

LA-UR-16-28139 (Accepted Manuscript)

Equilibrium Denaturation and Preferential Interactions of an RNA Tetraloop with Urea

Miner, Jacob Carlson
Garcia, Angel Enrique

Provided by the author(s) and the Los Alamos National Laboratory (2017-08-01).

To be published in: The Journal of Physical Chemistry B

DOI to publisher's version: 10.1021/acs.jpccb.6b10767

Permalink to record: <http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-16-28139>

Disclaimer:

Approved for public release. Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Equilibrium Denaturation And Preferential Interactions Of An RNA Tetraloop With Urea

Jacob C. Miner^{†,‡} and Angel E. García^{*,‡}

[†]*Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87545*

[‡]*Center for Nonlinear Studies, CNLS, MS B258, Los Alamos National Laboratory, Los Alamos, NM 87545*

E-mail: agarcia@lanl.gov

Phone: +1 505 665 3883. Fax: +1 505 665 2659

Abstract

Urea is an important organic cosolute with implications in maintaining osmotic stress in cells and differentially stabilizing ensembles of folded biomolecules. We report an equilibrium study of urea-induced denaturation of a hyperstable RNA tetraloop through unbiased replica exchange molecular dynamics. We find that, in addition to destabilizing the folded state, urea smooths the RNA free energy landscape by destabilizing specific configurations, and forming favorable interactions with RNA nucleobases. A linear concentration-dependence of the free energy (m -value) is observed, in agreement with the results of other RNA hairpins and proteins. Additionally, analysis of the hydrogen-bonding and stacking interactions within RNA primarily show temperature-dependence, while interactions between RNA and urea primarily show concentration-dependence. Our findings provide valuable insight into the effects of urea on RNA folding and describe the thermodynamics of a basic RNA hairpin as a function of solution chemistry.

Introduction

Urea is one of the most common denaturing cosolutes used in living systems to maintain osmotic stress and preserve biomolecular integrity.¹ A number of marine organisms, particularly fish species and some frogs, use urea for osmoregulation.^{2,3} In the context of proteins, urea has been used extensively to describe biomolecular stability as a function of denaturant concentration.⁴⁻⁸ From these studies a number of important insights into the behavior of urea have been gleaned including the general favorability of urea for certain chemical groups⁹ and correlations between urea concentration and the exposure of biomolecular surface area.^{10,11} Questions about whether urea unfolds proteins by 'direct' or 'indirect' mechanisms have been addressed through the application of all-atom molecular dynamics (MD) techniques that recapitulate the thermodynamics of urea-induced protein denaturation through enhanced phase space sampling.^{12,13} These studies have allowed researchers to assess the effects of cosolutes in solution with proteins, yielding information on how conditions more akin to intracellular environments affect the stability of biomolecules.^{14,15}

Similar questions in nucleic acid stability have emerged from experiments that show comparable thermodynamic behavior in the presence of urea.¹⁶⁻²⁰ These experiments reveal that urea can denature nucleic acids by close association with the nucleic acid surface, with favorable interactions at specific chemical groups in the nucleobases.^{21,22} Urea is often used to reduce both the heterogeneity of the structural ensemble of a particular nucleic acid sequence, and can also affect the native state melting temperature (T_m).¹⁷ The findings in these experiments also reveal that urea preferentially accumulates around nucleobases *without* disrupting intramolecular stacking interactions in single-stranded oligomers of RNA.²² This has led to an interest in using urea as a general probe of more complex tertiary structures, and a tool for reducing intramolecular contacts in nanopore sequencing.²³

Multiple simulation studies of urea have been conducted in conjunction with nucleic acid structures, including secondary structures (e.g. hairpins), and more complex tertiary structures (e.g. pseudoknots).^{24,25} These studies, however, have focused primarily on the kinetics

of unfolding and the localization of urea around RNA molecules over short time periods (~ 10 ns) in single MD simulations. The long time-scale stabilities of different preferential interactions of RNA with urea, or the landscape of an RNA molecule at equilibrium in urea have yet to be investigated. Recent advances in nucleic acid force fields have made it possible to model the equilibrium thermodynamics of RNA folding with unbiased MD simulations,^{26,27} and with advances in force fields for urea²⁸ it is now possible to extend these techniques to analyze the effect of urea on RNA denaturation and equilibrium thermodynamics.

The theoretical underpinnings of urea-induced unfolding of biomolecules are as relevant to RNA as they are to proteins, and many of the same principles that have been applied to study urea-induced unfolding of proteins in simulation can also be applied to RNA.

The change in free energy upon unfolding (ΔG_U) shows cosolute dependencies in the Gibbs-Duhem equation (Equation 1), where the product of the chemical potential (μ_i) and the number of particles (N_i) for each solution component (i) of all solution species (M) affect the free energy of unfolding in an additive manner along with the entropy (S), pressure (P), temperature (T) and volume (V).

$$\Delta G = -S \cdot \Delta T + V \Delta P - \sum_i^M N_i \cdot \Delta \mu_i. \quad (1)$$

The change in the free energy of unfolding relative to urea concentration follows a linear trend, and is quantified using the linear extrapolation method (LEM, Equation 2) with a concentration-dependent energy constant termed the m -value.⁴

$$\Delta G_U(T) = \Delta G_U^\circ(T) - m \cdot [C]. \quad (2)$$

Based on these equations, multiple theoretical models of macromolecular binding, site-specific interactions, and local-bulk preferential interactions (Γ) have been described for the interactions of urea with biomolecules in water (three-component solutions).^{10,29-32} The general form of these relations describe preferential interactions ($\Gamma_{3,2}$, Equation 3) for a cosolute

(3) near a macromolecule (2) relative to interactions of the macromolecule with the solvent (1).³² These interactions are measured in experiments using vapor pressure osmometry (VPO) and dialysis equilibrium³³ to determine the amount of cosolute (N_3) needed to keep a constant chemical potential (μ_3) upon addition of macromolecule (N_2).

$$\Gamma_{3,2} \equiv \left(\frac{\partial \mu_2}{\partial \mu_3} \right)_{N_2, T, P} = - \left(\frac{\partial N_3}{\partial N_2} \right)_{\mu_3, T, P}. \quad (3)$$

These three-component relations have been measured in MD simulations of proteins in solution with urea^{12,13} as well as RNA in solution with ions.³⁴ The cooperative effects of urea and ions have been assessed in experimental analyses of three-component solutions (water, KCL, urea) and have been compared with four-component solutions (water, nucleic acid, KCL, urea). From these studies, the dependence of the chemical potential of urea on the molality of salt ($\frac{\mu_{Urea}}{m_{KCL}}$) is observed to be unaffected by the presence of DNA.^{35,36} These results indicate that a constant salt molality is needed to isolate the effect of the nucleic acid on the urea chemical potential (Equation 3).

In this work we investigate the effects of urea on a hyperstable RNA tetraloop with a GCAA tetraloop-forming sequence (gcGCAAgc). This tetraloop is chosen due to a wealth of background literature describing its thermal stability and configurational heterogeneity.^{26,27,37-40} Additionally, a similarly hyperstable 8-nt DNA tetraloop was shown to be sensitive to the effects of urea, making this an appropriate test system.¹⁷ Our results describe the free energy dependence of unfolding and the global effects of urea on the free energy landscape. We also describe the preferential interactions between RNA and urea and the pairwise energies (van der Waals, electrostatics) involved in these interactions. The most salient structural interactions between RNA, urea and water are also elucidated, and the effects of urea concentration on hydrogen bonding and stacking interactions are described. Our results give new insight into how urea destabilizes RNA by reducing ensemble heterogeneity, yielding a folded ensemble that describes a two-state folding/unfolding relation. The complex interplay of interactions between RNA, water and urea in our simulations serves as

an informative model of this basic RNA motif and can be extended to describe larger RNA molecules in the presence of small cosolute molecules within cells.

Methods

Configurations of the gcGCAAgc sequence are initialized in a right-handed A-RNA stem based on structures formed in isothermal (NVT) replica-exchange molecular dynamics (REMD) simulations at 300 K, 0.1 MPa in 1.0 M KCL, 0 M urea (Figure 1). The molar salt concentration is chosen based on standard state used in the development of secondary structure models for base-pairing and hairpin formation of RNA.⁴¹⁻⁴³ The initial configuration (A_3) is similar to the NMR structure (PDB: 1ZIH), however the sheared base-pair between loop residues G_{L1} and A_{L4} possess a *syn*-glycosidic torsion for A_{L4} instead of the *anti* torsion shown in the crystal. This structure formed more frequently than the experimental structure in unbiased REMD simulations in the absence of urea, and at equilibrium, simulations begun from this structure still accurately described the free energy landscape.²⁷

The number of ions (100 K^+ , 93 CL^-) and water molecules (5168 H_2O) are kept constant to maintain salt molality, while urea molecules are added to produce 2, 4 and 6 M concentrations (Table 1).

Table 1: RNA-Urea Systems (REMD)

Urea (n)	Molarity (\underline{M})	Molality (m)	Edge (nm)
225	2.04	2.416	5.68103
500	4.08	5.369	5.88375
825	6.01	8.859	6.10853

Each RNA-Urea system is equilibrated first in an isothermal-isobaric ensemble (NPT) at 300 K and 0.1 MPa for 10 ns, and then in an NVT ensemble at temperature increments of 25 K from 275 to 500 K for an additional 10 ns. The NPT simulation employs a Berendsen barostat⁴⁴ and a velocity rescaling thermostat⁴⁵ each with coupling constants of 1.0 ps. The NVT simulation utilizes the same velocity-rescaling thermostat with a coupling constant of

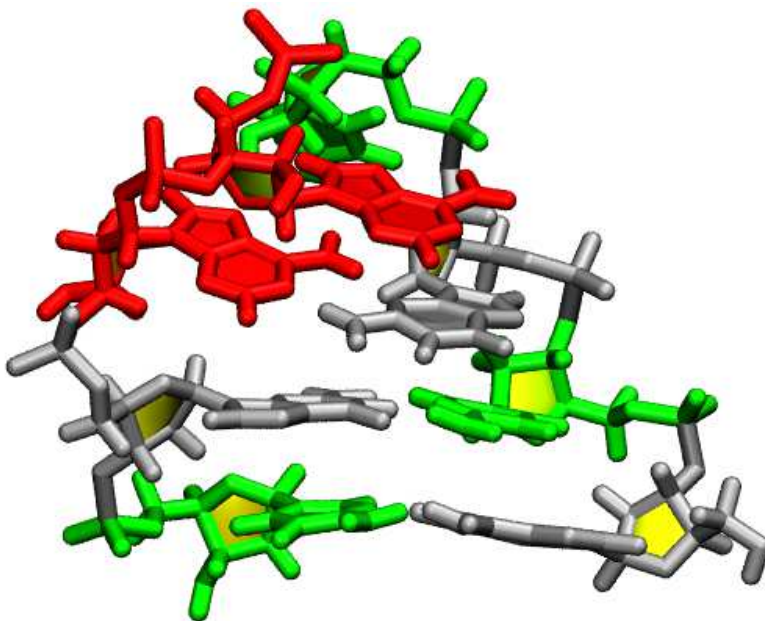


Figure 1: Initial A-RNA configuration of the gcGCAAgc tetraloop. Guanosines (gray), cytosines (green), and adenines (red) in the looping region form a network of hydrogen bonds and base stacks that are stabilized by a canonical stem region between two pairs of guanosine-cytosine base pairs at the 5' and 3' termini.

0.2 ps. The final 5 ns of the NVT simulations are used to generate REMD temperature schedules from 290.00 to 500 K with a 20% exchange probability using a protocol designed for REMD simulations of small biomolecules.⁴⁶ Each RNA-Urea REMD simulation is run at constant volume for 2 μ s using the velocity-rescaling thermostat with coupling constants of 0.2 ps. The first microsecond is treated as equilibration of the free energy landscape, the last microsecond is used for analysis.

All simulations are conducted using GROMACS v4.5.5, in cubic boxes with side lengths of 5.7 nm, 5.9 nm, and 6.1 nm for 2, 4 and 6 M concentrations respectively.⁴⁷ Urea molecules are modeled using a Kirkwood-Buff-derived force field.²⁸ The RNA oligomer is modeled using the parm99 AMBER force field⁴⁸ with modified Lennard-Jones parameters for RNA nucleobases.²⁶ All bonds and angles containing hydrogen are fixed using a fourth-order LINCS algorithm.⁴⁹ Parameters for monovalent ions (K^+ , CL^-) are described by Chen & Pappu,⁵⁰ and the TIP3P water model is used for the solvent.⁵¹ Particle mesh Ewald is used to describe long-range electrostatics with 0.12 nm grid spacing, fourth-order cubic interpolation,

and a 10^{-5} tolerance.⁵² Real-space electrostatics and van der Waals are modeled with cut-off distances of 10 Å, and neighbor lists are updated every ten time-steps. Equations of motion are integrated using a leap-frog integrator with a 2 fs time-step, and frames containing all atoms are saved every 2 ps.

A-RNA folding and dihedral angle principal component analysis

The folded configurations in the A-RNA ensemble are determined based on heavy-atom RMSD < 2.75 Å to the target experimental structure (PDB: 1ZIH),⁵³ and the formation of at least 5 hydrogen bonds between the stem base pairs.

A graphical representation of the heterogeneity in different RNA clusters is achieved through dihedral angle principal component analysis (dPCA).⁵⁴ Using a set of eigenvectors determined in a previous study of this RNA tetraloop²⁷ we describe the relative populations of each A-RNA and alternative configuration at all three urea concentrations using all backbone dihedral angles in the tetraloop ($\alpha, \beta, \gamma, \delta, \epsilon, \zeta$). Additional clustering within the dPCA clusters are defined using RMSD-based clustering.⁵⁵

Calculating preferential interactions (Γ) and m -values

Preferential interaction coefficients ($\Gamma_{3,2}$) are calculated from proximities of cosolutes (3) to the macromolecule (2) in solvent (1) for A-RNA and unfolded configurations. A two-domain representation of Equation 3 is achieved in Equation 4, where the number of solvent (1) and cosolute (3) particles are measured over a given distance from the macromolecule (2). The distances are measured from all heavy RNA atoms to the central atom of each solvent and cosolute molecule (O_W and C_U , respectively). In each case, the shortest pairwise distance is recorded. The difference between the observed and expected numbers of cosolute particles is determined using the 'bulk' ratio of total cosolute and solvent molecules $\left(\frac{N_3}{N_1}\right)_{bulk}$.

$$\Gamma_{3,2} = \left\langle N_{3,local} - N_{1,local} \cdot \left(\frac{N_3}{N_1} \right)_{bulk} \right\rangle. \quad (4)$$

Values for $\Gamma_{3,2}$ plateau between 5 and 6.5 Å, and it is the value of $\Gamma_{3,2}$ in this range that is used to measure differences between $\Gamma_{3,2}$ for unfolded and A-RNA configurations ($\Delta\Gamma_{3,2}$). From this value, a measure of the urea m -value (\tilde{m}) can be calculated for a given temperature (T) and urea concentration (C) using the relation:

$$\tilde{m} \simeq -\frac{R \cdot T \cdot (\Delta\Gamma)}{C}. \quad (5)$$

Another method for determining the m -value utilizes the temperature-pressure expansion of the free energy of unfolding (ΔG) for two-state biomolecular systems described by Hawley et. al.^{12,56,57}

$$\begin{aligned} \Delta G(T, P, C) = & \Delta G^\circ + m_1 \cdot [C] + \left(\Delta S^\circ + m_2 \cdot [C] \right) \cdot (T - T_0) \\ & + \left(\Delta V^\circ + m_3 \cdot [C] \right) \cdot (P - P_0) + \Delta\alpha^\circ \cdot (P - P_0) \cdot (T - T_0) \\ & - \left(\Delta C_p^\circ + m_4 \cdot [C] \right) \cdot \left(T \cdot \left(\ln \left(\frac{T}{T_0} \right) - 1 \right) + T_0 \right). \end{aligned} \quad (6)$$

Here ΔG is defined by changes in entropy (ΔS), volume (ΔV), linear expansion coefficient ($\Delta\alpha$) and heat capacity (ΔC_p) relative to temperature (T) and pressure (P). This relation can be used to calculate an m -value by taking the derivative of ΔG with respect to C:

$$\begin{aligned} m(T, P) = & -\frac{\partial \Delta G(T, P, C)}{\partial C} \\ = & -m_1 - m_2 \cdot (T - T_o) - m_3 \cdot (P - P_o) - m_4 \cdot \left(T \cdot \left(\ln \left(\frac{T}{T_o} \right) - 1 \right) + T_o \right). \end{aligned} \quad (7)$$

Notice that the m -value is a function of temperature and pressure ($m(T,P)$). In order

to assess the error of fitting to the Hawley equation, a bootstrapping procedure was implemented, whereby 50000 random data points were drawn with replacement from each temperature-pressure ensemble (the equivalent of 100-ns worth of timesteps), and repeated 500 times.^{58,59}

RNA interactions.

Hydrogen bonding

In order to measure the relative favorability of different interactions between the RNA and all molecules in solution the same criteria for hydrogen bonding was applied to RNA, water and urea. Hydrogen bonding employs a cut-off distance of 3.5 Å and an incident angle of 150° between the acceptor and donor groups. Increasing the acceptable angle only increases the raw number of interactions without affecting ΔN^{HB} , indicating robustness in the measurement.

Stacking

Stacking between any two RNA nucleobases (RNA:RNA), or an RNA nucleobase and a urea molecule (RNA:Urea) is defined by pairwise distances of their geometric centers and angles between the normal vectors.

For each nucleobase, a geometric center is defined as the center of a circle intersecting the C₂, C₄ and C₆ atoms (the pyrimidine ring). For urea molecules, the geometric center is taken as the center of a circle intersecting the O_U, N_{U1} and N_{U2} atoms. Vertical and horizontal components of the pairwise distances between geometric centers are defined as follows:

1. Vertical displacement (V_{disp}) is determined by defining a plane intersecting the three atoms of one nucleobase and determining the shortest distance between this plane and the geometric centers of another nucleobase or the nearest urea molecule. This takes the form of a normal vector from the nucleobase plane.

- Horizontal displacements (H_{disp}) are determined by taking the geometric center of the nucleobase atoms in the plane and measuring the distance from the geometric center to the V_{disp} vector where it points out of the plane and intersects the geometric center of the nearest nucleobase or urea molecule.

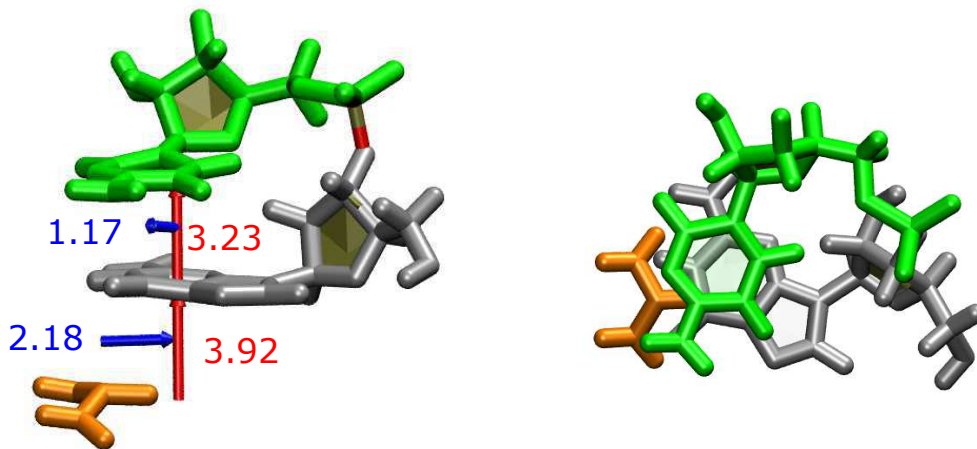


Figure 2: Stacking interactions between RNA nucleobases (gray, green) and urea (orange). The side view (left) shows RNA:RNA and RNA:Urea stacks with vertical (red) and horizontal (blue) offsets of geometric centers (given in Å). In the top view (right) RNA nucleobases are shown with pyrimidine rings overlaying one another and a urea molecule stacking to the guanosine.

Importantly, interactions between any two chemical moieties (nucleobases or urea) considered to be stacking cannot include hydrogen bonds, and angles between the normal vectors of both moieties cannot exceed 50° .

To determine appropriate stacking cut-offs for both RNA:RNA and RNA:Urea interactions, folded (A-RNA) and unfolded configurations are equilibrated using the same NPT and NVT methods for the REMD simulations. Quadruplicate simulations of both configurations are conducted using NVT simulations at 350 K, 0.1 MPa and 6 M urea for 100 ns each. Stacking between RNA nucleobases and urea is determined based on the pairwise vertical displacement (V_{disp}) exceeding the pairwise horizontal displacement (H_{disp}), and the total distance of the geometric centers not exceeding 5 Å. Intermediate stacking interactions

(one-half stack) follow the same rules of V_{disp} and H_{disp} , with total distances greater than 5 Å and less than 6 Å (see Supporting Information).

A check on the robustness of this metric was made by including the highly populated interactions where H_{disp} is less than V_{disp} and less than 4 Å. This region includes stacking interactions with the imidazole component of purine rings. While inclusion of this region doubles the total number of stacking interactions, the trends relative to temperature and concentration remain the same. Thus our metric captures the essential temperature and concentration-dependent details of RNA:RNA and RNA:Urea stacking.

All geometrically-defined interactions were measured using an in-house code developed from the MOSCITO4 framework.⁶⁰

Results and Discussion

Thermodynamic effects of urea on RNA tetraloop folding.

In the equilibrated ensembles of these simulations the majority of configurations have sampled the unfolded ensemble after the first microsecond and most of the A-RNA configurations in the equilibrated ensemble result from refolding of an unfolded precursor. This observation, and the fact that the ensembles reach their equilibrated populations at around 500 ns (see Figure S1 in Supporting Information), suggests that our simulations are of sufficient length to sample the free energy landscape and allow us to describe the equilibrium thermodynamics for the RNA tetraloop as a function of temperature and urea concentration.

A-RNA configurations are identified based on low heavy-atom RMSD (<2.75 Å) of the tetraloop stem (G_1, C_2, G_7, C_8) to the experimental structure (PDB: 1ZIH) and at least five hydrogen-bonds between the two base-pairs. The average number of folded and unfolded configurations is determined over the equilibrated ensemble, and the populations at each temperature are plotted over the temperature range (Figure 3A & B).

For the thermodynamics of the RNA tetraloop the most explicit effect of increasing

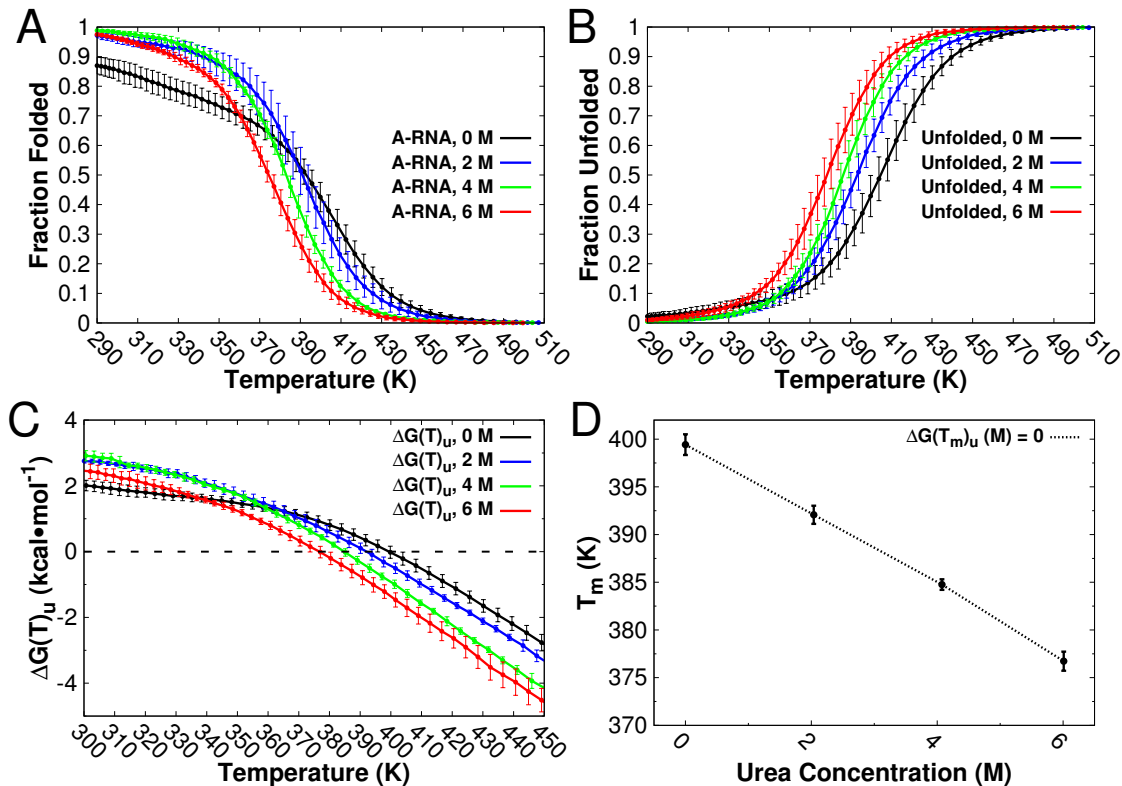


Figure 3: Tetraloop folding/unfolding as a function of urea concentration. **(A)** Ensemble fraction of A-RNA tetraloops and **(B)** Unfolded configurations vs. temperature are shown for all urea concentrations (errors are determined from block averages of 333 ns each). **(C)** Free energy change upon unfolding ($\Delta G(T)_u$). **(D)** Melting temperatures (T_m) derived from $\Delta G(T)_u$ show linear reduction relative to urea concentration.

urea concentration is the decrease in melting temperature and the increase in cooperativity for the melting curve (Figure 3A & B). We compare our results to a REMD simulation of the same GCAA tetraloop over the same temperature range at 0 M urea (shown in black). In this control simulation the formation of alternative folded configurations produces a non-cooperative folding profile at low temperatures.²⁷ As urea is added, the temperature-dependent population profiles become more cooperative, showing greater A-RNA populations at lower temperatures. The melting curves for the A-RNA configurations fit a two-state relation for the fraction folded ($f = \frac{e^{\Delta G_u(T)}}{1+e^{\Delta G_u(T)}}$) only in the presence of urea while the melting curves for the unfolded configurations consistently fit a two state relation even in 0 M urea.

We observe that the temperature at which half of the population adopts the A-RNA configuration ($T_{1/2}$) is essentially unchanged between 0 M and 2 M, even as the unfolded

population increases (Figure 3B). These observations indicate that the behavior of urea at low concentrations involves destabilization of alternative configurations, which matches with experiments in RNA electrophoresis that show narrowing of bands for single-stranded oligonucleotides.²⁰ As the concentration of urea increases, the sigmoidal curve is maintained while the value of $T_{1/2}$ for the folded ensemble is systematically reduced and the unfolded ensemble population increases.

The free energy of unfolding for the tetraloop can be described based on the populations of A-RNA (N_F) and unfolded (N_U) configurations using the relation $\Delta G_U = -R \cdot T \cdot \ln\left(\frac{N_U}{N_F}\right)$ (Figure 3C). By taking the temperature where the free energy is zero, the melting temperature (T_m) can be determined (Figure 3D, Table 2). A near linear trend is described with this data, showing a concentration-dependent decrease in T_m of $-3.75 \pm 0.09 \frac{K}{M_{urea}}$. These results are comparable to experimental measurements of DNA duplexes that show changes in melting temperatures as a function of urea concentration ($\sim -3 \frac{K}{M_{urea}}$).^{16,36}

Table 2: Concentration-dependent melting temperatures (T_m)

Urea Molarity (M)	Melting Temperature (T_m , K)
0	399 ± 1
2.04	392.1 ± 0.9
4.08	384.8 ± 0.5
6.01	376 ± 1

From these results we see that urea has the anticipated effects observed in experiment. The thermodynamics of the RNA tetraloop are reduced to a two-state system, and the alternative configurations are depleted.

Free energy landscapes and clustering analyses of RNA tetraloops at different urea concentrations.

The melting curves for the folded state indicate a reduction in the population of alternative configurations at low to intermediate temperatures, however the detailed effects of urea on the free energy landscape of the GCAA tetraloop require more in-depth structural analysis.

Clustering through dihedral angle principal component analysis (dPCA)⁵⁴ is employed using two eigenvectors that effectively parse the free energy landscape into six general configurations at the lowest ensemble temperature (Figure 4). These configurations, circled in blue at 0 M urea (Figure 4A) show four general types of A-RNA configurations (A_1 - A_4), as well as two alternative configurations: Z-RNA (Z) & Triplet (T). Examples of each configuration are shown in Figure S2 in the Supporting Information.

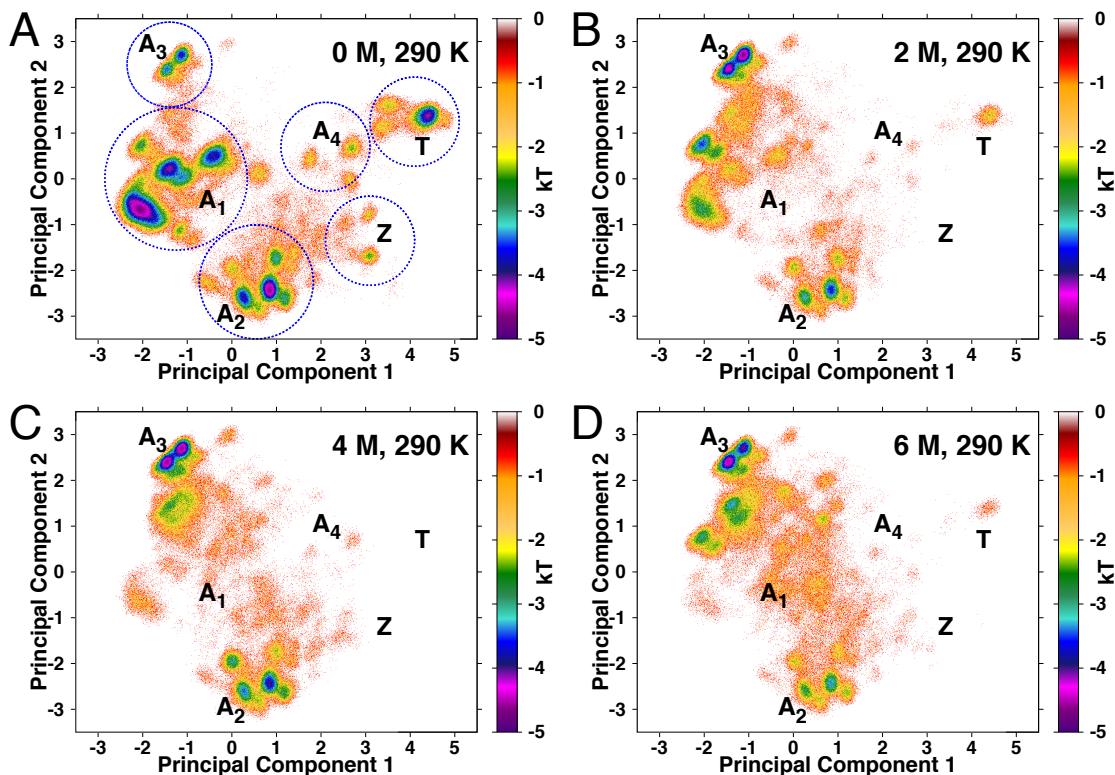


Figure 4: Free energy landscapes of GCAA tetraloop (290 K) described by dPCA. Labels represent the common configuration associated with each collection of peaks. (A) In 0 M urea six groups are well represented. (B)-(D) As urea concentration increases, the number of alternative (Z,T) peaks diminishes, and the favorability of several A-RNA peaks (A_1 , A_4) are reduced significantly.

In low urea concentrations (2 M) the two alternative configurations are almost completely eliminated from the folded ensemble, and two configurations (A_1 & A_4) are markedly reduced (Figure 4B). Despite these population shifts, A-RNA heterogeneity is retained: A_2 forms regularly and an additional cluster appears in the space between the A_1 and A_3 peaks, representing a variation of A_3 with the same sheared loop base-pair and only a slight change in

the backbone dihedral angles. The sampled configurations are in agreement with our analyses of global A-RNA and unfolded populations, but now include specific information about the distribution of different A-RNA substates. The heterogeneity of loop configurations is clearly reduced, and configurations with extended nucleobases (A_1) form less often than more native-like (A_3) substates.

While the majority of configuration types show some reduction as the urea concentration increases, A_1 shows very specific, concentration-dependent reduction. At 290 K, clustering reveals that A_1 tetraloops represent nearly 50% of all states in 0 M urea. The A_1 population is reduced to 25% in 2 M urea, and <10% in higher (>4 M) concentrations. A less-pronounced reduction is observed for A_2 structures, going from \sim 30% at 0 M, to \sim 20% at 6 M.

Alternative configurations, as shown in the trajectories of individual replicas (Figure S1) and the dPCA landscape, are the most significantly reduced. In 0 M, Z-RNA and Triplet configurations account for \sim 5% and \sim 13% of all configurations respectively, while Z-RNA accounts for <1% in all urea concentrations, and Triplets account for <5%.

Urea-induced exposure of RNA to solution

For the unfolded tetraloop configurations, increased urea concentration and temperature increase the level of RNA exposure to solution (Figure 5).

In the absence of urea (0 M) and at temperatures below 380 K, the unfolded configurations adopt more compact structures than A-RNA ($R_g < 7.8 \text{ \AA}$) and show reduced exposure to solution ($\Delta ASA_U < 0$). This indicates that low-temperature unfolding in the absence of urea produces highly condensed RNA structures. For systems containing urea there are markedly significant increases in both R_g and ΔASA_U . The differences between 0 and 2 M urea are quite significant ($+0.5 \text{ \AA } R_g$, $+75 \text{ \AA}^2 \Delta ASA_U$ at 320 K), as are those between 2 and 4 M urea ($+0.25 \text{ \AA } R_g$, $+40 \text{ \AA}^2 \Delta ASA_U$ at 320 K). Going from 4 to 6 M urea, low temperature unfolded states show only minimal increases in either metric.

The difference in ΔASA_U between 0 and 6 M urea ($+120 \text{ \AA}^2$ at 320 K) is approximately

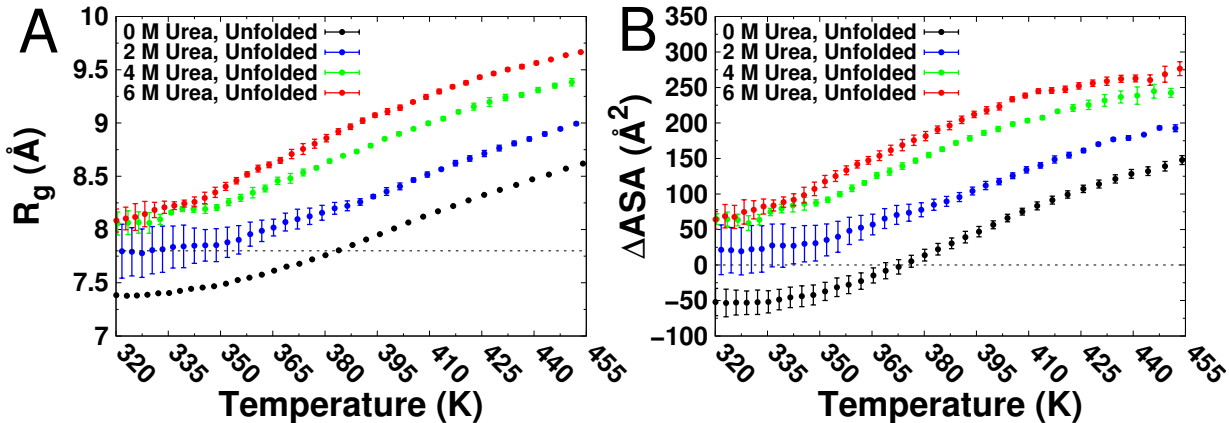


Figure 5: Changes in unfolded state structure as a function of temperature and urea concentration. **(A)** Radius of gyration for unfolded configurations (R_g) relative to temperature for different urea concentrations. The R_g for A-RNA ($\sim 7.8 \text{ \AA}$) is given for reference. **(B)** Change in accessible surface area upon unfolding (ΔASA_U) relative to temperature for different urea concentrations. The similarity in trend lines within each plot indicates that urea increases the baseline of exposure for unfolded RNA structures. The similarity between the two plots indicate that changes in R_g are commensurate with changes in exposed surface area of RNA upon unfolding.

half the area of an average nucleobase as determined by Lambert et. al.²¹ It is important to recognize that in the native RNA tetraloop configuration only the planar surfaces of a few RNA nucleobases are not exposed solution, so only a moderate value of ΔASA_U is expected.

To further characterize the effect of urea on the solution exposure of RNA, we next detail the direct, preferential interactions between urea and RNA using preferential interaction coefficients (Γ).

Preferential interactions of urea with RNA.

RNA structures in low temperature regions of all three urea concentrations are almost universally folded, but at $\sim 350 \text{ K}$ small populations of unfolded configurations begin to emerge. Utilizing the direct, preferential interactions between urea and RNA (Γ_U) based on Equation 4, the results at $\sim 350 \text{ K}$ are detailed in Figure 6. A cut-off of 6.5 \AA is chosen to calculate Γ_U from these plots in order to capture the effect of the first solvation layer and the most proximate urea molecules. Specific interactions of urea with RNA nucleobases and the back-

bone are analyzed separately, and the chosen cut-off helps to delineate directly-interacting urea from urea molecules that are interacting with one group and consequently fall into close proximity to another (e.g. urea that interacts with a nucleobase is in close proximity to the backbone).

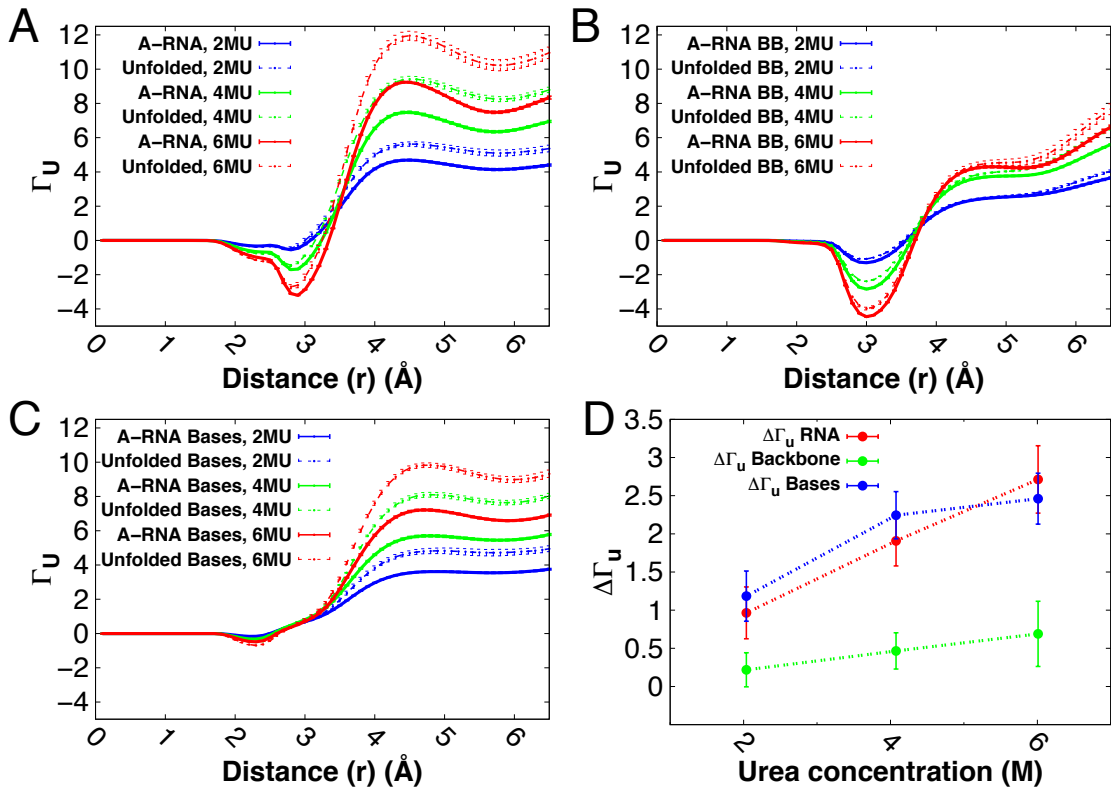


Figure 6: (A)-(C) Preferential interaction of urea (Γ_U) relative to distance (r) from different RNA groups (All, Backbone, Bases) at 2 M (blue), 4 M (green), and 6 M (red) are shown for A-RNA (solid), and unfolded (dashed) configurations. (A) Total RNA:Urea interactions for both A-RNA and unfolded configurations increase with concentration with a region of urea exclusion at 3 Å. (B) Backbone:Urea interactions increase less significantly and are minimally affected by unfolding. (C) Base:Urea interactions follow similar trends to total RNA, but show no significant regions of urea exclusion. (D) The change in Γ_U upon unfolding ($\Delta\Gamma_U$) increases almost linearly relative to concentration and is dominated by Base:Urea interactions.

Unfolded configurations consistently show a higher Γ_U compared to A-RNA configurations (Figure 6A). Preferential exclusion ($\Gamma_U < 0$) of urea is observed for the total RNA and the RNA backbone at 3 Å (Figure 6A & B) but favorable interactions for the RNA nucleobases are consistently observed (Figure 6C), and the difference in preferential interactions

increase with concentration (Figure 6D). These observations show that the concentration-dependent increase in RNA-urea interactions for the entire RNA tetraloop ($\sim 0.44 \pm 0.08 \frac{\Delta\Gamma_U}{M_{urea}}$) are driven primarily by interactions with RNA nucleobases. Similar, though less extreme, trends are observed for counterion interactions (Γ_{K^+}) suggesting that the preferential interactions of urea may also be exposing the RNA to increased preferential interactions with smaller cosolute particles (Figure S3).

From these thermodynamic and structural data it becomes possible to determine the free energy dependence of tetraloop unfolding on urea concentration.

Calculating m -values

RNA tetraloop m -values are an important indicator for how urea affects the energetic favorability of folded or unfolded configurations. To compare with experimental measures of RNA m -values our results are described using units of ($\text{kcal}\cdot\text{mol}^{-1}\cdot\text{m}^{-1}$).

We determine a rough temperature-dependent m -value using $\Delta\Gamma_U$ as described in Equation 5.⁶¹ Due to the similar populations of A-RNA and unfolded configurations over the range of melting temperatures defined in Figure 3D the average results were taken for 360-400 K in 10 K increments. Above 400 K and in high urea concentrations the A-RNA populations are too small and produce significant errors, while below 360 K and in low urea concentration the unfolded populations are too small and produce errors as well.

A more general m -value is determined by fitting the potential energies, temperature, pressure, and urea concentrations of a given temperature range to a Hawley-type free energy equation (Equation 7).^{12,56} Using an expanded temperature range (365-425 K) and their associated pressures we fit a smooth curve to the thermodynamic data that describes an overall weakening of the m -value at increased temperature and pressure, similar to the points for \tilde{m} . In Figure 7, the curve of the $-m$ -value, and its components are displayed: the reference m_1 , temperature-dependent $m_2\cdot(T-T_o)$ and $m_4\cdot\left(T\cdot\left(\ln\left[\frac{T}{T_o}\right]-1\right) + T_o\right)$, and pressure-dependent $m_3\cdot(P-P_o)$. Fitted values for Hawley-type equation parameters are provided in Table 3.

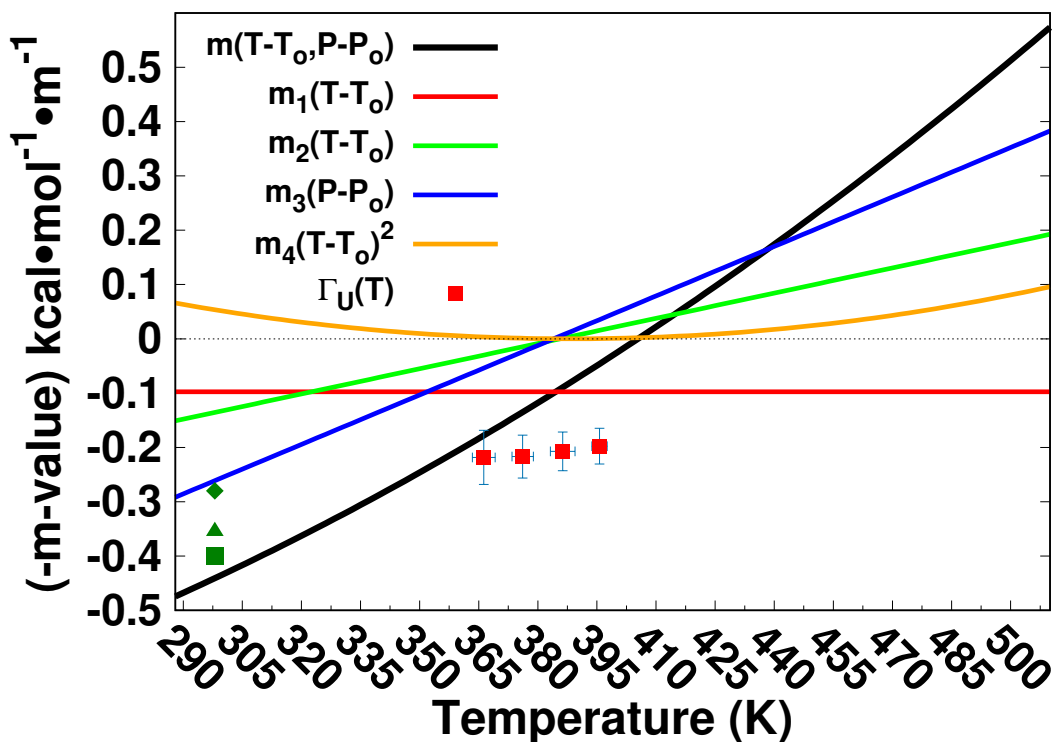


Figure 7: Temperature-dependent $-m$ -value calculations ($\frac{\partial(G)}{\partial(C)}$). Negative of m -values derived from $\Delta\Gamma_U$ in the T_m ranges for all urea concentrations (red) is compared to fitting the Hawley equation (black). The individual $-m_i$ values are included to show the contributions of temperature and pressure to $-m$. The pressure is the ensemble-averaged pressure obtained at the corresponding temperature in the REMD simulations ($P=\langle P(T)\rangle$). The $\Delta\Gamma_U$ range is limited due to increasing errors at high and low temperatures. Unscaled experimental and calculated $-m$ -values of a non-hyperstable RNA tetraloop with a UUUU loop sequence and a 24-nt stem (green) are included for reference at KCL concentrations of 100 mM (square), 200 mM (triangle) and 1 M KCL (diamond).²¹

For the fitted parameters, all pressure-dependent (ΔV° , $\Delta\alpha^\circ$, m_3) and second order (C_P , m_4) parameters show significant errors, which concur with results for small globular proteins analyzed by the same technique.⁵⁸

Table 3: Fitted parameters of the Hawley Equation. Reference (385.6 K, 165 MPa, 0 m).

ΔG_U°	$0.39 \pm 0.06 \frac{\text{kcal}}{\text{mol}}$	m_1	$0.098 \pm 0.006 \frac{\text{kcal}}{\text{mol}\cdot\text{m}}$
ΔS_U°	$0.028 \pm 0.003 \frac{\text{kcal}}{\text{mol}\cdot\text{K}}$	m_2	$-0.0015 \pm 0.0005 \frac{\text{kcal}}{\text{mol}\cdot\text{K}\cdot\text{m}}$
ΔV_U°	$-5 \pm 1 \frac{\text{mL}}{\text{mol}}$	m_3	$-0.013 \pm 0.27 \frac{\text{mL}}{\text{mol}\cdot\text{m}}$
ΔC_P	$0.068 \pm 0.47 \frac{\text{kcal}}{\text{mol}\cdot\text{K}}$	m_4	$-0.005 \pm 0.006 \frac{\text{kcal}}{\text{mol}\cdot\text{K}\cdot\text{m}}$
$\Delta\alpha$	$-0.079 \pm 0.46 \frac{\text{kcal}}{\text{mol}\cdot\text{K}\cdot\text{MPa}}$		

While no known experimental m -values exist for the GCAA tetraloop, a comparison can be made to experimental m -values for an RNA tetraloop with a longer stem (12-bp) and a loop sequence that is not hyperstable (UUUU, green points in Figure 7). To compare the two systems we need correct for the thermodynamic state (i.e. temperature and pressure at which the system unfolds), and systems size. There is significant overstabilization of the folded GCAA tetraloop in simulation compared to experiment ($T_m \approx 311$ K)¹⁷ which consequently increases the temperature of unfolding by approximately 75 K.²⁷ Therefore, we use a shifted reference state (385.6 K, 165 MPa) when comparing m -values. The calculated $-m$ -value (-0.098 kcal·mol⁻¹m⁻¹) is approximately one-third the $-m$ -value of the longer UUUU tetraloop in 1 M KCL (-0.28 kcal·mol⁻¹m⁻¹) at its reference state (298 K, 0.1 MPa). Since the length of the GCAA tetraloop sequence is also roughly one-third of the UUUU tetraloop, the results suggest that the m -value near the melting temperature of the GCAA tetraloop is of the same order of magnitude as the experimental value, given the corrections described above.

The fact that both the $\Delta\Gamma$ and Hawley measures of the $-m$ -value show similar values near the simulation melting temperature suggests that both methods offer a valid means of assessing the effect of urea on the free energy of RNA tetraloop folding through simulation, though a more significant temperature dependence is observed for the Hawley-based m -value.

From the Hawley fitting, the free energy of unfolding (G_U) can be split into enthalpic and entropic components (Figure 8). The enthalpy (H_U) is by far the dominant contributor to unfolding at the reference state, while the entropic contribution is nearly an order of magnitude weaker.

To better characterize the concentration dependence of the free energy landscape, the enthalpy (H_U) can be broken into internal energy (E_U) and pressure-volume work ($P \cdot (V_U + m_3 \cdot [C])$). Since these simulations are run at constant volume, and the concentration-dependence of the change in volume upon unfolding is weak, it is the internal energies where the greater effect is taking place and where further analysis must be directed.

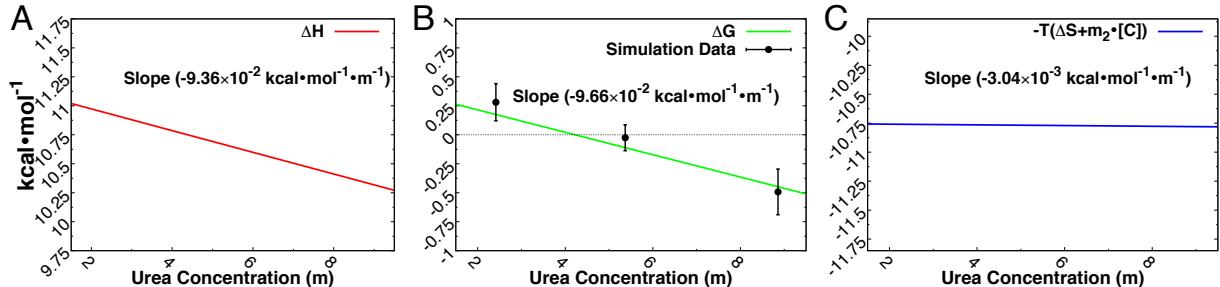


Figure 8: Thermodynamic partitioning in urea denaturation, Reference state (385.6 K, 165 MPa, 0 M). Urea-concentration dependence in terms of (A) Enthalpy (ΔH_U) shows strong concentration dependence. (B) Free energy (ΔG_U) shows a very similar concentration-dependence, with equilibrium ($\Delta G_U=0$) at 4.01125 m. (C) The product of temperature with entropy and its associating m_2 -value ($T \cdot (\Delta S_U + m_2 \cdot [C])$) shows only a weak dependence on urea concentration.

The system entropy (S_U) can be assessed through analysis of the interactions between RNA and different solution components. Structural interactions including hydrogen bonds and base-stacks all describe a loss of degrees of freedom for the system and thus a general reduction in system entropy. The concentration-dependence and temperature-dependence of these interactions will be necessary in order to provide insight into the effect of urea and temperature on entropy.

Analyses of both H_U and S_U will help elucidate how the free energy landscape of the RNA tetraloop is smoothed by the presence of urea.

Characterization of short-range energies upon tetraloop unfolding.

The differences in non-bonding (Lennard-Jones (LJ) & Coulombic) energies between the A-RNA and unfolded configurations provide valuable insight into the internal energy changes that accompany unfolding of RNA tetraloops as a function of concentration and temperature (Figure 9). From the slope of $\Delta G_U(m)$ it is clear that urea concentration makes unfolding an energetically favorable process and this trend is recapitulated for both urea-based and water-based interactions with RNA.

Population distributions in increasing temperature space universally favor unfolding, and

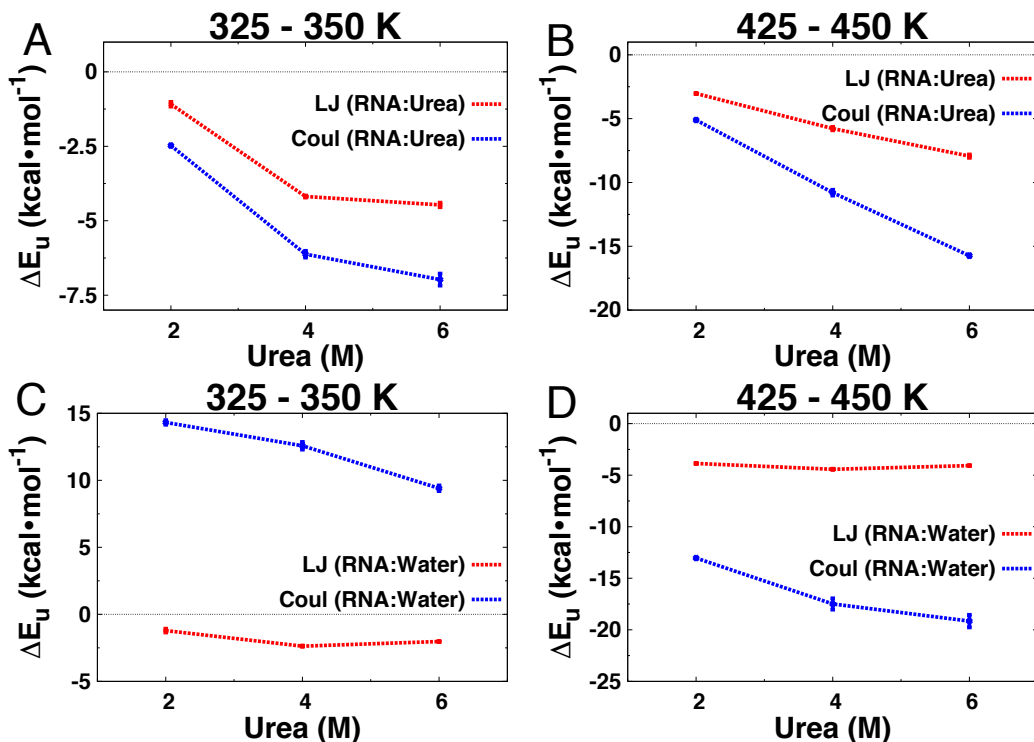


Figure 9: Changes in non-bonding (LJ & Coulombic) energies upon tetraloop unfolding taken in all three urea concentrations, over temperature ranges (325-350 K & 425-450 K). (A)-(B) RNA-Urea interactions show that urea universally favors interactions with unfolded RNA. (C)-(D) RNA-Water interactions show a greater dependence on temperature than concentration, where water favors the folded ensemble in low temperature and the unfolded ensemble in high temperatures. Energy is dominated by Coulombic interactions in all cases.

this is recapitulated in the increased favorability of interactions between RNA and different solution components at higher temperatures (Figure 9B & D). Temperature also plays a significant role in the trade-off of favorable ΔE_u . At low temperatures, all RNA:Urea interactions (-1 to -7.5 kcal·mol⁻¹) are significantly more favorable than all RNA:Water (15 to -3 kcal·mol⁻¹), whereas in high temperatures the Coulombic interactions for RNA:Water (-12.5 to -18 kcal·mol⁻¹) tend to dominate. Only at 6 M are RNA:Urea and RNA:Water Coulombic interactions nearly the same (-16 kcal·mol⁻¹ and -18 kcal·mol⁻¹, respectively).

Increases in temperature also make LJ interactions more favorable for RNA with both water and urea (-2.5 to -7.5 kcal·mol⁻¹) though RNA:Urea maintains a strong concentration dependence while RNA:Water does not show significant concentration dependence, except for Coulombic interactions (Figure 9D).

The concentration of urea increases the favorability of unfolding for all observed interactions except RNA:Water LJ, which is dominated by temperature and relatively invariant to concentration.

From these analyses we see that increasing the concentration of urea manifests favorability of unfolding based on the energy of short-range internal energies, and is dominated by Coulombic interactions. This suggests that addition of urea, and unfolding of RNA, are likely affecting the orientation of the dipole moments for both water and urea with respect to RNA. These changes can manifest in the formation of hydrogen-bonding interactions between RNA and both water and urea, and this will consequently affect the entropy of solution at the RNA surface.

Changes to solution interactions upon unfolding.

Hydrogen-bonding interactions

The precise interactions made between RNA and different solution components are important when considering how cosolutes might perturb the RNA free energy landscape and how the interactions with unfolded configurations create favorable free energies of unfolding. While Γ_U calculations show that RNA favors proximity of urea over water, the precise interactions of urea and RNA cannot be ascertained from Γ_U alone. Hydrogen bonds (HBs) between RNA, water and urea lend insight into how temperature and urea concentration can affect the favorability of these contacts (Figure 10A & B). Additionally, deconvolving these interactions into nucleobase-specific and backbone-specific interactions can provide better site-specific details about where water and urea associate on the RNA (Figure 10C & D).

Upon unfolding, RNA:RNA HBs are reduced significantly and show both temperature-dependent and concentration-dependent loss. At low temperatures RNA:Urea HBs nearly compensate for the loss in RNA:RNA HBs and show a very similar concentration-dependent trend. By contrast, RNA:Water HBs are actually reduced upon unfolding, but show weaker and weaker reduction as the urea concentration increases. At higher temperatures, RNA:RNA

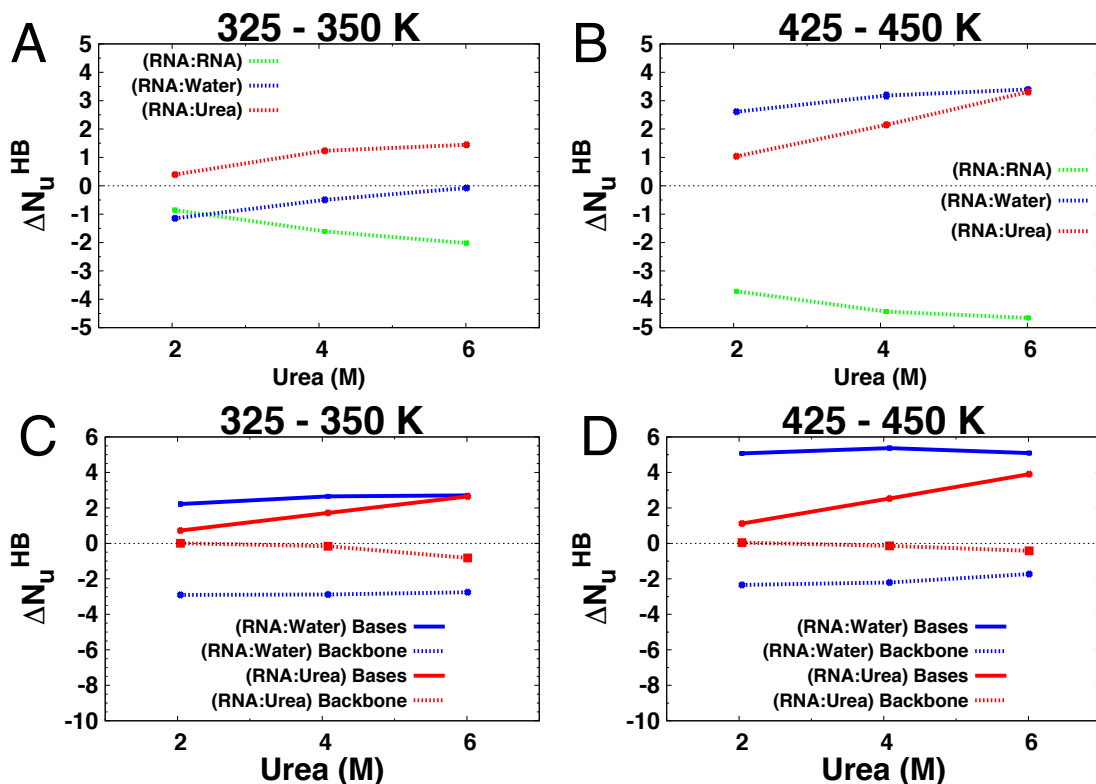


Figure 10: Concentration-dependent changes in HBs (ΔN_U^{HB}) upon unfolding. (A) Low temperature and (B) high temperature measurements show that RNA:Water and RNA:Urea interactions always increase with concentration while intra-RNA HBs decrease. (C)-(D) Deconvolution of the RNA:Water and RNA:Urea interactions into nucleobase (solid) and backbone (dashed) interactions show that upon unfolding, nucleobase interactions always increase, and backbone interactions remain constant or decrease.

HBs decrease and RNA:Water HBs actually exceed RNA:Urea at all but 6 M urea where both RNA:Water and RNA:Urea HBs are equal. RNA:Urea HBs show the sharpest concentration dependence, though at high temperatures the scale of broken RNA:RNA HBs exceeds that of RNA:Urea, and the concentration-dependent trend of RNA:RNA more closely matches RNA:Water. Both of these observations agree well with the energy calculations that show RNA:Water dominating high temperature unfolding.

The observed divergence in concentration-dependence in RNA:RNA HBs indicates that RNA:Urea HBs are not necessarily displacing RNA:RNA or RNA:Water HBs. Overall, the results suggest that temperature plays a more significant role in RNA:RNA HB breaking and RNA:Water HB formation while urea concentration plays a more important role in

RNA:Urea HB formation.

When the RNA:Urea and RNA:Water HBs are decomposed into backbone and nucleobase contacts it is found that the nucleobase interactions are always positive, and backbone interactions are always negative or near zero (Figure 10C & D). Water interactions and Backbone:Urea interactions are somewhat invariant to concentration and temperature.

Given that Γ_U for Backbone:Urea shows a very small change upon unfolding, the near-zero values for Backbone:Urea HBs are understandable. Nucleobase HBs show differential dependencies on the environment. Nucleobase:Water HBs are predominantly affected by temperature and accumulate ~ 3 additional HBs as temperatures increase.

In contrast, Nucleobase:Urea HBs are predominantly affected by concentration and increase linearly with the amount of urea in solution. It is worth noting that Nucleobase:Urea HBs also show a systematic increase as a function of temperature. This last observation suggests that the increase in ASA at elevated temperatures is complimented by an increase in the number of direct interactions between the RNA and urea.

Stacking interactions

Metrics of RNA stacking (vertical and horizontal displacements) were generated from analyses of 100-ns NVT simulations of A-RNA and unfolded configurations in 6 M urea and show similar cut-offs for both RNA:RNA and RNA:Urea interactions (Figure S4). The changes in stacking upon unfolding reveal weak dependence on urea concentration and greater dependence on temperature for RNA:RNA stacks (Figure 11).

RNA:Urea stacking show a minor increase with concentration, but almost no difference as a function of temperature. At high temperatures, where unfolded configurations have large ΔASA_U , the loss of RNA:RNA stacks is nearly double the gain in RNA:Urea stacks regardless of concentration. Interestingly, the raw numbers of RNA:RNA stacks decrease for both A-RNA and unfolded configurations as a function of temperature, but the raw numbers of RNA:Urea stacks change only minimally (Figure 11C & D).

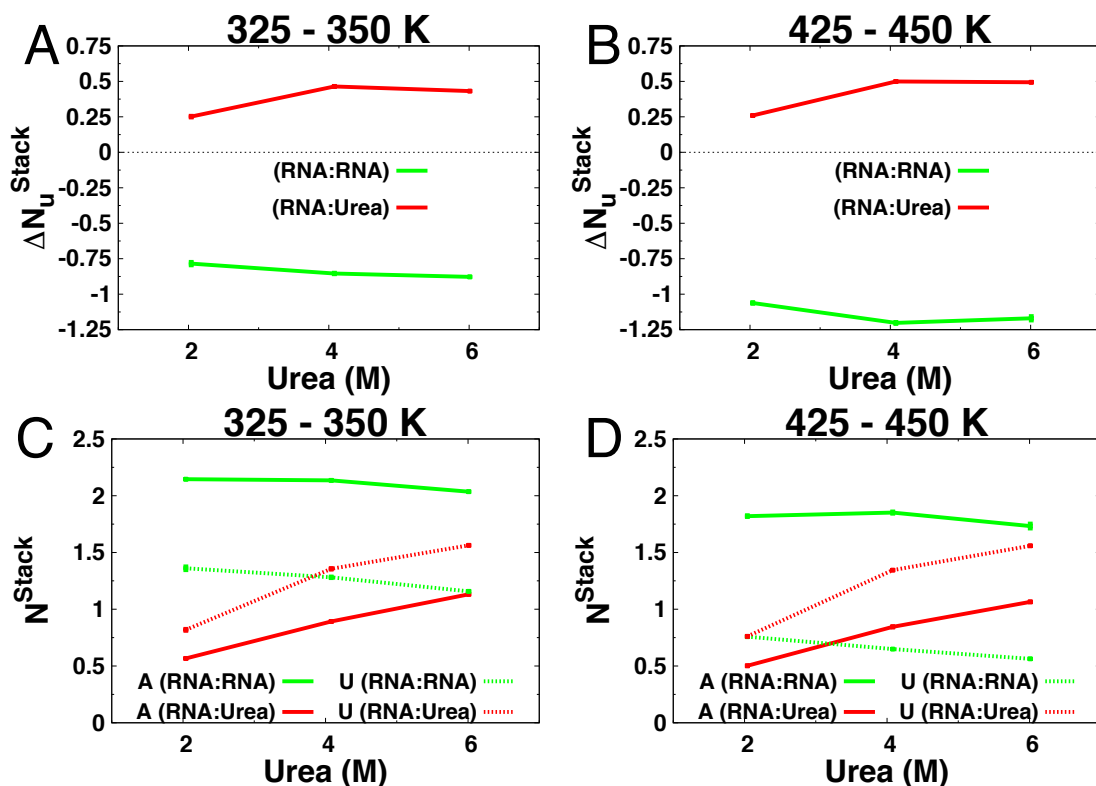


Figure 11: Changes in stacking upon unfolding (ΔN_U^{Stack}) as a function of temperature and urea concentration. (A) Low temperatures show RNA:RNA stacking decreasing and RNA:Urea stacking increasing marginally. (B) Higher temperatures show RNA:RNA stacking decreases even further while RNA:Urea remains relatively unchanged. (C) & (D) Stacking in A-RNA (A: solid) and unfolded (U: dashed) configurations show similar trends for both RNA:RNA and RNA:Urea interactions relative to urea concentration. High temperatures diminish the total number of RNA:RNA stacking interactions for both A-RNA and unfolded configurations, but only weakly affect total RNA:Urea stacking interactions.

These observations suggest that while RNA:Urea stacks are forming, and form more frequently in high urea concentration, the interactions are not necessarily supplanting RNA:RNA stacks. Retention of RNA:RNA stacking contacts in unfolded configurations is not significantly affected by urea concentration. These observations agree with numerous experiments of different RNA structures with helical motifs.^{21,22}

Conclusions

The results of our simulations concur with a plethora of experiments that describe multiple phenomena of RNA in the presence of urea. Our use of molecular dynamics allows us to determine how these phenomena manifest on the microscopic scale and permit quantitative measurements of reduced RNA heterogeneity, free energy dependence on urea, and interactions of RNA and urea that are responsible for RNA denaturation.

At the lowest temperature in our simulations (290 K) urea primarily denatures alternative configurations increasing the relative A-RNA population from $\sim 87\%$ at 0 M urea to $\sim 98\%$ at 6 M urea. Within the free energy landscapes described using dPCA the alternative configurations are depopulated while the heterogeneity and relative favorability of different configurations within the native ensemble is shifted. The most favored A-RNA configuration at 0 M urea (A_1) is reduced from 50% to $<10\%$ with the addition of 6 M urea and the least populated configuration (A_4) is reduced to near zero population.

The urea m -value for the tetraloop at the reference state (385.6 K, 165 MPa) is determined to be $\approx 0.1-0.25 \text{ kcal}\cdot\text{mol}^{-1} \cdot m^{-1}$ based on calculations using a Hawley-type equation, and preferential interactions of urea relative to concentration. While an experimental m -value for the gcGCAAgc tetraloop-forming sequence is not available in the literature, the range of m -values is consistent with other RNA hairpins and duplex structures, particularly those in high salt concentrations²¹ or with sequences of high GC content.²² Future simulations of systems over a broader range of salt concentrations with varying stem lengths will be needed to better elucidate the effect of urea on RNA stability.

Our free energy analysis reveals that enthalpy and internal energy are the dominant contributors to the free energy of unfolding, particularly the Coulombic interactions of RNA with water and urea. Similar observations have been made in simulations of the preQ₁ riboswitch by Yoon et. al. where electrostatic interactions dominated interactions between RNA and urea.²⁵ Their work suggested that penetration of water into RNA cavities precedes intercalation of, and unfolding by, urea. From our results high urea concentrations yield

increased ΔASA_U as well as higher ΔN_U^{HB} and more favorable ΔE_U for both water and urea. Thus both solvent and cosolute contribute to stabilize the unfolded ensemble.

Experimental m -values in several recent studies have been used to infer a high degree of residual RNA:RNA stacking in unfolded RNA oligomers.^{21,22} However, these studies relied on idealized oligomer geometries (single-strands from A-RNA duplex structures) and neglected the heterogeneity of even the simplest RNA oligomers.^{27,38,62} Stacking interactions have been monitored in previous simulations of RNA tetraloops and small duplexes and suggest that RNA:Urea stacking maintains unfolded configurations, possibly even supplanting RNA:RNA stacking.²⁴

Our own stacking metric, which defines both RNA:RNA and RNA:Urea stacking interactions with similar structural geometries, suggests that RNA:RNA stacking is temperature-dependent while RNA:Urea stacking is predominantly affected by concentration. The change in stacking upon unfolding shows RNA:RNA $\Delta N_U^{Stack} \sim -1.0$, and RNA:Urea $\Delta N_U^{Stack} \sim +0.5$. From these results we can claim that in an unfolded configuration, the temperature, rather than urea, depletes residual RNA:RNA stacking, and that residual RNA:RNA stacking is maintained despite changes in urea concentration.

Folding of an RNA oligomer in solution with urea depends on a complex set of intramolecular interactions⁵³ as well as a balance of interactions between the RNA and the surrounding solution.⁶³⁻⁶⁸ One important caveat to these results is that the force field parameters used in these and similar simulations are calibrated for different binary mixtures of solvent (H₂O) and different solutes (RNA, ions, urea). The accuracy of these solvent-solute interactions do not guarantee quantitative agreement for interactions between different solute components.^{69,70} The interactions of RNA and urea are complex and minimally affected by temperature, however the qualitative effect of urea as an RNA denaturant is apparent both in terms of its ability to hydrogen-bond and stack with RNA, and in terms of concentration-dependent changes to the free energy of unfolding.

In this work we have characterized the internal geometries and energies of salient RNA-

cosolute interactions and assessed how they affect the free energy landscape of a hyperstable RNA tetraloop. Our results provide a solid thermodynamic description of the effect of urea on the free energy landscape of the RNA tetraloop and provide explicit, atomic-level detail relating the many interactions responsible for stabilizing the unfolded ensemble in the presence of urea.

In addition to the conclusions of this study, we have established a framework for thermodynamic studies of RNA folding as a function of cosolute mixtures. These methods can be extended to analyses of other denaturants (e.g. glycine betaine) or osmolytic agents (e.g. trimethylamine-N-oxide) using this tetraloop or other RNA molecules. Further application of these methods will allow MD to provide structural and energetic descriptions of how RNA and other biomolecules fold in cellular environments where cosolutes affect the equilibrium of the configuration ensemble.⁷¹

Beyond our thermodynamic analysis of RNA tetraloop unfolding by urea, these methods have the potential to give additional insight into the kinetic behaviors of larger, catalytic RNAs. Two separate studies of the self-splicing reaction of the *Tetrahymena* group I intron revealed that urea increases the ribozyme reaction rate 12-fold (at 2.5 M urea)⁷² by increasing the initial rate of reaction, and 30-fold (at 5 M urea) by increasing the slower of the biphasic rates and reducing energy barriers for intermediate states.⁷³ These observations, calculated through gel electrophoresis and circular dichroism respectively, suggested alternative folding pathways induced by urea and reduction in the lifetimes of intermediate states.

Reduction in the number of alternative configurations and increased surface exposure upon unfolding in our results supports these observations and suggests that the effect of urea on the *Tetrahymena* group I intron and other tertiary RNA structures can be analyzed with these methodologies to determine the effects of urea concentration of reducing the variability of phase space and certain kinetic intermediates, thus potentially accelerating kinetic transitions as a function of urea concentration.

Supporting Information Available

Figure S1: Time-dependent folding trajectories and ensemble averages of folded and unfolded configurations at 2, 4 and 6 M Urea. Figure S2: Dominant configurations of the free energy landscape in the absence of urea. Figure S3: Preferential interactions of potassium ions with RNA as a function of urea concentration (Γ_{K^+}). Figure S4: Population distributions of pairwise vertical and horizontal displacements between all RNA nucleobases and urea. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

Acknowledgement

We thank Dr. Deepak Canchi for useful discussions on simulations of urea. We acknowledge the Extreme Science and Engineering Discovery Environment (XSEDE) for access to the Stampede Supercomputer as well as NSF Grants MCB-137078 and MCB-1050906, and by US DOE LDRD funds from Los Alamos National Laboratory.

References

- (1) Yancey, P.; Clark, M.; Hand, S.; Bowlus, R.; Somero, G. Living With Water Stress: Evolution of Osmolyte Systems. *Science* **1982**, *217*, 1214–1222.
- (2) Brown, G.; Brown, S. Urea and Its Formation in Coelacanth Liver. *Science* **1967**, *155*, 570–573.
- (3) Griffith, R. Guppies, Toadfish, Lungfish, Coelacanths and Frogs: A Scenario for the Evolution of Urea Retention in Fishes. *Environ. Biol. of Fishes* **1991**, *32*, 199–218.
- (4) Pace, C. Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves. *Methods Enzymol.* **1986**, *131*, 266–280.

- (5) Makhatadze, G.; Privalov, P. Protein Interactions with Urea and Guanidinium Chloride. *J. Mol. Biol.* **1992**, *226*, 491–505.
- (6) Makhatadze, G. Thermodynamics of Protein Interactions with Urea and Guanidinium Hydrochloride. *J. Phys. Chem. B.* **1999**, *103*, 4781–4785.
- (7) Bennion, B.; Daggett, V. The Molecular Basis for the Chemical Denaturation of Proteins by Urea. *Proc. Nat. Acad. Sci. U.S.A.* **2003**, *100*, 5142–5147.
- (8) Guinn, E.; Pegram, L.; Capp, M.; Pollock, M.; Record, M. Quantifying Why Urea is a Protein Denaturant, Whereas Glycine Betaine is a Protein Stabilizer. *Proc. Nat. Acad. Sci. U.S.A.* **2011**, *108*, 16932–16937.
- (9) Cannon, J.; Anderson, C.; Record, M. Urea-Amide Preferential Interactions in Water: Quantitative Comparison of Model Compound Data with Biopolymer Results Using Water Accessible Surface Areas. *J. Phys. Chem. B.* **2007**, *111*, 9675–9685.
- (10) Courtenay, E.; Capp, M.; Anderson, C.; Record, M. Vapor Pressure Osmometry Studies of Osmolyte-Protein Interactions: Implications for the Action of Osmoprotectants in Vivo and for the Interpretation of "Osmotic Stress" Experiments in Vitro. *Biochemistry* **2000**, *39*, 4455–4471.
- (11) Courtenay, E.; Capp, M.; Record, M. Thermodynamics of Interactions of Urea and Guanidinium Salts with Protein Surface: Relationship Between Solute Effects on Protein Processes and Changes in Water-Accessible Surface Area. *Protein Sci.* **2001**, *10*, 2485–2497.
- (12) Canchi, D.; Paschek, D.; García, A. Equilibrium Study of Protein Denaturation by Urea. *J. Am. Chem. Soc.* **2010**, *132*, 2338–2344.
- (13) Canchi, D.; García, A. Backbone and Side-Chain Contributions in Protein Denaturation by Urea. *Biophys. J.* **2011**, *100*, 1526–1533.

- (14) Canchi, D.; García, A. Cosolvent Effects on Protein Stability. *Annu. Rev. Phys. Chem.* **2013**, *64*, 273–293.
- (15) Borgohain, G.; Paul, S. Model Dependency of TMAO’s Counteracting Effect Against Action of Urea: Kast Model versus Osmotic Model of TMAO. *J. Phys. Chem. B.* **2016**, *120*, 2352–2361.
- (16) Klump, H.; Burkart, W. Calorimetric Measurements of the Transition Enthalpy of DNA in Aqueous Urea Solutions. *Biochim. Biophys. Acta* **1977**, *475*, 601–604.
- (17) Hirao, I.; Nishimura, Y.; Tagawa, Y.-I.; Watanabe, K.; Miura, K.-I. Extraordinarily Stable Mini-Hairpins: Electrophoretical and Thermal Properties of the Various Sequence Variants of d(GCGAAAGC) and Their Effect on DNA Sequencing. *Nuc. Acids Res.* **1992**, *20*, 3891–3896.
- (18) Shelton, V.; Sosnick, T.; Pan, T. Applicability of Urea in the Thermodynamic Analysis of Secondary and Tertiary RNA Folding. *Biochemistry* **1999**, *38*, 16831–16839.
- (19) Todorov, T.; de Carmejane, O.; Walter, N.; Morris, M. Capillary Electrophoresis of RNA in Dilute and Semidilute Polymer Solutions. *Electrophoresis* **2001**, *22*, 2442–2447.
- (20) Todorov, T.; Yamaguchi, Y.; Morris, M. Effect of Urea on the Polymer Buffer Solutions Used for the Electrophoretic Separations of Nucleic Acids. *Anal. Chem.* **2003**, *75*, 1837–1843.
- (21) Lambert, D.; Draper, D. Denaturation of RNA Secondary and Tertiary Structure by Urea: Simple Unfolded State Models and Free Energy Parameters Account for Measured *m*-values. *Biochemistry* **2012**, *51*, 9014–9026.
- (22) Guinn, E.; Schweinfus, J.; Cha, H.; McDevitt, J.; Merker, W.; Ritzer, R.; Muth, G.; Engelsgerd, S.; Mangold, K.; Thompson, P. et al. Quantifying Functional Group Inter-

- actions That Determine Urea Effects on Nucleic Acid Helix Formation. *J. Am. Chem. Soc.* **2013**, *135*, 5828–5838.
- (23) Japrun, D.; Henricus, M.; Li, Q.; Maglia, G.; Bayley, H. Urea Facilitates the Translocation of Single-Stranded DNA and RNA Through the α -Hemolysin Nanopore. *Biophys. J.* **2010**, *98*, 1856–1863.
- (24) Priyakumar, U.; Hyeon, C.; Thirumalai, D.; MacKerell, A. Urea Destabilizes RNA by Forming Stacking Interactions and Multiple Hydrogen Bonds with Nucleic Acid Bases. *J. Am. Chem. Soc.* **2009**, *131*, 17759–17761.
- (25) Yoon, J.; Thirumalai, D.; Hyeon, C. Urea-Induced Denaturation of PreQ1-Riboswitch. *J. Am. Chem. Soc.* **2013**, *135*, 12112–12121.
- (26) Chen, A.; García, A. High-Resolution Reversible Folding of Hyperstable RNA Tetraloops Using Molecular Dynamics Simulations. *Proc. Nat. Acad. Sci. U.S.A.* **2013**, *110*, 16820–16825.
- (27) Miner, J.; Chen, A.; García, A. Free-Energy Landscape of a Hyperstable RNA Tetraloop. *Proc. Nat. Acad. Sci. U.S.A.* **2016**, *113*, 6665–6670.
- (28) Weerasinghe, S.; Smith, P. A Kirkwood-Buff Derived Force Field for Mixtures of Urea and Water. *J. Phys. Chem. B* **2003**, *107*, 3891–3898.
- (29) Record, M.; Anderson, C. Interpretation of Preferential Interaction Coefficients of Non-electrolytes and of Electrolyte Ions in Terms of a Two-Domain Model. *Biophys. J.* **1995**, *68*, 786–794.
- (30) Schellman, J. A. Fifty Years of Solvent Denaturation. *Biophys. Chem.* **2002**, *96*, 91–101.
- (31) Timasheff, S. N. In *Advances in Protein Chemistry*; DiCera, R. . E. . K. ., Ed.; Elsevier, 1998; Vol. 51; pp 355–432.

- (32) Timasheff, S. Protein-Solvent Preferential Interactions, Protein Hydration, and the Modulation of Biochemical Reactions by Solvent Components. *Proc. Nat. Acad. Sci. U.S.A.* **2002**, *99*, 9721–9726.
- (33) Leipply, D.; Lambert, D.; Draper, D. In *Biophysical, Chemical, and Functional Probes of RNA Structure, Interactions and Folding: Part B*; Abelson, J., Simon, M., Eds.; Method Enzymol.; Academic Press: New York, 2009; Vol. 469; Chapter 21, pp 433–463.
- (34) Chen, A. A.; Draper, D. E.; Pappu, R. V. Molecular Simulation Studies of Monovalent Counterion-Mediated Interactions in a Model RNA Kissing Loop. *J. Mol. Biol.* **2009**, *390*, 805–819.
- (35) Hong, J.; Capp, M.; Anderson, C.; Jr., M. R. Preferential Interactions in Aqueous Solutions of Urea and KCl. *Biophys. Chem.* **2003**, *105*, 517–532.
- (36) Hong, J.; Capp, M.; Anderson, C.; Saecker, R.; Felitsky, D.; Anderson, M.; Record, M. Preferential Interactions of Glycine Betaine and of Urea with DNA: Implications for DNA Hydration and for Effects of These Solutes on DNA Stability. *Biochemistry* **2004**, *43*, 14744–14758.
- (37) Proctor, D.; Ma, H.; Kierzek, E.; Kierzek, R.; Gruebele, M.; Bevilacqua, P. Folding Thermodynamics and Kinetics of YNMG RNA Hairpins: Specific Incorporation of 8-Bromoguanosine Leads to Stabilization by Enhancement of the Folding Rate. *Biochemistry* **2004**, *43*, 14004–14014.
- (38) Ma, H.; Proctor, D.; Kierzek, E.; Kierzek, R.; Bevilacqua, P.; Gruebele, M. Exploring the Energy Landscape of a Small RNA Hairpin. *J. Am. Chem. Soc.* **2006**, *128*, 1523–1530.
- (39) Stancik, A.; Brauns, E. Rearrangement of Partially Ordered Stacked Conformations

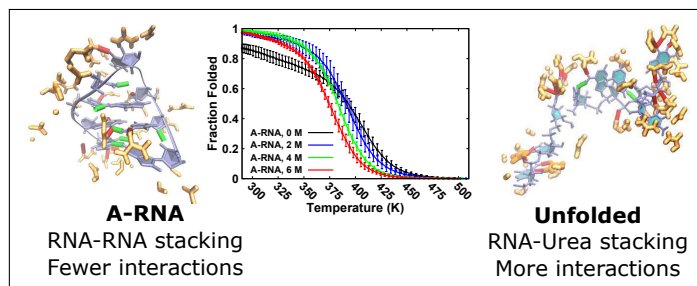
- Contributes to the Rugged Energy Landscape of a Small RNA Hairpin. *Biochemistry* **2008**, *47*, 10834–10840.
- (40) García, A.; Paschek, D. Simulation of the Pressure and Temperature Folding/Unfolding Equilibrium of a Small RNA Hairpin. *J. Am. Chem. Soc.* **2008**, *130*, 815–817.
- (41) Serra, M.; Axenson, T.; Turner, D. A Model for the Stabilities of RNA Hairpins Based on a Study of the Sequence Dependence of Stability for Hairpins of Six Nucleotides. *Biochemistry* **1994**, *33*, 14289–14296.
- (42) Walter, A.; Turner, D.; Kim, J.; Lyttle, M.; Müller, P.; Mathews, D.; Zuker, M. Coaxial Stacking of Helices Enhances Binding of Oligoribonucleotides and Improves Predictions of RNA Folding. *Proc. Nat. Acad. Sci. U.S.A.* **1994**, *91*, 9218–9222.
- (43) Serra, M. J.; Barnes, T. W.; Betschart, K.; Gutierrez, M. J.; Sprouse, K. J.; Riley, C. K.; Stewart, L.; Temel, R. E. Improved Parameters for the Prediction of RNA Hairpin Stability. *Biochemistry* **1997**, *36*, 4844–4851.
- (44) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W.; DiNola, A.; Haak, J. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3691.
- (45) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 1–7.
- (46) García, A.; Hecce, H.; Paschek, D. Chapter 5 Simulations of Temperature and Pressure Unfolding of Peptides and Proteins with Replica Exchange Molecular Dynamics. *Ann. Rep. Comput. Chem.* **2006**, *2*, 83–95.
- (47) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

- (48) Cornell, W.; Cieplak, P.; Bayley, C.; Gould, R.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Caldwell, J.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (49) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (50) Chen, A.; Pappu, R. Parameters of Monovalent Ions in the AMBER-99 Forcefield: Assessment of Inaccuracies and Proposed Improvements. *J. Phys. Chem. B* **2007**, *111*, 11884–11887.
- (51) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (52) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Hsing, L.; Pedersen, L. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (53) Jucker, F.; Heus, H.; Yip, P.; Moors, E.; Pardi, A. A Network of Heterogeneous Hydrogen Bonds in GNRA Tetraloops. *J. Mol. Biol.* **1996**, *264*, 968–980.
- (54) Riccardi, L.; Nguyen, P.; Stock, G. Free Energy Landscape of an RNA Hairpin Constructed Via Dihedral Angle Principal Component Analysis. *J. Phys. Chem. B* **2009**, *113*, 16660–16668.
- (55) Daura, X.; Suter, R.; van Gunsteren, W. Validation of Molecular Simulation by Comparison with Experiment: Rotational Reorientation of Tryptophan in Water. *J. Chem. Phys.* **1999**, *110*, 3049–3055.
- (56) Hawley, S. Reversible Pressure-Temperature Denaturation of Chymotrypsin. *Biochemistry* **1971**, *10*, 2436–2442.

- (57) Smeller, L. Pressure-Temperature Phase Diagrams of Biomolecules. *Biochim. et. Biophys. Acta.* **2002**, *1595*, 11–29.
- (58) English, C.; García, A. Folding and Unfolding Thermodynamics of the TC10b Trp-Cage Miniprotein. *Phys. Chem. Chem. Phys.* **2014**, *16*, 2748–2757.
- (59) Ganguly, P.; Vegt, N. v. d. Convergence of Sampling Kirkwood-Buff Integrals of Aqueous Solutions with Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 1347–1355.
- (60) Paschek, D.; Geiger, A. MOSCITO 4. Department of Physical Chemistry, University of Dortmund, 2002.
- (61) Myers, J.; Pace, C.; Scholtz, J. Denaturant m -values and Heat Capacity Changes: Relation to Changes in Accessible Surface Areas of Protein Unfolding. *Protein Sci.* **1995**, *4*, 2138–2148.
- (62) Bergonzo, C.; Henriksen, N.; Roe, D.; Cheatham, T. Highly Sampled Tetranucleotide and Tetraloop Motifs Enable Evaluation of Common RNA Force Fields. *RNA* **2015**, *21*, 1578–1590.
- (63) Sorin, E.; Rhee, Y.; Pande, V. Does Water Play a Structural Role in the Folding of Small Nucleic Acids? *Biophys. J.* **2005**, *88*, 2516–2524.
- (64) Auffinger, P.; Westhof, E. Water and Ion Binding Around RNA and DNA (C,G) Oligomers. *J. Mol. Biol.* **2000**, *300*, 1113–1131.
- (65) Tikhomirova, A.; Taulier, N.; Chalikian, T. Energetics of Nucleic Acid Stability: The Effect of ΔC_P . *J. Am. Chem. Soc.* **2004**, *126*, 16387–16394.
- (66) Lambert, D.; Leipply, D.; Draper, D. The Osmolyte TMAO Stabilizes Native RNA Tertiary Structures in the Absence of Mg^{2+} : Evidence for a Large Barrier to Folding from Phosphate Dehydration. *J. Mol. Biol.* **2010**, *404*, 138–157.

- (67) Strulson, C.; Boyer, J.; Whitman, E.; Bevilacqua, P. Molecular Crowders and Cosolutes Promote Folding Cooperativity of RNA Under Physiological Ionic Conditions. *RNA* **2014**, *20*, 331–347.
- (68) Gao, M.; Arns, L.; Winter, R. Modulation of the Thermodynamic Signatures of an RNA Thermometer by Osmolytes and Salts. *Angew. Chem. Int. Ed.* **2017**, 10.1002/anie.201611843.
- (69) Horinek, D.; Netz, R. Can Simulations Quantitatively Predict Peptide Transfer Free Energies to Urea Solutions? Thermodynamic Concepts and Force Field Limitations. *J. Phys. Chem. A* **2011**, 6125–6136.
- (70) Schneck, E.; Horinek, D.; Netz, R. Insight into the Molecular Mechanisms of Protein Stabilizing Osmolytes from Global Force-Field Variations. *J. Phys. Chem. B* **2013**, *117*, 8310–8321.
- (71) Breaker, R. Riboswitches and the RNA World. *Cold Spring Harb. Persp. Biol.* **2012**, *4*, a003566.
- (72) Pan, J.; Thirumalai, D.; Woodson, S. Folding of RNA Involves Parallel Pathways. *J. Mol. Biol.* **1997**, *273*, 7–13.
- (73) Pan, T.; Sosnick, T. Intermediates and Kinetic Traps in the Folding of a Large Ribozyme Revealed by Circular Dichroism and UV Absorbance Spectroscopies and Catalytic Activity. *Nat. Struct. Mol. Biol.* **1997**, *4*, 931–938.

Graphical TOC Entry



Equilibrium Denaturation and Preferential Interactions of an RNA Tetraloop with Urea

Jacob C. Miner^{†,‡} and Angel E. García^{*,‡}

[†]*Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87545*

[‡]*Center for Nonlinear Studies, CNLS, MS B258, Los Alamos National Laboratory, Los Alamos, NM 87545*

E-mail: agarcia@lanl.gov

Phone: +1 505 665 3883. Fax: +1 505 665 2659

Simulation Trajectories

All simulations (Figure S1) were initialized in an A-RNA configuration that possesses a sheared base-pair between loop residues G_{L1} and A_{L4} , and forms a *syn*-glycosidic torsion in the χ -angle of A_{L4} (A_3 , Figure S2C). Previous REMD simulations of this tetraloop sequence at 0 M urea show that the A_3 configuration is the most favorable structure possessing the G_{L1} - A_{L4} interaction, though the folded tetraloop is described by an ensemble of A-RNA configurations (Figure S2A-D). Further, REMD simulations that are run to an equilibrated population of A-RNA structures can sample the free energy landscape whether the ensemble is initialized in A_3 or unfolded configurations.¹

In each ensemble the majority of A_3 configurations sample the unfolded state at least once before undergoing refolding.

The populations of alternative configurations (Triplet & Z-RNA, Figure S2E-F) are small,

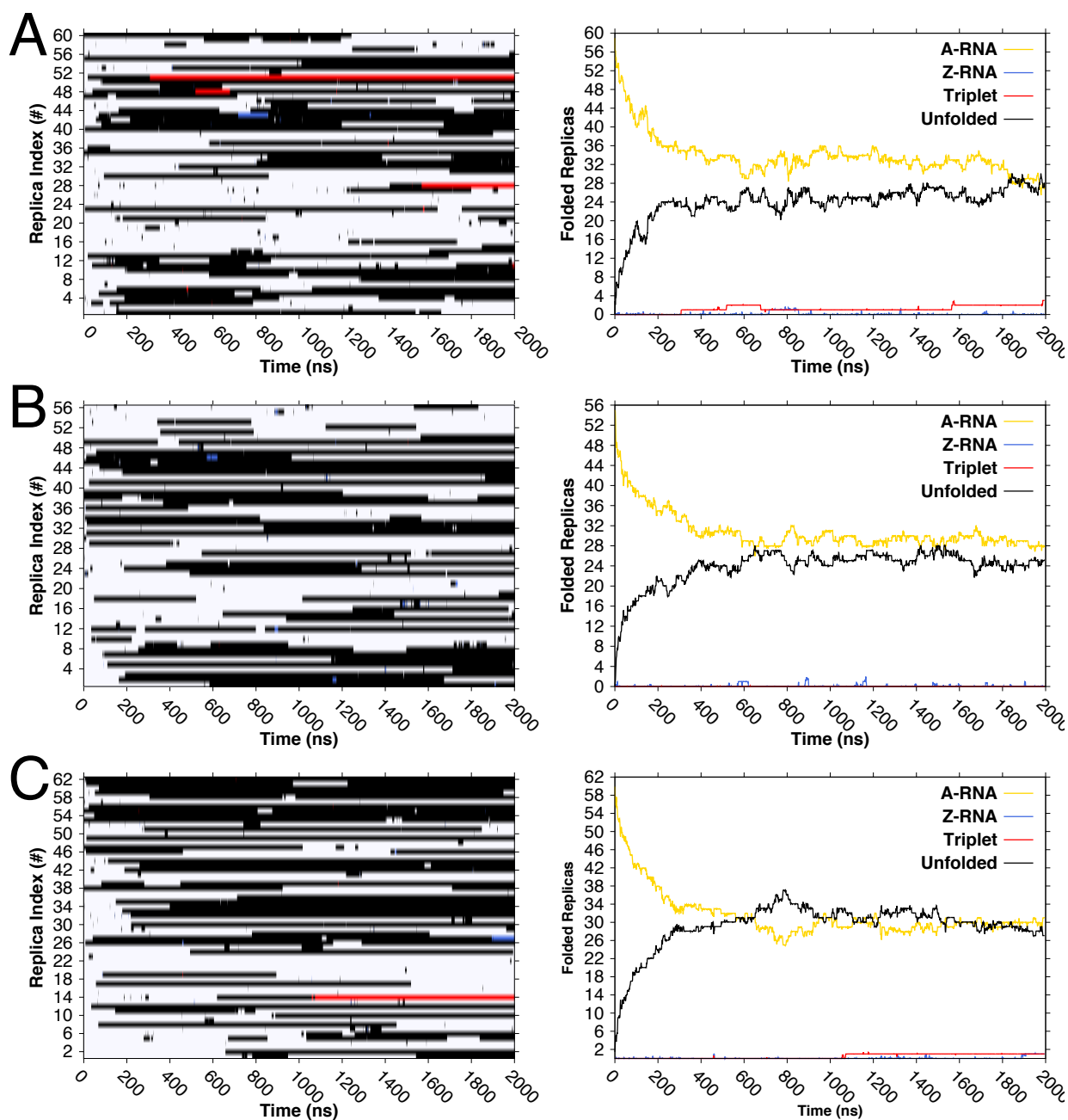


Figure S1: Time-dependent folding trajectories and ensemble averages of folded, alternative, and unfolded configurations. (A) 2 M Urea, (B) 4 M Urea, (C) 6 M Urea. A-RNA, Z-RNA, Triplet and unfolded configurations shown.

and individual configurations form sporadically. The Triplet configuration (Figure S2E) is the more stable of the two alternative configurations and maintains longer-lived folds than Z-RNA (Figure S2F). The earliest instances of folding for either of these configurations also shows some dependence on urea concentration. At 2 M, the earliest instances are in the first

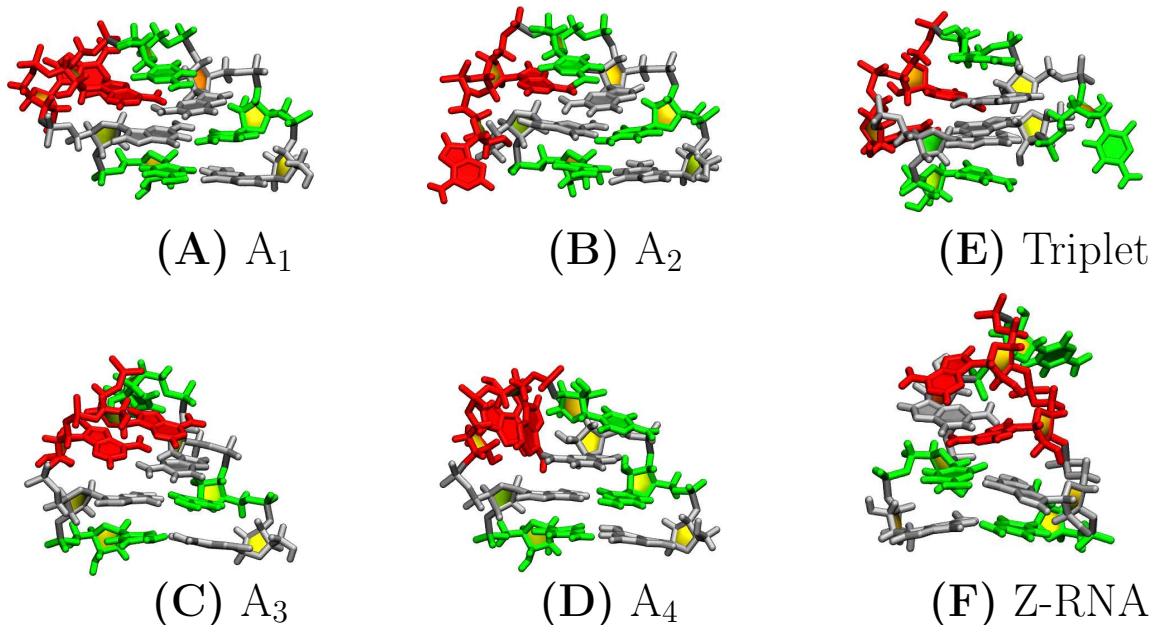


Figure S2: Representative structures of the tetraloop free energy landscape. Nucleobases are shown in gray (guanosines), green (cytosine), and red (adenine). **(A)-(D)** A-RNA configurations presented in order of most populated clusters at 0 M urea. Simulations were initialized in A_3 **(C)** configurations. **(E)** Triplet configuration is a misfolded trap formed by residues G_1 , A_{L4} and G_7 forming a tight hydrogen-bond cluster. **(F)** Z-RNA forms by the phosphate backbone adopting a left-handed (Z) configuration and base-pairing to form a two-base-pair duplex with a tetralooping region.

300 ns, while 4 M shows a long-lived Z-RNA forming at ~ 600 ns, and a long-lived Triplet forming at around $1 \mu\text{s}$ in 6 M urea. The populations of either configuration are usually ≤ 2 , even in 2 M urea.

Preferential Interactions of Counterions with RNA

The neutralizing counterions and excess salt used in our simulations do not become coordinated within any configuration on the free energy landscape of this particular RNA oligomer. However, the cosolute urea is observed to preferentially interact with RNA, and its effect on the interactions of other cosolutes is worth investigating, particularly for counterions, which have been observed to preferentially coordinate around this and other RNA in submolar to molar concentrations of KCl.¹⁻³

Using the same methods for measuring the preferential interactions of urea and the same

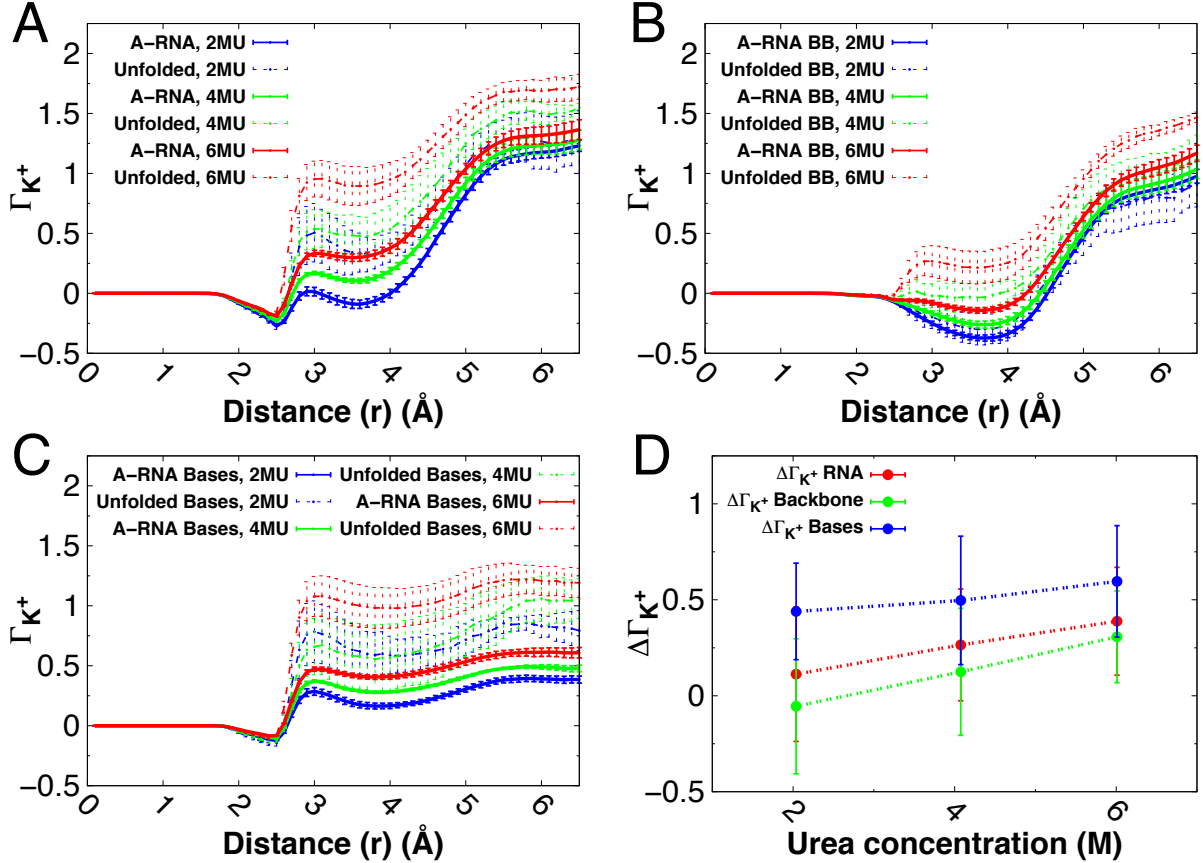


Figure S3: (A)-(C) Preferential interactions of counterions (Γ_{K^+}) relative to distance (r) from different RNA groups (all, backbone, nucleobases) at 2 M (blue), 4 M (green), and 6 M (red) at 350 K. A-RNA (solid), and unfolded (dashed) configurations are represented. (A) Total RNA:K⁺ interactions for both A-RNA and unfolded configurations increase with urea concentration. (B) Backbone:K⁺ interactions are low for distances < 5 Å, but show similar trends to total RNA at longer ranges. (C) Nucleobase:K⁺ interactions follow trends similar to total RNA within the first solvation shell, but not beyond. (D) The change in Γ_{K^+} upon unfolding ($\Delta\Gamma_{K^+}$) increases linearly relative to urea concentration and is dominated by Nucleobase:K⁺ interactions, but Backbone:K⁺ are within the error of the total RNA:K⁺ measurements.

temperature ensembles (~ 350 K) we investigate how A-RNA and unfolded configurations formed in simulation interact with counterions (K⁺) (Figure S3). We find that the concentration of urea actually drives the accumulation of counterions around the full RNA, and that unfolded configurations show higher values of Γ_{K^+} , indicating that urea is not occluding RNA:K⁺ interactions, but that by virtue of greater surface exposure, K⁺ have a higher preferential interaction with RNA.

When the interactions are decomposed into backbone and nucleobase interactions the backbone shows almost no exclusion or accumulation of K^+ within the first solvation shell ($<5 \text{ \AA}$) but shows accumulation of 1 K^+ at greater distances. Coupled with our observations of low accumulations of urea at 3 \AA from the backbone, we determine that the first solvation layer around the RNA backbone is primarily occupied by water.

The preferential interactions around the nucleobases tell a different story. While interactions begin to increase within the first solvation layer, they remain relatively constant at increasing distances up to the 6.5 \AA cutoff, suggesting that the relative concentration of K^+ remains constant. The A-RNA maintain a lower Γ_{K^+} ($\leq 0.5 K^+$) while the unfolded configurations show $\sim 1 K^+$ accumulated in 6 M urea.

An important point is that the interactions for the nucleobases define much of the Γ_{K^+} within the first solvation shell ($<5 \text{ \AA}$), while the backbone defines the accumulations beyond. The concentration-dependence of these interactions suggests that molecules of urea are doing more to expose RNA surface area for favorable counterion interactions than they are to inhibit these smaller cosolutes from interacting with RNA.

RNA:RNA & RNA:Urea Stacking

The stacking criteria defined in our work makes use of the geometric centers of pyrimidine rings (a common structure in all four RNA nucleobases that includes C_2 , C_4 , and C_6 atoms) and the polar groups around the central carbon of urea (O_U , N_{U1} , N_{U2}). We describe the vertical and horizontal pairwise displacements between each pair of nucleobases (Figure S4A) and between all nucleobases and urea molecules (Figure S4B).

A general cutoff of 6 \AA is determined based on unoccupied regions in the RNA:RNA plots for both A-RNA and unfolded configurations as well as similar long-range distributions in the RNA:Urea plots. This cutoff is set as the delineating mark for all stacking interactions. An inner, cutoff is observed at 5 \AA based on the most populated peaks at close proximity to both RNA:RNA and RNA:Urea in the unfolded configurations. Based on these observations,

regions of displacement less than 5 Å are considered stacked and given a stacking weight of 1, while regions greater than 5 Å and less than 6 Å are counted as half-stacked and weighted 0.5.

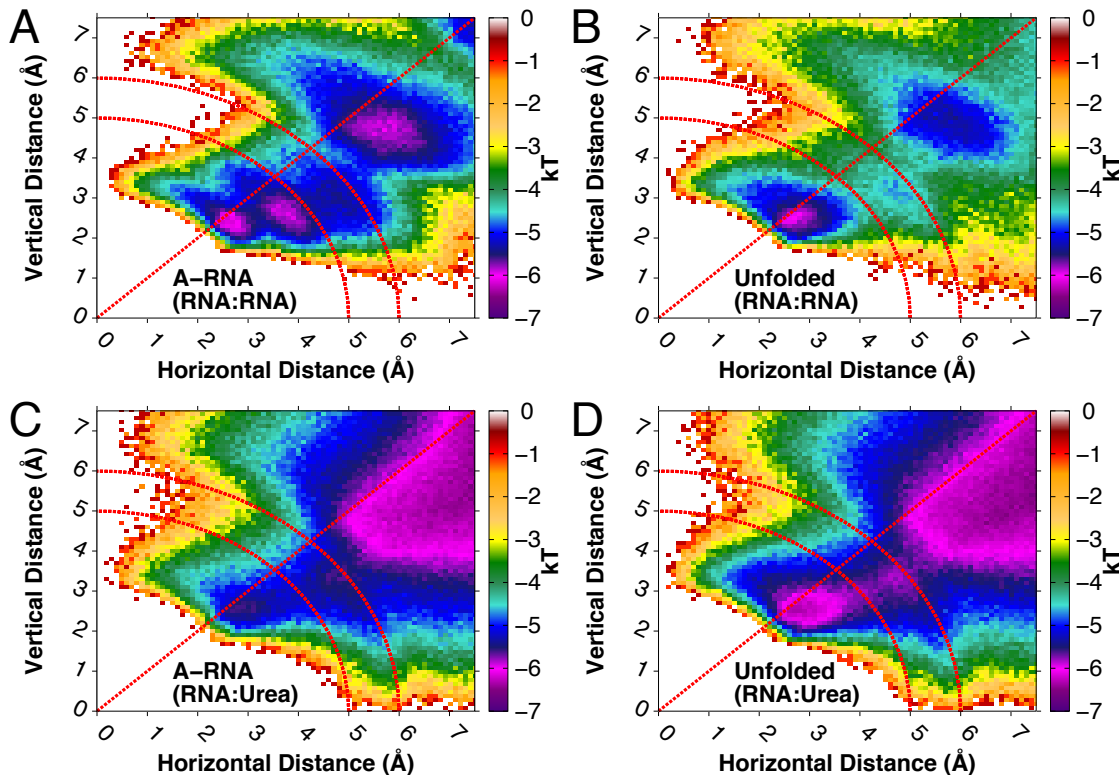


Figure S4: Vertical vs. horizontal displacements of RNA nucleobases and urea in RNA stacking interactions. RNA:RNA interactions for (A) A-RNA and (B) unfolded configurations show dense populations of RNA:RNA interactions <5 Å and an intermediate region with clear cutoff at 6 Å (red arc lines). For RNA:Urea interactions, (C) & (D), the same cutoffs show a similar cutoff at 5 Å. In each plot a diagonal red line is included to show where $V_{disp} = H_{disp}$ and regions above this line are considered stacked if the previous criteria are met.

Considerable populations are found in regions where horizontal displacements exceed vertical displacements, however most of these interactions describe urea in regions where hydrogen-bonding is more probable than stacking. Most stacking in these regions occur over the imidazole ring of the purine nucleobases (A & G). Inclusion or exclusion of these interactions does not significantly change the temperature-dependent and concentration-dependent trends in stacking for RNA:RNA or RNA:Urea interactions, thus stacking is counted only for regions where vertical displacements (V_{disp}) exceed horizontal displacements

(H_{disp}) .

References

- (1) Miner, J.; Chen, A.; García, A. Free-Energy Landscape of a Hyperstable RNA Tetraloop. *Proc. Nat. Acad. Sci. U.S.A.* **2016**, *113*, 6665–6670.
- (2) Chen, A. A.; Draper, D. E.; Pappu, R. V. Molecular Simulation Studies of Monovalent Counterion-Mediated Interactions in a Model RNA Kissing Loop. *J. Mol. Biol.* **2009**, *390*, 805–819.
- (3) Lambert, D.; Draper, D. Denaturation of RNA Secondary and Tertiary Structure by Urea: Simple Unfolded State Models and Free Energy Parameters Account for Measured m -values. *Biochemistry* **2012**, *51*, 9014–9026.