JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMMUNICATION

pubs.acs.org/JACS

# Atomistic Analysis of Pseudoknotted RNA Unfolding

Yujie Zhang, Jian Zhang,* and Wei Wang*

National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093, China

Supporting Information

**ABSTRACT:** As important functional structures, RNA pseudoknots provide excellent models for studying the interplay between secondary and tertiary structures and the roles of triplexes, noncanonical interactions, and coaxial stacking in the folding/unfolding process. Here we report a first atomistic and statistical analysis of the unfolding of the pseudoknot within gene 32 mRNA of bacteriophage T2. Multiple unfolding pathways, diverse transition states, and various intermediate structures were observed. Water molecules were found to be coupled with the unfolding process via the expulsion or concurrent mechanism.

Evidence is accumulating that the majority of the genomes of mammals and other complex organisms is transcribed into noncoding RNAs (ncRNAs) and that these ncRNAs perform an astonishing variety of functions.[1,2] Nowadays the universe of ncRNAs is still expanding rapidly.[3] To fully appreciate their functions, it is necessary to study the folding process by which they achieve three-dimensional structures. In this direction, much effort and progress have been made.[4−10]

Among all ncRNAs, those containing pseudoknotted structures are of particular interest, not only because of their functional importance in, for example, control of ribosomal frameshifting by viral RNAs but also because they provide an excellent model for studying the folding/unfolding process of complex structures. RNA pseudoknots have both secondary and tertiary structures, which serve a template for studying their interplay. They also contain complicated interactions/structures such as noncanonical base pairs, triplexes, coaxial stackings, and sharp turns. The study of pseudoknots may benefit predictions of RNA secondary structure as well, since their pseudoknotted topology and the noncanonical interactions therein present major obstacles to the development of both sampling algorithms and free-energy estimation rules.[11−13] The knowledge of noncanonical interactions may also improve predictions of RNA tertiary structures, as the presence of a few key noncanonical base pairs is sufficient to blur such predictions.[14]

There have been excellent experimental works on RNA pseudoknots. In a study of the folding/unfolding process of a pseudoknot in human telomerase RNA using optical tweezers at different loads, direct evidence of formation of nonnative structures and complex folding pathways was observed.[15,16] In an all-atom molecular dynamics (MD) study of the frame-shifting pseudoknot from beet western yellow virus, a short (5 ns) simulation at an elevated temperature (400 K) revealed early unfolding events.[17] Using coarse-grained Go-type models, Thirumalai and colleagues simulated the folding of three pseudoknots.[18] They found that there are significant sequence-dependent variations for RNAs with similar folds; the RNAs can fold by a hierarchical mechanism with parallel paths, in a highly cooperative manner, or through multiple pathways, depending on the sequence. On the basis of the simulations, they related the folding mechanism to the stabilities of the secondary structures.[18]

Although much progress has been made to date, a statistical understanding of the folding/unfolding process of RNA pseudoknots at the atomistic level is still lacking. Here the term "statistical" refers to the characterization of the entire free-energy landscape rather than just the investigation of a few trajectories. In this work, we performed an all-atom MD simulation of the unfolding process of the pseudoknot within the gene 32 mRNA of bacteriophage T2, which, to the best of our knowledge, provides for the first time an atomistic and statistical picture of the unfolding process of RNA pseudoknots.

This 36 nucleotide (nt) pseudoknot has two double helices (denoted as H1 and H2). This is an astonishingly tight structure that has many interesting features. The first is the sharp turn at A8, which connects the inner end of H1 and the far end of H2 (Figure 1). Presumably there is strong structural tension associated with this sharp turn, whose role in the folding/unfolding process is of great interest. The second feature is the presence of noncanonical base pairs between the shallow groove of H1 and loop 2, which makes the first helix virtually a triplex. The two helices are coaxially stacked on each other and form a long quasi-continuous helix in space. To facilitate the following discussions, we denote the base pairs between H1 and loop 2 as NC (denoting noncanonical), which is virtually a tertiary interaction. The fractions of base pairs within the H1, H2, and NC regions are denoted as Q1, Q2, and QA, respectively.

Because of the stochastic nature of MD, we ran 50 all-atom simulations to gain a statistical description of the unfolding reactions. All of the simulations were performed using Amber version 9.0[20] with the parm99 force field.[21] The GB/SA model[22] was used to accelerate the sampling efficiency in phase space. The temperature of the unfolding simulations was set to 350 K, which is slightly higher than the experimentally determined melting temperature (∼330 K)[23] for unfolding of the RNA within a practically feasible time. The first model in the NMR structures[19] was used as the initial conformation for all unfolding simulations. We further performed 20 unfolding simulations at 430 K with the explicit TIP3P water model. The force field was the same, and the salt concentration was 0.6 M. The total simulation time was ∼4 $\mu$s (1 $\mu$s implicit plus ∼3 $\mu$s explicit water model). Two control simulations with implicit and explicit water showed that the native structure was rather stable at 300 K within 80 ns.

Figure 2 gives four typical unfolding trajectories showing the unfolding process of H1, H2, and the NC region [all 50

**Received:** December 6, 2010
**Published:** April 18, 2011





dx.doi.org/10.1021/ja1109425 | J. Am. Chem. Soc. 2011, 133, 6882–6885

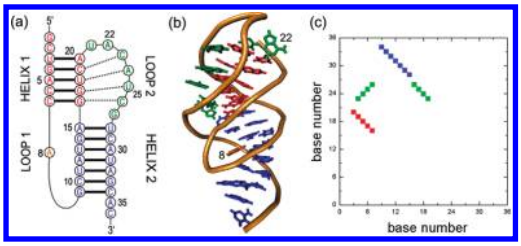


**Figure 1.** (a) Secondary structure, (b) tertiary structure, and (c) base-pair map of the pseudoknot (PDB entry 2TPK).[19] The two double helices are coaxially stacked, and loop 2 runs across the shallow groove of helix 1, forming noncanonical base pairs between them. Helix 1, helix 2, and loop 2 are colored red, blue, and green, respectively, and the base pairs within helix 1 and helix 2 and those between loop 2 and helix 1 are colored using the same code.

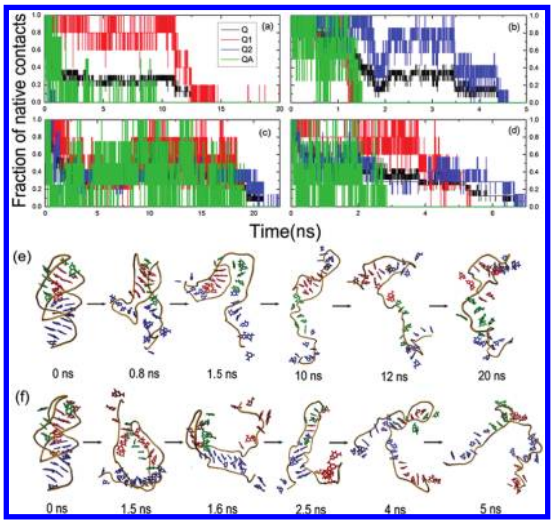


**Figure 2.** (a−d) Four typical unfolding trajectories. (e, f) Snapshots taken at the labeled times from the trajectories (a) and (b), respectively. The color code is the same as in Figure 1.

trajectories are documented in the Supporting Information (SI)]. From Figure 2a, the snapshots in Figure 2e, and the visual inspection of the unfolding animation, it was found that H2 unfolded first, starting at the 3′ terminus and roughly following an unzipping mechanism (Figure S24).[24,25] This mechanism then continued to drive the unfolding process and immediately unzipped the noncanonical interactions in the NC region at 1.5 ns. At ∼2 ns, the RNA reached an intermediate state constructed by H1 and a hairpin loop with a specific pattern of base stacking (from nucleotides A8 to A15; see the snapshot at 10 ns shown in Figure 2e and Figure S20). Previously, a similar intermediate state stabilized by base stacking was also seen in RNA hairpins,[24−27] indicating that the existence of intermediates in RNA unfolding may be general. The intermediate state observed here was quite stable and persisted for 10 ns. At 12 ns, H1 unzipped from the 3′ terminus and proceeded inward (Figure S24), after which the unfolding process went on to completion. The unfolded state was not a fully extended chain but had residual structures, characterized mainly by a hairpin loop closed by the C7−G16 base pair, that persisted until the end of our simulation (20 ns; Figure S21).

Figure 2b and the corresponding snapshots (Figure 2f) present a different unfolding sequence. At first, large structural

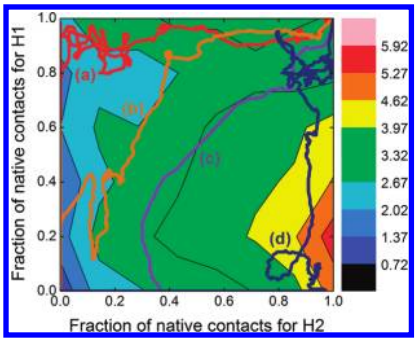


**Figure 3.** Nonequilibrium free-energy landscape projected on the two order parameters Q1 and Q2. The unit of the contour scale is kcal/mol. Also shown are four selected unfolding trajectories, which suggest that the unfolding pathways are diverse. The trajectories start from the upper-right corner (the native state) and end at the bottom-left. All of the superimposed trajectories can be seen in the SI.

fluctuations at both the 5′ terminus of H1 and the sharp-turn region (near nucleotide A8) were observed. H1 broke from the outer region first (i.e., the U3−A20 pair) but did not proceed to unfold the entire helix. Instead, the U3−A20 pair underwent a reversible breaking/reforming process. The actual unfolding of H1 started from its inner region, possibly as a result of the structural tension associated with the sharp turn (see the snapshot at 1.5 ns in Figure 2f). It roughly followed an unzipping mechanism (Figure S24) and proceeded from the inner region outward, finishing at ∼1.6 ns. The RNA then reached an intermediate state that resembled a long helical structure, consisting of the almost intact H2 and a long hairpin loop (see the snapshot at 2.5 ns shown in Figure 2f and Figure S22). This hairpin loop contained rather stable base-stacking interactions, partly attributed to the protection of closing H2. The intermediate structure persisted for several nanoseconds and then unfolded from the 3′ terminus of H2 following an almost exact unzipping mechanism (Figure S24). The whole process finished at ∼5 ns.

In the two trajectories discussed above, the unfolding events of H1 and H2 were well-separated in time. However, we also observed cases where they unfolded almost simultaneously, as shown in Figure 2c (also see Figures S4a and S5d). Furthermore, these two events were sometimes interwoven with each other. For example, in Figure 2d, H2 first unfolded partially and then remained there until H1 and the NC region completed their unfolding, after which H2 proceeded to unfold completely.

The 50 trajectories shown in Figure 2 and Figures S2−S18 demonstrate that the unfolding pathways are diverse in terms of the sequence in which microevents occur and the time periods over which they last. To obtain an overall view of the unfolding kinetics, we generated a projection of the free-energy landscape upon the two coordinates Q1 and Q2 and superimposed the trajectories on it (Figure 3 and Figures S2−S19). It should be noted that the bottom-right corner in the projection corresponds to high free energy, so most of the unfolding trajectories took pathways on the top-left half of the free-energy surface, as exemplified by pathways (a−c). Kinetically, this means that the unfolding of H2 preceded or was synchronized with that of H1, whereas the pathways whereby H2 unfolded after complete unfolding of H1 [exemplified by pathway (d)] were relatively rare. This aspect can be clearly seen from the superimposed trajectories in Figures S2−S18 and is further confirmed by Figure S23, which shows that the Q1 and Q2 values were comparable in the


dx.doi.org/10.1021/ja1109425 |J. Am. Chem. Soc. 2011, 133, 6882–6885

early unfolding stage and that the value of Q1 was higher than that of Q2 during the remainder of the unfolding process.

This aspect of unfolding pathways can be explained by the stability rule proposed by Thirumalai and colleagues for RNA folding.[18] To be consistent with this rule, the unfolding should start from structural elements that are least stable. We calculated the stabilities of H1 and H2 and found values of −10 and −12.4 kcal/mol, respectively, as estimated from Turner's free-energy rule.[28] However, our simulations showed that the stability of H1 is actually greater than H2 (Figure 3 and Figure S23). This apparent inconsistency suggests that the contribution of the noncanonical interactions (between H1 and loop 2) is very important to the stability of H1. When the totality of interactions was considered, the unfolding order was found to be in agreement with the stability rule.[18] The increased stability of H1 can also be seen as a result of the interplay between secondary structure (H1) and tertiary structure (the NC region).

The interplay between canonical and noncanonical interactions was observed not only in the thermodynamics of the unfolding process but also in its kinetic behavior. The noncanonical base pairs in the NC region were found to be closely correlated with the canonical ones within H1. In 30 of the 50 trajectories, the breaking of the NC region and H1 occurred simultaneously, while in the other cases, the former was closely followed by the latter. In a few cases, this coupling broke down (e.g., Figure S7e).

A loop 1 having a length of 1 nt is a typical structure often seen in H-type pseudoknots; it makes a sharp turn and presumably stores structural tension. We frequently observed sudden increases in the virtual angle in this region, indicating a rapid release of structural tension. Most interestingly, this event was often followed immediately by the partial unfolding of H2 (Figures S26 and S27), suggesting that the structural tension at least partially drives the initial unfolding.

Coaxial stacking usually contributes 1−2 kcal/mol to the overall RNA stabilities,[29,30] and its incorporation into the free-energy estimation rules should improve the prediction of RNA secondary structure.[31,32] Surprisingly, the simulation showed that it often broke very early in the unfolding process, possibly because of the structural tension within the sharp turn (Figures S28 and S29). More work must be done to see whether a pseudoknot with a longer loop 2 will show different kinetics.

Knowledge of the structure of the transition state (TS) is paramount for understanding the underlying kinetics of a molecular reaction. Unfortunately, TSs are particularly difficult to characterize by experiments or simulations. To extract the structures of the TSs, we utilized the method developed by Daggett and colleagues[33] (see the SI) followed by a clustering analysis. we obtained a total of 918 structures and 39 clusters, with the largest three clusters having 110, 62, and 55 structures, respectively (Figure S31).

The structures in the largest cluster resembled that of the native state but had lower formation probabilities for all base pairs, particularly for the noncanonical base pairs in the NC region (Figure 4a,d). This feature is consistent with the well-appreciated fact that for protein folding, the TS is usually a structurally looser version of the native state. However, the structures in the second-largest cluster were different: the first helix H1 and the NC region were roughly intact but H2 had completely disappeared (Figure 4b). This structural aspect indicates initial unfolding of H2 followed by unfolding of H1 and the NC region. The third largest cluster was in sharp contrast with the second: all of the base pairings in H1 and half in the NC region were missing while H2 remained almost intact (Figure 4c).

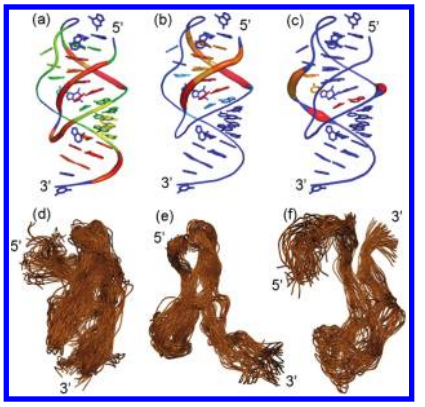


**Figure 4.** (a−c) Native structures colored according to the base-pair formation probability in the three largest clusters of TSs. The probability decreases in the order red > green > blue and is also reflected by the thickness of the backbone trace. (d−f) Corresponding conformations of the TS in each cluster. The figures were plotted using PyMOL and VMD.

More different TS structures were revealed by the cluster analysis, as shown in Figures S32−S35. The diversity of the TSs, together with the pathway complexity discussed above, demonstrates that the unfolding process of RNAs, like that of proteins, is actually a diffusion process on the free-energy surface, like water flowing down mountainsides along multiple routes;[34] there is no specific pathway, microscopically speaking. Furthermore, the structural diversity of the TSs indicates a broad transition and the possible failure of the widely used two-state model for describing the folding/unfolding of this pseudoknot.

To verify the results obtained using the GB/SA model, we performed 20 additional unfolding simulations with the explicit TIP3P water model at 430 K (see the SI for details). In these simulations, the unfoldings also followed multiple pathways (Figures S36−S42) and the TSs were diverse (Figures S43−S49), consistent with the results obtained using the implicit water model. However, we did observe differences. The most obvious one was that the probability to form the noncanonical base pairs in the TSs was much lower than with the implicit water model (Figure S50). Presumably, this may lead to a slight shift of the unfolding pathways toward the H2 side, as it destabilizes H1 and affects the pathways by the stability rule.[18] The physical reason for this difference merits further study.

Because of the relatively small number of trajectories using the explicit water model and the difference in temperature, it was not feasible to compare the extent of unfolding cooperativity and the diversity of TSs for the two water models quantitatively. Previously, Pande and colleagues reported that for protein BBA5, the structure of TSs from the implicit model was closer to the native structure than that from the explicit model because of the exaggeration of the destabilizing energy from the solvent−solute contact.[35] A similar problem has also been revealed for RNA hairpins.[36] Therefore, it may be expected that for the gene 32 mRNA pseudoknot, the TSs and unfolding pathways are more diverse than observed using the implicit model.

We also investigated the coupling between the water molecules and unfolding. By monitoring the radius of gyration ($R_g$) of the hydrophobic cores and the number of waters around them, we found that either (1) the water molecules first infiltrate the hydrophobic core with $R_g$ almost unchanged, with the expansion of the core occurring at a later stage (Figure S51a,c,g), or (2) the number of waters and $R_g$ increase concurrently (Figure S51b,d,h).


dx.doi.org/10.1021/ja1109425 |J. Am. Chem. Soc. 2011, 133, 6882–6885

These two scenarios are actually the unfolding version of the expulsion and concurrent mechanisms, respectively, which were proposed to describe the folding of proteins.[35] These two mechanisms were also observed by Pande and colleagues in RNA hairpins.[36] We did not observe the dewetting mechanism, which states that the rate-limiting step for folding is the formation of a sufficiently large vapor bubble around the nucleation sites.[37] This is tentatively attributed to the small size of the hydrophobic core in this pseudoknot.

From the simulations with explicit water model, the unfolded states were found to be more compact as a result of bridging interactions between nucleotides by water molecules through hydrogen-bonding networks (Figure S52). Hundreds of water molecules were found to have long residence times near the RNA backbone or side chains. The average residence time for these water molecules was ~1 ns (compared with ~500 ps in the first hydration shell[38]). Even longer residence times (>2 ns) were also observed (Figure S53).

Overall, our simulations have provided an unprecedentedly detailed atomistic picture of the unfolding process of the gene32 mRNA pseudoknot. The simulations revealed large diversities in both unfolding pathways and transition states, indicating the possible failure of the widely used two-state model in describing the folding/unfolding of this RNA pseudoknot. They also demonstrated the close interplay between the canonical and noncanonical interactions in determining the thermodynamic stabilities of structural elements and altering the kinetic pathways. Moreover, they provided an atomistic picture for the unzipping mechanism, the early break of the coaxial stacking, and the possible driving role of structural tension in the initial stage. Water molecules were found to be coupled with unfolding through the expulsion or concurrent mechanism in the early unfolding stage and also to mediate interactions between nucleotides and contribute to the stability of compact unfolded structures. All of this information will be useful for developing new free-energy rules for RNA structure prediction.

Caution should be taken when making inferences about folding from unfolding simulations, since folding and unfolding may not follow the same kinetics. However, the main conclusions made here, such as the diffusive nature of unfolding and the close interplay between canonical and noncanonical interactions in determining the unfolding thermodynamics and kinetics, are believed to be present in the folding process as well. The new discoveries concerning structural tension, coaxial stacking, and the role of water molecules are also expected to be seen in the folding process. Further direct simulations of the folding of RNA pseudoknots are ongoing and will be reported in due course.

## ■ ASSOCIATED CONTENT

**Supporting Information.** Models, methods, complete ref 20, and additional figures. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

**Corresponding Author**
jzhang@nju.edu.cn; wangwei@nju.edu.cn

## ■ ACKNOWLEDGMENT

This work was supported by the NNSF (10974088, 10834002, and 10704033), JiangSu Province (BK2009008), and the NBRPC (2007CB814800). Computational support from Shanghai Supercomputer Center and HPCC of Nanjing University is acknowledged.

## ■ REFERENCES

(1) Mattick, J. S.; Makunin, I. V. *Hum. Mol. Genet.* **2006**, *15*, 17.
(2) Kampa, D.; Cheng, J.; Kapranov, P.; Yamanaka, M.; Brubaker, S.; Cawley, S.; Drenkow, J.; Piccolboni, A.; Bekiranov, S.; Helt, G.; Tammana, H.; Gingeras, T. R. *Genome Res.* **2004**, *14*, 331.
(3) Storzs, G. *Science* **2002**, *296*, 1260.
(4) Treiber, D. K.; Williamson, J. R. *Curr. Opin. Struct. Biol.* **1999**, *9*, 339.
(5) Sosnick, T. R.; Pan, T. *Curr. Opin. Struct. Biol.* **2003**, *13*, 309.
(6) Thirumalai, D.; Hyeon, C. *Biochemistry* **2005**, *44*, 4957.
(7) Woodson, S. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 324.
(8) Laederach, A.; Shcherbakova, I.; Liang, M. P.; Brenowitz, M.; Altmen, R. B. *J. Mol. Biol.* **2006**, *358*, 1179.
(9) Chen, S. J. *Annu. Rev. Biophys.* **2008**, *37*, 197.
(10) Tinoco, I., Jr.; Bustamante, C. *J. Mol. Biol.* **1999**, *293*, 271.
(11) Solomatin, S. V.; Greenfield, M.; Chu, S.; Herschlag, D. *Nature* **2010**, *463*, 681.
(12) Zhang, J.; Dundas, J.; Lin, M.; Chen, R.; Wang, W.; Liang, J. *RNA* **2009**, *15*, 2248.
(13) Zhang, J.; Lin, M.; Chen, R.; Wang, W.; Liang, J. *J. Chem. Phys.* **2008**, *128*, No. 125107.
(14) Parisien, M.; Major, F. *Nature* **2008**, *452*, 51.
(15) Chen, G.; Wen, J. D.; Tinoco, I., Jr. *RNA* **2007**, *13*, 2175.
(16) Chen, G.; Chen, G.; Chang, K. Y.; Chou, M. Y.; Bustamante, C.; Tinoco, I., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12706.
(17) Csaszar, K.; Špačková, N.; Štefl, R.; Šponer, J.; Leontis, N. B. *J. Mol. Biol.* **2001**, *313*, 1073.
(18) Cho, S. S.; Pincus, D. L.; Thirumalai, D. *Proc. Natl. Acad. Sci. U. S.A.* **2009**, *106*, 17349.
(19) Holland, J. A.; Hansen, M. R.; Du, Z. H.; Hoffman, D. W. *RNA* **1999**, *5*, 257.
(20) Case, D. A.; *Amber 9*; University of California: San Francisco, 2006.
(21) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. *J. Comput. Chem.* **2003**, *24*, 1999.
(22) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(23) Nixon, P. L.; Theimer, C. A.; Giedroc, D. P. *Biopolymers* **1999**, *50*, 443.
(24) Chen, S. J.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 646.
(25) Sorin, E. J.; Rhee, Y. M.; Nakatani, B. J.; Pande, V. S. *Biophys. J.* **2003**, *85*, 790.
(26) Hyeon, C.; Thirumalai, D. *J. Am. Chem. Soc.* **2008**, *130*, 1538.
(27) Ma, H.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. *J. Am. Chem. Soc.* **2006**, *128*, 1523.
(28) Xia, T.; SantaLucia, J., Jr.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. *Biochemistry* **1998**, *37*, 14719.
(29) Walter, A. E.; Turner, D. H. *Biochemistry* **1994**, *33*, 12715.
(30) Kim, J.; Walter, A. E.; Turner, D. H. *Biochemistry* **1996**, *35*, 13753.
(31) Walter, A. E.; Turner, D. H.; Kim, J.; Lyttle, M. H.; Müller, P.; Mathews, D. H.; Zuker, M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9218.
(32) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. *J. Mol. Biol.* **1999**, *288*, 911.
(33) Jong, D. D.; Riley, R.; Alonso, D. O. V.; Daggett, V. *J. Mol. Biol.* **2002**, *319*, 229.
(34) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10.
(35) Rhee, Y. M.; Sorin, E. J.; Jayachandran, G.; Lindahl, E.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6456.
(36) Sorin, E. J.; Rhee, Y. M.; Nakatani, B. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2516.
(37) Wolde, P. R.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6539.
(38) Levy, Y.; Onuchic, J. N. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389.