

Thermodynamic insights into 2-thiouridine-enhanced RNA hybridization

Aaron T. Larsen¹, Albert C. Fahrenbach^{1,2}, Jia Sheng³, Julia Pian¹ and Jack W. Szostak^{1,2,*}

¹Howard Hughes Medical Institute, Center for Computational and Integrative Biology, and Department of Molecular Biology, Simches Research Center, Massachusetts General Hospital, Boston, MA 02114, USA, ²Earth-Life Science Institute, Tokyo Institute of Technology, 2-12-1-IE-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan and ³University at Albany, State University of New York, Department of Chemistry, The RNA Institute, 1400 Washington Avenue, Albany, NY 12222, USA

Received March 30, 2015; Revised June 24, 2015; Accepted July 14, 2015

ABSTRACT

Nucleobase modifications dramatically alter nucleic acid structure and thermodynamics. 2-thiouridine (s^2U) is a modified nucleobase found in tRNAs and known to stabilize U:A base pairs and destabilize U:G wobble pairs. The recently reported crystal structures of s^2U -containing RNA duplexes do not entirely explain the mechanisms responsible for the stabilizing effect of s^2U or whether this effect is entropic or enthalpic in origin. We present here thermodynamic evaluations of duplex formation using ITC and UV thermal denaturation with RNA duplexes containing internal s^2U :A and s^2U :U pairs and their native counterparts. These results indicate that s^2U stabilizes both duplexes. The stabilizing effect is entropic in origin and likely results from the s^2U -induced preorganization of the single-stranded RNA prior to hybridization. The same preorganizing effect is likely responsible for structurally resolving the s^2U :U pair-containing duplex into a single conformation with a well-defined H-bond geometry. We also evaluate the effect of s^2U on single strand conformation using UV- and CD-monitored thermal denaturation and on nucleoside conformation using ¹H NMR spectroscopy, MD and umbrella sampling. These results provide insights into the effects that nucleobase modification has on RNA structure and thermodynamics and inform efforts toward improving both ribozyme-catalyzed and nonenzymatic RNA copying.

INTRODUCTION

RNA plays an essential and diverse role in living systems, acting as genetic information carrier, catalyst and regulator (1–4). Functional RNAs adopt many well-defined 3D struc-

tures resulting from specific base–base interactions including normal Watson–Crick base pairs and a variety of other associations (5). Understanding the structure and thermodynamics of base–base interactions provides a foundation for elucidating RNA structure/function relationships, engineering novel applications such as RNA-based therapeutics and addressing questions related to the origins of life (6).

Nucleobase modification diversifies nucleic acid structure and function (7). The significance of this role is reflected by the fact that certain modified nucleobases are among the most highly conserved features of RNA and, for this reason, are regarded as chemical fossils of molecular evolution. Of the approximately 140 known post-transcriptional RNA modifications, 60 are specific to uridine (U) and 16 feature thiolation at the C2 position of U (8). These modified nucleobases include s^2U and various 5-modified derivatives of s^2U . While it is possible that some of these modifications are relics of the RNA world (9), the extent to which 2-thiolation is conserved at the anticodon site of tRNA suggests that this modification is of critical functional importance in translation (10). It has been demonstrated that s^2U and its 5-modified derivatives enhance the efficiency and accuracy of codon-anticodon recognition, prevent frame-shifting during translation and improve tRNA aminoacylation kinetics (11). Improved codon-anticodon recognition resulting from 2-thiolation in the human tRNA^{Lys3}_{UUU} has also proved to be important for reverse transcription of the HIV-1 viral genome (12).

Modified nucleobases (such as s^2U) also address complications related to the RNA world hypothesis including protein-free nucleic acid copying (13). Despite the fact that *in vitro* selection has afforded an RNA-dependent RNA polymerase capable of producing a polymer longer than itself, this feat has been performed only on purposefully designed and highly optimized templates (14). Poor rate and fidelity currently prevent the complete copying of non-optimized or highly structured RNA templates, including the RNA polymerase ribozyme itself (6). Similar challenges

*To whom correspondence should be addressed. Tel: +1 617 726 5102; Fax: +1 617 643 3328; Email: szostak@molbio.mgh.harvard.edu

also prevent the realization of nonenzymatic, template-directed nucleic acid replication systems capable of copying long templates. In these cases, self replication is limited in part by the frequent incorporation of mismatched base-pairs, occurring at rates as high as once for every 5–6 nucleotides incorporated in some systems (15). In addition to compromising the fidelity of the resulting genetic message, some mismatches also strongly inhibit subsequent polymerization, potentially halting further extension (16).

Issues of rate and fidelity can be addressed by the use of modified nucleobases such as s^2U . 2-Thiolation has been demonstrated to stabilize Watson–Crick U:A base pairs and destabilize U:G wobble-pairs, the most commonly observed mismatch in nonenzymatic RNA copying (17–19). We recently exploited these effects to enhance the rate and fidelity of nonenzymatic, template-directed RNA copying (20).

Although several thermodynamic (21) and NMR (17) studies have evaluated the effects of s^2U on base-pairing in RNAs, the actual mechanism by which s^2U affects RNA structure and thermodynamics remains unclear. Current mechanistic proposals include (i) enhanced base stacking, particularly S/N1 stacking due to the highly polarizable nature of sulfur (22); (ii) stronger N3 H-bonding as a result of the increased acidity of the N3 imino proton (17,23); (iii) sulfur-induced stabilization of the 3'-*endo* sugar pucker leading to enhanced conformational ordering (24); and (iv) a reduced desolvation penalty incurred during duplex formation owing to the relatively poor H-bonding ability of sulfur (25) (Figure 1). It is conceivable that any and all of these mechanisms may contribute to s^2U -induced duplex stabilization to differing degrees, and in particular, mechanisms (i) and (iii) are both likely to enhance preorganization (26) of the single strand prior to duplex formation.

Despite a statement from Davis and Kumar in 1997 regarding the necessity of addressing these questions by examining high-resolution structures of s^2U -containing double stranded RNA (dsRNA) (17), such structures have been reported only very recently (27). The overall structures of the dsRNA containing an internal s^2U :A pair (s^2U :A) and its native counterpart (U:A) differ by a root-mean-squared deviation (RMSD) of $< 0.2 \text{ \AA}$ (Figure 2A). In contrast, the structure of the dsRNA containing an internal s^2U :U pair (s^2U :U_c) differs significantly compared to its native counterpart (U:U_c) (Figure 2B). The unit cell of the s^2U :U_c crystal features only a single duplex with standard Watson–Crick H-bond geometry at all other base pairs (Figure 2C). The unit cell of the U:U_c crystal, however, features four different duplex structures (denoted by U:U_c1 through U:U_c4) featuring distinct H-bond geometries at the U:U wobble pair, two of which display irregular H-bond lengths and angles that correspond to relatively weak H-bond strengths. These observations suggest that U:U_c exists in several conformations of comparable energy, a feature which may be related to the low observed duplex stability resulting from the internal U:U mismatch.

Each of the above proposed mechanistic explanations for s^2U -enhanced duplex stabilization can be expected to have different effects on the thermodynamic parameters of duplex formation, namely ΔG , ΔH and ΔS . Here, we present the thermodynamic parameters of duplex formation measured from isothermal titration calorimetry (ITC) and ther-

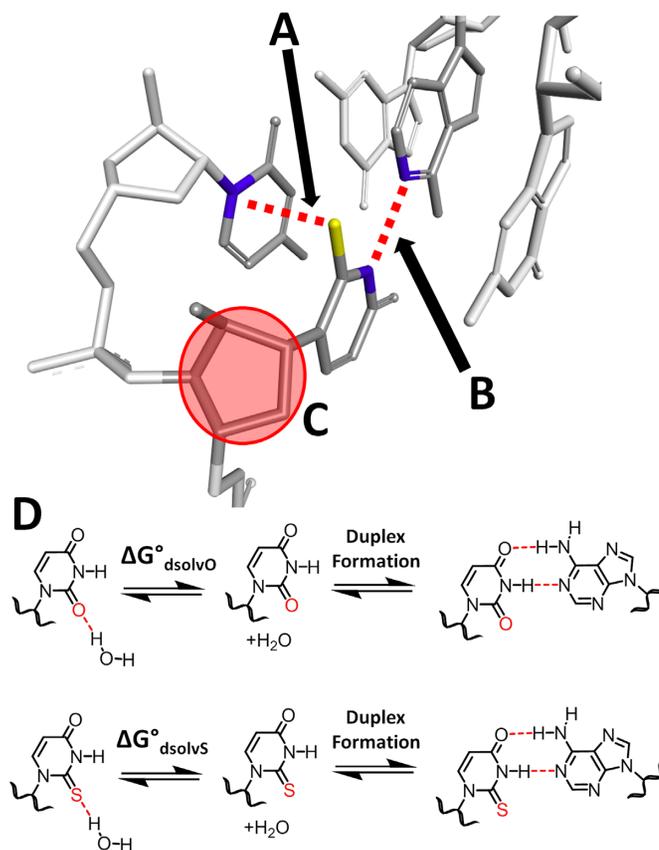


Figure 1. The proposed mechanisms of s^2U -induced stabilization of duplex RNA: S may (A) enhance S/N1 stacking due to the more polarizable nature of S; (B) increase N3 H-bonding due to the increased acidity of the N3 imino proton; (C) stabilize the 3'-*endo* sugar pucker (circled in red); the structure was adapted from PDB ID: 4U34 (27); and/or (D) reduce the desolvation penalty incurred during duplex formation.

mal denaturation experiments of two s^2U -containing heptamer RNA duplexes: s^2U :A and s^2U :U_c (5'-uagcs²Ucc-3' paired with 3'-aucgAgg-5' and 3'-aucgUgg-5', respectively) and their corresponding native structures, U:A, and U:U_c. We also evaluate the conformational stability of the single-stranded RNA (ssRNA) containing s^2U in comparison to its native counterpart using thermal denaturation monitored by UV absorbance and circular dichroism (CD) spectroscopies. Next, we evaluate the conformational effect of 2-thiolation on the sugar pucker of the free U nucleoside using ¹H NMR spectroscopy, MD simulations and umbrella sampling. Finally, we determine and compare the desolvation penalties for s^2U and U nucleosides. Our findings provide insights into the mechanisms responsible for the effect of s^2U on duplex formation and guide future studies focused on the structural and energetic effects of nucleobase modifications, particularly in the context of nonenzymatic RNA copying.

MATERIALS AND METHODS

Materials

All RNA oligonucleotides were chemically synthesized at 1.0- μ mol scales by solid-phase synthesis. The 2'-TBDMS-

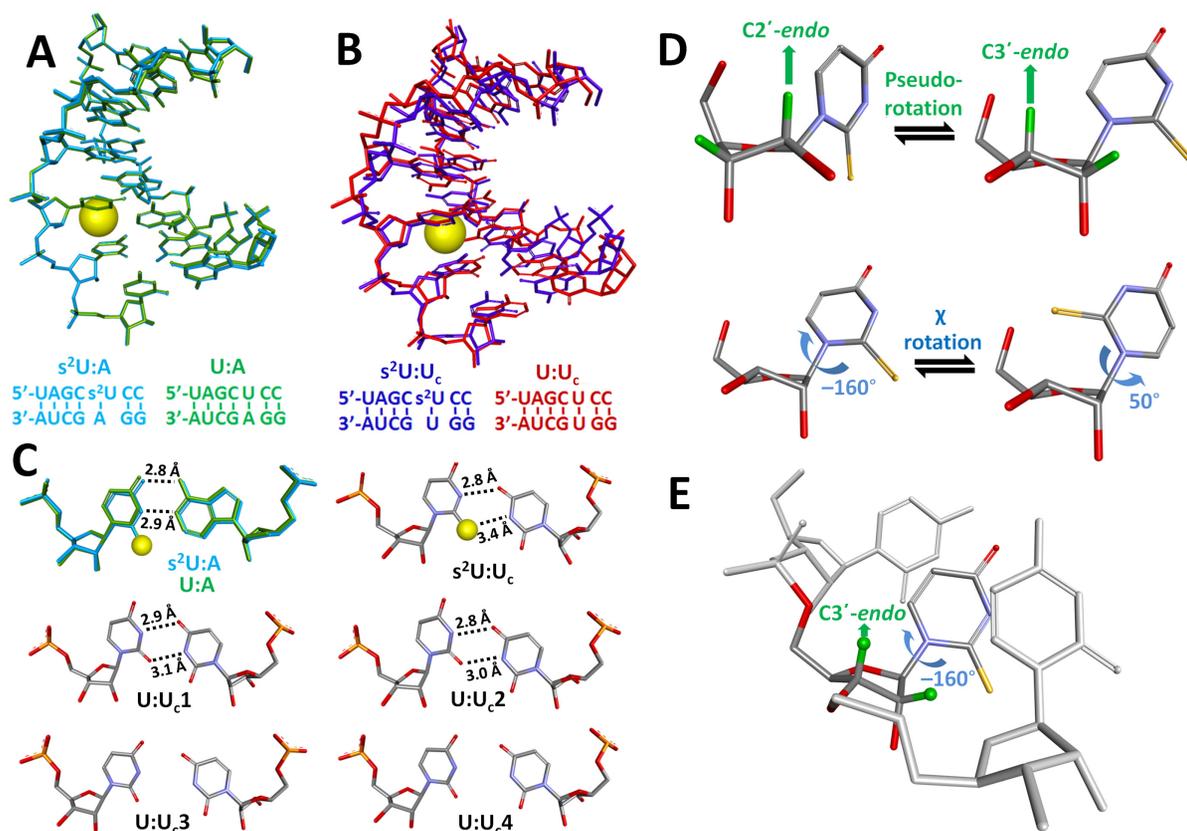


Figure 2. Crystal structures, sequences and structural features of the RNA duplexes investigated here; (A) Structural comparison of $s^2U:A$ (red, PDB ID: 4U34) and $U:A$ (blue, PDB ID: 4U37), RMSD is < 0.2 Å; (B) Structural comparison of $s^2U:U_c$ (light blue, PDB ID: 4U35) and $U:U_c1$ (pink, PDB ID: 4U38), the most structurally similar duplex from $U:U_c$, RMSD is 1.3 Å. Sulfur atoms are displayed as yellow spheres scaled to their van der Waals radii; (C) A comparison of H-bond geometries of base pairs at the site of s^2U -substitution. The geometries for $s^2U:A$ and $U:A$ overlap well and the geometry for $s^2U:U_c$ is most similar to that of $U:U_c1$; (D) Pseudorotation phase angle (P) and chi torsion angle (χ) rotation of a representative s^2U nucleotide; (E) s^2U nucleotide in PDB ID: 4U34 adopting the $C3'-endo$ sugar pucker and a χ value of -160° . These values are normally found in structured RNA as they maximize base–base stacking and minimize strain and clashes.

protected RNA phosphoramidites were obtained from ChemGenes and were dissolved in acetonitrile to a concentration of 0.1 M. Coupling was performed using 5-(benzylmercapto)-1H-tetrazole (5-BMT) (0.25 M) in acetonitrile over 10 min. 0.02 M I_2 in THF/Py/ H_2O solution was used as oxidizing reagent to maintain the oxidation state of the 2-thio-modification during synthesis. All other reagents were obtained from Glen Research. Synthesis was performed on the appropriate nucleoside immobilized via a succinate linker to controlled pore glass (CPG-500). All oligonucleotides were prepared in DMTr-off form. After synthesis, RNAs were cleaved from the solid support and fully deprotected with $NH_4OH:EtOH$ (3:1 v/v) at $55^\circ C$ overnight. The solvent was completely removed by SpeedVac concentration and the dried material was treated with 1 ml of $Et_3N \cdot 3HF$ at room temperature for 8 h. The reaction was quenched with 1 ml of water and the RNA was precipitated by adding 0.2 ml of 3 M sodium acetate and 6 ml of *n*-butanol. The solution was cooled to $-30^\circ C$ for 1 h before the RNA was recovered by centrifugation and finally dried *in vacuo* before purification by reverse-phase HPLC (Agilent 1100 series LC) on a Zorbax ExtendC18 column 9.4 mm \times 250 mm, 5 μ particle size (Agilent Technologies) equilibrated with 30 mM triethylammonium bicarbonate/2%

acetonitrile, pH 8.0, and eluted with an acetonitrile gradient. Purified nucleotides were lyophilized to dryness and stored at $-80^\circ C$.

Isothermal titration calorimetry (ITC)

All ITC experiments were performed using a TA Instruments Nano ITC. Samples were prepared by diluting stock solutions of HPLC-purified oligonucleotides into buffers containing 200 mM NaHEPES, pH 7.5 and 100 mM of NaCl. Oligonucleotide concentrations were verified by UV absorbance at 260 nm using extinction coefficients predicted by primary sequence. 50 μ l samples of oligonucleotide solutions at various concentrations were titrated into the sample cell containing 170 μ l of the complementary oligonucleotide at concentrations typically 5–10 fold lower. Samples were stirred at 350 rpm and $25^\circ C$ with 220 s between injections of 1.5–2.5 μ l. The values of ΔH , K_d and stoichiometry (n) were measured for the $s^2U:A$ and $U:A$ duplexes by fitting the data to a model described previously (28). Because of significant self-dimerization observed for 3'-aucgUgg-5' by optical melting, a more sophisticated model was required to afford thermodynamic parameters for $s^2U:U_c$ and $U:U_c$ duplexes. Details of this model can be

found in the SI. In all cases, ΔS was calculated using the relationship: $\Delta G = RT \ln(K_d) = \Delta H - T\Delta S$ (where R is the gas constant and T is temperature). For ease of comparison to values of ΔH , ΔS is expressed along with temperature ($T\Delta S$), where T is 298.15 K. Experiments were performed in triplicate and standard errors are reported as standard deviations. Control experiments with the inverted direction of titration revealed no significant differences.

Optical melting

Samples were prepared by diluting stock solutions of HPLC-purified oligonucleotides into buffers containing 200 mM NaHEPES, pH 7.5 and 100 mM of NaCl. Oligonucleotide concentrations were verified by UV absorbance at 260 nm using extinction coefficients predicted by primary sequence. All samples were heated to 98°C for 3 min and subsequently cooled to room temperature for 1 h before data were collected. Experimental RNA concentrations ranged between 200–6.75 μM and thermal melting curves were collected using an Agilent Cary 60 UV-Vis. Absorbance was recorded at 260 nm as the temperature was ramped between 4°C and 89°C at a rate of 1°C min⁻¹. All data were collected in duplicate. Values of T_m were obtained by directly fitting the resulting melting curves to a double baseline, two-state model (29). These values were used to construct plots of T_m^{-1} versus $\ln(C_T/4)$, where C_T is the total RNA concentration. A linear least-squares fit was applied to the resulting plots and the van 't Hoff ΔH and ΔS values were calculated from the slopes and intercepts according to the relationship:

$$\frac{1}{T_m} = \frac{R}{\Delta H} \ln(C_T/4) + \frac{\Delta S}{\Delta H}$$

Values of ΔG were calculated according to the relationship: $\Delta G = \Delta H - T\Delta S$, where T is temperature in K. Errors on the reported values of T_m correspond to half the difference between replicates. In the case of the $U_c:U_c$ homodimer, $C_T/4$ was replaced by C_T .

Circular dichroism spectroscopy of ssRNA

Samples were prepared by diluting stock solutions of HPLC-purified nucleotides to 100 μM in buffer containing 200 mM NaHEPES, pH 7.5 and 100 mM of NaCl. Nucleotide concentrations were verified by UV absorbance at 260 nm using extinction coefficients predicted by primary sequence. RNA samples were heated to 98°C for 3 min and subsequently cooled to room temperature for 1 h before data were collected. CD spectra were collected using an Aviv Instruments CD Spectrometer Model 202.

¹H NMR spectroscopy

NMR samples were prepared by dissolving nucleosides to a concentration of 10 mM in ²H₂O at pH 7.5. ¹H NMR spectra were collected at 25°C using a Varian 400 MHz NMR spectrometer (Oxford AS-400) equipped with a Varian 5 mm broadband PFG (z-gradient) probe. The data were analyzed using the Mnova NMR software suite. The 9 parameter parameterization (30) mode of the Matlab Pseudorotation GUI (31) was used to analyze the vicinal coupling

constants of the ¹H NMR spectra to afford the fractional populations of C3'-endo and C2'-endo states.

Molecular dynamics simulations

MD simulations were performed using the program NAMD 2.9 (32) with the CHARM27 parameter set modified to include parameters for s²U (33). All simulations used periodic boundary conditions with Langevin dynamics and a Langevin piston (34) to maintain the temperature at 298.15 K (with a damping constant of 5 ps⁻¹) and the pressure at 1 atm. The cutoff distance was 16 Å and the Particle Mesh Ewald Method (35) grid density was 1 Å⁻³.

The structures of nucleosides (U and s²U) were solvated in water boxes with periodic boundary conditions in VMD 1.9.1 using standard parameters. The resulting structures were minimized over 10 000 steps of 1 fs each. Unrestrained simulations were performed on each structure. Dihedral phase angles were calculated from the resulting trajectory files using the formula of Cremer and Pople (36). To increase sampling, 5 replica runs were performed with different minimized structures and initial velocities. Nucleosides were equilibrated for 10 ns followed by 50 ns production runs.

Umbrella sampling

Umbrella sampling calculations were performed on nucleosides with a harmonic restraint centered on the dihedral angle describing the rotation of the nucleobase relative to the sugar (O4'-C1'-N1-C2, or the 'chi' angle), varying successively from 0° to 360° every 10° with a force constant of 0.007 kcal·mol⁻¹deg⁻². Each simulation was 400 ps and five replicates were performed for both U and s²U. The free energy landscape was calculated using the Weighted Histogram Analysis Method (WHAM) (37).

Desolvation penalty investigations

Crippen's Fragmentation estimations of LogP were generated in ChemBioDraw Ultra 13.0. Values of LogP were empirically determined by dissolving samples in 200 μl of un-ionized water at 25, 5 and 1.25 mM (verified by UV absorbance) before adding 200 μl of 1-octanol and vigorously vortexing the biphasic solutions before allowing them to separate overnight. LogP was calculated as the log of $[\text{solute}]_{1\text{-octanol}}/[\text{solute}]_{\text{water}}$ (38).

Desolvation energies were calculated using the General Born Ion model in the Adaptive Poisson-Boltzmann Solver (APBS) suite of software (39). First, PDB files were converted to PQR files using the PDB2PQR server version 1.8 with the CHARMM27 force field and the protonation states assigned at pH 7.0. Calculations were then performed using the APBS version 1.4 executable. The input file for this calculation appears in the supporting information.

RESULTS

Isothermal titration calorimetry (ITC) studies and thermal denaturation of ssRNA

We evaluated the thermodynamic contributions of s²U toward RNA duplex stability by ITC. In all our experiments,

a known amount of a single-stranded RNA is titrated into a solution containing a known amount of its complement or single mismatch, and the heat released upon binding is measured. A series of such additions is plotted to yield a curve of heat released as a function of titration progress (Figure 3A). All titration experiments were carried out in 200 mM Na-HEPES, pH 7.5, 100 mM NaCl at 25°C.

It is important to consider that the presence of stable homodimers formed from the ssRNA in either the cell or the syringe during an ITC titration will decrease the observed favorability of the ΔG of duplex formation and will result in less favorable observed values of ΔH . We first carried out a series of variable concentration thermal denaturation experiments on solutions containing only the U, s²U, A or U_c strands, respectively, in order to evaluate this possibility. Thermal denaturation studies performed on the single strand U_c using standard UV absorbance methods indicate that significant homodimerization (formation of U_c:U_c) does occur at concentrations relevant to our ITC studies (Supplementary Figure S5). The resulting melting curves of U_c resemble those of typical double-stranded melting profiles, and the T_m^{-1} values scale linearly with the natural log of concentration, an observation which is consistent with the formation of stable U_c:U_c homodimers. We propose that this homodimer is stabilized by two Watson–Crick G:C and four G:U wobble pairs. The thermodynamic parameters, namely the changes in enthalpy ($\Delta H = -67.6$ kcal mol⁻¹) and the equilibrium constant ($K_d = 6.5$ μM), obtained from the van 't Hoff analysis of this melting data were used in a multistep binding model in order to obtain the true values of ΔG_{ITC} and ΔH_{ITC} for U:U_c and s²U:U_c duplex formation, the derivation of which we explain below in brief.

In contrast, thermal denaturation studies of the other three ssRNAs (U, A and s²U) produced curves of a far less sigmoidal nature not suitable for the reliable calculation of T_m (Supplementary Figure S4). This observation is consistent with only weak intermolecular associations, if any. Although we are unable to quantify the T_m values for these melting processes, we can gain further insight into whether the observed process is inter- or intramolecular by normalizing the melting curves obtained at different concentrations and overlaying them on a single graph. For an intramolecular association, the curves theoretically should overlap completely, while for an intermolecular interaction, the midpoint of the curves should shift monotonically with respect to changes in concentration, indicating a change in melting temperature. The curves for U and s²U, respectively, overlap almost completely (Supplementary Figure S4 A and B) over the range of concentrations employed after normalizing the absorbance by concentration, suggesting that the melting behavior of these single strands arise primarily from an intramolecular process. In the case of the A strand, the melting behavior is significantly different. In particular, there is a concentration-dependent shift of the normalized A260 value at 0.75 (approximately the vertical midpoint) from higher temperature to lower temperature as the concentration of A is decreased (Supplementary Figure S4E). This shift points to an intermolecular interaction, likely the formation of a more weakly bound homodimer, in comparison to that of the U_c strand with the same number of G:C

Watson–Crick but only two G:U wobble pairs. The presence of both the U_c:U_c and A:A dimers in solution during the ITC experiments will cause a decrease in the favorability of both the observed values of ΔH_{ITC} and ΔG_{ITC} to varying degrees.

Having an understanding of the single-stranded RNA dynamics, we went ahead and carried out the four sets of ITC experiments for s²U:A, U:A, s²U:U_c and U:U_c. The observed values of ΔG_{ITC} for U:A and s²U:A are -10.0 and -10.5 kcal mol⁻¹, while the observed values of ΔH_{ITC} are -47.7 and -45.5 kcal mol⁻¹, respectively. These measurements lead to calculated values of $T\Delta S_{ITC}$ (25°C) of -37.7 and -35.0 kcal mol⁻¹. At face value, these ITC results indicate that the presence of s²U results in enhanced duplex stability. The greater duplex stability of s²U:A compared to U:A results from a reduced entropic penalty, the effect of which is partially compensated by a reduction in heat released upon binding. In the case of the duplexes containing a single mismatch, the observed values of ΔG_{ITC} for U:U_c and s²U:U_c are -7.24 and -8.65 kcal mol⁻¹, while the observed values of ΔH_{ITC} are -9.14 and -14.7 kcal mol⁻¹, respectively. Note that these uncorrected observed values of ΔH_{ITC} are drastically more positive in comparison to the values observed for s²U:A and U:A (Table 1). These dramatically less favorable observed values of ΔH_{ITC} are most likely the result of U_c:U_c homodimer formation.

Correcting for the effect of homodimerization on the thermodynamic parameters obtained by ITC demanded the use of a multistep mechanistic model of duplex formation. A complete derivation and description of this model can be found in the SI. The model is based on a mechanism in which the U:U_c (or s²U:U_c) duplex can only form from the fraction of U_c not bound up in the U_c:U_c homodimer. Explicitly, the following multistep equilibrium is assumed:



where L is the equilibrium constant governing the dimerization of U_c and K is that for U:U_c. The value of L was measured independently from the single-strand melts described earlier. From this multistep equilibrium, we show that the value of the observed binding constant K_{OBS} when measured by fitting to a single-step binding isotherm becomes related to K by the following expression:

$$K_{OBS} = \frac{K}{1 + 2L[U_c]} \quad (2)$$

by knowing the value of L and the concentration of free U_c in the cell, we can calculate the true value of K .

From this multistep equilibrium, how the observed enthalpy ΔH_{OBS} is affected by the presence of the U_c:U_c homodimers can be described. Specifically, the observed change in enthalpy is equal (28) to the change in mole fractions of both the U_c:U_c dimer and the U:U_c duplex with respect to the beginning and end of the titration, with each term multiplied by their respective molar enthalpies. This formalism leads to the following result:

$$\Delta H_{OBS} = \Delta H_{U:U_c} - \Delta H_{U_c:U_c} \left(\frac{2L[U_c]}{1 + 2L[U_c]} \right) \quad (3)$$

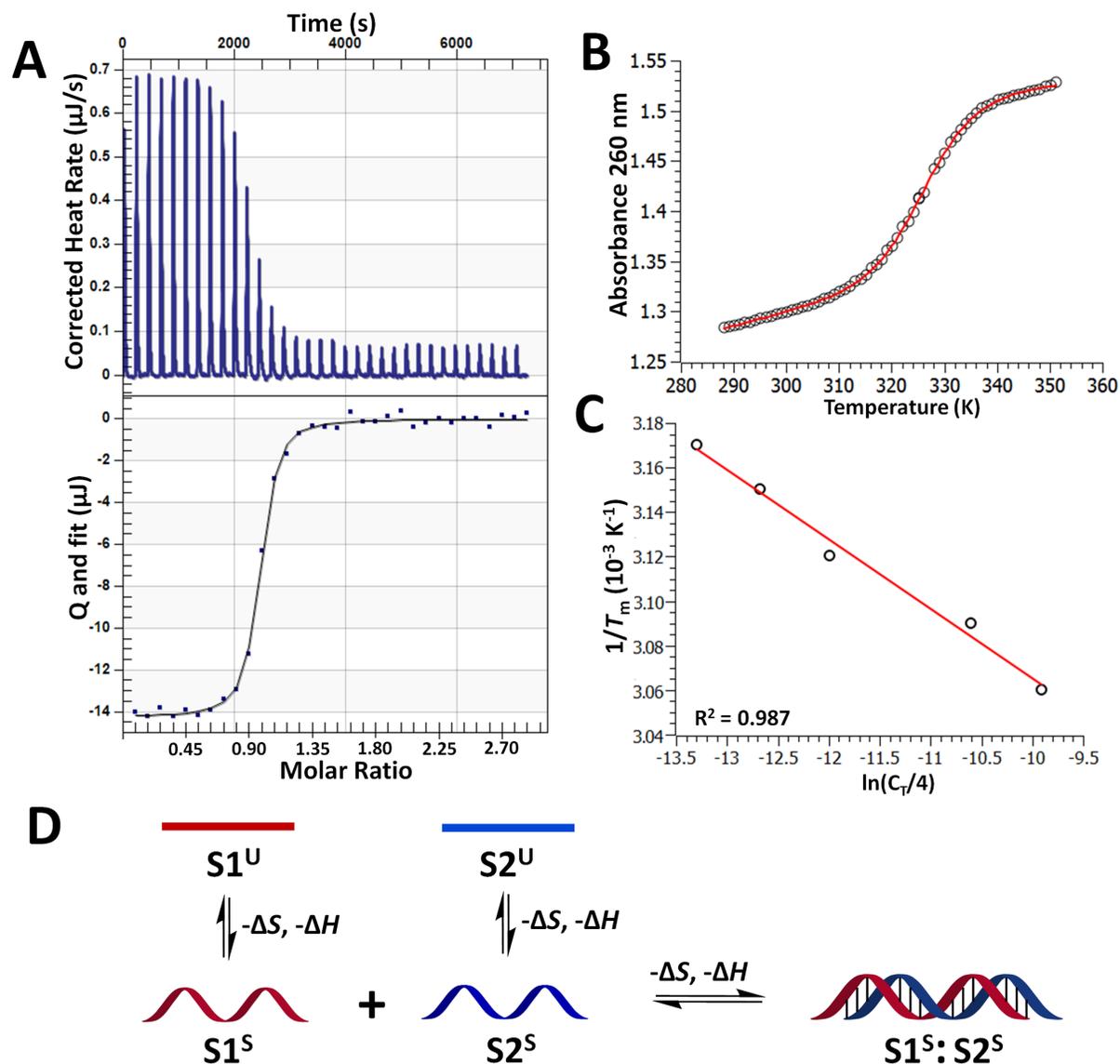


Figure 3. Experimental and mechanistic thermodynamic evaluations of duplex formation. (A) Representative ITC data measuring the thermodynamics of RNA duplex formation of $s^2U:U_c$. The s^2U -containing ssRNA ($100 \mu\text{M}$) was titrated into a solution containing its complement ($10 \mu\text{M}$) in 200 mM NaHEPES, pH 7.5, 100 mM NaCl at 25°C . Top: Raw curve of power versus time. Bottom: Fit of integrated peak areas (heat released per injection) yielding the thermodynamic parameters presented in Table 1. Isotherms for the remaining duplexes appear in Supplementary Figure S1; (B) Melting curve (black circles) collected during the thermal denaturation of $200 \mu\text{M}$ $U:A$ in 100 mM NaCl and 200 mM NaHEPES, pH 7.5 and the double-baseline, two-state fitted model (red line). Melting curves for the remaining duplexes appear in Supplementary Figure S2; (C) Linear least-squared fit (red line) of a van 't Hoff plot of inverse melting temperatures (T_m^{-1}) collected from optical melts at different total oligonucleotide concentrations (C_T) of $U:A$. Additional fits appear in Supplementary Figure S3; (D) Mechanistic scheme depicting the thermodynamic steps of duplex formation, beginning with the transition of single strands from an unstacked ($S1^U$ and $S2^U$) to a stacked ($S1^S$ and $S2^S$) state before the formation of the duplex ($S1^S:S2^S$). Both the stacking of the single strands and the duplex formation steps are expected to result in the release of heat and a negative change in entropy.

where $[U_c]$ is the initial concentration of free U_c at the beginning of the titration.

With this model in hand, we can correct the observed values of ΔG_{ITC} and ΔH_{ITC} for the $U:U_c$ and $s^2U:U_c$ duplexes. The values of ΔG_{OBS} are only slightly affected, becoming -8.18 and $-9.05 \text{ kcal mol}^{-1}$, respectively. The values of ΔH , however, undergo much larger changes to -64.3 and $-55.0 \text{ kcal mol}^{-1}$ (Table 1). Like in the case of the $s^2U:A$ and $U:A$ duplexes, 2-thiolation again results in greater duplex stability, even in the context of a mismatch. Duplex formation is more favorable by $0.5 \text{ kcal mol}^{-1}$ for $s^2U:A$ and by

$0.9 \text{ kcal mol}^{-1}$ for $s^2U:U_c$, compared to their native counterparts, $U:A$ and $U:U_c$. These values are consistent with previously reported optical melting studies of these (27) and other s^2U -containing RNAs (21). The greater duplex stability of $s^2U:U_c$ compared to $U:U_c$ resulted from a reduced entropic penalty ($T\Delta S_{\text{ITC}}$ at $25^\circ\text{C} = -46.0$ and $-56.1 \text{ kcal mol}^{-1}$, respectively), the effect of which is partially diminished by a reduction in heat released upon binding ($\Delta H_{\text{ITC}} = -55.0$ and $-64.3 \text{ kcal mol}^{-1}$, respectively). These results indicate that 2-thiolation increases the favorability of du-

Table 1. The Thermodynamic Parameters of RNA Duplex Formation by Isothermal Titration Calorimetry and Thermal Denaturation

entry	duplex	base pair	K_d^a (nM)	ΔH_{ITC}^a (kcal mol ⁻¹)	$T\Delta S_{ITC}^b$ (kcal mol ⁻¹)	ΔG_{ITC}^c (kcal mol ⁻¹)	ΔG_{vH}^d (kcal mol ⁻¹)	T_m^e (°C)
1	5'-uagcUcc-3' 3'-aucgAgg-5'	U:A	45.2(5)	-47.7(2)	-37.7(3)	-10.0(1)	-11.3 (-12.0)	53.6(1)
2	5'-uagcs ² Ucc-3' 3'-aucgAgg-5'	s ² U:A	21.1(2)	-45.5(2)	-35.0(2)	-10.5(1)	-12.8	64.7(4)
3	5'-uagcUcc-3' 3'-aucgUcg-5'	U:U	1010(120)	-64.3(1)	-56.1(1)	-8.18(1)	-7.6 (-7.7)	35.5(1)
4	5'-uagcs ² Ucc-3' 3'-aucgUcg-5'	s ² U:U	232(26)	-55.0(1)	-46.0(1)	-9.05(1)	-8.4	44.3(7)

^a ΔH_{ITC} and K_d values were evaluated directly from ITC titration data for U:A and s²U:A using a least-squares nonlinear regression analysis described previously (28) and for U:U_c and s²U:U_c using a model designed to compensate for significant self interactions observed by thermal melting for 3'-aucgUcg-5' (for details of this model and stoichiometry values (n), see the SI).

^b $T\Delta S_{ITC}$ was calculated according to $RT \ln(K_d) = \Delta H_{ITC} - T\Delta S_{ITC}$ where R is the gas constant and T is temperature.

^c ΔG_{ITC} was calculated using values of K_d from the ITC data according to $\Delta G_{ITC} = RT \ln(K_d)$. All ITC titrations were performed in triplicate at 25°C.

^d ΔG_{vH} was calculated from thermal denaturation data collected in duplicate and according to $\Delta G_{vH} = \Delta H_{vH} - T\Delta S_{vH}$, where values of ΔH_{vH} and ΔS_{vH} (Supplementary Table S1) were derived from linear fits of van 't Hoff plots of T_m^{-1} versus $\ln(C_T/4)$ where C_T is the total oligonucleotide concentration. The R^2 values of the plots for entries 1, 3 and 4 are approximately 0.99. The R^2 value of entry 2 is slightly lower at 0.94 (Supplementary Figure S3), possibly due to a premelting transition that persists at all oligonucleotide concentrations tested here. Values in parentheses are predictions based upon a nearest neighbor model at 1 M NaCl (40).

^e T_m was calculated from a double baseline, two-state fit of raw optimal melt data at a C_T of 200 μ M (see Materials and Methods for additional details). T_m values are averages and the errors correspond to half of the differences between the measured values.

plex formation for both s²U:A and s²U:U_c and that these effects may both result from analogous mechanisms.

Thermal denaturation of Duplex RNA

To increase confidence in the thermodynamic parameters derived from ITC, we measured the melting temperatures (T_m) of the four dsRNAs under the same conditions as the ITC experiments (100 mM NaCl and 200 mM NaHEPES at pH 7.5) over a range of RNA concentrations using variable temperature UV absorbance methods (Table 1) (41). The resulting melting curves were fit with a double-baseline, two-state model (42) (Figure 3B) and the values of T_m^{-1} were plotted as functions of $\ln(C_T/4)$, where C_T is the total RNA concentration (Figure 3C). The slope and intercept of the resulting van 't Hoff plots afford estimations of ΔG_{vH} , ΔH_{vH} and ΔS_{vH} for duplex formation (29). The T_m and ΔG_{vH} values appear in Table 1.

The values of T_m for the s²U-containing duplexes exceed those of their native counterparts by 11.1°C for s²U:A and 8.8°C for s²U:U_c at 200 μ M total oligonucleotide concentration. These results are consistent with those we reported previously for identical sequences (27) as well as literature values for similar sequences (21) and confirm that s²U enhances the thermal stability of dsRNA. We also observed significant premelting transitions in the case of s²U:A (Supplementary Figure S2), that may arise from changes in secondary structure or torsional rigidity, the fraying of terminal basepairs, or other structural changes (43) and can complicate van 't Hoff analyses. Despite these complications, the values of ΔG_{vH} , ΔH_{vH} and ΔS_{vH} obtained from our thermal denaturation studies exhibit the same trend as those obtained from ITC and, in each case, agree that 2-thiolation enhances the stability of dsRNA. The differences in ΔG_{vH} , ΔH_{vH} and ΔS_{vH} for U:A in comparison to s²U:A are consistent with similar sequences studied by Kierzek *et al.* (21) (Supplementary Table S1). The values of ΔG_{vH} , ΔH_{vH} and ΔS_{vH} for U:A and for U:U_c also conform to the predictions

of nearest-neighbor models, the latter of which takes into consideration the single, internal U:U mismatch (40).

Circular dichroism studies of single-stranded RNA

CD spectra afford information about the chirality and secondary structure of biomolecules, including RNA. The CD spectra of both dsRNA and ssRNA feature a prominent maximum near 280 nm. In both cases, this peak recedes toward the baseline as secondary structure is disrupted by thermal melting or other processes such as chemical denaturation (44). We performed CD melts on the s²U-containing ssRNA and its native counterpart. These experiments were performed under identical conditions to the UV melts. As expected, we observed an absorption maximum near 280 nm which diminished as the temperature was ramped from 4°C to 88°C (Figure 4A). The same trend was observed for both of the ssRNAs, but the CD signal of the s²U-containing ssRNA at 4°C exhibits a prominent feature near 240 nm and a negative peak near 330 nm (Figure 4B). The maximum near 280 nm was greater for the s²U-containing ssRNA over the entire temperature range tested. The signal for the s²U-containing ssRNA recedes at a lesser rate with respect to increasing temperature compared to its native counterpart over the temperature range examined here (Figure 4C). CD signals for U and s²U nucleosides were found to be nearly constant over a similar temperature range (Supplementary Figure S4), suggesting that the helical conformations of the single strands are responsible for the observations described here. For a discussion of the denaturation of the single strands observed by UV, see the SI.

¹H NMR studies of U and s²U nucleosides

The structural effects of s²U on ssRNA potentially originate from the conformational bias of the nucleoside itself. Any given nucleoside will preferentially adopt a sugar pucker conformation in solution that can be described as a distribution between C2'-endo and C3'-endo

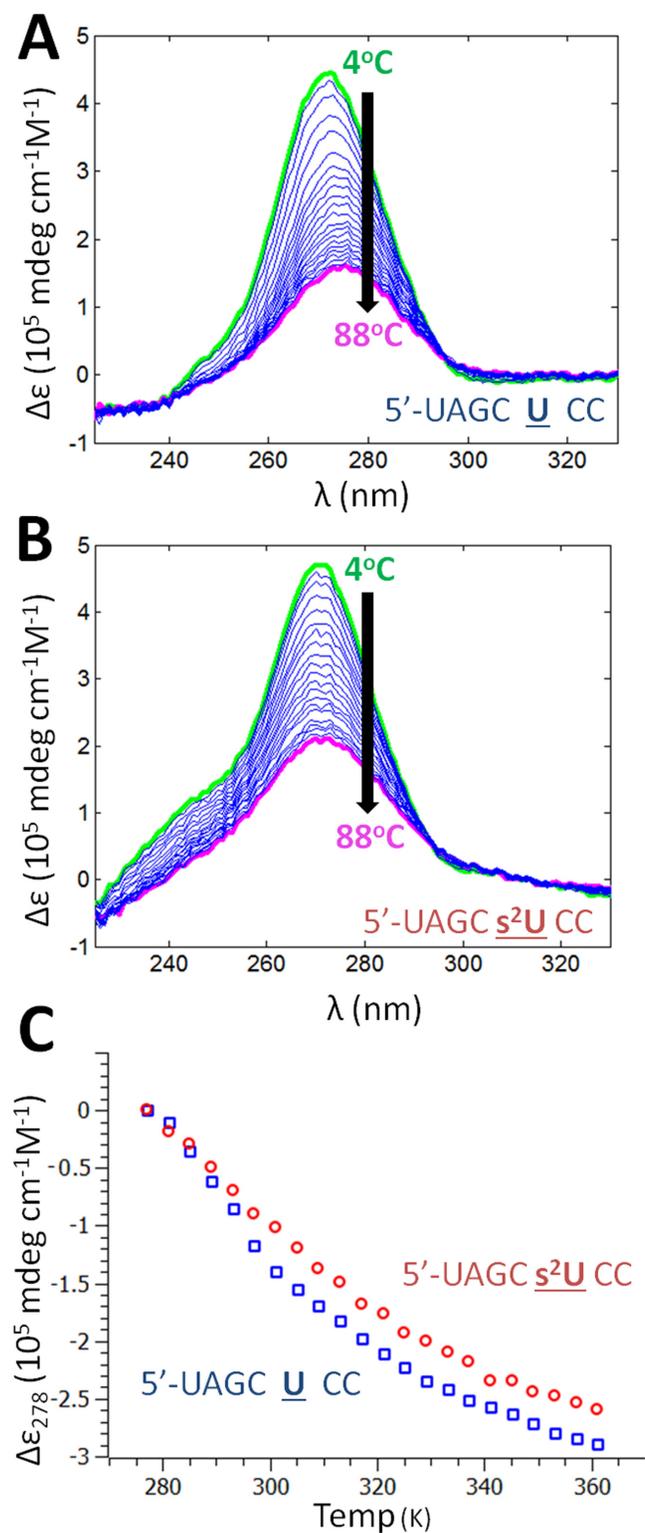


Figure 4. Circular dichroism studies of single-stranded RNA. (A) Overlapped spectra obtained by UV-CD spectroscopy of 5'-uagcUcc-3' at temperatures ranging from 4°C to 88°C using 50 μM oligonucleotide in buffer containing 100 mM NaCl and 200 mM NaHEPES at pH 7.5; (B) The identical experiment performed using 5'-uagcs²Ucc-3'; (C) Normalized UV-CD signal of the s²U-containing ssRNA and its native counterpart at 278 nm.

states (Figure 2). The ribose groups of structured RNA normally prefer C3'-endo conformation (A-form) as it prevents steric clashes between 2'-OH groups and the backbone and results in maximal base stacking interactions (45). It has also been demonstrated that non-enzymatic, template-directed RNA copying is most rapid when all three components—template, primer and activated monomer—are in the A-form conformation (46). To evaluate the effect of 2-thiolation on nucleoside sugar-pucker conformation, we evaluated the NMR spectra of U and s²U nucleosides at 25°C. Although this has been previously investigated, we decided to revisit this approach using a greater magnetic field strength (400 versus 270 MHz) and more precise temperature control, enabling a more reliable assessment of coupling constants. We also employed a more sophisticated version of the Karplus equation, affording more accurate estimations of the fractional populations. The fractional populations of the nucleoside in the C2'-endo and C3'-endo forms were calculated from the vicinal spin-coupling constants using the Matlab Pseudorotation GUI (31), which utilizes Diez's modification of the Karplus equation (47). The quotients of the fractional populations at 25°C give rise to equilibrium constants, enabling the determination of the free energy change separating conformational states using the relationship $\Delta G = -RT \ln(K_{\text{eq}})$, where R is the gas constant and T is the temperature in K. This method confirmed previous results (48) demonstrating that U and s²U nucleosides both preferentially adopt C3'-endo conformations in solution, with the s²U nucleoside favoring C3'-endo to a greater extent. Our results indicate that the fractional population of nucleoside in the C3'-endo conformation was significantly higher for s²U than U (80% and 54%, respectively, Table 2) at 25°C. The free energy difference between C2'-endo and C3'-endo at 25°C is approximately 0.7 kcal mol⁻¹ greater for s²U than U (Table 2).

Molecular dynamics (MD) simulations and umbrella sampling

Simulating molecular dynamics enables the investigation of higher energy conformations and the intermediate states separating them. To assess the effect of 2-thiolation on conformational dynamics, we performed MD simulations and umbrella sampling on U and s²U nucleosides.

Concisely and accurately describing RNA conformations requires a set of carefully chosen collective variables (CVs) that distinguish all significant conformers. The CVs chosen here are the phase angle of pseudorotation (P) and the chi torsion angle (χ) (45). P is a convenient descriptor of nucleoside sugar pucker; values of $\sim 30^\circ$ and $\sim 180^\circ$ correspond to C3'-endo and C2'-endo conformations, respectively. χ characterizes the relative orientation of base and sugar. χ values of approximately -160° are most often observed in structured RNA and correspond to maximal base stacking interactions.

We performed five replicate MD simulations of s²U and U over 50 ns at 25°C with explicit water. The observations of the simulations match those of the NMR studies: s²U occupied the C3'-endo conformation for $\sim 90\%$ of the simulation. This preference was much stronger than that of U which

Table 2. Nucleoside Conformation and Solvation Data

Nucleoside	Fractional population ^a C3'-endo	Fractional population ^a χ (-160°)	ΔG^b C3'-endo (kcal mol ⁻¹)	ΔG^b χ (-160°) (kcal mol ⁻¹)	Desolvation Free Energy (kcal mol ⁻¹) ^c	LogP ^d (calculated)	LogP ^e (measured)
U	54% (71%)	NA (67%)	-0.09 (-0.54)	NA (-0.41)	32.1	-2.28	-1.78(5)
s ² U	80% (92%)	NA (>99%)	-0.81 (-1.43)	NA (NA) ^f	28.7	-1.29	-1.17(7)

^aThe fractional populations of nucleoside in the C2'-endo and C3'-endo form were calculated from ¹H NMR vicinal spin coupling constants using Diez's modification of the Karplus equation. Values in parentheses are from MD simulations (Figure 5).

^bCalculated using the relationship $\Delta G = -RT\ln(K_{eq})$, where R is the gas constant, T is the temperature in K, and K_{eq} is based upon the fractional populations of each conformation. Values in parentheses are from MD simulations.

^cCalculated using the APBS software suite (see Materials and Methods).

^dCalculated using Crippen's Fragmentation, a structure-based estimation.

^eDetermined by partitioning the nucleoside of interest between deionized water and 1-octanol. The error for LogP (measured) is the standard deviation from separate experiments.

^fThere is no observable population in the C2'-endo conformation, invalidating the calculation of ΔG .

spent only ~70% of the simulation in the C3'-endo (Figure 5). These fractional populations indicate that the free energy difference separating C3'-endo and C2'-endo conformations is roughly 0.9 kcal mol⁻¹, a value almost identical to that calculated from ¹H NMR (Table 2). The MD simulations also predict that 2-thiolation has a significant effect on χ . The χ value of s²U remained constant at -160° while U alternated between the conformation corresponding to a χ value of -160° and a second conformation corresponding to 50°. The χ value of 50° appears to be a low energy conformation for the free nucleoside, but it is unlikely to result in optimal base-base stacking in the structured RNA (Figure 2E).

To further investigate the high energy conformations not sampled by MD and to probe the effect of 2-thiolation on the free energy landscapes of P and χ , we proceeded to perform umbrella sampling on the nucleosides. The resulting free energy landscapes (Figure 5C,D) are consistent with the results of the MD simulations. The free energy landscapes of P for both nucleosides feature a global minimum at ~30°, corresponding to the C3'-endo conformation, and a local minimum near ~180°, corresponding to the C2'-endo conformation. The free energy barrier between these states is ~0.5 kcal mol⁻¹ larger for s²U than U. Values of P 'left' of -30° and 'right' of -120° were not sampled as these regions correspond to very high energy conformations not observable by this simulation.

The free energy landscapes of χ for both nucleosides feature prominent minima near -160° and 50°, corresponding to the values of χ observed by MD. For s²U, the global minimum at -160° is ~1 kcal mol⁻¹ lower in energy than the local minimum at 50°. However, for U, the energy of the same two minima is almost indistinguishable.

Evaluation of the desolvation penalty

Solvent-oriented H-bond acceptors are often desolvated during duplex formation (25). The greater the strength of the H-bond acceptor, the greater the free energy penalty for this process (49). Removing the H-bond acceptor entirely or substituting it with a weaker acceptor, such as S, has been shown to partially alleviate this penalty (25). In order to evaluate the effect of desolvation in our system, we

measured the difference in partition coefficients (LogP) between s²U and U nucleosides and compared the measured differences to theoretical predictions (Table 2). The values of LogP for s²U and U, measured directly by partitioning s²U and U nucleosides in a biphasic system of deionized water and octanol, were found to be -1.17 and -1.78, respectively. We employed Crippen's Fragmentation (50), a structure-activity relationship-based approach, to afford theoretical partition coefficient estimations for comparison. These theoretical values of LogP for s²U and U were calculated to be -1.29 and -2.28, respectively, numbers that agree well with the experimentally determined values. Next, we computationally evaluated the Born Ion Solvation Energy (51) using the APBS software suite, and this level of theory agreed that desolvation penalty for s²U is 3.4 kcal mol⁻¹ less than for U (Table 2).

DISCUSSION AND CONCLUSION

Nearly all nucleic acid complexations, including the formation of duplexes, triplexes and tetraplexes, are associated with overall favorable changes in enthalpy ($-\Delta H$) and unfavorable changes of entropy ($-\Delta S$) (52). For short RNAs, duplex formation can be conceptualized thermodynamically as a two-step process (Figure 3D) (29). The single strands must first proceed from an unstacked random coil to a stacked helical conformation before forming the duplex. A favorable change in enthalpy is associated with both steps: heat is released as the bases stack during the ordering of the single strands; and next, as H-bonds are formed during duplex formation. The enthalpic favorability of this process is partially countered by a desolvation penalty (53), but overall, each step can be expected to be associated with a negative change in enthalpy. Hybridization also results in an overall increase in order, as reflected by the negative sign of ΔS associated with RNA duplex formation. The ITC results for each duplex examined here conform to these expectations and the trends in duplex stability are consistent with those determined from the thermal denaturation data and with the predictions of a nearest neighbor analysis (Table 1), with one exception. The inability to quantitatively account for the possible homodimerization of the A strand has likely resulted in values of ΔH_{ITC} observed for U:A and

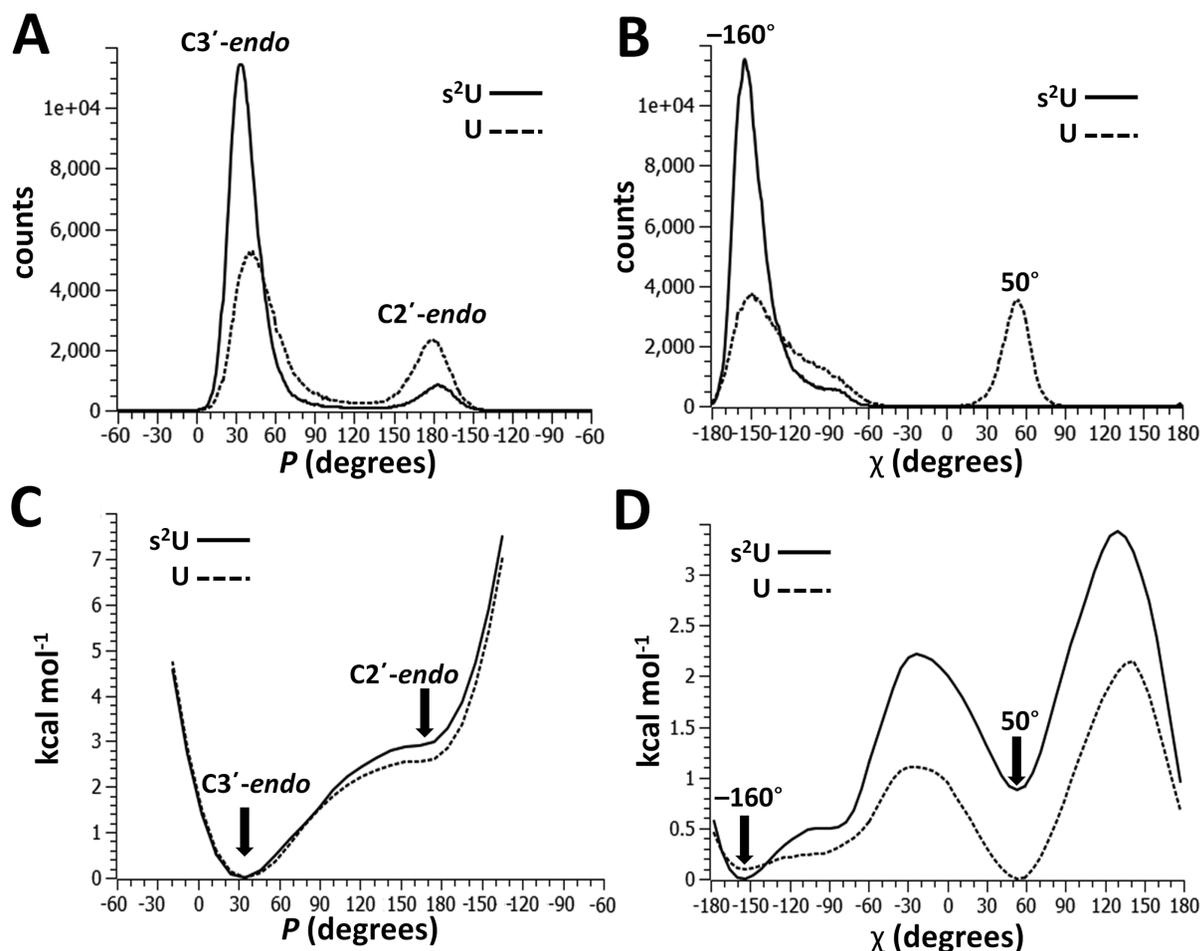


Figure 5. MD simulations and umbrella sampling of U and s²U nucleosides at 25°C. (A) Conformation histogram of the pseudorotation phase angle (P). s²U spends a larger proportion of time in the preferred C3'-endo conformation; (B) Conformation histogram of chi angle (χ). s²U almost exclusively occupies a χ of $\sim -160^\circ$ while U spends a significant portion of the simulation occupying a χ of $\sim 50^\circ$; (C) Free energy landscape for P . Energy minima appear at values corresponding to C3'-endo and C2'-endo conformations. The barrier separating these conformations is larger for s²U; (D) Free energy landscape for χ . Energy minima appear at values corresponding to $\sim -160^\circ$ and $\sim 50^\circ$. The barrier separating these conformations is, again, larger for s²U.

s²U:A (-47.7 and -45.5 kcal mol⁻¹, respectively) that are less negative than their true values. This reasoning likely explains why the observed values of ΔH_{ITC} are less favorable than their van 't Hoff counterparts. In addition, if a correction to the observed values could be applied, we hypothesize the values of ΔH_{ITC} for s²U:A and U:A would become more negative than the corrected values for s²U:U_c and U:U_c. Nevertheless, the main finding that 2-thiolation increases duplex stability by lessening the entropic cost of formation is not expected to change by such a correction, although the magnitude of the difference between the enthalpies of formation for s²U:A and U:A might.

Because ΔH and ΔS are related in part to the formation of H-bonds and stacking interactions during hybridization, mechanistic information can be obtained from the differences in these changes ($\Delta\Delta H$ and $\Delta\Delta S$) resulting from 2-thiolation. The values of $\Delta\Delta H_{\text{ITC}}$ and $\Delta\Delta S_{\text{ITC}}$ are 2.2 kcal mol⁻¹ and 9.1 cal mol⁻¹ K⁻¹, respectively, for s²U:A compared to U:A and 9.3 kcal mol⁻¹ and 33.9 cal mol⁻¹ K⁻¹, respectively, for s²U:U_c compared to U:U_c. The source of this reduced entropic penalty is likely the contribution of

2-thiolation to the structural preorganization (26) of the ss-RNA. Similar preorganizing effects have been implicated in biasing anticodon structure toward the A-form which has the effect of improving anticodon/codon interactions (54). This may partially explain the high conservation of s²U at the tRNA anticodon site.

The defining feature of a preorganized structure is the loss of some amount of its conformational flexibility (26), which is often assumed to result in smaller entropic penalties during binding (ex. hybridization) (52). By this reasoning, any change that shifts the equilibrium of a single stranded RNA toward the highly stacked and conformationally ordered state—*i.e.* single-strand preorganization—will result in a reduced entropic penalty for duplex formation. Recalling the definition of Gibbs free energy, $\Delta G = \Delta H - T\Delta S$, a change that leads to enhanced preorganization of the ssRNA will most likely correspond to a positive $\Delta\Delta S$. A more preorganized single strand also requires less additional stacking to facilitate duplex formation, corresponding to less heat being released and a positive $\Delta\Delta H$. These differences are precisely what are observed by ITC when comparing duplex

formation for the s^2U -containing duplexes against their native counterparts ($s^2U:A$ vs $U:A$ and $s^2U:U_c$ vs $U:U_c$) and is consistent with the results of Kumar (17). The results of the single-strand thermal denaturation experiments monitored by CD are consistent with the hypothesis that 2-thiolation leads to greater single-strand order between 4 and 88°C and greater resistance to melting, further supporting the hypothesis that 2-thiolation contributes to a more highly ordered single strand.

Further evidence for the hypothesis that s^2U preorganizes the single strand can be found by comparing the crystal structures of $s^2U:U_c$ and $U:U_c$ (Figure 2). The crystal structure of $U:U_c$ contains four duplexes with four distinct H-bond geometries at the site of the $U:U$ base pair, two of which appear ‘weak’ based upon $[N-H\cdots O]$ bond lengths and angles (Figure 2C). These distinct geometries suggest that $U:U_c$ exists in solution in several conformations of comparable energy with disparate H-bond strengths. The crystal structure of $s^2U:U_c$, however, contains only a single duplex conformation with a single ‘strong’ H-bond geometry at the site of the $s^2U:U$ base pair, in which the S atom is directly involved as an H-bond acceptor. Based upon these structures, it appears that 2-thiolation favors a single strong H-bond geometry that resolves the duplex structure into a single conformation. This explanation is consistent with the observed thermodynamic parameters and is compatible with the hypothesis that 2-thiolation preorganizes the s^2U -containing single strand.

To evaluate the hypothesis that 2-thiolation contributes to a stronger N3 H-bond due to the increased acidity of the N3 imino proton, we repeated the determination of pKa values for both U and s^2U nucleosides using pH titrations monitored by UV and NMR spectroscopies (Supplementary Figure S7). Both methods agree that the N3 imino proton of s^2U is more acidic than that of U by ~1 pKa unit. This observed difference likely results from the much lower pKa of an aryl thiol versus an aryl hydroxyl, which would stabilize the thiolate resonance structure of the conjugate base. Although the thermodynamic data collected here cannot be directly used to confirm a stronger N3 H-bond resulting from 2-thiolation, this hypothesis remains plausible.

We propose that the source of the observed single strand stabilization is, to a large degree, the effect of 2-thiolation on the nucleoside conformation. The 1H NMR results demonstrate that 2-thiolation stabilizes the preferred $C3'$ -*endo* conformation of the nucleoside in solution. The free energy difference observed by NMR at the nucleoside scale is of similar magnitude to the difference observed for duplex formation by ITC (Table 1, all at 25°C). The NMR results are well complemented by MD and umbrella sampling efforts, both of which agree that 2-thiolation produces a greater preference for the $C3'$ -*endo* conformation and a χ angle comparable to that observed in structured RNA. The predicted differences in free energy are similar to those observed by NMR, ITC and duplex thermal denaturation. Previous NMR studies (48) have suggested that 2-thiolation results in a steric interaction between the 2-thiocarbonyl and the bulky 2'-OH, potentially resulting in the heightened free energy barriers predicted here.

With respect to desolvation, the importance of this effect may be highly dependent on the duplex under consideration. The O2 atom of U5 in the crystal structure of $U:A$ does not participate directly in a base pair H-bond but is situated within H-bond distance (2.9 Å, measured from the center of the S and O atoms) of a water oxygen atom (Supplementary Figure S8). In addition to possessing a highly similar overall structure to that of $U:A$ (overlap RMSD of < 0.2 Å), the crystal structure of $s^2U:A$ shares both of these features—the S2 atom does not participate directly in a base pair H-bond but is situated within H-bond distance (3.4 and 3.5 Å) of two water oxygen atoms. These distances agree well with the known geometric characteristics of H-bonds involving S atoms (55). In this case, 2-thiolation does not appear to result in desolvation of the S atom and, therefore, we hypothesize that a reduced desolvation penalty for s^2U is unlikely to contribute greatly toward the enhanced free energy of duplex formation for $s^2U:A$.

In contrast, the S atom in the crystal structure of $s^2U:U_c$ participates directly in a base pair H-bond and appears to be fully desolvated as no water oxygen atoms are resolvable within 4 Å from the S2 atom. The crystal structure resolution of $U:U_c$ is insufficient to determine whether or not the corresponding O2 atom is solvated. In the case of $s^2U:U_c$, it is at least plausible that a reduced desolvation penalty for s^2U may contribute to the enhanced free energy of duplex formation.

s^2U and other modified nucleobases may prove vital to improving the fidelity and rate of template-directed, nucleic acid replication systems, enzymatic or otherwise. The ubiquity and conservation of modified nucleobases strongly suggests that this chemistry has been commonly employed by nature to strengthen weak interactions and enhance accurate base–base recognition. From either a chemical biology or origins of life perspective, it is conceivable that the use of modified nucleobases may assist in the realization of a self-replicating RNA system. From a purely chemical perspective, it is impressive to consider the significant effect that a single, seemingly innocuous, atomic substitution can have on a large and complex system. By augmenting and informing ongoing structural and *in silico* studies of other modified nucleobases, this work may inform future studies of exotic nucleic acids and assist in their growing utility.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

The Autonomous Province of Trento, Call ‘Grandi Progetti 2012’, project ‘Characterizing and improving brain mechanisms of attention—ATTEND [CHE-0809413]; Simons Foundation [290363 to J.W.S.]. A.T.L. is partially supported by a fellowship from Fonds de recherche du Québec—Nature et technologies. A.C.F. would like to thank ELSI for a postdoctoral fellowship award. J.W.S. is an Investigator of the Howard Hughes Medical Institute. Funding for open access charge: Simons Collaboration on the Origins of Life.

Conflict of interest statement. None declared.

REFERENCES

- Fedor, M.J. (2009) Comparative enzymology and structural biology of RNA self-cleavage. *Annu. Rev. Biophys.*, **38**, 271–299.
- Doudna, J.A. and Lorsch, J.R. (2005) Ribozyme catalysis: not different, just worse. *Nat. Struct. Mol. Biol.*, **12**, 395–402.
- Henkin, T.M. (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev.*, **22**, 3383–3390.
- Roth, A. and Breaker, R.R. (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.*, **78**, 305–334.
- Leontis, N.B., Stombaugh, J. and Westhof, E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
- Szostak, J.W. (2012) The eightfold path to non-enzymatic RNA replication. *J. Syst. Chem.*, **3:2**, 1–14.
- Sheng, J., Gan, J., Soares, A.S., Salon, J. and Huang, Z. (2013) Structural insights of non-canonical U•U pair and Hoogsteen interaction probed with Se atom. *Nucleic Acids Res.*, **41**, 10476–10487.
- Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendex, F.A.P., Fabris, D. and Agris, P.F. (2011) The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.*, **39**, D195–D201.
- Gilbert, W. (1986) Origin of life: The RNA world. *Nature*, **319**, 618.
- Agris, P.F. (2004) Decoding the genome: a modified view. *Nucleic Acids Res.*, **32**, 223–238.
- Rodriguez-Hernandez, A., Spears, J.L., Gaston, K.W., Limbach, P.A., Gamper, H., Hou, Y.-M., Kaiser, R., Agris, P.F. and Perona, J.J. (2013) Structural and mechanistic basis for enhanced translational efficiency by 2-thiouridine at the tRNA anticodon wobble position. *J. Mol. Biol.*, **425**, 3888–3906.
- Vendex, F.A.P., Murphy IV, F.V., Cantara, W.A., Leszczyńska, G., Gustilo, E.M., Sproat, B., Malkiewicz, A. and Agris, P.F. (2012) Human tRNA^{Lys3}_{UUU} is pre-structured by natural modifications for cognate and wobble codon binding through keto–enol tautomerism. *J. Mol. Biol.*, **416**, 467–485.
- Eigen, M. (1971) Selforganization of matter and evolution of biological macromolecules. *Naturwissenschaften*, **58**, 465–523.
- Attwater, J., Wochner, A. and Holliger, P. (2013) In-vivo evolution of RNA polymerase ribozyme activity. *Nat. Chem.*, **5**, 1011–1018.
- Hagenbuch, P., Kervio, E., Hochgesand, A., Plutowski, U. and Richert, C. (2005) Chemical primer extension: efficiently determining single nucleotides in DNA. *Angew. Chem. Int. Ed.*, **44**, 6588–6592.
- Rajamani, S., Ichida, J.K., Antal, T., Treco, D.A., Leu, K., Nowak, M.A., Szostak, J.W. and Chen, I.A. (2010) Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *J. Am. Chem. Soc.*, **132**, 5880–5885.
- Kumar, R.K. and Davis, D.R. (1997) Synthesis and studies on the effect of 2-thiouridine and 4-thiouridine on sugar conformation and RNA duplex stability. *Nucleic Acids Res.*, **25**, 1272–1280.
- Strobel, S.A. and Cech, T.R. (1995) Minor groove recognition of the conserved G-U pair at the Tetrahymena ribozyme reaction site. *Science*, **267**, 675–679.
- Shohda, K.-I., Okamoto, I., Wada, T., Seio, K. and Sekine, M. (2000) Synthesis and properties of 2'-O-Methyl-2-thiouridine and oligoribonucleotides containing 2'-O-Methyl-2-thiouridine. *Bioorg. Med. Chem. Lett.*, **10**, 1795–1798.
- Heuberger, B.D., Pal, A., Del Frate, F., Topkar, V.V. and Szostak, J.W. (2015) Replacing uridine with 2-thiouridine enhances the rate and fidelity of nonenzymatic RNA primer extension. *J. Am. Chem. Soc.*, **137**, 2769–2775.
- Testa, S.M., Disney, M.D., Turner, D.H. and Kierzek, R. (1999) Thermodynamics of RNA–RNA duplexes with 2- or 4-thiouridines: implications for antisense design and targeting a group I intron. *Biochemistry*, **38**, 16655–16662.
- Mazumdar, S.K., Saenger, W. and Scheit, K.H. (1974) Molecular structure of poly-2-thiouridylic acid, a double helix with non-equivalent polynucleotide chains. *Mol. Biol.*, **85**, 213–219.
- Pal, B.C. (1971) Synthesis of 2-thiouridine and 2-thioisouridine by Mercuri procedure. *J. Org. Chem.*, **36**, 3026–3028.
- Agris, P.F., Sierzputowska-Gracz, H., Smith, W., Malkiewicz, A., Sochacka, E. and Nawrot, B. (1992) Thiolation of uridine carbon-2 restricts the motional dynamics of the transfer RNA wobble position nucleoside. *J. Am. Chem. Soc.*, **114**, 2652–2656.
- Siegfried, N.A., Kierzek, R. and Bevilacqua, P.C. (2010) Role of unsatisfied hydrogen bond acceptors in RNA energetics and specificity. *J. Am. Chem. Soc.*, **132**, 5342–5344.
- Cram, D.J. (1986) Preorganization - from solvents to spherands. *Angew. Chem. Int. Ed.*, **25**, 1039–1057.
- Sheng, J., Larsen, A., Heuberger, B.D., Blain, J.C. and Szostak, J.W. (2014) Crystal structure studies of RNA duplexes containing s²U:A and s²U:U base pairs. *J. Am. Chem. Soc.*, **136**, 13916–13924.
- Sigurskjold, B.W. (2000) Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal. Biochem.*, **277**, 260–266.
- Mikulecky, P.J. and Feig, A.L. (2006) Heat capacity changes associated with DNA duplex formation: salt- and sequence-dependent effects. *Biochemistry*, **45**, 604–616.
- van Wijk, J., Huckriede, B.D., Ippel, J.H. and Altona, C. (1992) Furanose sugar conformations in DNA from NMR coupling-constants. *Methods Enzymol.*, **211**, 286–306.
- Hendrickx, P.M.S. and Martins, J.C. (2008) A user-friendly Matlab program and GUI for the pseudorotation analysis of saturated five-membered ring systems based on scalar coupling constants. *Chem. Cent. J.*, **2**, 20.
- Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L. and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
- Sarzyńska, J. and Kuliński, T. (2005) Parametrization of 2-thiouracil and 4-thiouracil in CHARMM all-atom empirical force field. *Comput. Methods Sci. Technol.*, **11**, 49–55.
- Feller, S.E., Zhang, Y., Pastor, R.W. and Brooks, B.R. (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.*, **103**, 4613–4621.
- Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
- Cremer, D. and Pople, J.A. (1975) General definition of ring puckering coordinates. *J. Am. Chem. Soc.*, **97**, 1354–1358.
- Kumar, S., Bouzida, D., Swendsen, R.H., Kollman, P.A. and Rosenberg, J.M. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, **13**, 1011–1021.
- Sangster, J. (1989) Octanol-water partition-coefficients of simple organic-compounds. *J. Phys. Chem. Ref. Data*, **18**, 1111–1229.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10037–10041.
- Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
- Koštutić, M., Jud, L., Da Veiga, C., Frener, M., Fauster, K., Kreutz, C., Ennifar, E. and Micura, R. (2014) Surprising base pairing and structural properties of 2'-trifluoromethylthio-modified ribonucleic acids. *J. Am. Chem. Soc.*, **136**, 6656–6663.
- Tuner, D.H., Sugimoto, N. and Freier, S.M. (1988) RNA structure prediction. *Ann. Rev. Biophys. Biophys. Chem.*, **17**, 167–192.
- Schurr, J.M., Delrow, J.J., Fujimoto, B.S. and Benight, A.S. (1997) The question of long-range allosteric transitions in DNA. *Biopolymers*, **44**, 283–308.
- Rinnenthal, J., Klinkert, B., Narberhaus, F. and Schwalbe, H. (2010) Direct observation of the temperature-induced melting process of the *Salmonella* fourU RNA thermometer at base-pair resolution. *Nucleic Acids Res.*, **38**, 3834–3847.
- Li, L. and Szostak, J.W. (2014) The free energy landscape of pseudorotation in 3'-5' and 2'-5' linked nucleic acids. *J. Am. Chem. Soc.*, **136**, 2858–2865.
- Zhang, N., Zhang, S. and Szostak, J.W. (2012) Activated ribonucleotides undergo a sugar pucker switch upon binding to a single-stranded RNA template. *J. Am. Chem. Soc.*, **134**, 3691–3694.
- Diez, E., Esteban, A.L., Guilleme, J. and Bermejo, F.L. (1981) Relationships between torsional angles and ring-puckering coordinates. *J. Mol. Struct.*, **70**, 61–64.

48. Yokoyama,S., Watanabe,T., Murao,K., Ishikura,H., Yamaizumi,Z., Nishimura,S. and Miyazawa,T. (1985) Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 4905–4909.
49. Auffinger,P. and Westhof,E. (1998) Hydration of RNA Base Pairs. *J. Biomol. Struct. Dyn.*, **16**, 693–707.
50. Wildman,S.A. and Crippen,G.M. (1999) Prediction of physicochemical parameters by atomic contributions. *J. Chem. Inf. Comput. Sci.*, **39**, 868–873.
51. Atkins,P.W. and MacDermott,A.J. (1982) The Born equation and ionic solvation. *J. Chem. Educ.*, **59**, 359–360.
52. Kool,E.T. (1997) Preorganization of DNA: design principles for improving nucleic acid recognition by synthetic oligonucleotides. *Chem. Rev.*, **97**, 1473–1487.
53. Klostermeier,D. and Millar,D.P. (2002) Energetics of hydrogen bond networks in RNA: hydrogen bonds surrounding G+1 and U42 are the major determinants for the tertiary structure stability of the hairpin ribozyme. *Biochemistry*, **41**, 14095–14102.
54. Takai,K. and Yokoyama,S. (2003) Roles of 5-substituents of tRNA wobble uridines in the recognition of purine-ending codons. *Nucleic Acids Res.*, **31**, 6383–6391.
55. Zhou,P., Tian,F., Lv,F. and Shang,Z. (2009) Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins*, **76**, 151–163.