997**, *273*, 898–912.
(40) Constantine, K.; Mueller, L.; Andersen, N. H.; Tong, C. W.; Friedrichs, M.; Bruccoleri, R. *J. Am. Chem. Soc.* **1995**, *117*, 1084110854.
(41) Espinosa, J. F.; Gellman, S. H. *Angew. Chem., Int. Ed.* **2000**, *39*, 2330–2333.
(42) Kobayashi, N.; Honda, S.; Yoshii, H.; Munekata, E. *Biochemistry* **2000**, *39*, 6564–6571.
(43) Roccatano, D.; Amadei, A.; Nola, A. D.; Berendsen, H. J. *Protein Sci.* **1999**, *8*, 2130–2143.
(44) Bonomi, M.; Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Am. Chem. Soc.* **2008**, *130*, 13938–13944.

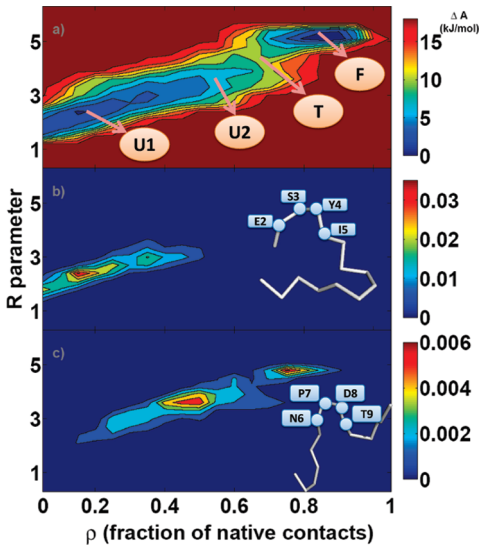

Figure 5. (a) Contour maps of the free energy change, ΔA, as a function of the position in the $\rho-R$ plane (fraction of native contacts within the peptide - R parameter). ΔA is calculated with respect to the state with the highest probability, i.e., the β-hairpin structures. The free energy values are given in kJ/mol. (b and c) Probability distributions of unfolded structures possessing the non-native E2-S3-Y4-I5 turn (b) and the native N6-P7-D8-G9 turn (c) along the $\rho-$R plane (see Methods for details).

that in four out of six trajectories the hydrogen bonds form first, whereas in the remaining two they form after the side-chain packing (data not shown). Hence, the folding is initiated at the turn and is highly cooperative.

How does the polypeptide chain find the NPDG turn during its search through configurational space? To address this question, the free energy landscape of the peptide was constructed as a function of two representative reaction coordinates, namely the R parameter and the fraction of native contacts, ρ (see Figure 5). Two main basins occur, one corresponding to the fully folded state, F ($R \geq 4.8$ and $\rho \geq 0.7$) and the other to the unfolded state ($R < 4.8$ and $\rho < 0.7$), consistent at low resolution with a two-state model for the folding kinetics. The minimum in the unfolded state, U1, is ~3 kJ/mol higher in free energy than F. In the unfolded state a broad plateau, U2, is present at a free energy value of ~6−9 kJ/mol. A transition state, T, can be clearly identified with a free energy value of ~12 kJ/mol.

To analyze the structural features in the unfolded ensemble, we performed a rmsd (root-mean-square deviation)-based clustering analysis[45] of the unfolded configurations ($R < 4.8$ and $\rho < 0.7$) from the 15 trajectories. In the highest populated cluster (30.9 ± 6.1% of the unfolded conformations) partial structuring is found, involving the formation in the N-terminal segment of a turn (E2-S3-Y4-I5), different from the native N6-P7-D8-G9. The E2-S3-Y4-I5 structures populate mostly the U1 minimum (see Figure 5b). In contrast, only 7.9 ± 2.7% of the unfolded structures (the second highest populated cluster) possess the native N6-P7-D8-G9 turn, and these populate the U2 plateau (see Figure 5c) (for the estimate of the errors see Methods). The remainder of the unfolded structures do not possess persistent structural features, i.e. each of the other clusters is populated by less than 5%.

Analysis of the conformational transitions occurring along trajectories 1−6 reveals a common folding mechanism. Before

(45) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236–240.

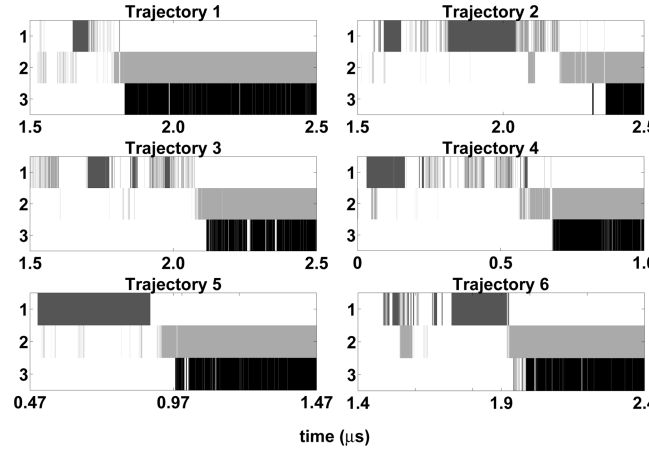

Figure 6. Order parameters (rows) during the folding windows of trajectories 1−6. The rows represent the existence of the following: (1) E2-S3-Y4-I5 turn, (2) N6-P7-D8-G9 turn, (3) fully folded hairpin. See Methods for details on the definition of the different parameters.

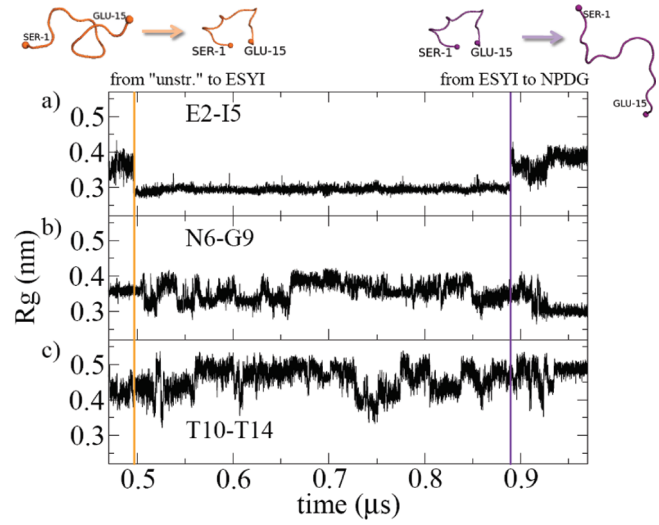

Figure 7. Time evolution of the main chain R_g of residues E2-S3-Y4-I5 (a), residues N6-P7-D8-G9 (b), and residues T10-W11-T12-V13-T14 (c) along a representative trajectory (trajectory 5). With the vertical lines are indicated the transition from an "unstructured" ensemble to the ESYI-conformation and from the ESYI-conformation to the NPDG-conformation.

forming the native NPDG turn, which then triggers the formation of the fully folded hairpin, the peptide is always found to populate structures with the non-native ESYI turn (Figure 6). On going from an "unstructured" ensemble to the ESYI-conformation, compaction of the N-terminal part of the chain is observed, as indicated by the decrease of the radius of gyration (R_g) of the main chain in this region (Figure 7a), while the rest of the chain fluctuates between more extended and more compact conformations (see Figure 7b and 7c). Subsequently, on going from the ESYI- to the NPDG-conformation, the ESYI turn becomes disrupted, and an extension of the chain is observed, as indicated by the high R_g in all the regions of the chain (Figure 7a, b, and c), followed by the formation of the NPDG turn concomitant with a decrease of the R_g in this region (Figure 7b). On the free energy landscape (Figure 5) this step corresponds to a transition from the U1 minimum to the U2 plateau, where the peptide forms the native NPDG turn before passing through the transition state, T. Finally, at T the NPDG

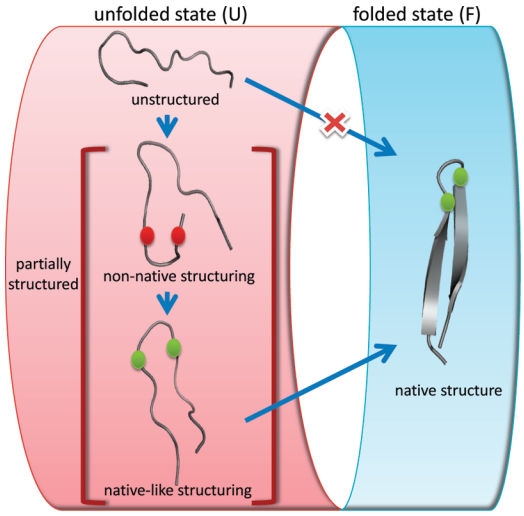

Figure 8. Schematic representation of the Peptide 1 folding mechanism. The folded state is not reached directly from fully unstructured conformations. Instead, it is formed via the formation of obligatory non-native contacts within the peptide (ESYI turn) followed by formation of partial native contacts (NPDG turn) from which the peptide folds to the folded β-hairpin.

turn partially opens to allow for a rearrangement of the side-chains and the backbone, and makes its way to the fully folded state, F.

A clear mechanism emerges from the above results. In this mechanism, the direct transition to the folded state from fully unstructured conformations does not take place. Instead, the hairpin is always observed to form by obligatory steps from partially structured conformations to the final folded structure, as shown in Figure 8. First, a non-native turn is formed, followed by the formation of the native turn, and finally, the folded hairpin is reached through the cooperative effect of the interstrand hydrogen bonding and side-chain interactions.

Conclusions

Characterization of the conformational transitions that trigger the folding of peptides, such as β-hairpins, at atomic resolution is of fundamental importance for the understanding of protein folding (and misfolding) processes. Although many reports exist on peptides simulated at atomic resolution on the nanosecond time scale,[6,7,43,46−48] simulations on the microsecond time scale are few.[31,49−53] The aim of the present microsecond MD simulations was to reveal a reliable and robust mechanism of β-hairpin folding by examining the kinetics *in silico*. This was accomplished by performing multiple unbiased all-atom MD simulations of a model peptide, Peptide 1, starting from unfolded conformations, with each trajectory covering a time frame at least 3 times longer than the expected folding time determined by experiment.[13] The simulations, performed in aqueous solution and at room temperature to reproduce the experimental conditions, reached a total simulation time of 37.5 μs, and several folding events to the native NMR structure were observed.

The computed folding time of 1−1.5 μs obtained from the simulations is similar to the experimental folding time of ∼0.8 μs. The extensive sampling achieved allowed dissection of the folding process, and a single mechanism was found. In this, the NPDG turn of the hairpin is the major determinant driving the folding. After its formation, the NPDG turn provides the opportunity for the interstrand hydrogen bonds and the native side-chain packing to form in a cooperative manner. These results resolve the existing ambiguity regarding the peptide folding mechanism.

Furthermore, the analyses of the simulations reveal that the transition to the folded state never occurs directly from fully unstructured conformations. Instead, the polypeptide chain follows a stepwise pathway through partially structured unfolded conformations: a non-native turn at the N-terminal segment of the peptide chain, i.e., the ESYI turn, is initially and obligatorily formed, followed by the formation of the native NPDG turn which, finally, triggers the transition to the folded hairpin.

Experiments on the collapsed denatured state of several proteins under near-native conditions have shown the presence of preorganized structures.[54−56] These structures have potentially dramatic consequences for the thermodynamics and kinetics of folding.[57] The reduction of the accessible conformational space is expected to speed up folding relative to the case in which unfolded polypeptide chains are structureless random coils. The results presented here reveal structuring in the unfolded state of the peptide, that is characterized by the sequential, reptation-like formation of a non-native turn and the native turn, prior to folding. It remains to be seen whether this represents a frequent mechanism for the folding of β-sheets.

The ability of MD simulation to reproduce folding processes in an unbiased way starting from random coils has been long awaited. A challenge now is to examine whether the type of process here can be generalized to other hairpin sequences and to the folding of larger peptides and proteins.

Acknowledgment. We acknowledge the Deutsche Forschungsgemeinschaft (DFG) for financial support under Grant SM 63/12-1 and the United States Department of Energy for a Laboratory-Directed Research and Development Grant to J.C.S. at ORNL. We thank Hannes Neuweiler and Sören Doose for helpful discussions. We acknowledge the NSF Teragrid for computational resources. We are also thankful to M. Angeles Jiménez for kindly providing us the NMR structure of Peptide 1.

JA9064365

(46) Wu, X.; Brooks, B. R. *Biophys. J.* **2004**, *86*, 1946–1958.
(47) Liang, C.; Derreumaux, P.; Wei, G. *Biophys. J.* **2007**, *93*, 3353–3362.
(48) Luo, Z.; Ding, J.; Zhou, Y. *J. Chem. Phys.* **2008**, *128*, 225103.
(49) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740–744.
(50) Daidone, I.; Simona, F.; Roccatano, D.; Broglia, R. A.; Tiana, G.; Colombo, G.; Nola, A. D. *Proteins* **2004**, *57*, 198–204.
(51) Daidone, I.; D'Abramo, M.; Nola, A. D.; Amadei, A. *J. Am. Chem. Soc.* **2005**, *127*, 14825–14832.
(52) Han, W.; Wu, Y.-D. *J. Am. Chem. Soc.* **2005**, *127*, 15408–15416.
(53) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. *Biophys. J.* **2008**, *94*, L75–L77.
(54) Laurence, T. A.; Kong, X.; Jäger, M.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17348–17353.
(55) Hoffmann, A.; Kane, A.; Nettels, D.; Hertzog, D. E.; Baumgärtel, P.; Lengefeld, J.; Reichardt, G.; Horsley, D. A.; Seckler, R.; Bakajin, O.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 105–110.
(56) Martinez, J. C.; Pisabarro, M. T.; Serrano, L. *Nat. Struct. Biol.* **1998**, *5*, 721–729.
(57) Rose, G. D.; Fleming, P. J.; Banavar, J. R.; Maritan, A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16623–16633.
(58) Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol.* **1989**, *206*, 759–777.
(59) Sibanda, B. L.; Thornton, J. M. *Methods Enzymol.* **1991**, *202*, 59–82.
(60) Richardson, J. S.; Getzoff, E. D.; Richardson, D. C. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 2574–2578.