Multiple Probes are Required to Explore and Control the Rugged Energy Landscape of RNA Hairpins

Changbong Hyeon*,†,‡ and D. Thirumalai*,§

Center for Theoretical Biological Physics, University of California, San Diego, California 92093, Department of Chemistry, Chung-Ang University, Seoul 156-756, Republic of Korea; and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742

Received September 15, 2007; E-mail: hyeoncb@umd.edu; thirum@glue.umd.edu

Recent experiments show that the kinetics of hairpin formation in RNA or ss-DNA is best described as a multistep process,[1] thus challenging the conventional premise that small nucleic acid hairpins fold in a two-state manner.[2] Kinetic data from high-resolution ultrafast T-jump experiments show that metastable intermediates are populated.[1] In contrast, single molecule force experiments showed that at the transition mid-force (f_m), the end-to-end distance (R) exhibits a bimodal distribution.[3] Motivated in part by these seemingly contradictory observations we probe the dynamics of hairpin formation, which is not only an elementary event in RNA folding but also plays a key role in transcription and translation.

The mechanism of RNA hairpin formation, which is thought to involve nucleation near the loop followed by base-pair formation in the stem, naturally suggests two plausible reaction coordinates, namely, local processes that describe native base-pair formation Φ, and global coordinate that reports on chain compaction R. Using the three interaction site (TIS) model for a 22-nucleotide (nt) RNA hairpin, P5GA (PDB ID: 1EOR), which has a GAAA tetraloop and 18-nts in the stem,[4] and exhaustive conformational sampling, we computed the diagram of states in terms of the order parameter R as a function of temperatures (T) and forces (f) (Figure 1). Along constant f or T, the hairpin forms in an apparent two-state manner, which is consistent with both the classical spectroscopic measurements[2,5] and the single molecule force experiments.[3] The free-energy landscape $F(R,\Phi)$ using R and Φ ($=\langle 1 - \cos(\phi - \phi^o)\rangle = 1/N_{loop}\Sigma_{i\in\{loop\}}\{1 - \cos(\phi_i - \phi_i^o)\}$), where ϕ_i is the ith dihedral angle (see Supporting Information (SI) and SI Figure 4 for the precise definition) and ϕ_i^o is the value in the native conformation, at different T, f values is profoundly different (Figure 1). If the bias toward the native state is strong ($T < T_m$, $f < f_m$) the native basin of attraction (NBA) with $R \approx 1.5$ nm and $\Phi \approx 0.0$ (Figure 1D) is preferentially populated. The structures in the unfolded basins of attraction (UBAs) are diverse depending on the T, f conditions (Figure 1A-F). The distribution in R is broad ($0 < R < 10$ nm) when $T > T_m$, $f \approx 0$ pN (Figure 1F) but its dispersion decreases substantially as f increases beyond the f_m (Figure 1A).

When the configurational states visited during the f-quench refolding are projected onto (R,Φ) plane, the frequently visited regions are different from the highly populated regions in $F(R,\Phi)$ (compare Figure 2A with Figure 1A,D). Surprisingly, the kinetically explored (R,Φ) regions (Figure 2A) are similar to the positions of the basins found in Figure 1E with $T = T_m \approx 345$ K and $f = 0$ pN. It follows that the ensemble of the stretched state conformations ((R,Φ) $\approx$ (11 nm, 0.5)) do not simply diffuse downhill as implied by the free-energy surface (Figure 1D). Collective fluctuations that are hard to capture using low dimensional coordinates drive the
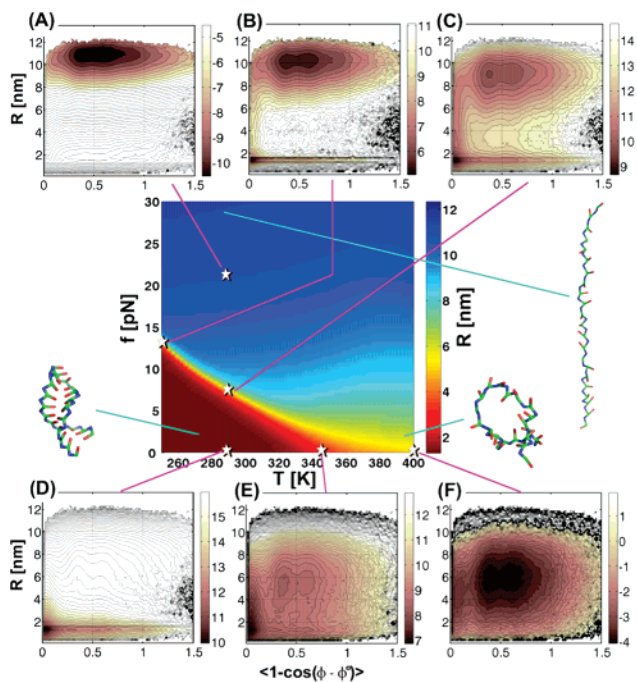
† University of California.
‡ Chung-Ang University.
§ University of Maryland.

Figure 1. Equilibrium phase diagram showing the end-to-end distance of RNA hairpins with varying temperatures (T) and forces (f) with two-dimensional free-energy surfaces at selected values of (T,f). The free energy is color-coded in k_BT unit, and its magnitude is shown in the scale on the right for each panel.

transition to the hairpin structure. Thus, using the equilibrium free-energy surfaces to interpret the kinetic pathways can lead to a misleading picture of the folding mechanism. The expectation that kinetics can be gleaned from $F(R,\Phi)$ may be valid if the RNA internal dynamics is rapid enough to establish quasi-equilibrium. For f or T-quench, such an assumption breaks down, especially, if the depth of quench is large.[6] Several examples in polymer dynamics[7] and protein folding[8] also show that the treatment based on quasi-equilibrium cannot account for the pathway diversity observed under nonequilibrium conditions.

Analysis of the paths followed by various molecules upon f-quench (from Figure 1A to Figure 1D) and T-quench (from Figure 1F to Figure 1D) not only illustrates the differences in the folding kinetics but also gives a quantitative description of the multistate kinetics. Upon f-quench, nucleation commences from a narrow distribution of fully extended conformations (**E**) and occurs via a small loop (**SL**) formation followed by zipping, $\mathbf{E} \rightarrow \mathbf{D_f} \leftrightarrow \mathbf{I_{SL}^f} \rightarrow \mathbf{N}$ (Figure 2B). In contrast, T-quench refolding starts from a broad distribution of thermally denatured ensemble (**D_T**), and occurs by multiple pathways either by small loop ($\mathbf{I_{SL}^f}$, $\mathbf{I_{SL}^T}$) or a large loop ($\mathbf{I_{LL}}$) formation followed by zipping (Figure 2B). The overall
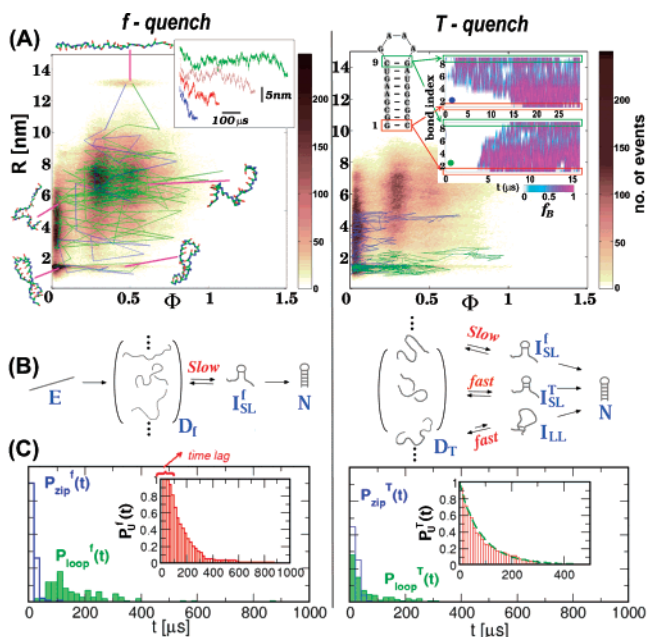
Figure 2. Kinetic analysis of the refolding trajectories upon *f*-quench and *T*-quench. (A) Configurational space navigated by the refolding trajectories projected onto the (R,Φ) plane. The trajectories of the individual molecules are overlapped on the (R,Φ) plane, and the corresponding trajectories monitored using a single parameter (R for *f*-quench and native base pair index for T-quench condition) are shown in the insets. For T-quench, dynamics of $\mathbf{D_T} \rightarrow \mathbf{I^T_{SL}} \rightarrow \mathbf{N}$ (blue) and $\mathbf{D_T} \rightarrow \mathbf{I_{LL}} \rightarrow \mathbf{N}$ (green) are shown. (B) Summary of the pathways to the NBA inferred from the dynamics (see SI Figures 5 and 6). (C) Statistical analysis of refolding kinetics. The refolding time for each molecule is decomposed into looping and zipping time as $\tau_{FP} = \tau_{loop} + \tau_{zip}$. The fraction of unfolded molecules ($P_U(t) = 1 - \int_0^t d\tau P_{FP}(\tau)$) where $P_{FP}(\tau)$ is the refolding or first passage time distribution) is plotted in the inset. $P_U^f(t)$ that is fit to $P_U^f(t) = \exp[-(t - 50\ \mu s)/138\ \mu s]$ for $t > 50\ \mu s$ shows a lag phase at $0 < t < 50\ \mu s$ suggesting an obligatory intermediate ($\mathbf{I^f_{SL}}$), whereas $P_U^T(t)$ is well fit to $P_U^T(t) = 0.44 \exp[-t/63\ \mu s] + 0.56 \exp[-t/104\ \mu s]$.

nucleation time upon *f*-quench is larger than under *T*-quench (compare $P^f_{loop}(t)$ and $P^T_{loop}(t)$ in Figure 2C), but the time scales for zipping are similar (compare $P^f_{zip}(t)$ and $P^T_{zip}(t)$ in Figure 2C). The differences in the ensembles of denatured state accessed during T ($\mathbf{D_T}$) and *f*-quench refolding ($\mathbf{D_f}$) results in significant variations in the nucleation times (Figure 2C). The extended conformations in the $\mathbf{D_f}$ ensemble increases are structurally more homogeneous than those in the $\mathbf{D_T}$ ensemble (Figure 2B). The proximity between the two complementary RNA strands in the $\mathbf{D_T}$ ensemble increases the probability of loop formation. Once a loop is stabilized by a single base pair, zipping of the remaining base pairs in the stem occurs rapidly in both *f*- and *T*-quench simulations. The dynamics of loop formation in the $\mathbf{D_f}$ ensemble is severely inhibited by an entropic barrier arising from the requirement that several loop dihedral angles in the GAAA tetraloop must adopt native-like values for nucleation to occur. Until the coordinated motions bring the loop into native conformation, the loop formation leading to the $\mathbf{I_{SL}}$ state is not realized[4b] (see also SI Figure 4). The structural diversity in the $\mathbf{D_T}$ ensemble is reflected in the multiple pathways in *T*-quench which are not observed in the *f*-quench. During the *T*-quench refolding, the ensemble of unfolded configurations belonging to ($\mathbf{D_T} \cap \mathbf{D_f}$) form a loop by sampling the loop dihedral

angle space (the upper inset in Figure 2A), whereas the ensemble of unfolded configurations belonging to ($\mathbf{D_T} \cap \mathbf{D_f}^C$), where the superscript C stands for complement, reach the native state by forming the stacking interaction between the complementary base pairs near the 5′ and 3′ ends (the lower inset in Figure 2A). Thus, upon *T*-quench RNA hairpins form by heterogeneous routes[9] which are quantified using at least three distinct refolding pathways (Figure 2). In contrast, the pathway diversity in *f*-quench refolding is greatly restricted because of entropic pinning of the 5′ and 3′ ends. As a result, refolding can be well-described by the classical picture of loop nucleation followed by zipping of base pairs in the stem.

In summary we find that the complex kinetics observed in the folding of ribozyme[10] is already reflected in hairpin formation. Our predictions should be applicable to other hairpins as well because they are based on the topology of the folded structure rather than the details of the sequence. Just as in β-hairpin formation,[11] there are multiple pathways to the structured state which shows that even simple structures in proteins and RNA have a rugged energy landscape. Exploring the details of the heterogeneous kinetics requires multiple experiments that control (easily achieved by using *f*) the conformations in the UBA. The prediction that there are profound differences between T and *f* jump quench refolding can be tested experimentally.

Supporting Information Available: Definition of dihedral angle in TIS model, computation details of free-energy surfaces, simulation method for thermodynamics and kinetics, and Figures 3−6. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Ma, H.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. *J. Am. Chem. Soc.* **2006**, *128*, 1523−1530. (b) Ma, H.; Wan, C.; Xu, A.; Zewail, A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 712−716. (c) Jung, J.; van Orden, A. *J. Am. Chem. Soc.* **2006**, *128*, 1240−1249.

(2) (a) Turner, D. H.; Sugimoto, N.; Freier, S. M. *Ann. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 167−192. (b) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I., Jr. *Nucleic Acids. Structures Properties and Functions*; University Science Books: Sausalito, CA, 2000. (c) Tinoco, I., Jr.; Sauer, K.; Wang, J. C.; Puglisi, J. D. *Physical Chemistry: Principles and Applications in Biological Sciences*; Prentice Hall: Englewood Cliffs, NJ, 2002.

(3) (a) Liphardt, J.; Onoa, B.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C. *Science* **2001**, *292*, 733−737. (b) Woodside, M. T.; Anthony, P. C.; Behnke-Parks, W. M.; Larizadeh, K.; Herschlag, D.; Block, S. M. *Science* **2006**, *314*, 1001−1004.

(4) (a) Hyeon, C.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6789−6794. (b) Hyeon, C.; Thirumalai, D. *Biophys. J.* **2006**, *90*, 3410−3427.

(5) Crothers, D. M. *Acc. Chem. Res.* **1969**, *2*, 225−232. (b) Crothers D. M.; Cole, P. E.; Hilbers, C. W.; Shulman, R. G. *J. Mol. Biol.* **1974**, *62*, 383−401.

(6) Li, M. S.; Hu, C. K.; Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 93−98.

(7) (a) Shivashankar, G. V.; Feingold, M.; Krichevsky, O.; Libchaber, A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 7916−7921. (b) Perkins, T. T.; Quake, S. R.; Smith, D. E.; Chu, S. *Science* **1994**, *264*, 822−826. (c) Rohbot-Raviv, Y.; Zhao, W. Z.; Feingold, M.; Wiggins, C. H.; Granek, R. *Phys. Rev. Lett.* **2004**, *92*, 098101.

(8) Bradley, C. M.; Barrick, D. *Structure* **2006**, *14*, 1303−1312.

(9) Ansari, A.; Kuznetsov, S. V.; Shen, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7771−7776.

(10) (a) Thirumalai, D.; Hyeon, C. *Biochemistry* **2005**, *44*, 4957−4970. (b) Woodson, S. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 324−330. (c) Treiber, D. K.; Williamson, J. R. *Curr. Opin. Struct. Biol.* **2001**, *11*, 309−314. (d) Chen, S.-J.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 646−651.

(11) Veishans, T.; Klimov, D. K.; Thirumalai, D. *Folding Des.* **1997**, *2*, 1−22.

JA0771641