JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMMUNICATION

pubs.acs.org/JACS

Cooperative and Directional Folding of the preQ$_1$ Riboswitch Aptamer Domain

Jun Feng,[†] Nils G. Walter,[†] and Charles L. Brooks, III[*,†,§]

[†]Biophysics and Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States
[§]Center for Theoretical Biological Physics, University of California San Diego, San Diego, California 92037, United States

Supporting Information

ABSTRACT: Riboswitches are *cis*-acting RNA fragments that regulate gene expression by sensing cellular levels of the associated small metabolites. In bacteria, the class I preQ$_1$ riboswitch allows the fine-tuning of queuosine biosynthesis in response to the intracellular concentration of the queuosine anabolic intermediate preQ$_1$. When binding preQ$_1$, the aptamer domain undergoes a significant degree of secondary and tertiary structural rearrangement and folds into an H-type pseudoknot. Conformational "switching" of the riboswitch aptamer domain upon recognizing its cognate metabolite plays a key role in the regulatory mechanism of the preQ$_1$ riboswitch. We investigate the folding mechanism of the preQ$_1$ riboswitch aptamer domain using all-atom Gō-model simulations. The folding pathway of such a single domain is found to be cooperative and sequentially coordinated, as the folding proceeds in the 5′ → 3′ direction. This kinetically efficient folding mechanism suggests a fast ligand-binding response in competition with RNA elongation.

Riboswitches represent a whole class of RNA structural motifs exerting a gene expression regulatory mechanism widely utilized among eubacteria.[1,2] Riboswitches are typically found in the 5′ untranslated regions (UTRs) of mRNAs and comprise a highly specific metabolite-binding aptamer domain and an adjoining expression platform.[3,4] The structural organization of the aptamer domain upon recognizing its cognate ligand directs adoption of a specific secondary structure of the expression platform, which in turn dictates the expression level of the downstream gene(s). Thus, as a metabolite sensor, a riboswitch provides a simple and direct strategy for the feed-back regulation of the expression of a (set of) related gene(s).

In bacteria, the class I preQ$_1$ riboswitch modulates the gene transcription of several queuosine synthetic enzymes, in response to the intracellular concentration of preQ$_1$ (7-aminomethyl-7-deazaguanine), a queuosine biosynthetic intermediate.[5] The class I, type II, *Bacillus subtilis queC* preQ$_1$ riboswitch aptamer domain (Figure 1), which resides in the 5′ UTR of the queuosine biosynthetic operon, is the smallest naturally occurring aptamer domain known to date, comprising minimally 34 residues. The secondary structure of this modular RNA element is predicted to simply consist of a stem−loop followed by a short adenine-rich single-stranded "tail".[5] The preQ$_1$ unbound aptamer domain appears to be largely unstructured.[6] When binding its metabolite, the aptamer domain folds into a classic H-type pseudoknot,[7] as evidenced by both X-ray crystallography and NMR spectroscopy.[6,8] In the folded state, the 5-base pair (bp) stem P1 consolidates into the hairpin region with the 3′ end of the tail forming a second 3-bp stem P2 with its upstream loop L1. P2 coaxially stacks on P1 via the associated ligand. The adenine-tract (A-tract) of the tail forms a stretch of A-minor motif contacts in the minor groove of the P1 stem. Folding of the aptamer domain down-regulates gene transcription as it sequesters part of a downstream anti-terminator sequence and favors formation of an alternative terminator hairpin.[9] The folding decision of the aptamer domain is essential to the regulatory mechanism of the riboswitch. Although high-resolution tertiary structures of the preQ$_1$−RNA complex are available, critical information on the conformational "switch" from the ligand-free to the folded ligand-bound state is lacking, and the folding pathway triggered by the ligand is yet to be investigated in any detail.

Computer simulation can be used to provide insights into the folding mechanisms and folding landscapes of biomacromolecules. Energy landscape theory[11,12] states that the free energy landscape of a protein is funnel shaped, minimally frustrated, and strongly biased toward the native state, which lies at the bottom of the funnel. Implementing this idea into a physical model, Gō-model simulations have been widely applied to study the folding landscape of proteins and nucleic acids.[13−19] The energy function of the topology-based Gō-model is strongly biased toward the native (or target) structure. In the nonbonded potential, attractive interactions are only assigned to contact pairs (contacts found in the native state) within a specific distance cutoff, whereas all non-native contacts are mutually repulsive. As a consequence, the biased potential models a smooth, funneled landscape that significantly reduces local thermodynamic traps and minimizes the energetic frustration on the folding pathway.

In this study, we use Gō model simulations to investigate the folding of the preQ$_1$ riboswitch aptamer. All heavy atoms of the RNA are represented explicitly. We present the folding kinetics, folding landscape, and detailed folding pathways of the aptamer domain. We employ the fraction of native contacts (Q) as a measure of folding progress to characterize secondary and tertiary structure formation.

We first use thermodynamic simulations in which the RNA is simulated over a wide temperature range to capture the aptamer (un)folding transition. The heat capacity C_v is calculated to monitor aptamer folding and unfolding, following $C_v = \sigma_E^2/k_BT$,

Received: November 19, 2010
Published: March 04, 2011

dx.doi.org/10.1021/ja110411m | *J. Am. Chem. Soc.* 2011, 133, 4196–4199

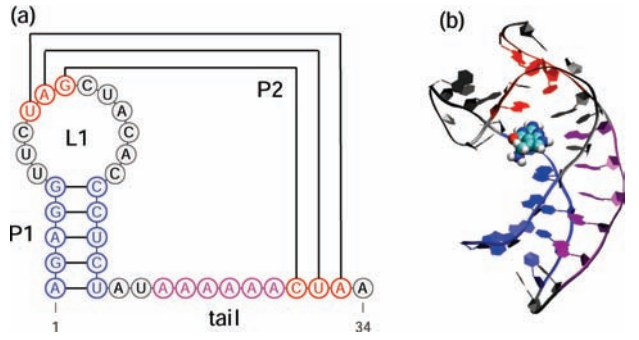

Figure 1. Representation of the preQ$_1$-bound aptamer domain structure. (a) Secondary structure (rendered using VARNA[10]). (b) Tertiary structure (rendered using VMD http://www.ks.uiuc.edu/Research/vmd/) in a ribbon cartoon. P1, P2, and the A-tract of the tail are represented in blue, red, and purple, respectively, and preQ$_1$ is rendered as a van der Waals volume.

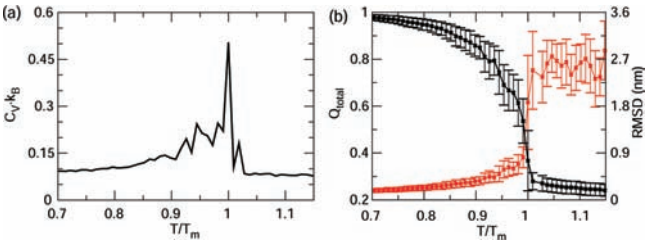

Figure 2. Melting of the aptamer−ligand complex. (a) Heat capacity profile as a function of temperature, scaled by T_m. (b) Average fraction of native contacts (black) and rmsd (red) of RNA as a function of scaled temperature. Standard deviation of average values at each temperature are shown as error bars.

where σ_E^2 is the energy fluctuation at a given temperature T and k_B is the Boltzmann constant. Thermal unfolding of the aptamer domain, shown in Figure 2, displays a single sharp peak in the temperature-dependent heat capacity profile, indicating a two-state folding transition. As expected, structural transitions indicated by significant changes in root-mean-square displacement and loss of native contacts occur collectively about the melting temperature (T_m).

Kinetic simulations of the aptamer folding are performed at 11 temperatures under T_m, each starting with 78 representative unfolded conformations. We observe fast conformational transitions among unfolded states. We do not find any rate-limiting intermediate states that could pose significant barriers along the folding reaction coordinates. The sharp peak in the heat capacity at T_m suggests a highly cooperative folding/unfolding transition for the aptamer domain. To quantify this behavior, we use a dimensionless index Ω to measure the folding cooperativity. The index Ω is defined as $\Omega = T_{max}^2/\Delta T \cdot |dQ/dT|_{max}$, where ΔT is the width at the half-maximum of the peak of $|dQ/dT|_{max}$ and T_{max} is the temperature at which $|dQ/dT|_{max}$ has a peak.[20] Ω goes to infinity for a sharp two-state transition and approaches zero for a completely noncooperative transition. The aptamer appears to fold in a highly cooperative manner as indicated by a large Ω, on the order of 10^6.

The folded aptamer domain resembles an H-type pseudoknot and is stabilized largely by the base-pairing interactions in the two helices, P1 and P2, respectively. We fold the aptamer domain at temperatures below T_m in the absence and presence of the cognate ligand. Figure 3 shows the ligand influence on the folding

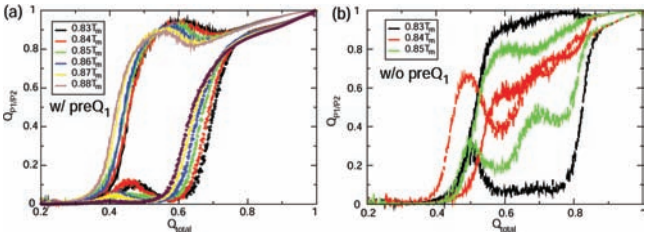

Figure 3. Fraction of native contacts in P1 (solid line) and P2 (dashed line) as a function of the fraction of total contacts in the folding simulations with and without ligand. Note the premature formation of structure in P2 in the absence of preQ$_1$. For clarity, folding at $0.86T_m$ − $0.88T_m$ in the absence of ligand can be found in the SI.

progress variables representing the two helices. Apparently, without the assistance of preQ$_1$ the aptamer domain encounters a topologically frustrated folding pathway. On the contrary, folding of the aptamer domain with its ligand appears to be smoothly coordinated, where folding of the P2 stem always follows the consolidation of the P1 stem. The small bumps in Figure 3a indicate that premature formation of the P2 stem can hamper the folding of the P1 stem, and that P2 has to unfold for the development of P1, which is known as "backtracking"[21] as opposed to misfolding. It is noteworthy that the structural changes that occur as the aptamer domain folds are reversed in the loss of secondary structure elements in the melting simulations (Supporting Information [SI]). However, it should be noted that in the absence of the ligand the 3-bp P2 stem is thermodynamically unstable, is not highly populated under physiological conditions,[6,22] and in our simulations is simply imposed by the parameters of the G model.[15,16] Henceforth, we will focus on folding in the presence of ligand.

Aptamer folding in the presence of ligand is hierarchical and is monitored by the native contacts shown in Figure 4. The tail of the aptamer domain consists of two nucleotides at the 5′ end connecting the upstream hairpin, the 6-adenine tract, and four bases at the 3′ end (Figure 1a). This tail folds after the formation of the P1 stem−loop (Q_{P1}). The A-tract region first establishes tertiary contacts with the P1 stem ($Q_{A\text{-}tract}$). The bases at the 3′ end then interact with their counterparts in the L1 loop to form the 3-bp P2 stem (Q_{P2}).

The sequential structural ordering is also evident in the free energy profile $F(Q_{tail}, Q_{P1})$ and $F(Q_{A\text{-}tract}, Q_{P2})$, calculated using WHAM (Figure 4c,d).[23] Figure 5 illustrates a representative folding pathway, which is observed in 85% of 858 independent folding simulations, regarding all structural aspects of the aptamer domain (P1, P2, tail, A-tract). The RNA structural ordering begins with the formation of the 5′ P1 stem, followed by the insertion of the A-tract segment, and finally, is completed with the consolidation of the P2 stem. The preQ$_1$ aptamer domain folds essentially along a single route from the 5′ to the 3′ end, in the same direction as RNA transcription proceeds.

RNAs are known to have rugged folding landscapes.[24] Folding of large and structurally complex RNAs is thought to be hierarchal, where the assembly of the thermodynamically more stable secondary structural elements precedes the collapse of the remote tertiary contacts.[25] However, an RNA pseudoknot is a rather simple and compact structural motif that is a widespread building block of RNA tertiary structures. In protein folding studies, small single-domain proteins have often been reported to fold according to a first-order rate law.[26−28] Analogously, G-model folding of the preQ$_1$ riboswitch aptamer appears to

dx.doi.org/10.1021/ja110411m |J. Am. Chem. Soc. 2011, 133, 4196–4199

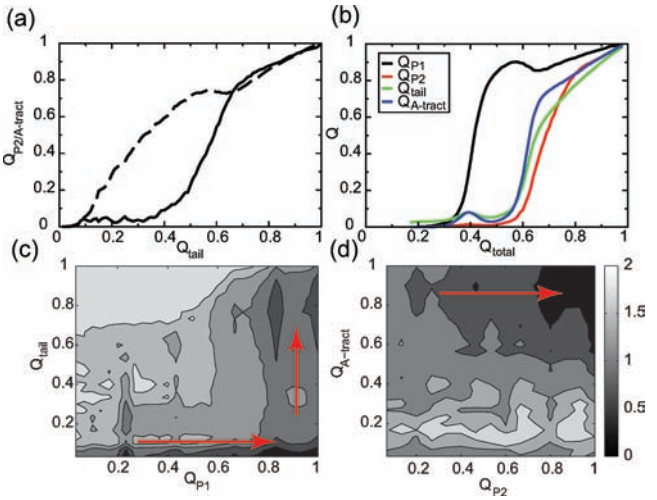

Figure 4. Detailed folding pathway of the aptamer in the presence of preQ$_1$ at 0.88T_m. (Folding at other temperatures behaves similarly and thus is not shown.) (a) Fraction of contacts in the A-tract (dashed line) and P2 (solid line) as a function of fraction of contacts in the tail. (b) Fraction of contacts in P1, P2, tail, A-tract as a function of fraction of total contacts. (c,d) Free energy profile $F(Q_{P1}, Q_{tail})$ and $F(Q_{A\text{-}tract}, Q_{P2})$. The dominant folding pathway follows the orderly establishment in Q_{P1}, $Q_{A\text{-}tract}$, and Q_{P2} (red arrows).

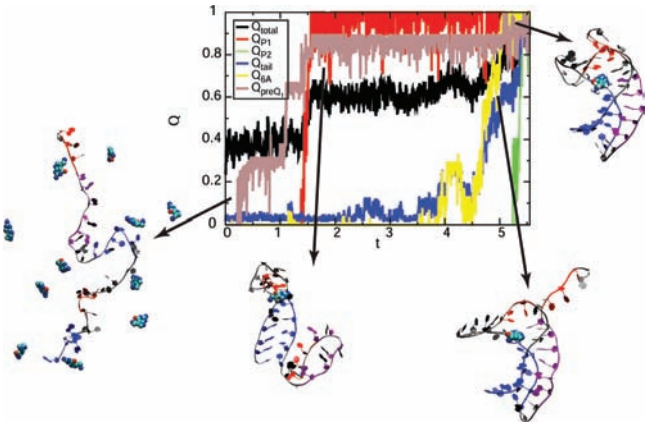

Figure 5. Representative folding pathway of the aptamer with ligand. The fraction of native contacts formed in the various segments of the RNA, P1, P2, tail, A-tract, and preQ$_1$ increases as time (arbitrary t) evolves. The surrounding structures are representative of the conformations observed during the folding process.

follow a simple two-state kinetics when forming the ligand-bound structure in a 5′ to 3′ direction. However, it should be noted that our Gō-like potentials are topologically biased. Gō-model simulations do not provide information related to any metastable intermediate states with non-native interactions that might be trapped in local minima, unless such interactions occur as a result of early formation of native contacts that must be lost for folding to proceed. In fact, both two-state and multistate transitions of RNA pseudoknots have been observed in thermal[29,30] and mechanical[31,32] folding/unfolding experiments. Our study investigates the folding mechanism of the isolated aptamer domain, which displays minimal topological frustration along the folding pathway when its cognate ligand is present.

The current view concerning the folding pathway of an RNA pseudoknot is often based on dissecting folding into the formation of its individual constitutes, which include two helical stems (P1 and P2) and tertiary stem-loop contacts. A recent folding study[33] proposed that relative stabilities of the helical stems are the main determinant of the folding mechanism, based on the observation of distinct assembly pathways of three structurally related pseudoknots. It was argued that, if there is sufficient difference in the stability of the two stems, the relatively more stable one forms first. Otherwise, parallel folding pathways could exist for folding the two stems with similar stability. In the preQ$_1$ aptamer domain study presented here then, it is perhaps not surprising that the thermodynamically more stable 5-bp P1 stem folds before the 3-bp P2 stem, but we also observe that the six consecutive adenines in the A-tract of the aptamer tail make unusually extensive tertiary contacts with the stem, even ahead of the secondary structure formation of P2. Thermodynamic stabilities of individual structural elements are directly coupled to the kinetics of the folding process, consistent with the prevailing hierarchical perspective of RNA folding.

Folding of the aptamer domain is key to the functional control of the preQ$_1$ riboswitch. Putting aptamer folding in the context of mRNA transcription is instrumental for our understanding of gene regulation by alternative RNA folding. It has been appreciated that many transcriptionally responsive riboswitches rely on kinetic rather than thermodynamic control, as evidenced by the large difference between the apparent binding affinity (K_D) of the ligand and the concentration (C_{50}) of ligand needed to reach half-maximal transcriptional control *in vitro*.[34,35] After all, once RNA polymerase passes through the riboswitch element, mRNA elongation will continue irrespective of upstream formation of a transcription terminator/antiterminator. To ascertain its impact, the ligand-binding kinetics must therefore be tightly coupled to the rate of transcription.[3] The 5′ to 3′ directional folding of the preQ$_1$ riboswitch aptamer domain observed here suggests that folding can occur concomitantly with aptamer synthesis by transcription. A picture emerges wherein, once ligand is encountered, a partially prefolded aptamer domain competes efficiently with mRNA elongation by quickly capturing its 3′ tail to signal the "off" state to the downstream gene expression platform. Interestingly, a similar 5′ to 3′ sequential folding strategy has been successfully harnessed in computational pseudoknot prediction.[36] The exceedingly conserved preQ$_1$ riboswitch aptamer domain[5] ensures the preservation of this exquisite folding design throughout evolution.

In summary, we here have employed G-model simulations to study the folding mechanism of the preQ$_1$ riboswitch aptamer domain and delineate the sequence of folding events. Our simulations suggest that folding of the preQ$_1$ riboswitch aptamer domain follows a highly cooperative and nearly single-routed pathway, where the pseudoknot in the presence of its cognate ligand assembles sequentially by formation of the P1 stem, stem−tail contacts, and finally the P2 stem. This folding mechanism is consistent with and in fact exploits the commonly observed hierarchical folding of RNA. The resulting directional folding of the preQ$_1$ aptamer is proposed to provide an efficient ligand-binding mechanism for conformational "switching" and riboswitch function.

■ ASSOCIATED CONTENT

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

AUTHOR INFORMATION

Corresponding Author
brookscl@umich.edu

ACKNOWLEDGMENT

We thank the center for Multiscale Modeling Tools in Structural Biology (MMTSB) for providing the computing facility. This work is supported by Grants from NIH (RR012255) and NSF (PHY0216576).

REFERENCES

(1) Winkler, W.; Nahvi, A.; Breaker, R. R. *Nature* **2002**, *419*, 952.
(2) Nudler, E.; Mironov, A. S. *Trends Biochem. Sci.* **2004**, *29*, 11.
(3) Garst, A. D.; Batey, R. T. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2009**, *1789*, 584.
(4) Roth, A.; Breaker, R. R. *Annu. Rev. Biochem.* **2009**, *78*, 305.
(5) Roth, A.; Winkler, W. C.; Regulski, E. E.; Lee, B. W.; Lim, J.; Jona, I.; Barrick, J. E.; Ritwik, A.; Kim, J. N.; Welz, R.; Iwata-Reuyl, D.; Breaker, R. R. *Nat. Struct. Mol. Biol.* **2007**, *14*, 308.
(6) Kang, M.; Peterson, R.; Feigon, J. *Mol. Cell* **2009**, *33*, 784.
(7) Pleij, C. W. A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 337.
(8) Klein, D. J.; Edwards, T. E.; Ferre-D'Amare, A. R. *Nat. Struct. Mol. Biol.* **2009**, *16*, 343.
(9) Rieder, U.; Kreutz, C.; Micura, R. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10804.
(10) Darty, K.; Denise, A.; Ponty, Y. *Bioinformatics* **2009**, *25*, 1974.
(11) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167.
(12) Brooks, C. L.; Onuchic, J. N.; Wales, D. J. *Science* **2001**, *293*, 612.
(13) Ueda, Y.; Taketomi, H.; Go, N. *Biopolymers* **1978**, *17*, 1531.
(14) Hills, R. D.; Brooks, C. L. *Int. J. Mol. Sci.* **2009**, *10*, 889.
(15) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Proteins: Struct., Funct., Bioinf.* **2009**, *75*, 430.
(16) Whitford, P. C.; Schug, A.; Saunders, J.; Hennelly, S. P.; Onuchic, J. N.; Sanbonmatsu, K. Y. *Biophys. J.* **2009**, *96*, L7.
(17) Shimada, J.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11175.
(18) Hyeon, C.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S. A.* **2005**, *102*, 6789.
(19) Lin, J. C.; Thirumalai, D. *J. Am. Chem. Soc.* **2008**, *130*, 14080.
(20) Klimov, D. K.; Thirumalai, D. *Fold. Des.* **1998**, *3*, 127.
(21) Gosavi, S.; Chavez, L. L.; Jennings, P. A.; Onuchic, J. N. *J. Mol. Biol.* **2006**, *357*, 986.
(22) Rieder, U.; Lang, K.; Kreutz, C.; Polacek, N.; Micura, R. *ChemBioChem* **2009**, *10*, 1141.
(23) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339.
(24) Thirumalai, D.; Lee, N.; Woodson, S. A.; Klimov, D. K. *Annu. Rev. Phys. Chem.* **2001**, *52*, 751.
(25) Tinoco, I.; Bustamante, C. *J. Mol. Biol.* **1999**, *293*, 271.
(26) Zwanzig, R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9801.
(27) Jackson, S. E. *Fold. Des.* **1998**, *3*, R81.
(28) Plaxco, K. W.; Simons, K. T.; Ruczinski, I.; David, B. *Biochemistry* **2000**, *39*, 11177.
(29) Nixon, P. L.; Giedroc, D. P. *Biochemistry* **1998**, *37*, 16116.
(30) Theimer, C. A.; Giedroc, D. P. *RNA* **2000**, *6*, 409.
(31) Chen, G.; Wen, J. D.; Tinoco, I. *RNA* **2007**, *13*, 2175.
(32) Green, L.; Kim, C. H.; Bustamante, C.; Tinoco, I. *J. Mol. Biol.* **2008**, *375*, 511.
(33) Cho, S. S.; Pincus, D. L.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 17349.
(34) Wickiser, J. K.; Cheah, M. T.; Breaker, R. R.; Crothers, D. M. *Biochemistry* **2005**, *44*, 13404.
(35) Wickiser, J. K.; Winkler, W. C.; Breaker, R. R.; Crothers, D. M. *Mol. Cell* **2005**, *18*, 49.
(36) Dawson, W. K.; Fujiwara, K.; Kawai, G. *PLoS One* **2007**, *2*, e905.