J|A|C|S
ARTICLES
Published on Web 01/14/2006

Exploring the Energy Landscape of a Small RNA Hairpin

Hairong Ma,[†] David J. Proctor,[‡] Elzbieta Kierzek,[⊥] Ryszard Kierzek,[⊥] Philip C. Bevilacqua,*[,‡] and Martin Gruebele*[,†,§]

Contribution from the Center for Biophysics and Computational Biology, Department of Chemistry and Department of Physics, University of Illinois, Urbana, Illinois 61801, Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, and Institute of Bioorganic Chemistry, Polish Academy of Sciences, Z. Noskowskiego 12/14, 61-704 Poznañ, Poland

Received August 6, 2005; E-mail: gruebele@scs.uiuc.edu; pcb@chem.psu.edu

Abstract: The energy landscape of a small RNA tetraloop hairpin is explored by temperature jump kinetics and base-substitution. The folding kinetics are single-exponential near the folding transition midpoint T_m. An additional fast phase appears below the midpoint, and an additional slow phase appears above the midpoint. Stem mutation affects the high-temperature phase, while loop mutation affects the low-temperature phase. An adjusted 2-D lattice model reproduces the temperature-dependent phases, although it oversimplifies the structural interpretation. A four-state free energy landscape model is generated based on the lattice model. This model explains the thermodynamics and multiphase kinetics over the full temperature range of the experiments. An analysis of three variants shows that one of the intermediate RNA structures is a stacking-related trap affected by stem but not loop modification, while the other is an early intermediate that forms some stem and loop structure. Even a very fast-folding 8-mer RNA with an ideal tetraloop sequence has a rugged energy landscape, ideal for testing analytical and computational models.

Introduction

The folding of RNA starts from a large ensemble of partially or wholly unfolded conformations and navigates toward the much smaller native ensemble. RNA folding is more hierarchically organized than protein folding.[1,2] In particular, RNA readily forms independently stable secondary structure elements at low counterion concentration; thus, the folding of secondary structure motifs can be studied in the absence of tertiary structure formation. The time scale for secondary structure formation (e.g., zipping) can be as fast as the 0.1−1 μs range.[3]

RNA hairpins consist of a base-paired stem capped with a loop (Figure 1). The hairpin motif is among the most common secondary structure elements in RNA and takes part in various important biological functions including ligand binding and tertiary folding initiation.[2,4,5] Although the function, thermodynamics, and kinetics of RNA hairpins have been extensively investigated,[2,6−10] our knowledge of the full free energy landscape of RNA hairpins is incomplete.

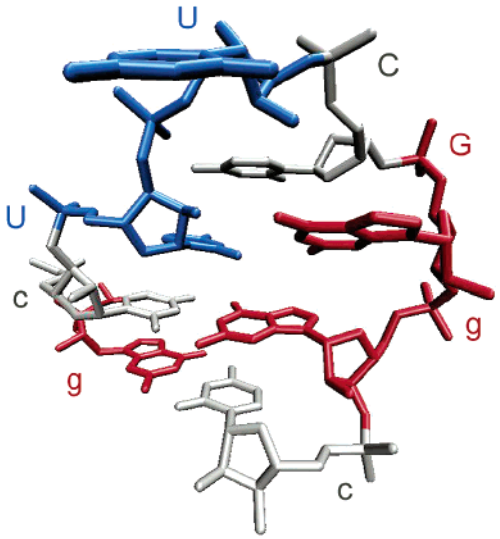

Figure 1. Structure of the gcUUCGgc hairpin (pdb ID: 1F7Y, truncated to residues 31−38). There are stabilizing hydrogen bonds between the first U and the fourth G in the loop, as well as the third C and a phosphate. Visualization was via VMD.[23]

In some cases two-state models can be used to describe RNA hairpin-folding kinetics.[11,12] However, recent experiments and theoretical calculations suggest that RNA/DNA hairpin kinetics cannot be described generally by two-state models.[9,10,13] A rough

[†] Center for Biophysics and Computational Biology, University of Illinois.
[§] Department of Chemistry and Department of Physics, University of Illinois.
[‡] Department of Chemistry, The Pennsylvania State University.
[⊥] Institute of Bioorganic Chemistry, Polish Academy of Sciences.

(1) Brion, P.; Westhof, E. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113−137.
(2) Tinoco, I.; Bustamante, C. *J. Mol. Biol.* **1999**, *293*, 271−281.
(3) Poerschke, D. *Biophys. Chem.* **1974**, *2*, 97−101.
(4) Uhlenbeck, O. C. *Nature* **1990**, *346*, 613−614.
(5) Varani, G. *Ann. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 379−404.
(6) Wilson, K. S.; Hippel, P. H. v. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8793−8797.
(7) Bonnet, G.; Krichevsky, O. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8602−8606.
(8) Shu, Z. Y.; Bevilacqua, P. C. *Biochemistry* **1999**, *38*, 15369−15379.

RNA energy landscape arises from energetic and topological frustration. Energetic frustration is caused by the conflict between base stacking (favoring collapsed states of the molecule) and electrostatic interactions (favoring extended states); topological frustration is caused by backbone connectivity.[14]

Temperature and mutation can be used to "tune in" different parts of a rough energy landscape. Here we report sequence-dependent folding kinetics of an RNA hairpin tetraloop family by using laser temperature-jumps. The sequences of these hairpins are chosen to differ either in the loop or the stem. We find that the hairpin has simple kinetics at the transition midpoint T_m as previously reported,[15] but multiphasic kinetics elsewhere. Stem and loop mutations have different effects on the observed multiphasic kinetics at different temperatures. The experiments, coupled with a simple lattice model and a four-state model simulation, provide a structural model for the RNA hairpin-folding mechanism.

Materials and Methods

Design of the RNA Hairpin Model System. The 8 nt sequence, a tetraloop with a 2 base pair stem, is the smallest RNA that folds into a stable and well-formed hairpin.[16] The short stem with only 2 base pairs was chosen to lower the melting temperature so that fully unfolded conditions could be explored, and to minimize alternative base pairings and the chance of falling into a misfolded trap. The short stem may also help increase the effect of loop composition on folding kinetics. To examine the sequence dependence of hairpin-folding kinetics, hairpins with several loop and stem sequences were synthesized. Our reference system was 5′-gcUUCGgc (stem nucleotides in lower case, loop nucleotides in upper case, see Figure 1). We also studied the loop variant UUUU, as well as the stem variant 5′-ggUUCGcc. A third variant, UUC^{8Br}G (^{8Br}G stands for 8-bromo substituted guanine), has a G to ^{8Br}G substitution in the loop and was previously shown to reduce the conformational entropy of the unfolded state without a large effect on the native state.[15] We also made gcCGA^{8Br}Ggc and ggacCCCCgucc sequences (not discussed in detail) that served to verify that the observed multiexponential behavior could also be seen for other loop sequences or stem lengths.

RNA Oligonucleotide. The unmodified 8 nt RNA oligonucleotides were purchased from Dharmacon Research. Synthesis of the 8-bromo-guanine RNA oligonucleotides was described previously.[15,17,18] RNA oligonucleotides were stored in low salt buffers of $P_{10}E_{0.1}$ (10 mM sodium phosphate, 0.1 mM Na_2EDTA, pH 7.1) to favor the hairpin conformation.[17,19−21]

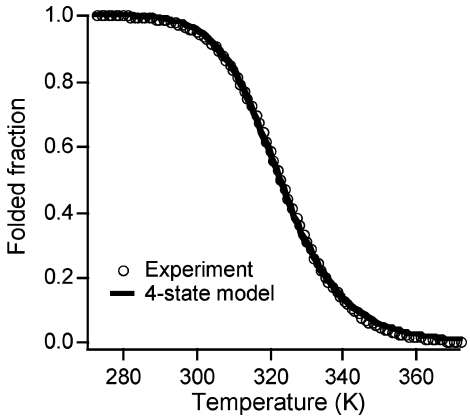

Figure 2. Heat denaturation populations (1 = native, 0 = denatured) of gcUUCGgc. The UV melting curve (○) can be described as a cooperative transition with T_m = 323 K. The four-state model calculated native population (—) is in good agreement with the experimental data.

The RNA concentration was measured by UV absorbance at 280 nm for the denatured state, prepared at 90 °C. The extinction coefficient is estimated from nearest-neighbor analysis[22] and assumed to be identical for the unmodified and modified hairpins as described.[17]

Thermal Denaturation Before the thermodynamic measurements at pH 7, oligomers were heat denatured at 90 °C for 1−2 min. UV absorption-detected melting experiments were performed in a 1 cm path length cuvette at 260 and 280 nm, at strand concentration of 1−25 μM. Repeated melting curves are independent of the concentration within this range, indicating that duplex formation is not a major component of the melting transition as shown in Figure 2.

Phosphorus NMR was used to measure the chemical shift of our reference RNA gcUUCGgc as a function of temperature. Spectra were concentration independent up to the 250 μM maximum concentration tested. Proton decoupled phosphorus NMR spectra also showed good chemical shift dispersion, seven resonances, and the unusual downfield-shifted resonance characteristic of a UUCG loop, confirming the hairpin species.[15,17,24] Because the NMR data lacked complete folding and unfolding baselines, the thermodynamic parameters were determined by globally fitting the temperature dependence of the NMR and UV melting data simultaneously using a two-state nonlinear least-squares fit.[15]

Temperature-Jump Setup. RNA folding/unfolding kinetics were detected by jumping the temperature of RNA solutions by 7.5 °C within 10 ns. The RNA concentration range was 0.9−4.6 mM. Except for the CGA^{8Br}G variant, which is not analyzed in detail here, no concentration dependence of the kinetics was detected over the ranges tested. UV absorption at 280 nm was used to detect the RNA relaxation that restores equilibrium after the temperature jump. The experiment was repeated for a range of temperatures to measure the full temperature dependence of the kinetics. The temperatures cited in the analysis are after the jump, when relaxation occurs. The laser T-jump apparatus and temperature calibration procedure have been described in detail.[25−27]

(9) Chen, S. J.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 646−651.
(10) Ansari, A.; Kuznetsov, S. V.; Shen, Y. Q. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7771−7776.
(11) Bonnet, G.; Krichevsky, O.; Libchaber, A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8602−8606.
(12) Ying, L. M.; Wallace, M. I.; Klenerman, D. *Chem. Phys. Lett.* **2001**, *334*, 145−150.
(13) Zhang, W. B.; Chen, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1931−1936.
(14) Thirumalai, D.; Lee, N.; Woodson, S. A.; Klimov, D. K. *Annu. Rev. Phys. Chem* **2001**, *52*, 751−762.
(15) Proctor, D. J.; Ma, H. R.; Kierzek, E.; Kierzek, R.; Gruebele, M.; Bevilacqua, P. C. *Biochemistry* **2004**, *43*, 14004−14014.
(16) Molinaro, M.; Tinoco, I. *Nucleic Acids Res.* **1995**, *23*, 3056−3063.
(17) Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2003**, *125*, 2390−2391.
(18) Xia, T. B.; SantaLucia, J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X. Q.; Cox, C.; Turner, D. H. *Biochemistry* **1998**, *37*, 14719−14735.
(19) Varani, G.; Cheong, C. J.; Tinoco, I. *Biochemistry* **1991**, *30*, 3280−3289.
(20) Antao, V. P.; Tinoco, I. *Nucleic Acids Res.* **1992**, *20*, 819−824.
(21) Proctor, D. J.; Schaak, J. E.; Bevilacqua, J. M.; Falzone, C. J.; Bevilacqua, P. C. *Biochemistry* **2002**, *41*, 12062−12075.
(22) Borer, P. N. *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, 3rd ed.; CRC Press: Cleveland, OH, 1975; Vol. I.
(23) Humhrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(24) Williams, D. J.; Hall, K. B. *J. Mol. Biol.* **2000**, *297*, 1045−1061.

The detection method (UV absorbance) differs from our previous fluorescence detection, and will be summarized briefly. A cylindrical quartz cell (Hellma) with a 100 μm path length was used to reduce distortion of the UV beam profile by the temperature jump, allowing for absorbance detection. Either no absorbance changes, or small <14 ns steps were observed in blank buffer solutions. The sample temperature was controlled by a thermoelectric element driven by a temperature controller (Lakeshore 330) with a diode feedback sensor.

The probe beam was a 280 nm, UV pulse train generated by a frequency tripled mode-locked Ti:Sapphire laser. This pulse train probed the sample every 14 ns. The UV beam was gated onto the sample cell by a mechanical shutter for only a few ms to avoid photobleaching of the RNA between measurements, absorbed by the sample, and focused by a lens onto a fast photomultiplier detector (Hamamatsu R5600) to measure the transmitted intensity every 14 ns. Band-pass filters were placed in front of the detector to filter out unwanted scattered light. The signal was digitized by a transient digitizer with 1 GHz bandwidth. Transmittance traces were converted into absorption traces (ΔA), and relaxation time constants were determined with the fitting program IGOR (WaveMetrics) using single- or double-exponential empirical models for all the absorbance decays.

$$S(t) = A_0 + A\,e^{-kt} + A'\,e^{-k't} \qquad (1)$$

Kinetic traces were reproduced multiple times and averaged and smoothed to provide comparable signal-to-noise ratios for all mutants at different temperatures. The highest signal-to-noise ratio in a single trace was achieved near the melting midpoint.

Absorbance changes accounting for up to one-third of the total signal were observed within the 14 ns dead time of the instrument and correspond either to intrinsic temperature dependence of RNA absorption, slight changes in beam profile induced by the temperature change, rapid base unstacking, or other rapid conformational changes that we cannot resolve. These earliest events are not considered further in our analysis of the data. For direct comparison of the changes in functional form, the relaxation data shown in the figures are normalized to lie between 1 (after 14 ns) to 0 (at 500 μs). The absorbance actually increases upon heating due to base destacking, as discussed in detail in ref 15.

Under certain conditions (near T_m), the kinetics appear to be two-state, and the major relaxation phase can be well described by a single rate constant k_{main}.[15] The forward (k_1) and reverse (k_{-1}) folding rate constants are then calculated using two-state model kinetics and thermodynamics:

$$k_{main} = k_1 + k_{-1} \qquad (2)$$

and

$$K_{eq} = \frac{k_1}{k_{-1}} \qquad (3)$$

Results

Thermal Stability of the Variants. Thermal melts can be approximated by a two-state model. Figure 2 shows that the folding temperature T_m of gcUUCGgc is near 323 K. The sigmoidal curve is well-fitted by a phenomenological two-state model (not shown). As discussed below, the kinetics are two-state only near T_m, so we will turn to more physically motivated models for the full temperature range. The stem variant, ggUUCGcc, melts 22° lower than the "wildtype".[15] The melting transition of gcUUC^{8Br}Ggc lies 7° higher than that of gcUUCGgc, and the melting transition of the UUUU variant is 23° lower. As discussed elsewhere, the 8-bromo substitution destabilizes the unfolded state by conformational restriction, whereas the UUUU variant destabilizes the native loop structure.[15]

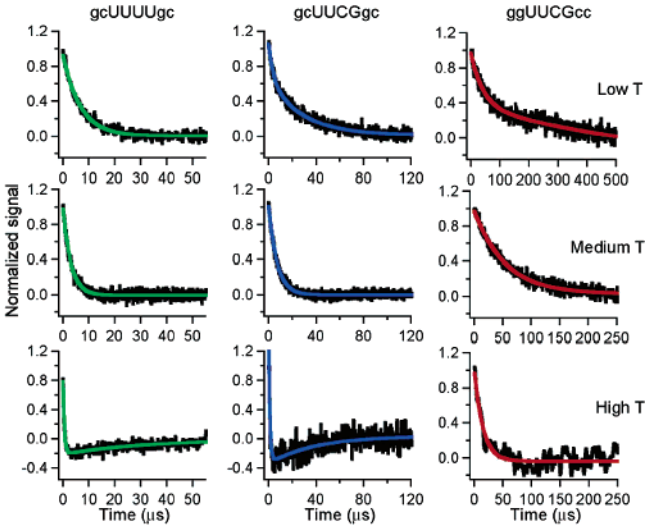

Figure 3. Representative kinetic relaxation traces of hairpins at low, medium, and high temperatures compared to T_m. Note the generally faster main phase toward the bottom, and slower time scales toward the right. (Left column) gcUUUUgc loop variant with the empirical fits to eq 1 shown as green curves: same behavior as the wild type at medium (30 °C) and high T (48 °C), but a single-exponential suffices at low T (13 °C). (Middle column) gcUUCGgc wild type with the empirical fits shown as blue curves. At low T (25 °C), a double exponential is required; at medium T (50 °C) a single-exponential suffices, and at high T (78 °C), a double exponential is again required. (Right column) ggUUCGcc stem variant with the empirical fits shown as red curves: same behavior as the wild type at low (18 °C) and medium T (28 °C), but a single-exponential suffices at high T (43 °C).

The wildtype loop shows three distinct temperature-tunable kinetic phases. At T_m, the relaxation of gcUUCGgc is described by a single exponential. When the temperature is tuned higher or lower than the melting temperature, the hairpin kinetics deviate significantly from a single-exponential decay (Figure 3). The kinetics contain three rate components, which are denoted as k_{fast}, $k_{main,}$ and k_{slow}. Below T_m, the relaxation is best fitted by a double exponential model. The fast phase k_{fast} is several times larger than the main phase k_{main} at all temperatures where the fast phase amplitude can be distinguished. The fast phase amplitude vanishes when the temperature is raised to 323 K. Above T_m, relaxation kinetics become biphasic again: a slow phase k_{slow} with negative amplitude appears.

Stem and Loop Mutations Affect the Kinetics Differently. Stem and loop mutations can be used to probe whether the additional phases observed originate from stem or loop dynamics on the free energy landscape. The UUUU tetraloop substitution (Figure 3) has little influence on the biphasic nature of the high-

(25) Turner, D. H.; Flynn, G. W.; Lundberg, S. K.; Faller, L. D.; Sutin, N. *Nature* **1972**, *239*, 215−217.
(26) Williams, A. P.; Longfellow, C. E.; Freier, S. M.; Kierzek, R.; Turner, D. H. *Biochemistry* **1989**, *28*, 4283−4291.
(27) Ballew, R. M.; Sabelko, J.; Gruebele, M. *Nat. Struct. Biol.* **1996**, *3*, 923−926.

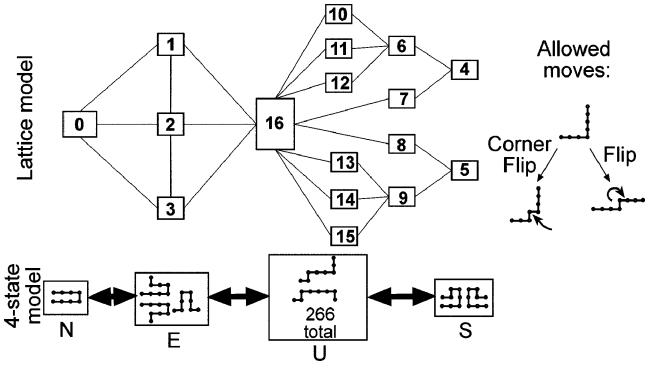

Figure 4. Connection diagram of the 272 microstates for the 8 nt tetraloop hairpin on a 2-D lattice. State 16 lumps together most of the 266 "unstructured" microstates. State 0 is the native state with a unique hairpin conformation. States 1−3 represent the conformations with nativelike loops but frayed stems. States 4 and 5 represent the conformations with stems, but nonnative loop structures. They are lumped together at bottom in correspondence with the four-state model.

temperature kinetics, but eliminates the additional fast phase observed at low temperature.

The stem modification ggUUCGcc has the opposite effect: the fast phase still manifests itself at low temperature, but the negative amplitude slow phase does not appear at high temperature (Figure 3). Thus, the relaxation unique to high temperature is connected with changes in stem sequence.

We also detected biphasic kinetics at high temperature in the UUC^{8Br}G tetraloop substitution, in the loop substitution CGA^{8Br}G at high temperature, and with the longer stem loop variant ggacCCCCgucc at low temperature (data not shown in Figure 3).

An Adjusted 2-D Lattice Model Can Reproduce the Temperature Effect Qualitatively. A minimalist 2-D lattice model with sequence-specific interactions can incorporate both topological and energetic frustration effects. Although such a model cannot do justice to the full conformational distribution, some useful insights about multiphasic kinetics can be obtained.

All 272 conformations of a sequence-specific 8-mer on a 2-D square lattice were enumerated to simulate the RNA folding thermodynamics and kinetics. The nucleoside beads are linked by "bonds" which geometrically constrain the hairpin conformations (Figure 4). Thermodynamics were calculated from the partition function obtained from the sequence-dependent energies and entropies. Kinetics were obtained by solving the multistate Master Equation with a physically motivated move set, as detailed below.

Sequence-dependent energies E_i were assigned to the 272 conformations using the "nearest-neighbor model" developed by Turner and co-workers.[28] A base is considered to be stacked only when it is adjacent to a Watson−Crick base pair and its backbone is perpendicular the base pairing axis; the native loop is considered formed when residues 3−6 occupy a square on the lattice and residues 3 and 6 of the loop are not part of a 90° kink. No loop initiation term was included. The resulting model energies and entropies are shown in Table 1 (state 0 is the native state, state 16 lumps together most of the unfolded states as shown in Figure 4).

(28) Serra, M. J.; Turner, D. H. Predicting thermodynamic properties of RNA. In *Energetics of Biological Macromolecules*; Johnson, M. L., Ackers, G. K., Eds.; Academic Press: San Diego, 1995; Vol. 259, pp 242−261.

In a 2-D lattice model, the conformational entropy of the RNA is seriously underestimated: the positions of the monomers on the lattice are limited by discrete sites, and the additional flexibility of loop or unstructured residues is neglected. To correct for this shortcoming at least on average, a degeneracy factor α is included in the partition function Z to add the missing conformational entropy:

$$Z = \sum_{i=0}^{271} \alpha^{n_i} e^{-\beta E_i} = \sum_{i=0}^{271} \exp[-\beta E_i + n_i \ln \alpha] = \sum_{i=0}^{271} \exp[-\beta g_i] \tag{4}$$

In this formula, n_i counts the number of "free beads" (not involved in stacking or loop formation). Because of the different sugar puckers, and the backbone and glycosidic torsion angles, each nucleoside can sample dozens of conformations per 2-D bead of the lattice simulation. From the partition function, all the thermodynamic parameters of the system can be calculated. The calculated population fraction in Figure 2 perfectly matches the experimental transition melting temperature and curvature with $\alpha = 55$.

The kinetics were simulated by solving the Master Equation

$$\frac{dP_j}{dx} = \sum_{i=1, i\neq j}^{n} k_{ij} P_i - \sum_{i=1, i\neq j}^{n} k_{ji} P_j \tag{5}$$

for the 272 microscopic states. The rate coefficients are derived from the lattice model such that only states that interconvert by flips or corner flips (Figure 4) have a nonzero rate coefficient, k_{ij}. This is because such local movements are expected to have the lowest barriers. For simplicity, all barriers between microscopic states (when referenced to the higher-energy state of each pair of connected microscopic states) were assumed identical to compute rate coefficients. The barriers thus represent an effective average barrier height among microstates, still allowing barriers among thermodynamic states in the four-state model to differ. The connectivity of microscopic states and their relation to four structural ensembles in the four-state model discussed later is shown in Figure 4. The native state 0 forms ensemble N. The unfolded states (most of which are lumped in to one box "16") form ensemble U. Ensemble E (for "early" forming) (conformations 1, 2, and 3) has incomplete stem structure. Ensemble E connects the native and unfolded states, thus it lies "on-pathway" with our move set. Ensemble S (for "stem" sensitive) (conformations 4 and 5) has stem structure but non-native loops. S is not directly connected to N, and lies "off-pathway" with our move set.

The lattice model simulations with Turner's parameter set have double exponential kinetics with rates k_{fast} and k_{main} at low temperature (Figure 5, middle). The fast-phase amplitude is sensitive to the energy of on-pathway ensemble E. Ensemble S is not sufficiently populated to produce nonexponential kinetics at high temperature.

When the entropy of the states in the ensemble S is adjusted to a larger value (Table 1), the inverted phase at high temperature is also reproduced by the lattice model (Figure 5, bottom; see Discussion). Both the adjusted and unadjusted 2-D lattice models also reproduce the thermodynamics of the RNA melting transition fairly accurately (Figure 5, top).

A Four-State Model Provides the Simplest Accurate Picture. The three time scales and four ensembles that emerge

Table 1. Summary of the Thermodynamic Parameters[a] Used for the Lattice Model in Figure 4

state	enthalpy (adjusted) kJ/mol	entropy (adjustable), kJ/(mol·K)	state	enthalpy (adjusted) kJ/mol	entropy (adjustable), kJ/(mol·K)
0	−100.96	0	9	0 (−3.97)	0.1998 (0.2331)
1	−51.67	0.1332	10	0	0.1998
2	−44.52	0.1665	11	0	0.1998
3	−41.34	0.1332	12	0	0.1998
4	−56.44 (5.96)	0.0666 (0.3164)	13	0	0.1998
5	−56.44 (5.96)	0.0666 (0.3164)	14	0	0.1998
6	0 (−3.97)	0.1998 (0.2331)	15	0	0.1998
7	0 (−3.97)	0.1998 (0.2331)	16	0	0.3122
8	0 (−3.97)	0.1998 (0.2331)			

[a] Parameters based on Serra et al. ref 24; adjusted parameters for the improved-fit model are shown in parentheses.

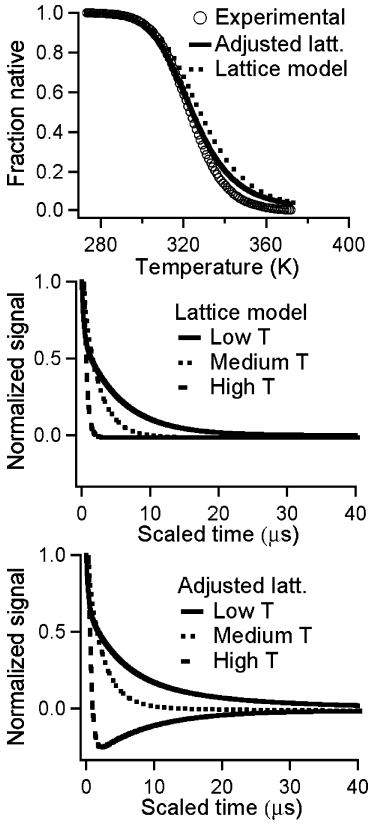

Figure 5. Simulated thermodynamics and kinetics of the gcUUCGgc tetraloop hairpin on a 2-D lattice. (Top) Fraction native as a function of temperature for gcUUCGgc. (○) experimental data; (- - -) lattice model; (—) adjusted lattice model. (Middle) Representative decays of the lattice model at (—) low *T*, (...) medium *T*, and (- - -) high *T*. (Bottom) Representative decays of the adjusted lattice model at (—) low *T*, (...) medium *T*, and (- - -) high *T*.

Table 2. Summary of Thermodynamic Parameters for the Four-State Model of gcUUCGgc

states	enthalpy,[a] kJ/mol	entropy,[a] kJ/mol·K
N	−100.96	0
E	−55.64	0.1332
U	13.51	0.3563
S	24.24	0.3729

[a] Enthalpies and entropies are on an absolute scale with the entropy of the native state set to 0 at $T = 0$ K.

from the adjusted lattice model motivated a simple four-state model in terms of N, E, U, and S states. This model has to reproduce both the thermodynamics (Figure 2) and the kinetics (Figure 3) for all variants at all 10 or more temperatures studied per hairpin. For thermodynamics, we started by averaging the enthalpies and entropies of the lattice model microstates over each ensemble. The enthalpy and entropy of each ensemble was then slightly adjusted to match the thermodynamics exactly (Table 2). Figure 2 shows the level of agreement obtained between calculated and experimental native fraction, with the population of folded vs unfolded states partitioned between N and E (folded) and U and S (unfolded), respectively.

Table 3. Summary of Kinetic Parameters[a] for the Four-State Model of gcUUCGgc

states	$\Delta G_0^{\ddagger}$, kJ/mol	$\Delta G_1^{\ddagger}$, kJ/mol·K	$\Delta G_2^{\ddagger}$, J/mol·K^2	$\Delta G_3^{\ddagger}$, mJ/mol·K^3
N−E	−10.92	−0.0666	−1.499	2.9024
E−U	49.32	−0.2448	−1.499	2.9024
U−S	99.26	−0.3646	−1.499	2.9024

[a] Activation free energies are on the same absolute scale as free energies computed from Table 2.

To reproduce the kinetics, the four states are connected as motivated by the lattice model in Figure 4:

$$\mathrm{S} \underset{k_{23}}{\overset{k_{32}}{\rightleftharpoons}} \mathrm{U} \underset{k_{12}}{\overset{k_{21}}{\rightleftharpoons}} \mathrm{E} \underset{k_{01}}{\overset{k_{10}}{\rightleftharpoons}} \mathrm{N} \qquad (6)$$

As in the lattice model, E can be accessed from both U and N, whereas S can only be accessed from U. To calculate kinetics with this model, we need to introduce barrier heights. As an initial guess, a simplified rate matrix was generated for the above reaction by adding/averaging the lattice model rate coefficients as appropriate when the microstates are lumped into ensembles S, U, E, and N. To obtain a quantitative fit, the dependence of the free energy barriers between S/U, U/E, and E/N was expanded as a polynomial in temperature (T in K),

$$\Delta G^{\ddagger} = \Delta G_0^{\ddagger} + (\Delta G_1^{\ddagger} \cdot T) + (\Delta G_2^{\ddagger} \cdot T^2) + (\Delta G_3^{\ddagger} \cdot T^3) \qquad (7)$$

The first term is individually adjusted for each barrier; the second term is assumed to be an average of the activation entropy of the two neighboring states; the higher-order terms were assumed identical for all of the barriers (Table 3). The thermodynamic fitting model in Table 3 requires eight parameters for the gcUUCGgc prototype, and six parameters for the remaining sequences ($\Delta G_2^{\ddagger}$ and $\Delta G_3^{\ddagger}$ are fixed at the prototype values). The four-state model thus greatly reduces the number of degrees of freedom compared to a raw fit of the data (over 200 parameters total: three lifetimes and two relative amplitudes for over ten temperatures each for four hairpins). Figure 6 shows the resulting free energy landscape at low *T*, medium *T*, and high *T* for the reference hairpin.

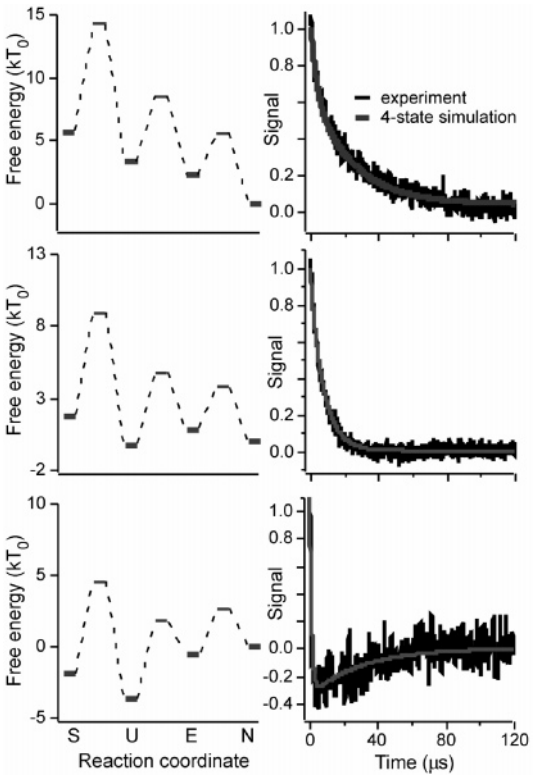

Figure 6. (Left) Free energy landscape of the RNA gcUUCGgc tetraloop hairpin at (top) low $T_0 = 298$ K, (middle) medium $T_0 = 323$ K, and (bottom) high $T_0 = 351$ K in the four-state model. (Right) Experimental and calculated relaxation curves of gcUUCGgc in the four-state model. (Black) experimental. (Grey) calculated decays. The temperatures correspond to the free energy diagrams and show the progression from double- to single- to double-exponential with inverted phase. The four state model in Table 2 fits all wild type and variant data as accurately as the empirical fit in Figure 3.

With the above activation free energy expression, the prefactor in a Kramers rate equation[29]

$$k(T) = \nu^{\dagger} \frac{\eta(25\ ^{\circ}\mathrm{C})}{\eta(T)} \mathrm{e}^{-\Delta G(T)^{\dagger}/kT} \qquad (8)$$

becomes $= 7.1\ \mu s^{-1}$ at 25 °C. The prefactor is inversely proportional to the solvent viscosity $\eta(T)$, which we model with the following empirical relationship adapted from ref 30:

$$\frac{\eta\left(\frac{T}{K}\right)}{cP} = (146 + 1090\ \exp[-0.2085(T - 273.1)] + 557\ \exp[-0.07(T - 273.1)])/[890] \qquad (9)$$

The four-state reaction model fully reproduces not only the thermodynamics and relative kinetic amplitudes but also the temperature dependencies of all kinetic components for all variants. Figure 6 illustrates this for the gcUUCGgc fit; the variants (not shown) fit equally well. Figure 7 shows the resulting free energy diagram for all hairpin variants at low temperature (ca. 25 °C below the melting midpoint). The absence of phases for the variants does not allow all states to be fully constrained (arrows in Figure 7). Rather, it sets lower limits on some free energies. This does not affect the interpretation in the next section.

(29) Kramers, H. A. *Physica* **1940**, *7*, 284.
(30) Weast, R. C. *CRC Handbook of Chemistry and Physics*; CRC Press: Cleveland, OH, 1997.

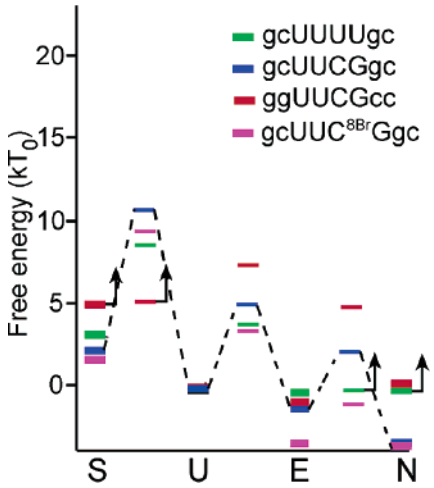

Figure 7. Free energy level diagram from the four-state fit at low temperature ($T_0 = 298$ K for gcUUCGgc, 291 K for ggUUCGcc, 301 K for gcUUC^{8Br}Ggc and 286 K for gcUUUUgc, to compare at similar relative temperatures below T_m). The free energies have been shifted to make U the reference state. The variants are color coded as in Figures 3 and 6. The arrows indicate states for which the fit provides only a limit on the free energy.

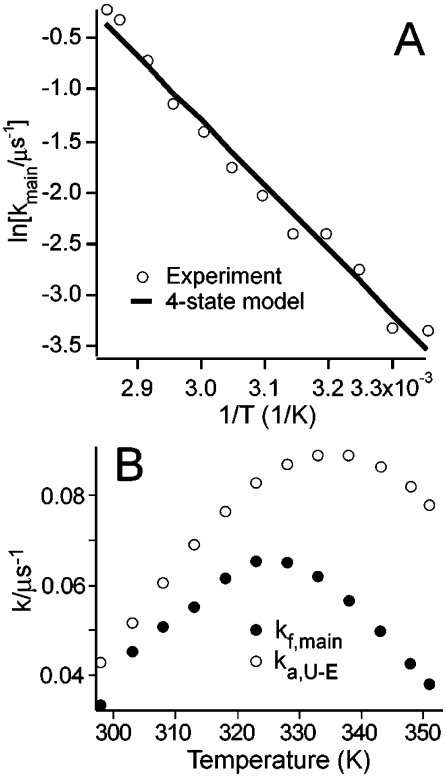

Figure 8. **(**A**)** Four-state model simulated rate coefficients for the gcUUCGgc main phase as a function of temperature (solid line) agree well with the observed rate coefficients (open circles) on an Arrhenius plot. (B) The forward rate $k_{f,main}$ extracted from the main phase by assuming two-state thermodynamics does not agree with the not directly observable U−E transition rate $k_{a,U-E}$ obtained from the model free energies in Tables 2 and 3.

Figure 8a shows that the absolute values of the experimental and simulated k_{main} are in good agreement on an Arrhenius plot. However, the main rate eigenvalue cannot be uniquely attributed to a single barrier over the entire temperature range. It is related to the change of at least two of the barrier heights in the four-state model, as discussed below.

Discussion

Temperature provides an adjustable window onto the RNA hairpin free energy landscape. Changing the temperature can tune kinetic phases in and out of existence, allowing selected local minima to be probed and differentiated. We detected such complex kinetics for variants with loops UUCG, UUUU, and UUC^{8Br}G, and stems gc_gc and gg_cc, discussed in detail here. We also detected multiexponential kinetics in additional variants with loop gcCGA^{8Br}Ggc and stem ggacCCCCgucc, so this complex behavior is not unique to the loops and stem combinations analyzed in detail here.

At low temperature the energy surface of gcUUCGgc is biased toward the native state (Figure 6), and thus the native state and the ensemble E are mostly populated. When the temperature is jumped up to destabilize N and E, the relaxation starts with a fast phase from N to E because that barrier is small, followed by the main phase toward the unfolded state, U. The N/E and E/U transitions dominate the kinetics, and the state E can be viewed under these conditions.

As the temperature is raised, E accumulates less population, and the fast phase diminishes. The kinetics are dominated by the major barrier crossing to U, nominally associated with the intermediate rate coefficient k_{main}. A single exponential provides a good approximate description of the kinetics.

At high temperature the energy surface is biased toward the unfolded side. The unfolded state U and the trapped state S are the most populated states; N and E are no longer viewed separately. The T-jump induced relaxation begins with the barrier crossing from N/E to U with rate coefficient k_{main}, followed by the slower relaxation between U and the "off-pathway" state S, creating an additional slow phase k_{slow} in the relaxation. Thus, the kinetics smoothly change from a fast and medium phase to a medium and inverted slow phase, as shown in Figures 5 and 6.

Figure 8 illustrates why the RNA hairpin free energy landscape must be modeled as a whole. The Arrhenius plot of the main relaxation phase, nominally attributed to the U−E barrier crossing, appears to be linear. Yet the folding rate $k_{f,main}$, extracted from k_{main} by using eqs 2 and 3, is not in agreement with the folding rate $k_{a,U-E,}$ calculated directly from the four-state model free energies in Tables 2 and 3. Although k_{main} is nominally related to the U−E barrier crossing, this rate eigenvalue has contributions from other barriers, particularly when they come within $3kT$ of the U−E barrier. The different structural ensembles and their connecting barriers in Figure 7 are not sufficiently separated in free energy for the individual steps of the folding reaction to be treated independently.

If temperature tuning reveals the energetics of the RNA hairpin states, then mutation reveals structural information about the ensembles S and E. The ggUUCGcc stem modification eliminates the negative phase at high temperature in Figure 3. This can be explained by a less stacked state S: the gg_cc stem substitution is a good stack-former, and would increase the S-state free energy relative to U as shown in Figure 7. The effect of the loop mutation gcUUUUgc on state S is small by comparison, showing that nativelike loop structure is not required to stabilize the state S. The 8-bromo substitution also causes only a small shift with respect to U, showing that S and U are similarly configurationally restricted by the bromo substitution.

The mutations have more subtle effects on the state E: both the stem and loop mutations significantly raise the native-state free energy compared to U, but they raise the free energy of state E by less than 20% of that amount. The on-pathway intermediate E has formed only a small amount of nativelike stem and loop structure, and thus lies early along the reaction coordinate from U to N; hence, its free energy is not disrupted as much as the native state's. In the gcUUUUgc variant model in Figure 7, both the E−N transition state and N-state free energies could lie substantially above the E state (arrows indicate lower limits). This explains the disappearance of the fast low-temperature phase for that variant. A "molten" hairpin with a nonnative loop may be the lowest free energy structure, and no E−N relaxation is observed.

Interestingly, the 8-bromo substitution causes a different apparent behavior of E. The native state N is slightly stabilized compared to that of U, especially at higher temperatures (not shown in Figure 7) where the penalty of decreased entropy in the unfolded state becomes larger. E is substantially stabilized compared to U and reaches nearly native stability. This indicates that the 8-bromo substitution forces equivalence or near-equivalence of the E and N states by locking the E-state loop into nativelike conformation. This explains why the 8-bromo substitution causes the disappearance of the fast phase at low temperature. Note that the reasons the fast phases disappear for the UUUU and 8-bromo variants are quite different: in the former case, the E/N states become like the wild-type E state; in the latter case the E/N states become more like the wild-type N state.

The simple lattice model is in approximate agreement with the four-state results. There are important similarities and differences. The S state has nativelike stem structure in the lattice model, and a nonnative loop. This is essentially the interpretation derived from the mutation study and four-state model. The E state has frayed stem structure in the lattice model (in agreement with the four-state model and mutation data), and intact loop structure (not in agreement). The disagreement is not surprising since the simplicity of the lattice model does not allow for partially folded loops, a necessity for describing the E state. With this grain of salt and the reinterpretation of states 1−3 in Figure 4 as frayed in both stem and loop, the lattice model calculations in Figure 5 provide a surprisingly quantitative description of the thermodynamics and kinetics.

It is worth considering why the entropy of the S ensemble in the lattice model had to be adjusted upward to obtain this agreement (Table 1). One possibility is that the lattice model overestimates the amount of structure formed in S. It is possible that the S state is composed of states that are primarily single-stranded, with little or no base pairing, which would lead to the much larger entropy needed to describe that state. Single-strand stacking is known to be highly dependent on stem sequence, with GG and CC steps being especially good stackers.[31] Variations in single-strand stacking are important in equilibrium studies and have been invoked to explain heat capacity changes in RNA and DNA folding.[32,33]

The lattice model and four-state model cannot exclude the possibility of unfolded → stem only → stem + loop transitions

(31) Turner, D. H.; Bevilacqua, P. C. Thermodynamic considerations for evolution by RNA. In *The RNA World: the nature of modern RNA suggests a prebiotic RNA world*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Press: Cold Spring Harbor, 1993; pp 447−464.

as a minor component of the kinetics. In the 2-D lattice model, this would correspond to adding "kinking" moves that interconvert states 0 (N) and 4/5 (S). However, our analysis suggests that the transition from highly unfolded structures to the native hairpin is more facile than the corresponding transition from misfolded compact structures that have stem contacts already formed, otherwise the slow phase would not be observed. An analogue of this behavior has been observed recently in a peptide hairpin. The peptide trpzip2 was probed by tryptophan fluorescence at blue-shifted (some hydrophobic contacts formed) and red-shifted (fully solvent-exposed) wavelengths. The relaxation of the red-shifted ensemble was much faster, indicating that fully unfolded states are more mobile and may fold more rapidly than compact but partially misfolded states which have to escape from a deeper local minimum before they can get to the native state.[34]

Stem formation is more prevalent than nativelike loop formation in the nonnnative free energy basins we observe, even in the small 8-mer RNA studied here. This is in agreement with multi-trajectory stochastic dynamics studies of RNA secondary structure formation, which indicate that stem structure comprises a stable subunit for tetraloop hairpins.[35]

Acknowledgment. This work was supported by NSF Grant MCB-0316925 and PRF Grant 37333-AC7 of the ACS Petroleum Research Fund (H.M. and M.G.); NSF Grant MCB-0527102 (D.P. and P.C.B.), NIH Grant 1 R03 TW 1068-01 (R.K. and D. H. Turner), a Fellowship from the Alfred P. Sloan Foundation, and a Camille Dreyfus Teacher-Scholar Award to P.C.B.

JA0553856

(32) Draper, D. E.; Bukhman, Y. V.; Gluick, T. C. Thermal Methods for the Analysis of RNA Folding Pathways. In *Current Protocols in Nucleic Acid Chemistry*; Beaucage, S. L., et al., Eds.; Wiley & Sons: New York, 2000; pp 11.13.11−11.13.13.
(33) Holbrook, J. A.; Capp, M. W.; Saecker, R. M.; M. T. Record, J. *Biochemistry* **1999**, 38.
(34) Yang, W. Y.; Gruebele, M. *J. Am. Chem. Soc.* **2004**, *126*, 7758−7759.
(35) Sorin, E. J.; Engelhardt, M. A.; Herschlag, D.; Pande, V. S. *J. Mol. Biol.* **2002**, *317*, 493−506.