

Context-sensitivity of isosteric substitutions of non-Watson–Crick basepairs in recurrent RNA 3D motifs

Emil F. Khisamutdinov^{1,2,*}, Blake A. Sweeney^{3,4} and Neocles B. Leontis¹

¹Department of Chemistry and Center for Photochemical Science, Bowling Green State University, Bowling Green, OH 43403, USA, ²Department of Chemistry, Ball State University, Muncie, IN 47306, USA, ³Department of Biological Sciences, Bowling Green State University, Bowling Green, OH 43403, USA and ⁴European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK

Received July 10, 2021; Editorial Decision July 26, 2021; Accepted July 29, 2021

ABSTRACT

Sequence variation in a widespread, recurrent, structured RNA 3D motif, the Sarcin/Ricin (S/R), was studied to address three related questions: First, how do the stabilities of structured RNA 3D motifs, composed of non-Watson–Crick (non-WC) basepairs, compare to WC-paired helices of similar length and sequence? Second, what are the effects on the stabilities of such motifs of isosteric and non-isosteric base substitutions in the non-WC pairs? And third, is there selection for particular base combinations in non-WC basepairs, depending on the temperature regime to which an organism adapts? A survey of large and small subunit rRNAs from organisms adapted to different temperatures revealed the presence of systematic sequence variations at many non-WC paired sites of S/R motifs. UV melting analysis and enzymatic digestion assays of oligonucleotides containing the motif suggest that more stable motifs tend to be more rigid. We further found that the base substitutions at non-Watson–Crick pairing sites can significantly affect the thermodynamic stabilities of S/R motifs and these effects are highly context specific indicating the importance of base-stacking and base-phosphate interactions on motif stability. This study highlights the significance of non-canonical base pairs and their contributions to modulating the stability and flexibility of RNA molecules.

INTRODUCTION

Compared to double stranded DNA, the interaction patterns of bases in structured RNA molecules are considerably more complex. Atomic resolution structures of large

RNAs have revealed that a large fraction of interacting bases do not form Watson–Crick (WC) base pairs. In fact, the standard *cis* Watson–Crick/Watson–Crick (cWW) AU and GC base pairs only account for ~70% of all base pairs in structured RNAs including the small and large ribosomal RNA subunits (1–3). The remaining one-third of all base pairs are non-WC base pairs. The non-WC base pairs occur primarily in hairpin, internal or junction loops. They serve to structure these loops and to expose functional groups, especially those on the Watson–Crick edges, to form specific binding sites for proteins, small molecule ligands, metals, or other RNAs (4).

Previous work on RNA adaptation to different environments has focused on the ratio of GC to AU Watson–Crick base pairs (5) as GC base pairs are generally more stable than AU pairs. These authors found that in ribosomal RNA (rRNA) the relationship between the proportion of GC base combinations at Watson–Crick base paired positions and the optimal growth temperatures is very weak up to ~60°C optimal growth temperature. Above this temperature, there is a roughly linear relationship between the percent of GC pairs and the optimal temperature, as GC base pairs generally confer more stability to Watson–Crick helices than AU pairs. The same positive correlation of GC content with optimal growth temperature was found in helices of the RNA component of the signal recognition particle among Archaea (6). In contrast, no such simple relationship was found to exist for genomic DNA; the average GC content varies widely from approximately 30% to more than 60% in bacteria irrespective of the optimal growth temperature (7). Many loop regions are important for the function and overall structure of RNA molecules. Therefore, we suspect that the sequences of loop regions may also adapt to temperature.

In our recent work, we compared the 3D structures of the rRNAs of the distantly related bacteria *Escherichia coli* (mesophile) and *Thermus thermophilus* (thermophile) by aligning the structurally conserved base-paired positions in

*To whom correspondence should be addressed. Tel: +1 765 285 8084; Fax: +1 765 285 6505; Email: kemil@bsu.edu

the 5S, 16S and 23S rRNAs (1). We aligned all non-WC as well as Watson–Crick base pairs, and found that (i) the geometric families of aligned base pairs were nearly 100% conserved, and (ii) almost all base substitutions resulted in isosteric or near isosteric base pairs. This result strongly supported the hypothesis that base pair isostericity is fundamental for understanding the rules of sequence adaptation during RNA evolution. Isostericity indicates which base pairs can potentially substitute for each other preserving the 3D structure of the motif. However, nothing is known about how the isosteric base substitutions impact the stabilities of well-structured RNA 3D motifs and, if there is a selection for particular base combinations depending on optimal growth temperature of the organism. These related questions were addressed in this paper.

The energetic values of RNA base pairs have been studied for several decades. Initially, the focus was on the canonical (i.e., G-C, A-U, and G-U) base pairs and later studies were included non-canonical base pairs (8–11). Particularly, thermodynamic studies of RNA internal loops have taken a ‘bottom up’ approach which attempts to cover the sequence space of internal loops starting systematically with small loops (1×1 , 1×2 , 2×2 , 1×3 and 2×3) (12). Such an approach quickly confronts the problem of combinatorial explosion. It is simply not possible to study all possible sequence variants of internal loops of size 4×4 or greater without a high throughput methodology.

In this work, we use a different approach. We chose a recurrent, well-structured and relatively large internal loop (5×4 nucleotides not counting the closing bases or 6×5 with closing bases) for detailed study. We have chosen the sarcin-ricin (S/R) motif as the focus of study of phylogenetic and ecological sequence variation and thermodynamic measurements for these reasons:

(i) The motif is recurrent and widespread in structured RNA molecules (13–15); (ii) It is highly conserved, yet different sequence compositions can give essentially the same 3D structure of S/R motif as it will be shown below; (iii) the S/R motif is modular and autonomous, it occurs independently in different molecules and in different locations in the same molecule (16); (iv) the motif is unique, the G-bulged region of the S/R motif was found to have unusually complex electronic structure arrangement due to the combination of S-turn, stacking and base-phosphate (BPh) interactions which was not observed in similar small RNA internal loops (17) making the S/R motif very difficult target for molecular modeling and simulation force fields.

We began by cataloguing the sequence variations observed in atomic-resolution 3D structures, and then we examined sequence variations of the motif at corresponding positions in aligned sequences. We collected these data to choose sequence variants for thermodynamic characterization. We also studied sequences containing single mutations that are predicted to disrupt the motif, guided by base pair isostericity matrices (18).

Thermodynamic analysis of S/R motif variations found in organisms occupying different temperature niches and present in 3D structures indicates that a single change in a non-WC pair can significantly affect the overall stability

of the motif. To verify whether sequence variants that have different thermodynamic stabilities still form the same motif, we probed their solution structures using single-strand specific nucleases T1 and A, double strand specific nuclease V1, as well as, DEPC. We found that isosteric base pair substitutions preserve the overall conformation of the motif, while single base substitutions that disrupt key non-WC base pairs destabilize the entire S/R motif.

MATERIALS AND METHODS

RNA oligonucleotides

RNA oligonucleotides used for thermodynamic studies were obtained from IDT DNA (www.idtdna.com) and purified using preparative denaturing polyacrylamide gel electrophoresis (PAGE) or preparative 20 cm \times 20 cm thin-layer chromatography (TLC) (*n*-propanol:ammonium hydroxide:water, 55:35:10). RNA was eluted from gel slices or TLC matrix using a buffer containing 0.3 M NaCl, 10 mM Tris pH 8.0, 0.5 mM EDTA, then ethanol precipitated with 2.5 volume of cold 100% ethanol, rinsed twice with cold 80% ethanol, dried, and dissolved in 30 μ l of double deionized (dd H₂O) water. Purified RNAs produced single bands on a 16.5 cm \times 20 cm analytical denaturing 20% polyacrylamide gel. For UV-melting measurements, the RNA stock solutions were extensively dialyzed first against 10 mM EDTA, 100 mM NaCl, and 10 mM sodium cacodylate and then against dd H₂O and dissolved in dd H₂O for storage. For each melting profile, a small volume of the oligomer stock was dissolved in 0.1 ml of buffer containing 10 mM sodium cacodylate (pH 6.9), 0.5 mM EDTA and either 10 mM MgCl₂ and 100 mM NaCl or just 1.0 M NaCl.

The RNA hairpin molecules for structural probing were prepared by runoff transcription of PCR amplified DNA templates. Synthetic DNA molecules coding for the antisense sequence of the desired RNA were purchased from IDT DNA (www.idtdna.com) and amplified by PCR using primers containing the T7 RNA polymerase promoter. Template strands for T7 transcription were designed using the online program written by Jesse Stombaugh (link: <http://rna.bgsu.edu/oldwebsite/rnatodna.html>). PCR products were purified using the QiaQuick PCR purification kit (Qiagen Sciences, Maryland 20874). RNA molecules were prepared by *in vitro* transcription using T7 RNA polymerase (Takara Bio Inc., <http://www.takara-bio.com>) and purified on denaturing PAGE as describe above.

5'-end ³²P ATP labeling of RNA oligonucleotides. RNA hairpins, prepared by *in vitro* transcription, were labeled with gamma ³²P at the 5' terminus. Molecules were first dephosphorylated using Calf Intestinal alkaline phosphatase—10 000 U/ml (CIP, <http://www.neb.com>) using the manufacture's procedure. The dephosphorylated RNAs were 5'-end labeled with [γ -³²P] ATP (PerkinElmer Co.) using T4 polynucleotide kinase (New England Biolabs Inc.) according to procedure adopted from Ambion Applied Biosystems INC (http://www.ambion.com/techlib/misc/RNA5_labeling.html). Labeled RNAs were purified on 20% denaturing PAGE before use.

Thermodynamic experiments

Optical melting experiments. Absorbance versus temperature profiles were measured to determine the melting transitions of the duplex RNA molecules and to characterize the temperature-dependent transitions of the single strands. To form duplexes, complementary single strands were mixed at equimolar concentrations. All melting curves were obtained in buffer containing 10 mM sodium cacodylate buffer (pH 6.9), 0.5 mM EDTA and either 0.1 M NaCl/10 mM MgCl₂ or just 1 M NaCl. Single-stranded extinction coefficients were calculated from the extinction coefficients for dinucleotide monophosphate and nucleosides using the literature parameters (19). Strand concentrations were determined from high-temperature (90°C) absorbance at 260 nm (20). Absorbances were measured in 1 cm path length quartz cuvettes (Starna Cells, Inc), using a 100 CARY-BIO UV-Visible spectrophotometer (Varian, Inc.) equipped with a thermoelectrically controlled cell holder. Absorbances were obtained as a function of temperature at 260 nm or 280 nm with a heating rate of 1°C/min. Absorbance readings were taken every 0.5°C over the range 15–90°C. At least five melting transitions were obtained for each sample.

The transitions of RNA duplexes melted in presence of 1 M NaCl were reversible for two temperature cycles and no significant changes in T_M were observed. However, transitions in the presence of Mg²⁺ were not reversible since Mg²⁺ catalyzes the hydrolysis of RNA backbone (21).

Data analysis of UV denaturation curve. Thermodynamic parameters (ΔG_{37}° , ΔH° , ΔS°) for duplex formation of RNA S/R motif constructs were calculated from the melting profiles by curve fitting using the MeltWin (version 3.0) software. This program fits the shape of each curve to the two-state model with sloping base lines given by Equation (1) using a nonlinear least-squares procedure (22).

$$A_T = (1 - f) \times A_{SS} + f \times A_{DS} \quad (1)$$

In Equation (1): A_T is the total absorbance, A_{SS} is the absorbance of RNA single strands, A_{DS} is the absorbance of the RNA double strand complex and f is the fraction of total RNA in the duplex form. Assuming the transitions from the duplex to single strands conform to the two-state model, then, the fraction of RNA duplexes can be related to equilibrium dissociation constant K_D according to Equation (2):

$$K_D = C_T \times (1 - f)^2 / (2 \times f) \quad (2)$$

where C_T is the total RNA concentration. This allows for the extraction of the thermodynamic parameters assuming linear sloping baselines and temperature independent ΔH° and ΔS° values (22):

$$K = \exp[-\Delta H^\circ/R + \Delta S^\circ/RT] \quad (3)$$

Correlation coefficients (R^2) of ≥ 0.99 were observed for the fits to the melting curves using Equation (1). Reported ΔH° and ΔS° values and their standard deviations were based on at least five separate experiments for each sample.

To confirm the validity of the two-state model for the duplexes we measured T_M as a function of C_T for the U2G/C12A variant at four concentrations ranging from 5

to 20 μ M. We constructed a Van't Hoff plot to calculate thermodynamic parameters of non-self complementary duplexes according to the method of Borer *et al.* (23):

$$1/T_m = (R/\Delta H^\circ) \times \ln(C_T/4) + \Delta S^\circ/\Delta H^\circ \quad (4)$$

where R is the gas constant, 1.987 cal mol⁻¹ K⁻¹; T_m is melting temperature of a duplex.

All transitions confirm the two-state model, ΔH° values from the two methods generally agree within 10% error (24,25).

The Gibbs free energy change at 37°C was obtained using the following equation:

$$\Delta G_{37}^\circ = \Delta H^\circ - (310.15K) \Delta S^\circ \quad (5)$$

RNA structure probing

Limited enzymatic digestion of labeled hairpins. Prior to digestions, samples of radiolabeled hairpins 1 μ l (1 μ l, \sim 2 pmol = 100 000 cpm) containing 1 μ l yeast tRNA (10 mg/ml) were heated to 94°C in 5 μ l of dd H₂O for 2 min (denaturing step) and snap-cooled on ice. To this we added 1 μ l of 10 \times Tris-HCl buffer (1 M Tris pH 8.2, 2 M KCl, 10 mM EDTA) and the resulting mixtures were incubated at 30°C for 10 min (annealing step). Finally, 1 μ l of 50 mM MgCl₂ was added and the mixture incubation continued for 30 min (assembling step). RNA samples were digested using 1 μ l of T1 (1 U/ μ l), or A (0.011 ng/ μ l) or V1 (0.0005 U/ μ l) for 1 min at 30°C. The digestion reactions were stopped by adding 10 μ l of colorless gel-loading solution (10 M urea, 1.5 mM EDTA). The products of digestions were analyzed on 8 M urea-15% sequencing PAGE.

Diethylpyrocarbonate modification and cleavage of RNA hairpin bases. RNA chemical modification by a diethylpyrocarbonate (DEPC) was performed under native conditions. Briefly, labeled RNA hairpins (1 μ l, \sim 100 000 cpm) were annealed in SC buffer (pH 6.9) in presence of 10 μ g yeast tRNA. Next, 15 μ l of DEPC (SigmaAldrich) was added to each RNA sample. Samples were incubated for 30 min at 20°C with mixing every 10 min. Treated RNA strands were precipitated with 2.5 volume of cold ethanol and dried RNA pellets were resuspended in 20 μ l aniline-acetate buffer (1 M Aniline, 13.9 N glacial acetic acid, pH 4.5). The cleavage reaction was performed by incubating the mixture for 15 min at 60°C in the dark. Samples were frozen on dry ice and then lyophilized in the SpeedVac. Pellets were resuspended in 50 μ l of dd H₂O, frozen, and dried again on the SpeedVac to remove organics. Finally, cleaved RNA products were resuspended in 6 μ l of denaturing loading dye solution (8 M urea, Tris-HCl pH 7.4, EDTA 1 mM, 0.1% (w/v) bromophenol blue) and analyzed on 8 M urea-15% sequencing PAGE.

RESULTS AND DISCUSSION

Searching for sarcin-ricin motifs in the 3D structure database

Like many other 'loops' the nucleotides of S/R motif are represented as single-stranded in 2D structure. The standard S/R motif is an asymmetrical 5 \times 4 internal loop (Figure 1, left), although some versions comprise two more

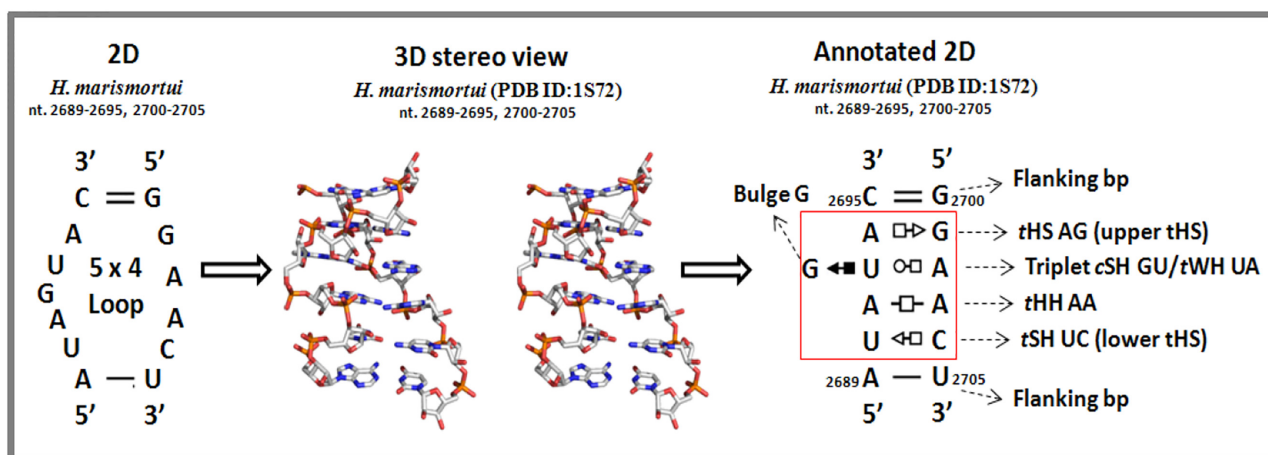


Figure 1. Different representations of the S/R motif from helix-95 of *H. marismortui* 23S rRNA. Left: 2D structure of 5×4 nt asymmetric internal loop, 3D stereo view from PDB ID 1S72 showing non-canonical base pairs. Right: 2D diagram with annotation of all five non-canonical interactions in structure of S/R motif.

nucleotides (6×5 internal loop). The 3D structure (Figure 1, center) shows that all of bases are, in fact, involved in non-WC base pairing, resulting in 5 bp and one base triple. The essential features of the 3D structure are depicted using symbols from the Leontis-Westhof base pair classification (Figure 1, right) (4).

To examine the sequence variation in the S/R motifs present in atomic resolution 3D structures, we carried out mixed symbolic and geometric searches, using WebFR3D (26). We searched a non-redundant set of annotated RNA 3D structures from the PDB with resolution better than 4 Å. (<http://rna.bgsu.edu/rna3dhub/nrlist>) using geometric discrepancy 0.8 Å/nucleotide as the similarity cutoff (27). The non-redundant set contained ribosomal structures from six different organisms. Most notably, the existing data allowed us to compare sequence variation between two related microbial organisms, the mesophile *Deinococcus radiodurans* and the thermophile *T. thermophilus*, that have adapted to different thermal environments, as well as between two phylogenetically diverged organisms that have adapted to similar thermal ranges, the mesophiles *E. coli* and *D. radiodurans*.

Figure 2 shows 45 instances of S/R and S/R-like motifs that have been identified in 15 distinct PDB files (listed in Table 1). Consistent with previous results (18), five recurrent S/R motifs are found in conserved positions in the large ribosomal subunit (LSU) of each of the six organisms and three in the small ribosomal unit (SSU). Throughout this paper we will refer to the S/R motif located in H95 of 23S rRNA from *Haloarcula marismortui* as the prototype, the sequence of which is indicated by the red rectangle in the upper row of Figure 2A. The H95 motif of bacterial 23S rRNA is sensitive to the toxins sarcin and ricin giving the motif its name (28).

The base pair families and the order in which they occur in S/R motif are conserved, but the sequence varies in different organisms and locations. For example, in almost all eubacterial 23S rRNA the tHH base pair in the H95 motif is AC, while in most Archaeal and Eukaryal LSUs it is AA. In addition, some S/R motifs have one more non-WC base

pair forming 6×5 loop. Overall, however the H95 motif is very conserved in sequence.

Generally, S/R motifs in other locations display more sequence variability, especially, in the tSH base pair, where we observe seven different base combinations in LSU, SSU and 5S rRNA 3D structures AG, AA, UC, CC, AC, GG and UA (Figures 1 and 2). All these sequence variants, with the exception of G-G, form isosteric tSH base pairs (29).

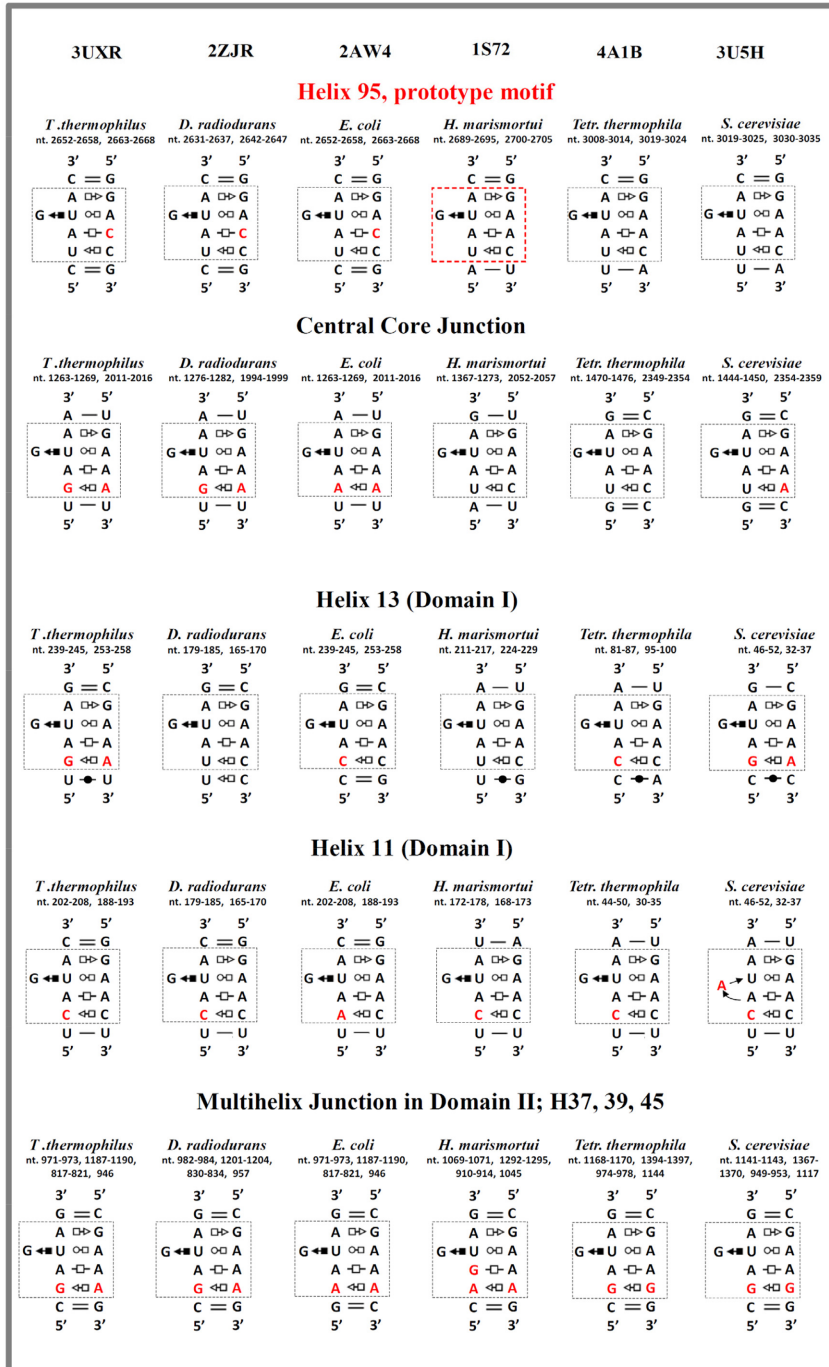
Unusual substitutions are observed in some S/R motifs. For example, in the *S. cerevisiae* version of the LSU H11 S/R motif, there is an unusual bulged A in place of highly conserved G which forms a cSH/tWH base triple with conserved U and A bases of the motif. Also, an unusual GA tHH base combination is observed in the *H. marismortui* version of 23S rRNA S/R motif that is embedded in the complex junction from domain II.

Within 16S rRNA, S/R-like motifs occur in three places: h27; the three-way junction comprising h28, 29 and 43; and h17 (Figure 2D). These motifs differ from the prototypical structure in that the h27 S/R motif contains a conserved bulged C, the 3WJ S/R motif is missing one tSH base pair and the S/R motif from h17 shares only tHS AG and tWH UA base pairs with the prototype. Nevertheless, the overall geometry of these motifs is similar to the prototypical structure, and we include them in the analysis of sequence variations.

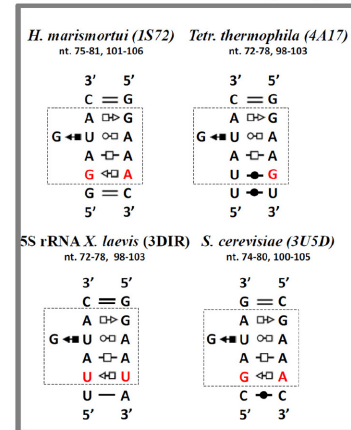
Base pair isosterism in S/R motifs

Two base pairs are said to be isosteric when they meet three criteria (1): (1) The C1'–C1' distances are nearly the same; (2) the paired bases are related by identical rotations in 3D space; and (3) H-bonds are formed between equivalent atoms of the bases. Consequently, isosteric base substitutions in RNA 3D motifs are structurally neutral and do not interrupt the 3D geometry, while non-isosteric mutations distort the structure or prevent it from forming. Table 2 represents the observed sequence variations from 3D structures of S/R motif tabulated in 4×4 Isostericity matrices. Squares representing isosteric base pairs in the same

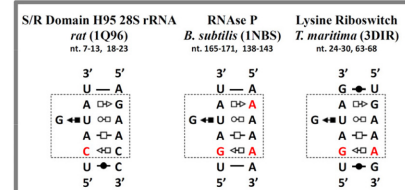
A. S/R motifs from 23S and 28S rRNA



B. Eukaryal and Archaeal 5S rRNA S/R motif (Loop E)



C. S/R motif in other structures



D. 16S/18S rRNA S/R and S/R-like

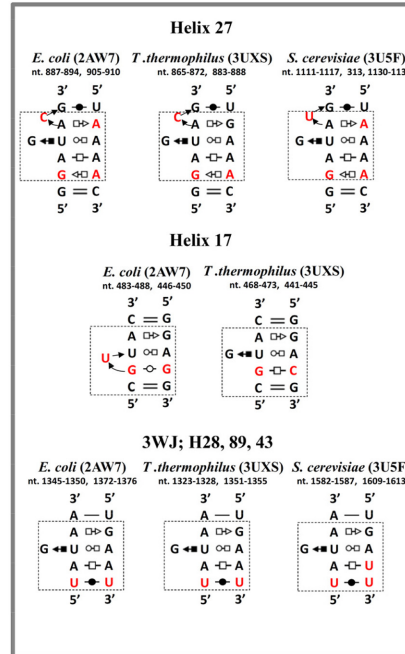


Figure 2. 2D diagrams of S/R motif from different organisms found in 3D rRNA data base. (A) S/R motifs from large ribosomal subunit, the pdb code of the structures is given on the top. Rows represent diagrams of S/R motif corresponding to particular place in the rRNA and name with the nucleotide numbers are given on the top of each diagram. Each column is structure of particular organism. (B) S/R motif diagrams from archaeal and eukaryal 5S rRNA, pdb codes are given in parentheses. (C) S/R motif diagram from other structures. (D) S/R-like motifs from 16S rRNA 3D structures of *E. coli* and *T. thermophilus* and 18S rRNA of *S. cerevisiae*.

Table 1. Representative PDB structure files containing S/R motifs

Organisms	Temperature preference	RNA type	PDB ID	References
Archaeon				
<i>H. marismortui</i>	Mesophile	50S (LSU)	1S72	(49)
Bacteria				
<i>E. coli</i>	Mesophile	50S (LSU) 30S (SSU)	2AW4 2AW7	(50)
<i>T. thermophilus</i>	Thermophile	50S (LSU) 30S (SSU)	3UXR 3UXS	(51)
<i>D. radiodurans</i>	Mesophile	50S (LSU)	2ZJR	(52)
<i>B. subtilis</i>	Mesophile	Ribonuclease P	1NBS	(53)
<i>T. maritima</i>	Thermophile	Lysine riboswitch	3DIR	(54)
Eukarya				
<i>Tetr. thermophila</i>	Thermophile	60S (LSU)	4A1B	(55)
<i>S. cerevisiae</i>	Mesophile	60S (LSU)	3U5H	(56)
<i>X. laevis</i>	Mesophile	5S rRNA	1UN6	(57)
RAT	Mesophile	S/R Domain from 28S rRNA	1Q96	(58)

base pair family are shaded the same color. Different colors indicate different isosteric groups. For example, there are two isosteric groups in the tHS family and almost all sequence variants belong to the same family for both tHS base pairs of the motif. Supplementary Figure S1 compares structural and geometrical parameters of isosteric and near isosteric tHS and tHH base pairs that appear frequently in S/R motifs.

While substitution of isosteric base pairs does not distort the 3D structure of the motif, it may affect the thermodynamic stability or the rigidity/flexibility of the structure. This aspect of isosteric substitutions in conserved motifs has not been addressed adequately in previous work.

Guided by the structural analysis, we suspected that some of the observed sequence variation in non-WC base pairs in the S/R motifs may be due to thermal adaptation. For example, Figure 2 demonstrates that in both thermophilic organisms (*T. thermophilus*, and *T. maritima*) the upper tHS base pair is always AG, whereas in mesophilic groups AA also occurs (e.g. *E. coli* h27 16s rRNA in Figure 2D and *B. subtilis* RNase P in Figure 2C). The lower tSH base pair accommodates multiple sequence combinations; however, it is not known to what extent, if any, these variations affect the stability of the S/R motif. For example, there are two N-H \cdots N bonds in the tSH GA base pair but only one amine-carbonyl N-H \cdots O=C H-bond in the isosteric tSH UC pair. Does this mean that motifs in which GA substitutes for UC at this position are significantly more stable? Are there correlations between the organism's optimal growth temperature and the sequence variations? The present study aims to address these questions.

Thermodynamic study of RNA duplexes containing sarcin-ricin motifs

Having analyzed sequence variability in the S/R motifs found in the 3D structural database, we designed RNA duplexes (Figure 3) that systematically incorporate the key sequence variations identified by these comparisons. Next, we describe the design of these duplexes and the results of UV-melting experiments that were carried out to quantify their stabilities.

Design of sequence variants and nomenclature. For consistency, we numbered each nucleotide in the S/R motif sequences from the 5' to 3' direction starting with the leftmost strand and including the flanking base pair, as shown in Figure 3 (the red rectangle). Numbered nucleotides are enclosed in red rectangles and nucleotides that differ from prototype in Figure 3 are colored in red in all figures.

Each duplex has its unique number 1 through 16 and is named and labeled to reflect the following:

- Nucleotide substitution(s) relative to the prototype duplex. For example, the '#2_U2C' duplex U at position 2 changed to C. Each nucleotide substitution is colored in red (Figure 3).
- Motif location in rRNA, indicated by the helix number and organisms name, e.g. 23S-H13 *E. coli*.
- Optimal growth temperature class (psychrophilic, mesophilic, thermophilic, hyperthermophilic).

We searched for S/R motif instances in the non-redundant RNA 3D structural database using WebFR3D (26) and identified 14 unique S/R motif sequences (Figure 2). These 4 \times 5 internal loops were inserted in RNA duplexes flanked by two canonical base pairs. To prevent sequence end fraying of the flanking helices prior to melting of the motif we terminated each helix with GC Watson-Crick pairs. The sequences were checked with Mfold (www.mfold.bioinfo.rpi.edu, Zuker 2003) to avoid the formation of self-complementary secondary structures, internal hairpin structures or misaligned duplexes. We used duplexes for melting experiments rather than hairpin constructs to avoid specific interactions of hairpins with divalent metal ions (30,31). Moreover, base substitutions are easier and cheaper to design in duplex RNAs and data analysis in duplexes is more robust.

We also added two control molecules that were not observed in 3D structures. These controls contain base substitutions intended to prevent the formation of the S/R motif by disrupting one of the two tHS base pairs. For example, in the #4_U2G molecule U2, which forms a tSH base pair with C12, is substituted by G. The resulting GC juxtaposition is not expected to form a tSH pair since GC is never observed to form this type of base pair in any 3D structure. On the

Table 2. RNA base pair variations observed in 3D structures of S/R motifs in nonredundant data set^a

Upper tHS	A	C	G	U	Lower tHS	A	C	G	U
A	2		36		A	3		13	1
C					C	1	8		10
G					G			2	
U					U				

tWH	A	C	G	U	cHS	A	C	G	U
A					A				
C					C				
G					G				
U	42				U			40	

tHH	A	C	G	U	cWW	A	C	G	U
A	35	3		1	A				12
C			1		C	1	2	22	
G	1				G		20		3
U					U	5	1	1	15

^aWithin each family, the base combinations that form isosteric base pairs are shaded the same color. For example in tWH base combinations UA and CC are isosteric and shaded light blue. In addition, each color represents specific isosteric group as follows:

Blue color indicate i1 group; Green boxes—i2 group; Violet, Orange and Red boxes indicate i3, i4 and i5 groups respectively. Purple boxes indicate i6 group of isosteric pairs. Yellow colored shades indicate combined i1/i2 isosteric pairs within the family. Grey colored cells are modeled interactions, not yet observed in high resolution X-ray (59). Unshaded cell in each table indicate base combinations that do not form that type of base pair.

other hand, the GC combination can form a canonical and stable Watson–Crick base pair, which is expected to disrupt the formation of the S/R motif. Likewise, the A6C substitution in the #16_U2G/A6C/C12A is expected to disrupt the formation of the other tHS base pair and thus the whole motif by producing a CG between positions 6 and 9 of the motif.

All 16 designed RNA duplexes, fourteen of which are observed in 3D structures and two of which are unobserved controls are shown in Figure 3 with differences from the prototype indicated by red rectangles.

Selecting S/R motif flanking base pairs

The flanking Watson–Crick base pairs of S/R motifs can vary. In addition to the canonical cWW GC and AU flanking pairs, in 3D structures cWW GU, CC, UC, UU and CA base pairs also occur (Figure 2 and summarized in Table 2). It is not possible to perform thermodynamic studies on all variations of flanking pairs. Thus, we designed a controlled study varying one flanking cWW pair of the motif (molecule #10_U2G/C12A). We kept the G7C8 closing base pair constant and tested four different variants of the other closing

base pair (C1G13, G1C13, U1G13 and U1A13). The results of these UV-melting studies are shown in Table 3. Except for the wobble U1G13 flanking pair all variants showed distinct transition curves. For duplex design we chose the C1G13 cWW base pair for two reasons. First, this variant has a high melting temperature of 37.3°C, and secondly, it is easier to purify this molecule on TLC plates because it has fewer guanines on the left strand.

UV-melting of duplexes containing S/R motifs

UV-melting experiments of S/R motif RNA duplexes were performed by monitoring of the absorbance change at 260 nm. Some RNA melting curves were also recorded at 280 nm as a check and gave similar results. Melting controls of single strands did not exhibit distinct transitions. Table 4 shows the thermodynamic parameters for duplex formation that were obtained from averaging the independently fitted melting curves to the two-state model, adjusted to 100 μM Ct with the assumption that extinction coefficients for single strands depend linearly on temperature. To confirm two state behavior we generated a Van't Hoff plot of T_M^{-1} versus $\log(Ct/4)$ of the #10_U2G/C12A duplex and obtained

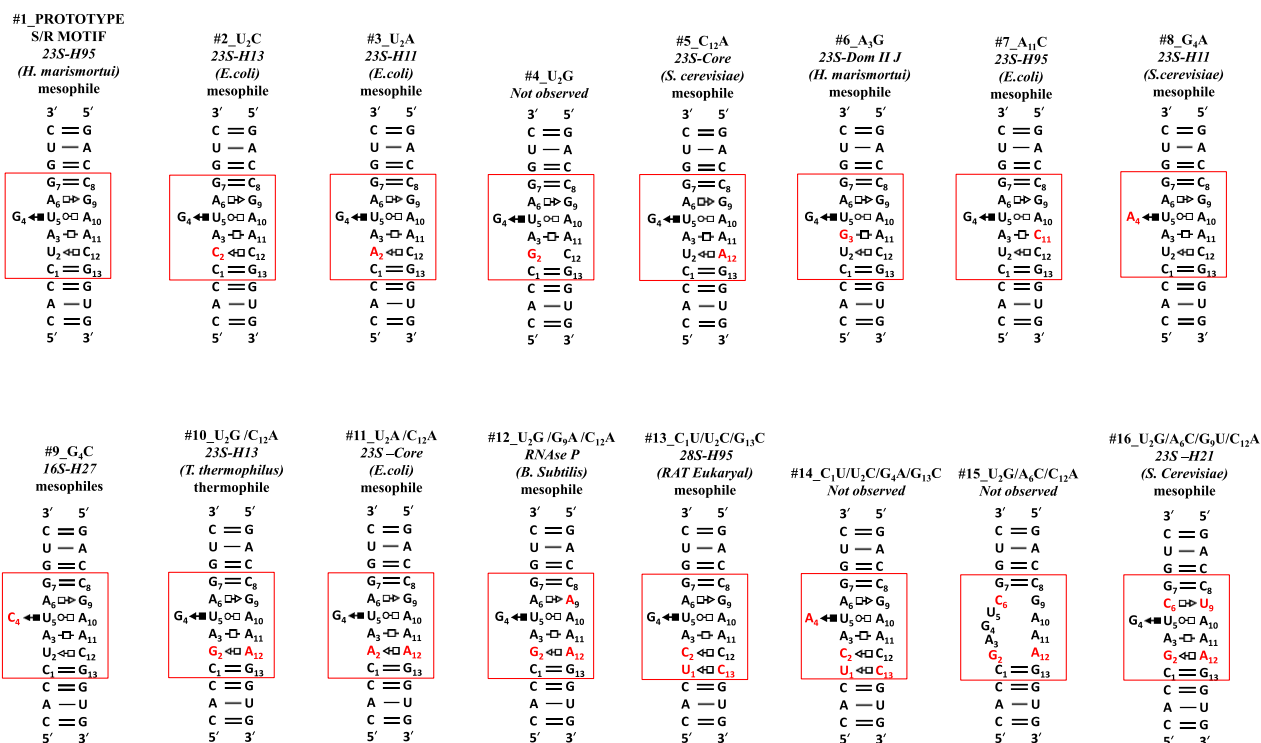


Figure 3. Sequences and schematic structures of the RNA sarcin-ricin motifs used in thermal denaturation studies. Names are given on the top of each duplex as follows: nucleotide mutation in respect to prototype molecule (red colored rectangle), appearance in rRNA with indication of helix number (e.g. 23S-H13), name of bacteria where sequence appears (e.g. *E. coli*), and optimal growth temperature organism adapted (e.g. mesophile). The symbols of base pair interactions are given according to N.B. Leontis and E. Westhof nomenclature

Table 3. Effect of flanking base pairs on stability of S/R motif

RNA duplexes with different flanking base pairs	G = C		G = C		G = C		G = C		G = C	
	$G_7 = C_8$	$G_7 = C_8$	$G_7 = C_8$	$G_7 = C_8$	$G_7 = C_8$	$G_7 = C_8$	$G_7 = C_8$	$G_7 = C_8$	$G_7 = C_8$	$G_7 = C_8$
	$A_6 \triangleright G_9$	$A_6 \triangleright G_9$	$A_6 \triangleright G_9$	$A_6 \triangleright G_9$	$A_6 \triangleright G_9$	$A_6 \triangleright G_9$	$A_6 \triangleright G_9$	$A_6 \triangleright G_9$	$A_6 \triangleright G_9$	$A_6 \triangleright G_9$
	$G_4 \leftarrow U_5 \circ A_{10}$	$G_4 \leftarrow U_5 \circ A_{10}$	$G_4 \leftarrow U_5 \circ A_{10}$	$G_4 \leftarrow U_5 \circ A_{10}$	$G_4 \leftarrow U_5 \circ A_{10}$	$G_4 \leftarrow U_5 \circ A_{10}$	$G_4 \leftarrow U_5 \circ A_{10}$	$G_4 \leftarrow U_5 \circ A_{10}$	$G_4 \leftarrow U_5 \circ A_{10}$	$G_4 \leftarrow U_5 \circ A_{10}$
	$A_3 \square A_{11}$	$A_3 \square A_{11}$	$A_3 \square A_{11}$	$A_3 \square A_{11}$	$A_3 \square A_{11}$	$A_3 \square A_{11}$	$A_3 \square A_{11}$	$A_3 \square A_{11}$	$A_3 \square A_{11}$	$A_3 \square A_{11}$
	$G_2 \triangleleft A_{12}$	$G_2 \triangleleft A_{12}$	$G_2 \triangleleft A_{12}$	$G_2 \triangleleft A_{12}$	$G_2 \triangleleft A_{12}$	$G_2 \triangleleft A_{12}$	$G_2 \triangleleft A_{12}$	$G_2 \triangleleft A_{12}$	$G_2 \triangleleft A_{12}$	$G_2 \triangleleft A_{12}$
	$C_1 = G_{13}$	$C_1 = G_{13}$	$C_1 = G_{13}$	$C_1 = G_{13}$	$U_1 \bullet G_{13}$	$C_1 = G_{13}$	$U_1 \bullet G_{13}$	$C_1 = G_{13}$	$U_1 \bullet G_{13}$	$U_1 - A_{13}$
	$G = C$	$G = C$	$G = C$	$G = C$	$G = C$	$G = C$	$G = C$	$G = C$	$G = C$	$G = C$
	5' 3'	5' 3'	5' 3'	5' 3'	5' 3'	5' 3'	5' 3'	5' 3'	5' 3'	5' 3'
T_M (°C) ^a	37.3	39.9	unable to fit	30.5	30.8					
$-\Delta G_{37}^\circ$ (kcal/mol)	6.6 ± 0.1	7.0 ± 0.4		5.6 ± 0.7	5.6 ± 0.7					
$-\Delta H^\circ$ (kcal/mol)	52.0 ± 2.3	53.1 ± 2.9		43.7 ± 1.4	45.4 ± 3.5					
$-\Delta S^\circ$ (e.u.)	146.5 ± 7.3	148.6 ± 12.8		122.8 ± 18.7	128.4 ± 15.2					

^aRNA duplexes were melted in 10 mM Sodium Cacodylate buffer pH = 6.9, 0.5 mM EDTA, 100 mM NaCl, 10mM MgCl₂. T_m values calculated at 100 μM total RNA concentration.

Table 4. Thermodynamic parameters for S/R motif duplex formation at 10 mM MgCl₂, 100 mM NaCl^c and at 1 M NaCl

RNA DUPLEXES	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16
	Prototype SRM 23S-IHS (H. marismortui) mesophile	U ₂ C 23S-H11 (E. coli) mesophile	U ₂ A 23S-H11 (E. coli) mesophile	U ₂ G Notobrevi mesophile	C ₁₂ A 23S-Center I (N. Screeni) mesophile	A ₁ G 23S-Down III (H. marismortui) mesophile	A ₁ C 23S-IHS (E. coli) mesophile	G ₄ A 23S-H11 (E. coli) mesophile	G ₄ C 16S-JH2* mesophile	U ₂ G-C ₁₂ A 23S-H11 (E. coli) thermophile	U ₂ A-C ₁₂ A 23S-C10 (E. coli) mesophile	U ₂ G-G ₄ A-C ₁₂ A RNAP P (E. coli) mesophile	C ₁ U ₂ C-G ₄ C (R. IT. Eubacter) mesophile	C ₁ U ₂ C G ₄ A-G ₄ C (E. coli) mesophile	C ₁ U ₂ C G ₄ A-G ₄ C (E. coli) mesophile	U ₂ G-A ₁ C 23S-H11 (E. coli) Notobrevi mesophile
T _M (°C) ^a	72.7 ± 0.1	70.6 ± 0.2	73.3 ± 0.1	69.7 ± 0.1	69.9 ± 0.3	66.0 ± 0.2	65.5 ± 0.3	67.1 ± 0.4	64.8 ± 0.1	77.9 ^d ± 0.4	74.2 ± 0.3	58.1 ± 0.4	54.4 ± 0.2	44.5 ± 0.5	69.7 ± 0.2	66.5 ± 0.6
-ΔG ₃₇ ^o (kcal/mol)	16.6 ^c ± 0.2	15.6 ± 0.5	16.5 ± 0.3	11.5 ± 0.1	12.2 ± 0.2	10.9 ± 0.1	12.5 ± 0.1	10.6 ± 0.2	11.5 ± 0.1	15.9 ± 0.1	15.3 ± 0.4	9.1 ± 0.3	10.9 ± 0.1	7.9 ± 0.1	12.4 ± 0.4	12.9 ± 0.4
-ΔS ^o (e.u.)	260.5 ^c ± 6.4	247.7 ± 16.5	253.8 ± 7.9	130.4 ± 4.3	151.1 ± 6.3	131.9 ± 1.2	187.4 ± 5.1	114.2 ± 8.5	158.3 ± 3.4	208.9 ± 3.7	187.0 ± 13.7	100.5 ± 17.3	234.1 ± 0.9	155.5 ± 4.3	150.5 ± 16.5	194.7 ± 4.6
-ΔH ^o (kcal/mol)	97.4 ^c ± 2.2	92.4 ± 5.6	95.2 ± 2.7	51.9 ± 1.4	59.06 ± 2.2	51.87 ± 0.4	70.6 ± 1.7	46.2 ± 2.9	60.6 ± 1.1	80.7 ± 1.3	72.3 ± 4.6	40.3 ± 5.6	83.6 ± 0.2	56.1 ± 1.2	58.8 ± 5.5	73.3 ± 4.9
T _M (°C) ^b	67.3 ± 0.8	62.9 ± 0.1	65.4 ± 1.0	61.7 ± 1.1	67.7 ± 0.4	60.2 ± 0.8	64.9 ± 0.2	64.4 ± 0.9	61.8 ± 0.5	70.3 ± 0.8	62.4 ± 1.1	59.4 ± 0.7	54.6 ± 0.4	42.9 ± 0.5	71.3 ± 0.1	70.9 ± 0.9
-ΔG ₃₇ ^o (kcal/mol)	12.8 ± 0.2	11.8 ± 0.01	13.1 ± 0.4	9.6 ± 0.2	10.4 ± 0.1	9.1 ± 0.1	10.5 ± 0.1	8.9 ± 0.2	8.9 ± 0.1	12.5 ± 0.3	9.6 ± 0.2	9.9 ± 0.2	11.3 ± 0.2	7.6 ± 0.1	13.8 ± 0.2	15.1 ± 1.3
-ΔS ^o (e.u.)	186.2 ± 11.7	180.24 ± 2.9	208.7 ± 3.3	102.7 ± 14.9	104.2 ± 1.1	91.4 ± 3.2	120.3 ± 4.1	64.4 ± 7.1	75.6 ± 2.2	156.9 ± 9.3	98.1 ± 5.7	131.7 ± 4.6	248.9 ± 5.0	158.2 ± 7.7	191.1 ± 4.3	230.4 ± 17.6
-ΔH ^o (kcal/mol)	70.6 ± 4.0	67.6 ± 1	77.8 ± 1.4	41.5 ± 4.5	42.7 ± 0.4	37.5 ± 1.1	47.8 ± 1.3	28.9 ± 2.4	32.4 ± 0.7	61.1 ± 3.2	40.0 ± 1.6	50.8 ± 5.6	88.5 ± 1.7	56.6 ± 2.4	73.1 ± 1.5	86.5 ± 16.0

^aRNA duplexes melted in 10 mM Na⁺ Cacodylate buffer pH 6.9 in presence of 10 mM MgCl₂ and 100 mM NaCl, T_M values calculated at 10⁻⁴ M oligomer concentration (Ct) using average curve fits of at least 4 replicate experiments.
^bRNA duplexes melted in 10 mM Na⁺ Cacodylate buffer pH 6.9 in presence of 1 M NaCl, T_M values calculated at 10⁻⁴ M oligomer concentration (Ct) using average curve fits of at least four replicate experiments.
^cDuplexes which are different only in upper tHS pairs are in yellow shaded cells; lower tSH pairs are in blue colored cells; 'bulges' nt in brown colored cells; tHH pairs are in violet colored cells.
^dThe green shaded cells correspond to differences of upper and lower tHS A-G base pair combination.

agreement with calculated parameters within 10% for the two methods, as shown in Supplementary Figure S2, validating the two state assumption.

Mg²⁺ and Na⁺ buffers. It is standard to perform UV melting experiments of RNA duplexes in the presence of 1 M NaCl, to compensate for the absence of divalent ions. However, 3D motifs like S/R do not fold correctly in the absence of divalent ions. Defined conformation of RNA bases prior to melting is crucial and often depends on the addition of divalent ions. Serra et. al. showed that the stability of RNA structural motifs depends on Mg²⁺ ions, in particular, a melting temperature increase in the presence of 50 mM MgCl₂ has been demonstrated for eukaryotic 5S loop E (32). Moreover, Brownian dynamic simulation studies suggested three potential but delocalized metal-binding sites in the S/R motif (33). This was confirmed by extensive molecular dynamic simulation (34). Therefore, to investigate the role of Mg²⁺ on the S/R motif RNA duplex stability we performed UV-melting studies in two distinct buffer compositions, one with Mg²⁺ and one without:

- 1) 10mM Sodium Cacodylate buffer pH 6.9, 0.5 mM EDTA with 0.1 M NaCl and 10 mM MgCl₂.
- 2) 10mM Sodium Cacodylate buffer pH 6.9a, 0.5 mM EDTA and 1 M NaCl.

The differences in thermodynamic parameters of RNA duplexes from 10 mM Mg²⁺ and in 1M NaCl are summarized in Table 3 (grey shaded cells correspond to experiment with no Mg²⁺ ions). The data suggest that majority of RNA duplexes generally have higher stability in Mg²⁺ containing buffer. The differences in thermodynamic parameters obtained between two buffers are summarized in Table 4.

To find the optimal Mg²⁺ concentration for the S/R motif formation we measured T_M of the #10_U2G/C12A RNA duplex as a function of [Mg²⁺] from 1 to 50 mM. Figure 4 shows that at 10 mM MgCl₂ the T_M curve reaches its plateau and further increase in [Mg²⁺] does not significantly affect the T_M. Thus, we used 10 mM MgCl₂ in all experiments involving magnesium. Typical thermal denaturation curves of five representative RNA duplexes containing S/R motif sequence taken at 10 mM MgCl₂ are shown in Figure 5. The samples containing Mg²⁺ were only melted once due to Mg²⁺ driven RNA cleavage.

Influence of tHS base pair combinations on S/R motif duplex stability

The tHS base pair family frequently occurs in functional RNA motifs including S/R motif, Kink-turn, loop E, multi-way junctions as well as hairpin loops (e.g. GNRA hairpin loop). The most well known member of this family is the

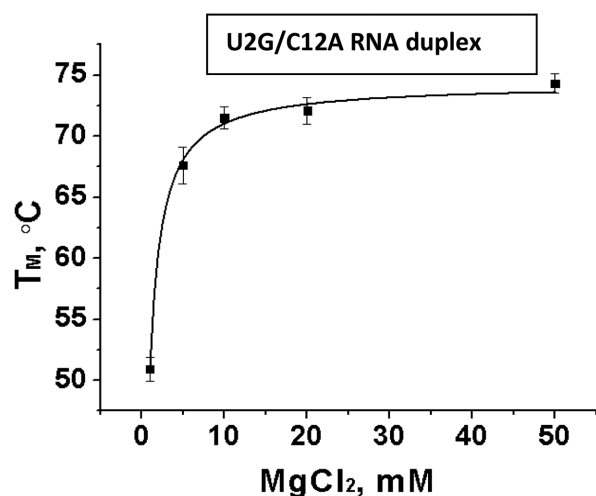


Figure 4. Melting temperature of reference duplex as a function of $[Mg^{2+}]$. At 10 mM $MgCl_2$, the T_M curve reaches its plateau.

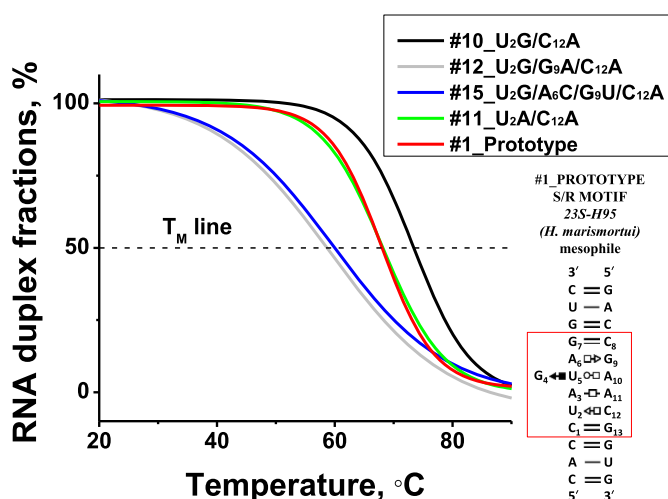


Figure 5. Representative absorbance versus temperature profiles of five RNA duplexes melted in sodium cacodylate buffer pH 6.94, 10 mM $MgCl_2$ and 0.1 M NaCl. Corresponding names are given in legend. 2D structures of these molecules could be found in Figure 3.

tHS A–G pair, also known as sheared AG base pair which often contributes to RNA stability by extensive cross-strand purine stacking.

Sarcin/Ricin motifs contain two conserved tHS base pairs with opposite orientations. We refer to the tHS pair between nts 6 and 9 as the ‘upper tHS’ pair and the pair between 2 and 12 as the ‘lower tSH’ pair.

Upper tHS position. Four variants of the upper tHS pairing position (nt 6 and 9) were studied:

A6–G9, A6–A9, C6–U9 and C6–G9 corresponding to molecules #10, #12, #15 and #16 respectively. In all these duplexes the lower tSH position was G2–A12 (see Table 4, yellow shaded cells). Three of these combinations are isosteric with the prototype and should preserve the conformation of S/R motif. The fourth variant juxtaposes C6 and G9, which cannot form the tHS pair and likely

forms cWW, which is expected to disrupt the motif structure. Comparison of their thermodynamic parameters reveals the following results: the most stable RNA duplex is #10_U2G/C12A, which has A6–G9 with ΔG°_{37} of -15.9 kcal/mol in the presence of Mg^{2+} ions. This tHS base combination is the most common in S/R motifs according to WebFR3D results (Summarized in Table 2). The single mutation in the upper tHS position in molecule G9A generates the isosteric A6–A9 tHS base pair. This substitution dramatically decreased the stability of the RNA duplex by ~ 6 kcal/mol compared to tHS A6–G9 (see entry for the #10_U2G/G9A/C12A duplex in Table 2). The intrinsic energies for these isosteric tHS AG and tHS AA pairs were compared previously (35) using quantum chemical calculations and it was found that the tHS AG base pair is substantially more stable than tHS AA consistent with our observations. Molecule #15_U2G/A6C/G9U/C12A with isosteric C6–U9 has a ΔG°_{37} value of -12.9 kcal/mol which is also less stable than A6–G9, but more stable than A6–A9.

The stability of the control #16_U2G/A6C/C12A S/R motif variant which juxtaposes C6–G9 pair has a ΔG°_{37} value of -12.4 kcal/mol, which is similar to the #15_U2G/A6C/G9U/C12A duplex containing tHS C6–U9. However, the melting experiment performed in 1 M NaCl resulted in lower ΔG°_{37} of -13.8 kcal/mol for this juxtaposed C6–G9 control pair (Table 4, lower panel yellow shaded cells).

Interestingly, presence of 1 M NaCl also stabilizes the formation of RNA duplexes containing tHS C6–U9 and A6–A9 by 2.2 kcal/mol and 0.8 kcal/mol respectively (Table 5). The presence of Mg^{2+} ions favors only the tHS A6–G9 pair by $\Delta \Delta G^\circ = -3.4$ kcal/mol. To summarize, the isosteric upper tHS base pair combinations can be ordered by stability as follows:

upper tHS with Mg^{2+} : A6–G9 > C6–U9 > A6–A9
 upper tHS no Mg^{2+} : C6–U9 > A6–G9 > A6–A9

Lower tSH position. The lower tSH base pair position exhibits much more sequence variability than the upper tHS (see Table 2). We studied seven RNA duplexes (#1, #2, #3, #4, #5, #10 and #11) that are distinct in only lower tSH base pair combinations. They are tSH U2–C12 (prototype), C2–C12, C2–A12, C2–G12, A2–U12, A2–G12 and A2–A12. The calculated ΔG°_{37} parameters for these duplexes are shown in the Table 4 and are indicated as blue colored cells for clarity.

Data shows that many of these duplexes have relatively similar ΔG°_{37} of ~ 15.5 kcal/mol and the order of stability follows:

tSH U2–C12 ($\Delta G^\circ_{37} = -16.6$ kcal/mol) > tSH A2–C12 ($\Delta G^\circ_{37} = -16.5$ kcal/mol) > tSH G2–A12 ($\Delta G^\circ_{37} = -15.9$ kcal/mol) > tSH C2–C12 ($\Delta G^\circ_{37} = -15.6$ kcal/mol) > tSH A2–A12 ($\Delta G^\circ_{37} = -15.3$ kcal/mol) >> tSH U2–A12 ($\Delta G^\circ_{37} = -12.2$ kcal/mol).

All lower tSH pairs are isosteric (see above). And only tSH U2–A12 is ambiguous. The A at position C12 could form a Watson–Crick base pair with U2, which is expected to disrupt the motif. However, UA is also an allowed base combination for tSH, and so it is possible that the tSH pairing geometry is conserved. This sequence variant tests the

Table 5. Effect of Mg²⁺ ions in stabilization of S/R motif duplex^a

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16
RNA DUPLEXES	Prototype SRM 23S-1193 (<i>H. marismortui</i>) mesophile	U ₂ C 23S-1113 (<i>E. coli</i>) mesophile	U ₂ A 23S-1111 (<i>E. coli</i>) mesophile	U ₂ G Notobrevia	C ₁₂ A 23S-Core-1 (<i>S. servus</i>) mesophile	23S-Dom III (<i>H. marismortui</i>) mesophile	A ₁₁ C 23S-1195 (<i>E. coli</i>) mesophile	G ₄ A 23S-1111 (<i>C. crescentii</i>) mesophile	G ₄ C 16S-1127 (<i>E. thermophilus</i>) mesophile	U ₂ G/C ₁₂ A 23S-1113 (<i>E. thermophilus</i>) thermophile	U ₂ A/C ₁₂ A 23S-Core (<i>E. coli</i>) mesophile	U ₂ G/G ₉ A/C ₁₂ A RNAP P (<i>E. coli</i>) mesophile	C ₁₂ U ₂ C/G ₉ A/C 23S-1193 (<i>R.112</i>) mesophile	C ₁₂ U ₂ G/C G ₉ A/C 23S-1111 (<i>C. crescentii</i>) mesophile	U ₂ G/A ₁₁ C C ₁₂ A Notobrevia	U ₂ G/A ₁₁ C G ₉ A/C 23S-1127 (<i>C. crescentii</i>) mesophile
ΔΔT _M (°C)	5.4	7.70	7.9	8.0	2.2	5.8	0.6	2.7	3.0	7.6	11.8	-1.3	-0.2	1.6	-1.6	-4.4
ΔΔG ₃₇ ^o (kcal/mol)	-3.8	-3.8	-3.4	-1.9	-1.8	-1.8	-2.02	-1.7	-2.6	-3.4	-5.7	0.8	0.4	-0.3	1.4	2.2
ΔΔS ^o (e.u.)	-74.3	-67.5	-45.1	-27.7	-46.9	-40.5	-67.1	-49.8	-82.7	-52	-88.9	31.2	14.8	2.7	40.6	35.7
ΔΔH ^o (kcal/mol)	-26.8	-24.8	-17.4	-10.4	-16.36	-14.4	-22.8	-17.3	-28.2	-19.6	-32.3	10.5	4.9	0.5	14.3	13.2

^a Values obtained by subtracting S/R motif data (see Table 5) melted in presence of MgCl₂ and its absence. For example for prototype S/R motif molecule: ΔΔG₃₇^o = ΔG₃₇^o Mg²⁺ - ΔG₃₇^o 1 M NaCl.

resilience of the motif to ambiguous base substitutions. In fact, this sequence variant occurs in the central junction of yeast 26S rRNA as shown in Figure 2. In the 3D structure (PDB ID: 3U5H) we observe the U and A forming the tSH geometry, so we expect this variant to form the motif.

On the other hand, the control molecule #4. U₂G comprises tSH G₂-C₁₂. The tSH GC combination has never been observed in any 3D structure and thus is not expected to form this combination. UV-melting data shows that this RNA duplex has a ΔG₃₇^o of -11.5 kcal/mol and suggests that this substitution significantly reduces the stability of the S/R motif.

Comparison of the lower tSH isosteric sequence combinations of the S/R motif RNA duplexes melted in the presence of 1 M NaCl shows a slightly different order of stability as compared to the experiment with 10 mM Mg²⁺:

tSH A₂-C₁₂ (ΔG₃₇^o = -13.1 kcal/mol) > tSH U₂-C₁₂ (ΔG₃₇^o = -12.8 kcal/mol) > tSH G₂-A₁₂ (ΔG₃₇^o = -12.5 kcal/mol) > tSH C₂-C₁₂ (ΔG₃₇^o = -11.8 kcal/mol) > tSH U₂-A₁₂ (ΔG₃₇^o = -10.4 kcal/mol) > tSH A₂-A₁₂ (ΔG₃₇^o = -9.6 kcal/mol).

Interestingly, all isosteric combinations at the lower tSH depend on the presence of Mg²⁺, and, generally Mg²⁺ ions stabilize the S/R motif of these duplexes, while we do not observe this effect at the upper tHS position. ‘Moreover, while isosteric base pair substitutions at the lower tSH position do not significantly impact the stability of S/R motif, substitution in the upper tHS may depending upon the sequence. For example, comparison the free energy parameters between RNA duplex #10_U₂G/C₁₂A with upper and lower tHS A-G and #11_U₂A/C₁₂A RNA duplex with upper tHS A₆-G₉ but lower tSH A₂-A₁₂ shows the difference in ΔΔG^o is small, only -0.6 kcal/mol (15.9–15.3). However, comparison between molecules #10_U₂G/C₁₂A and #12_U₂G/G₉A/C₁₂A with upper tHS A₆-A₉ and lower tSH G₂-A₁₂ shows a much larger difference in ΔΔG^o of -6.8 kcal/mol (see Table 4, their T_M are in green colored cells)

So why do we observe such big changes in stabilities of RNA duplexes mutated in the upper tHS and not in the lower tSH? The stability of S/R motif must be defined

by the context of tHS base pair family within the structure. ‘For example, while lower tSH base pair stacked between tHH and cWW pairs, the upper tHS stacked between cWW on one side and tWH pair of the triple on the other. These different nearest neighbor nucleotides of the upper and lower tHS pair, in fact, could explain the overall duplex stability.’ In addition, the stability of S/R motif structure depends not only on the stacking interactions but also on base-phosphate (BPh) interactions (36,37). 3D structures of S/R motifs with different upper tHS base pair combinations revealed formation of different BPh hydrogen bonding as demonstrated in Figure 6. This figure shows that in the prototype S/R motif, nucleotide U₂693, a part of the triplet, interacts with nucleotide G₂701 forming 5BPh interaction. The same hydrogen bond between corresponding nucleotides U₈91 and A₉06 in S/R motif containing upper tHS AA belongs to 6BPh family. This suggests that BPh interactions might be crucial in stabilizing the S/R motif. In fact, the importance of BPh interactions was pointed out by Zirbel *et al.* (37), and revealed their evolutionary roles in ribosomal RNA. It was shown that ~12% of nucleotides in large RNAs are involved in direct BPh interactions, and these interactions are the most conserved during evolution. This may explain why we see such large differences in stability when the upper tHS varies. Surprisingly, while there is no BPh interaction between nucleotides U₂69 and U₂94 when the upper tHS is C-U, as shown in Figure 6, these duplexes are more stable than tHS A-A.

Bulged nucleotides stabilize the S/R motif duplex by high entropy contribution

We studied three S/R duplexes (#1, #8 and #9) that have differences only in the bulged nt G₄, A₄ and C₄ (Table 4 in brown shaded cells). The prototype RNA duplex with highly conserved bulge G₄ revealed the lowest ΔG₃₇^o value of -16.6 kcal/mol, while ΔG₃₇^o for the S/R motifs with bulges A₄ and C₄ were found to be -10.6 kcal/mol and -11.5 kcal/mol respectively. The free energy difference of the bulged G₄ compared to A₄ and C₄ is ~-5.5 kcal suggesting that the prototype duplex much more stable than A₄

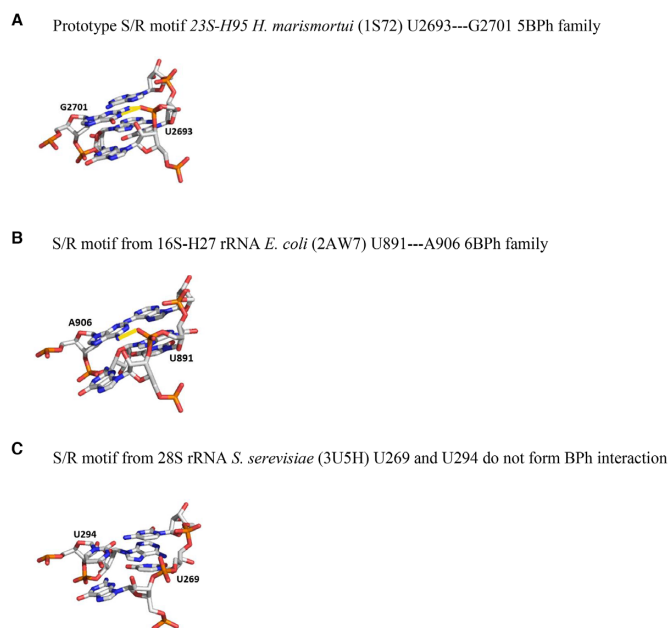


Figure 6. Examples of BPh interactions between upper tHS and triplet U. When the upper tHS is AG there is a hydrogen bonding between oxygen of phosphate group of U and aminogroup of G which belongs to 5BPhfamily (A). When the upper tHS is AA then interaction between the phosphate group of U and aminogroup of A belongs to 6BPh family (B). There is no BPh interaction when the upper tHS has CU base pair combination (C).

and C4 (duplexes G4A and G4C). However, mutations in bulged nucleotides result in significant differences in calculated entropy values. The calculated entropy values for these duplexes are shown in Table 4 in brown colored cells for clarity. The prototype motif with the bulged G4 has a ΔS° of -260.5 e.u. while substitution to bulged A4 or C4 resulted in significantly higher ΔS° values of -114.2 e.u. and -158.3 e.u. Even larger entropy change was observed in presence of 1 M NaCl for A4 and C4 (see Table 4). Evidently, the high entropy values for S/R motifs containing bulged a A4 or C4 contributes to the overall stability of the entire duplex. Why there is a big difference in entropy parameters of A4 and C4 compare to prototype G4? The survey of 3D structure demonstrates that A and C bulged nucleotides in the S/R motif context are very rare (Figure 2). As shown in the crystal structure of 23S rRNA in H-11 of *S. cerevisiae*, the bulged A does not participate in the triple interaction and the nucleotide is bulged out from the motif entirely suggesting the flexibility of entire S/R motif. The same pattern is observed for the bulged U (not used in our studies) at equivalent position from S/R motif of 16S rRNA of H17 (Figure 7). Moreover, the previously reported NMR data also suggest that the replacement of bulged G to A destabilizes the internal loop while overall conformation appears to be similar (38). Our data is consistent with reported observations in the literature and demonstrates that even if the bulged A or C do not participate in triplet interaction, they help to stabilize the formation of S/R motif by a large entropy contribution.

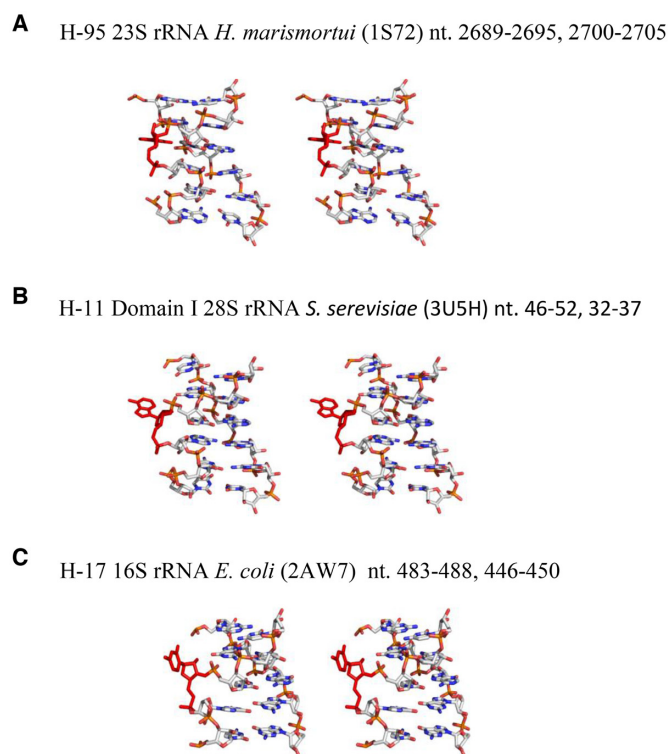


Figure 7. Comparison of structures of S/R motifs varying the bulged base. In most S/R motifs the bulged base is G, as exemplified in the S/R motif from H-95 of 23S rRNA (Panel A). The G is always observed to form a base triplet and base phosphate interaction. When this G is replaced by A (Panel B) or U panel (C), the base is no longer observed to form these interactions.

Contribution of *trans* Hoogsteen/Hoogsteen base pair combinations to stability of S/R motif

Three RNA duplexes (#1, #6 and #7) were studied in which only tHH the base combination differed: tHH A3-A11, tHH G3-A11 and tHH A3-C11 (Table 4, violet shaded cells). The calculated ΔG°_{37} revealed that the tHH A3-A11 base pair combination is more stable than tHH A3-C11 and the least stable is tHH G3-A11. The order of stability is well correlated with the occurrences of tHH base combinations, as tHH AA is the most frequent in existing 3D structures (Figure 2, and summarized in Table 3) and most stable. These data clearly demonstrate that the tHH base pair is important in the S/R motif, as a single mutation of the tHH A3-A11 pair combination to tHH A3-C11 or tHH G3-A11 influences the overall motif stability. In addition, the stability of the S/R motif with tHH G3-A11 pair is entropically driven as its ΔS° value of -131.9 e.u. is significantly higher than that of tHH A3-A11, $\Delta S^\circ = -260.5$, and tHH A3-C11, $\Delta S^\circ = -187.4$ in presence of $MgCl_2$.

Structural probing of S/R motifs

Next, to determine which of the sequence variants studied by thermodynamic methods likely form the S/R motif, we employed established enzymatic and chemical probes, including RNases T1, A and V1 and the chemical probe DEPC (diethylpyrocarbonate).

To directly probe the RNA duplexes used in UV-melting studies is not feasible, so we designed longer hairpin shaped molecule which included a control motif for probing experiments. The sequences of twelve of the sixteen RNA constructs designed for structure probing are shown Figure 8. Each contains a stable GAGA hairpin tetra-loop, which was intended to form a standard GNRA structure. All constructs are 62 nucleotides long. Using the hairpin design we were able to implement an efficient method for S/R motif structure probing: Each probing construct has a constant (control) and a variable region, corresponding to S/R motif sequences from the melting experiment as shown in Figure 8. The S/R motif closest to the 5' and 3' ends of each construct is the control motif (enclosed in the dotted-line rectangle in Figure 8) and corresponds to the sequence of the #10_U2G/C12A S/R variant. It is the same in all molecules and serves as a control for establishing the relative intensities of the cleavage when probing different molecules under different conditions. The motif closest to the hairpin loop is variable and is highlighted in a red rectangle in Figure 8. It corresponds to the different sequence variants inserted in RNA duplexes for UV-melting studies. This design strategy allows us to analyze the probing results for all RNAs simultaneously as the same control S/R motif is presented in each construct.

Structure probing by nucleases

RNA structure probing was carried out with the helix specific RNase V1 and two single-strand specific nucleases T1 and A. RNAs were 5'-end labeled with ^{32}P phosphate. All RNAs were probed in their native conformations in the presence of 10 mM MgCl_2 . The buffer conditions and annealing procedures were identical to the UV-melting experiments described above to ensure that the RNA molecules adopt the same conformation. The cleavage data averaged over several experiments are summarized in Figure 8 and corresponding autoradiogram images are provided in Figure 9.

RNase T1. The cleavage pattern of RNA hairpins treated with RNase T1 is indicated by green arrow in Figure 8 (for cleavage pattern see Figure 9A). These data show that in all constructs, nucleotides G33 and G31 in the hairpin loop are the most sensitive to RNase T1 digestion and likely are most accessible to enzymes. These nucleotides are located in the hairpin loop region and the GAGA sequence could fold into GNRA-type structure. The difference in sensitivity between G33 and G31 to RNase T1 suggests that G33 remains unpaired, while G31 makes some interaction, consistent with 3D structures of known GNRA motifs, which shows that G33 is unpaired while G31 forms a tSH non-WC base pair with A34 (39,40). Thus, the probing data is consistent formation of the GNRA motif as intended in our design.

All other G's in the probed constructs remained unreactive to RNase T1 except for G24 in the molecule that contains the #12_U2G/G9A/C12A sequence variant. Nucleotide G24 corresponds to the bulged G that forms a cSH pair with U25. In this hairpin, the base pair that stacks on

the cSH G24–U25 is tHS A26–A39 instead of tHS A26–G39. The increased sensitivity observed for G24 suggests that tHS A26–A39 pairing makes the S/R motif more flexible and thus lowers its thermostability. This is consistent with the data obtained from UV-melting analysis for the duplex containing the tHS A6–A9 and tHS A6–G9 pairs, demonstrating that the A6–G9 pair favors formation of the S/R motif by $\Delta G_{37}^\circ \sim -6$ kcal/mol compared to the tHS A6–A9 pair (see Table 4).

RNase A. To further characterize the flexibility of S/R motifs we employed RNase A which is specific for unpaired pyrimidine nucleotides. The RNase structure probing data are summarized in Figure 8, using blue arrows to indicate the cleavage sites, and Figure 9B. The cleavage pattern obtained after subsequent treatment of hairpins with the RNase revealed consistent cleavage of U12 across all S/R motifs, suggesting structural similarity in the constant region of all constructs. U12 participates in the base triple and is very conserved among all S/R variants obtained from crystal structures (Figure 2). However, we observed different cleavage patterns in the variable regions of hairpins. The conserved U25 in the variable region was sensitive to RNase A among all hairpins with the exception of the hairpin #16_U2G/A6C/C12A (lane #16_U2G/A6C/C12A in figure 9B). As discussed above, this hairpin sequence was designed to disrupt the S/R formation thus we assume that the U 25 forms a canonical cWW pair with A40, which prevents the cleavage of U25. This indicates that mutation of tHS A26–G39 to the non-isosteric C26–G39 pair distorts the S/R motif conformation and thus the stability is reduced (Table 4).

U's are observed to be sensitive to RNase A when they are present in tHS basepairs (Figure 8). Interestingly, the C's in the tSH C22–C42 and U22–C42 pairs are not sensitive to RNase A, except when there are two adjacent Y–Y tSH pairs (molecules #13_C1U/U2C/G13C and #14_C1U/U2C/G4AC/G13C). The RNase A data provides evidence that the G22–C42 juxtaposition does not affect the formation of the base triple, because U25 shows the same strong cleavage as in the other sequence variants. These results are consistent with those obtained by RNase A treatment of 5S rRNA from *X. laevis*, where loop E also forms an S/R motif (41). The U that participates in the base triple UAG interaction was found to be sensitive to RNase and is very conserved in S/R motifs (42).

RNase V1. Strong RNase V1 cuts are observed in all regions of the RNA constructs that we expected to form Watson–Crick helices (Figure 8 red arrows and Figure 9C). The regions of constructs that contain the S/R and hairpin motifs are resistant to nuclease V1 cuts. Only in the #16_U2G/A6C/C12A construct are the nucleotides U25 and C26 susceptible to cleavage. This is additional evidence that the hairpin containing disrupting mutations has a distinctly different structure in variable sequence region.

Overall, these results are consistent with all 12 designed molecules folding into the intended long hairpin conformation with two internal loops and with U25 forming a non-Watson–Crick base pair in all cases except for the disrupting

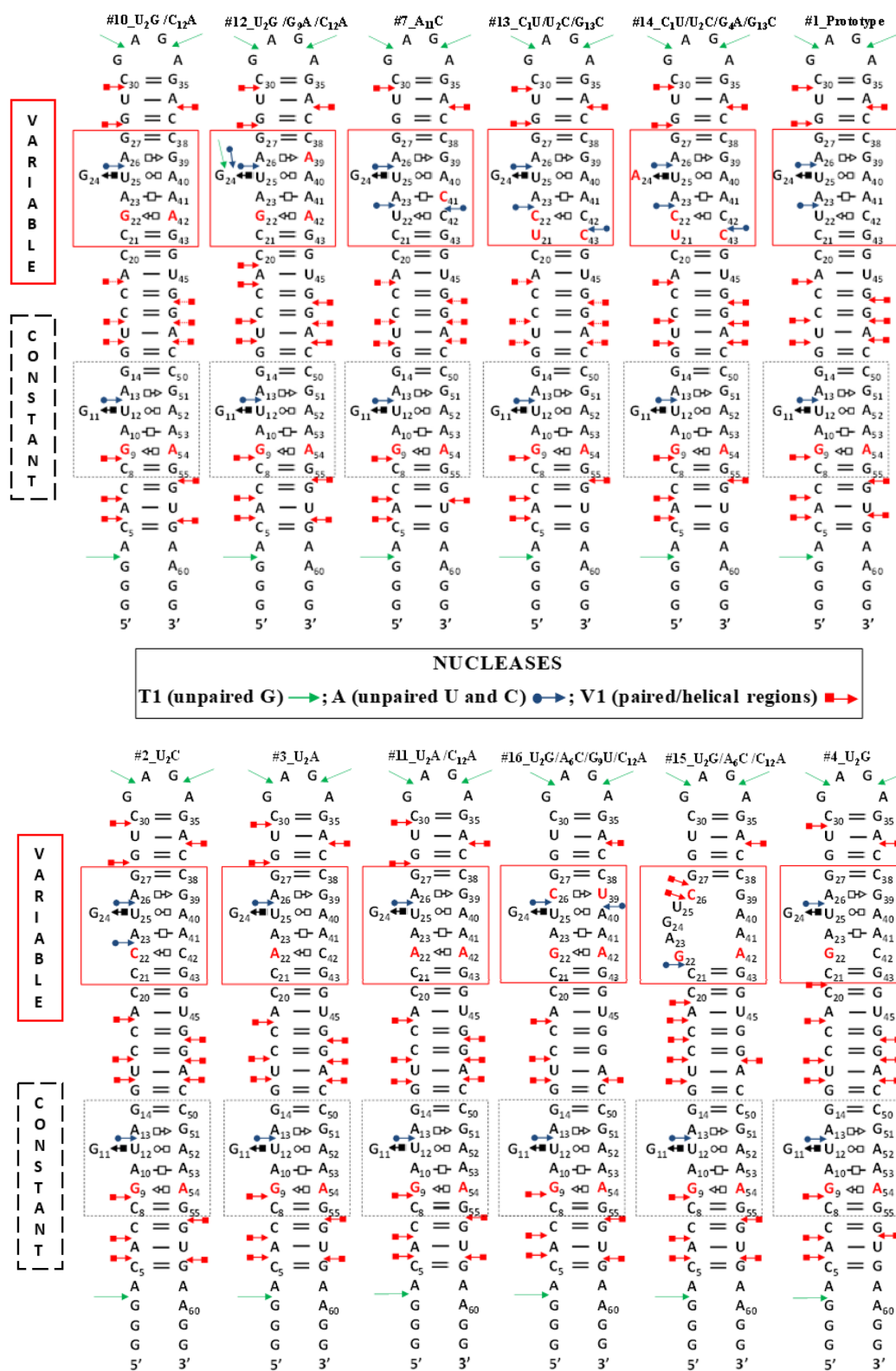


Figure 8. 2D diagrams of RNA hairpins used in structure probing experiments. These molecules consist of the same sequences of GAGA hairpin loop and have two distinct sequences regions: (i) constant region (highlighted as a puncture black rectangle) corresponds to sequence of molecule #10.U2G/C12A which conserved among all molecule and (ii) variable regions (highlighted as a red rectangle) implemented as a signature of individual molecule. The variable region contains sequences of duplexes used in thermal denaturation experiment. Probe RNA molecules are named according to the names of short duplexes inserted in variable region. For instance, prototype S/R motif has this name because the sequence of prototype S/R motif short duplex was implemented to the variable region. The summary of the nuclease digestion experiments is also shown. Cleavage patterns of RNA hairpin secondary structures is represented by symbols corresponding to the legend on center.

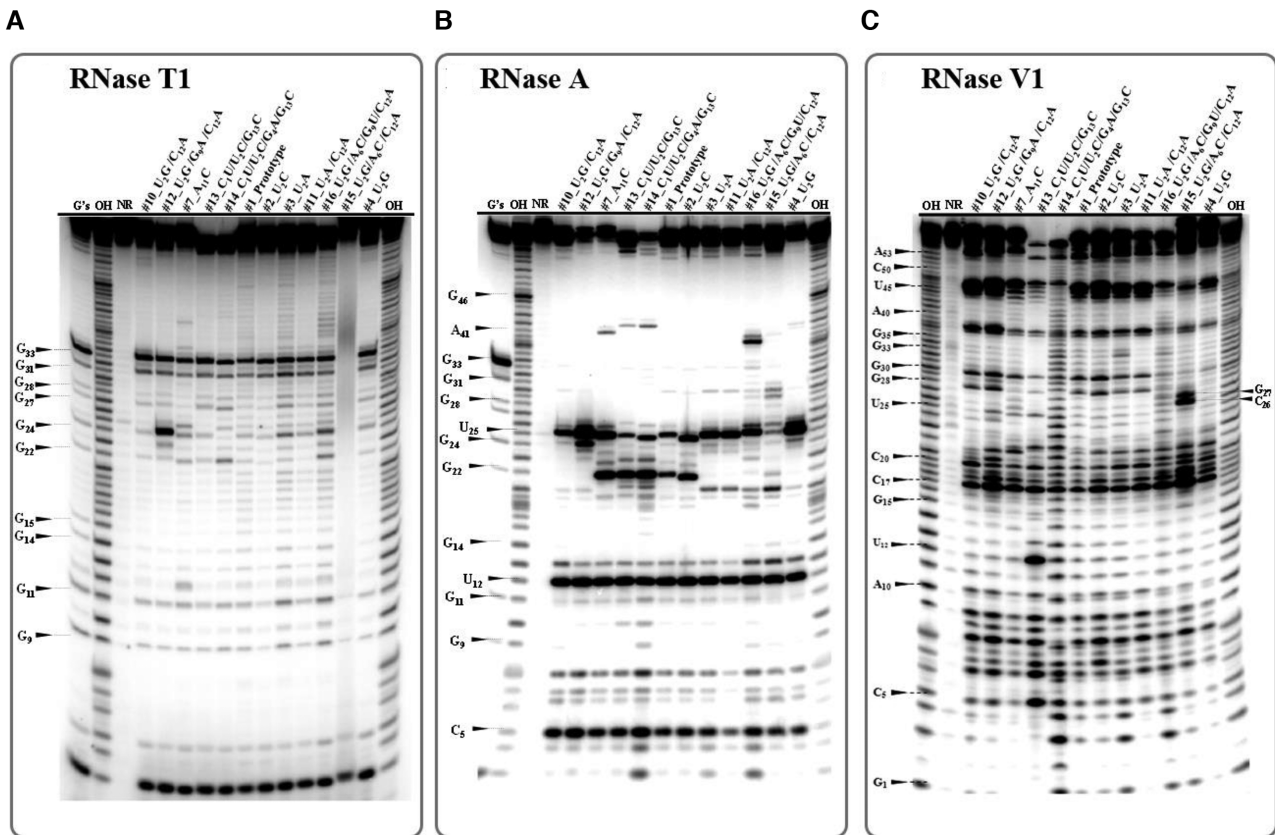


Figure 9. Representative autoradiograms of 15% denaturing gel and cleavage pattern of RNA hairpins by nucleases. The data represents cleavage pattern of 5'-end labeled RNA hairpins treated with specific nucleases shown on the top of each panel: (A) RNase T1, (B) RNase A, and (C) RNase V1. Lanes: 'G's' is RNase T1 sequencing reaction in presence of 8M urea; 'OH' are - base hydrolysis ladders; 'NR' is nuclease free RNA molecules (control); other lanes correspond to the names of individual RNA molecules treated with nucleases.

molecule where it is likely forming a Watson–Crick conformation instead.

Chemical probing

The reactivity of the N7 purine positions of three sequence variants #10_U2G/C12A, #12_U2G/G9A/C12A and #16_U2G/A6C/C12A was probed with DEPC. This probe reacts more strongly with A/N7 than G/N7. The reactivity of the adenine residues was determined by subsequent aniline cleavage of 5'-end labeled RNA hairpins. An autoradiogram image and the results from several such experiments are provided in Figure 10.

Comparison of chemical reactivity data for variable and constant regions of the hairpin constructs indicates that the conformation depends on sequence. DEPC accessibility of A32/N7 and A34/N7 in the hairpin loop is very similar for all studied RNAs. This is also true for Gs at position 11 and 24 located in the core base triple. Moreover, we do not observe a significant difference in cleavage patterns of RNAs between native and semi-denaturing conditions. Under the semi-denaturing condition, where the Mg^{2+} ions were eliminated from reaction buffer, only slightly increased cuts were observed.

The variable region in the variant #16_U2G/A6C/C12A is designed to disrupt the S/R motif by preventing the

upper tHS base pair from forming, because C26-G39 juxtapositions do not form tHS pairs. As expected, the #16_U2G/A6C/C12A hairpin behaves very differently in the variable region (Figure 10B, lanes 6 and 9). In addition to G24 reactivity, nucleotides G39 and A40 become fully reactive with DEPC in native and semi-denaturing buffers indicating that this internal loop does not have a stable conformation (Figure 10B, rightmost molecule). These data are also consistent with the nuclease digestion experiments discussed above.

In conclusion, our structure probing experiments demonstrate that all S/R sequences fold into the 3D structure characteristic of this motif with an exception of one control molecule which is not expected to form the motif. Although exact orientations of all nucleotides cannot be established based on structure probing alone, G24 and U25 are cleaved in a manner consistent with forming the core base triple with A40. The structure probing data validates our observations made during the UV-melting experiments.

Sequence variation in S/R motifs observed in 16S rRNA sequences from Eubacteria with known optimal growth temperatures

Thus far our analysis was based on the sequence variation data derived from the RNA X-ray structures. These

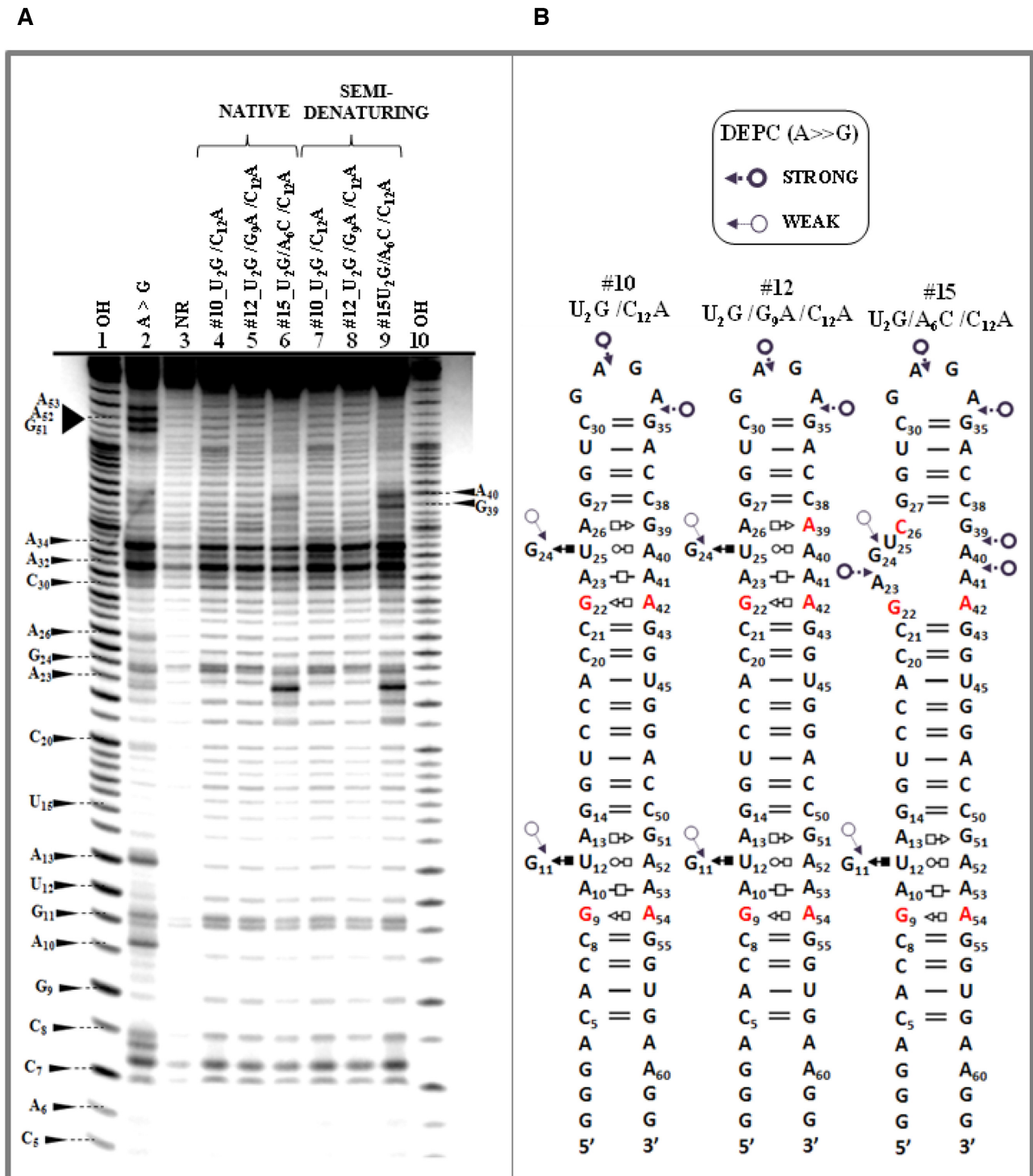


Figure 10. Autoradiogram and summary of chemical probing analysis. (A) Structure probing using DEPC to probe accessibility of N-7 positions of Hoogsteen edges of A and G residues. Lanes: 1 and 10 are 'OH⁻': alkali hydrolysis step ladder of U₂G/C₁₂A hairpin, 2 is 'A > G'- Maxam-Gilbert sequencing reaction, 3 is non-reacted ('NR') U₂G/C₁₂A RNA, treated with aniline as control; NATIVE 4,5 and 6 are RNAs U₂G/C₁₂A, U₂G/G₉A/C₁₂A, and U₂G/A₆C/C₁₂A - samples treated with DEPC under native conditions (80mM NaCB buffer pH 6.94, 100mM KCl, 10 mM MgCl₂, 0.5 mM EDTA); SEMI-DENATURING 7, 8 and 9 are RNAs U₂G/C₁₂A, U₂G/G₉A/C₁₂A and U₂G/A₆C/C₁₂A -samples treated with DEPC under semi-denaturing condition (80mM NaCB buffer pH 6.94, 100 mM KCl, 0.5 mM EDTA). (B) Summary of DEPC sensitivity given by legend on the top of the 2D structures.

structures come from relatively few organisms and provide limited information on sequence variability. To expand our knowledge of S/R motif sequence variation we analyzed ribosomal sequence alignments. We hypothesize that the organisms living at higher temperatures have more rigid RNA 3D motifs while organisms living at lower temperatures have less rigid structures. In order to test this hypothesis, we constructed a special dataset containing 16S alignments with sequences annotated by the optimal growth temperature and phylogenetic group of the source organism.

Finding organisms with known optimal growth temperatures. Bacterial type strains associated with the optimal growth temperature were manually extracted from the literature, with a particular effort devoted to identifying extremophiles. Manual extraction was used because the previous attempt at extracting this data, the PGTdb (43), is outdated and no longer maintained. Species which grow optimally <20°C were classified as psychrophilic; species growing from 20°C to <50°C were classified as mesophilic, species growing between 50°C and less than 70°C were thermophilic while species growing optimally at or >70°C were hyperthermophilic (44). The distribution of organisms in different phylas by their optimal growth temperatures can be found in Supplementary table.

Selecting a source of sequence alignments. There are several alternative sources of aligned rRNA sequences. We examine 16S alignments because the 16S is often sequenced and therefore provides a large dataset to examine. Unfortunately, very few 23S sequences for extremophiles are available, so we did not attempt to analyze the large ribosomal subunit. In order to choose one 16S alignment, we aligned two 16S 3D structures from distantly related bacteria *E. coli* (PDB 2AVY) and *T. thermophilus* (PDB: 1J5E) using R3DAlign (45). Next, we compared the resulting structural alignment with sequence alignments downloaded from Green Genes, RDP and SILVA (46–48). As a result, Green Genes alignment has been chosen as it had the most agreement with the R3DAlign structural alignment.

Selecting a representative sequence alignment dataset. The representative dataset was created by extracting all sequences corresponding to each species and all of its subspecies with known optimal growth temperatures from the Green Genes 16S alignment updated on October 6, 2010 (<http://greengenes.lbl.gov>). Each species and its subspecies were represented with a single sequence selected based on the highest A, C, G and U content. This was done in order to reduce the bias due to over representation of some organisms in the dataset. This dataset contains 561 sequences; within these sequences 82 belong to psychrophilic, 328 mesophilic, 108 thermophilic and 44 hyperthermophilic organisms (Supplementary table). We used Uniprot's taxonomy to obtain the phylogenetic information for each organism.

Correlating structures with sequence alignments. Since some nucleotides in crystal structures are not resolved, the Green Genes alignment and the sequence from the 3D structure may not match. Therefore, we need to determine

the correlation between the columns of the sequence alignments and the structure positions. To this end, we extracted all rows from Green Genes, which came from *E. coli*, and removed all gaps, then aligned each sequence to the sequence from the *E. coli* 3D structure (PDB: 2AVY). The Green Genes sequence with the largest number of aligned columns was used to correlate positions in the Green Genes alignment with the PDB sequence.

AG and AA trans Hoogsteen-Sugar base pairs correlate with organism growth temperature. Using the correlations between sequence alignments and 3D structures we now examine the variations in S/R-like motifs H17, H27 and 3WJ found in the 16S rRNA. We extracted the columns from the alignment which corresponded to each motif. We then grouped each sequence by the phylum of its source organism and optimal growth temperature class and then counted unique sequences. The resulting tables can be found in Supplementary Table S1.

Examining the S/R motif in the H27 of the 16S shows a strong correlation between AG and AA in the upper tHS base pair and growth temperature. As temperature increases the fraction of tHS AG also increases (Figure 11 top panel). This does not appear to be due to phylogeny as the variation is found in Bacteriodes, Firmicutes, and Betaproteobacteria. Other positions remain conserved or nearly conserved. This correlation is noticeable only in one tHS base pair of one S/R motif, but it is supported by our observations based on the UV-melting experiments.

The bottom panel in Figure 11 shows the sequence variations of the S/R motif in H17. The bulged base in the triplet is a G in thermophiles and hyperthermophiles, while in psychrophiles it may be a U. All other examined motifs do not show a strong correlation between sequence and optimal growth temperature.

Overall, we do see some variations which may be due to temperature adaptation. However, our current methods are not precise enough to detect subtle changes. Further statistical analysis on a larger dataset is needed to validate these initial, limited findings.

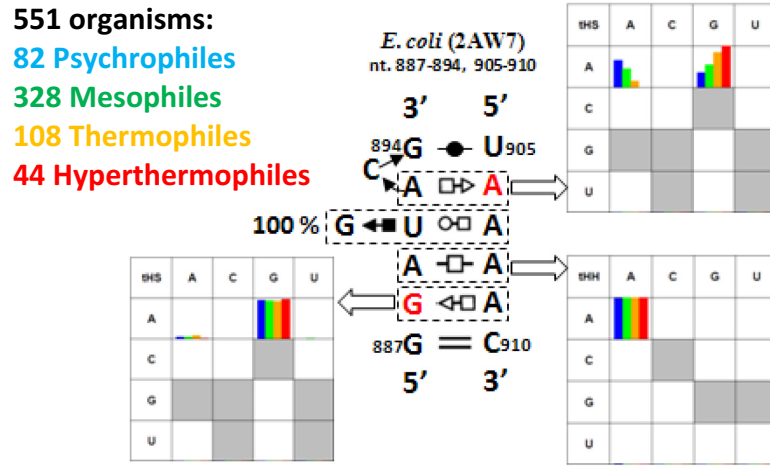
CONCLUSION

In the present work, we systematically study sequence variation at structurally conserved positions of S/R motif in rRNA from different organisms. Herein, our effort is aimed at gaining insight into the thermodynamics of isosteric and non-isosteric substitutions within the S/R motif. Besides the stability investigation, a structural probing analysis is performed to observe which mutations would disturb the motif.

This study demonstrates that isosteric substitutions do not disturb the S/R motif as it was expected from survey of 3D structures and survey of 16S rRNA from organisms adapted to different temperatures.

The most frequent variation was observed at the 'lower' tSH base pair and any neutral substitutions at this place do not significantly impact on the overall stability of the motif. On the other hand, non-isosteric substitutions destabilize the motif but the conformation appears to be modest as reflected from structural probing experiments.

S/R motif in the H27 of the 16S rRNA



S/R motif in the H17 of the 16S rRNA

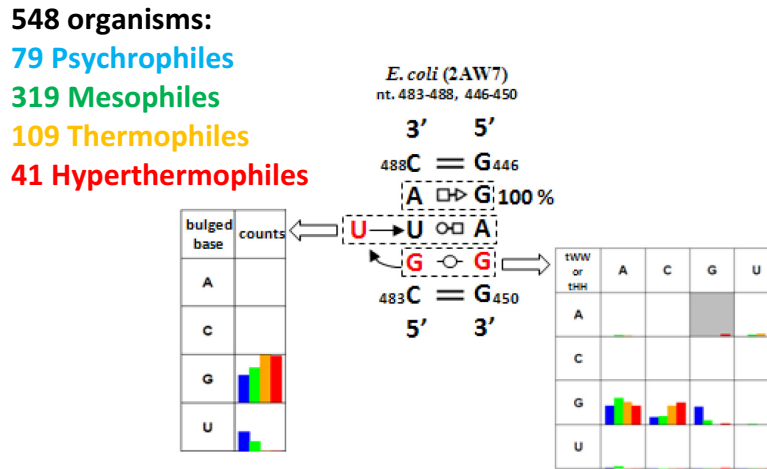


Figure 11. Survey of non-canonical base pairs of S/R motifs in 16S rRNA of Eubacteria. Figure contains 2D diagrams of two S/R motifs one from helix 27 on top and on the bottom is S/R from helix 17. Each bp family has a table with colored bars containing sequence variations of psychrophiles (blue color bar), mesophiles (green color bar), thermophiles (orange color bar) and hyperthermophiles (red color bar). Data in each cell is plotted according to counts of organisms representing sequence of particular non-canonical bp.

Substitutions at the ‘upper’ tHS pair are the most crucial for the stability and conformation of the motif. Non-isosteric mutations destroy the geometry of S/R motif. Isosteric mutations tHS A6–A9 or C6–U9 destabilize the motif but the geometry is preserved. Moreover, tHS A6–A9 makes the motif more flexible as reflected by RNase T1 digestion pattern.

Any mutations at bulged nucleotides participating in triplet interaction significantly impact on the stability of the motif as shown by UV-melting experiments.

Our phylogenetic survey suggests a correlation between non-canonical base pairs and optimal growth temperatures. In high temperature environments tHS AG is the most frequent base pair in S/R motifs, whereas tHS AA is prevalent in cold environments. In addition, the bulged G, which is conserved in most organisms, mutates to U, A or C mostly in psychrophiles and mesophiles. This suggests that organisms may adapt to different temperature environments not only by adjusting the number of GC and AU Watson–Crick pairs in the helical regions as demonstrated previously (5),

but also by selecting optimal isosteric sequence combinations for their non-canonical base pairs.

Using S/R motif as an example, we have demonstrated that non-canonical interactions in structured RNA motifs are very important to overall structure and stability. By changing base pair at certain non-canonical positions we can control stability, structure, and perhaps function of large RNAs.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This manuscript is dedicated to our Ph.D. mentor and friend prof. Neocles B. Leontis for his tireless work toward understanding 3D structure of RNA. We also thank Anton Petrov for critical reading of this manuscript and Aleksandra Vasilevskaya for compiling sequences and optimal growth temperatures of organisms from literature sources.

FUNDING

National Institute of General Medical Sciences [2 R15GM055898-05 to N.B.L.]; National Institute of Biomedical Imaging and Bioengineering [R03EB027910 to E.F.K.] of the National Institutes of Health. Funding for open access charge: Ball State University and Bowling Green State University.

Conflict of interest statement. None declared.

REFERENCES

- Stombaugh, J., Zirbel, C.L., Westhof, E. and Leontis, N.B. (2009) Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.*, **37**, 2294–2312.
- Gutell, R.R., Cannone, J.J., Shang, Z., Du, Y. and Serra, M.J. (2000) A story: unpaired adenosine bases in ribosomal RNAs. *J. Mol. Biol.*, **304**, 335–354.
- Gutell, R.R., Weiser, B., Woese, C.R. and Noller, H.F. (1985) Comparative anatomy of 16-S-like ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, **32**, 155–216.
- Leontis, N.B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
- Galtier, N. and Lobry, J.R. (1997) Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J. Mol. Evol.*, **44**, 632–636.
- Miralles, F. (2010) Compositional properties and thermal adaptation of SRP-RNA in bacteria and archaea. *J. Mol. Evol.*, **70**, 181–189.
- Hurst, L.D. and Merchant, A.R. (2001) High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. *Proc. Biol. Sci. Roy. Soc.*, **268**, 493–497.
- Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. U. S. A.*, **83**, 9373–9377.
- 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.
- Turner, D.H. and Mathews, D.H. (2010) NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.*, **38**, D280–D282.
- Xia, T., SantaLucia, J. Jr, Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C. and Turner, D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry*, **37**, 14719–14735.
- Schroeder, S.J. and Turner, D.H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA. *Biochemistry*, **39**, 9257–9274.
- Lancaster, L., Lambert, N.J., Maklan, E.J., Horan, L.H. and Noller, H.F. (2008) The sarcin-ricin loop of 23S rRNA is essential for assembly of the functional core of the 50S ribosomal subunit. *RNA*, **14**, 1999–2012.
- Shi, X., Khade, P.K., Sanbonmatsu, K.Y. and Joseph, S. (2012) Functional role of the sarcin-ricin loop of the 23S rRNA in the elongation cycle of protein synthesis. *J. Mol. Biol.*, **419**, 125–138.
- Szewczak, A.A. and Moore, P.B. (1995) The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.*, **247**, 81–98.
- Leontis, N.B. and Westhof, E. (1998) A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J. Mol. Biol.*, **283**, 571–583.
- Mladek, A., Spöner, J.E., Kulhanek, P., Lu, X.J., Olson, W.K. and Spöner, J. (2012) Understanding the sequence preference of recurrent RNA building blocks using quantum chemistry: the intrastrand RNA dinucleotide platform. *J. Chem. Theory Comput.*, **8**, 335–347.
- Leontis, N.B., Stombaugh, J. and Westhof, E. (2002) Motif prediction in ribosomal RNAs: Lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie*, **84**, 961–973.
- Cavaluzzi, M.J. and Borer, P.N. (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.*, **32**, e13.
- Borer, P.N., Kan, L.S. and Ts'o, P.O. (1975) Conformation and interaction of short nucleic acid double-stranded helices. I. Proton magnetic resonance studies on the nonexchangeable protons of ribosyl ApApGpCpUpU. *Biochemistry*, **14**, 4847–4863.
- AbouHaidar, M.G. and Ivanov, I.G. (1999) Non-enzymatic RNA hydrolysis promoted by the combined catalytic activity of buffers and magnesium ions. *Z. Naturforsch. C*, **54**, 542–548.
- McDowell, J.A. and Turner, D.H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of (rGAGGUCUC)₂ by two-dimensional NMR and simulated annealing. *Biochemistry*, **35**, 14077–14089.
- Borer, P.N., Dengler, B., Tinoco, I. Jr and Uhlenbeck, O.C. (1974) Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.*, **86**, 843–853.
- Marky, L.A. and Breslauer, K.J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers*, **26**, 1601–1620.
- SantaLucia, J. Jr, Kierzek, R. and Turner, D.H. (1990) Effects of GA mismatches on the structure and thermodynamics of RNA internal loops. *Biochemistry*, **29**, 8813–8819.
- Petrov, A.I., Zirbel, C.L. and Leontis, N.B. (2011) WebFR3D—a server for finding, aligning and analyzing recurrent RNA 3D motifs. *Nucleic Acids Res.*, **39**, W50–W55.
- Sarver, M., Zirbel, C.L., Stombaugh, J., Mokdad, A. and Leontis, N.B. (2008) FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol.*, **56**, 215–252.
- Endo, Y., Chan, Y.L., Lin, A., Tsurugi, K. and Wool, I.G. (1988) The cytotoxins alpha-sarcin and ricin retain their specificity when tested on a synthetic oligoribonucleotide (35-mer) that mimics a region of 28 S ribosomal ribonucleic acid. *J. Biol. Chem.*, **263**, 7917–7920.
- Abu Almakarem, A.S., Petrov, A.I., Stombaugh, J., Zirbel, C.L. and Leontis, N.B. (2012) Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res.*, **40**, 1407–1423.
- Alam, S., Grum-Tokars, V., Krucinska, J., Kundracik, M.L. and Wedekind, J.E. (2005) Conformational heterogeneity at position U37 of an all-RNA hairpin ribozyme with implications for metal binding and the catalytic structure of the S-turn. *Biochemistry*, **44**, 14396–14408.
- Greenbaum, N.L., Mundoma, C. and Peterman, D.R. (2001) Probing of metal-binding domains of RNA hairpin loops by laser-induced lanthanide(III) luminescence. *Biochemistry*, **40**, 1124–1134.
- Serra, M.J., Baird, J.D., Dale, T., Fey, B.L., Retatagos, K. and Westhof, E. (2002) Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *RNA*, **8**, 307–323.

33. Hermann, T. and Westhof, E. (1998) Exploration of metal ion binding sites in RNA folds by Brownian-dynamics simulations. *Structure*, **6**, 1303–1314.
34. Sponer, J.E., Reblova, K., Mokdad, A., Sychrovsky, V., Leszczynski, J. and Sponer, J. (2007) Leading RNA tertiary interactions: structures, energies, and water insertion of A-minor and P-interactions. A quantum chemical view. *J. Phys. Chem. B*, **111**, 9153–9164.
35. Mladek, A., Sharma, P., Mitra, A., Bhattacharyya, D., Sponer, J. and Sponer, J.E. (2009) Trans Hoogsteen/sugar edge base pairing in RNA. Structures, energies, and stabilities from quantum chemical calculations. *J. Phys. Chem. B*, **113**, 1743–1755.
36. Zgarbova, M., Jurecka, P., Banas, P., Otyepka, M., Sponer, J.E., Leontis, N.B., Zirbel, C.L. and Sponer, J. (2011) Noncanonical hydrogen bonding in nucleic acids. Benchmark evaluation of key base-phosphate interactions in folded RNA molecules using quantum-chemical calculations and molecular dynamics simulations. *J. Phys. Chem. A*, **115**, 11277–11292.
37. Zirbel, C.L., Sponer, J.E., Sponer, J., Stombaugh, J. and Leontis, N.B. (2009) Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res.*, **37**, 4898–4918.
38. Seggerson, K. and Moore, P.B. (1998) Structure and stability of variants of the sarcin-ricin loop of 28S rRNA: NMR studies of the prokaryotic SRL and a functional mutant. *RNA*, **4**, 1203–1215.
39. Correll, C.C. and Swinger, K. (2003) Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 Å resolution. *RNA*, **9**, 355–363.
40. Sorin, E.J., Engelhardt, M.A., Herschlag, D. and Pande, V.S. (2002) RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.*, **317**, 493–506.
41. Andersen, J., Delihans, N., Hanas, J.S. and Wu, C.W. (1984) 5S RNA structure and interaction with transcription factor A. I. Ribonuclease probe of the structure of 5S RNA from *Xenopus laevis* oocytes. *Biochemistry*, **23**, 5752–5759.
42. Chan, Y.L. and Wool, I.G. (2008) The integrity of the sarcin/ricin domain of 23 S ribosomal RNA is not required for elongation factor-independent peptide synthesis. *J. Mol. Biol.*, **378**, 12–19.
43. Huang, S.L., Wu, L.C., Liang, H.K., Pan, K.T., Horng, J.T. and Ko, M.T. (2004) PGTdb: a database providing growth temperatures of prokaryotes. *Bioinformatics*, **20**, 276–278.
44. Rothschild, L.J. and Mancinelli, R.L. (2001) Life in extreme environments. *Nature*, **409**, 1092–1101.
45. Rahrig, R.R., Leontis, N.B. and Zirbel, C.L. (2010) R3D Align: global pairwise alignment of RNA 3D structures using local superpositions. *Bioinformatics*, **26**, 2689–2697.
46. Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M. et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.*, **37**, D141–D145.
47. DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G.L. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.*, **72**, 5069–5072.
48. Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J. and Glockner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.*, **35**, 7188–7196.
49. Klein, D.J., Moore, P.B. and Steitz, T.A. (2004) The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.*, **340**, 141–177.
50. Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M. and Cate, J.H. (2005) Structures of the bacterial ribosome at 3.5 Å resolution. *Science*, **310**, 827–834.
51. Bulkeley, D., Johnson, F. and Steitz, T.A. (2012) The antibiotic thermorubin inhibits protein synthesis by binding to inter-subunit bridge B2a of the ribosome. *J. Mol. Biol.*, **416**, 571–578.
52. Harms, J.M., Wilson, D.N., Schluenzen, F., Connell, S.R., Stachelhaus, T., Zaborowska, Z., Spahn, C.M. and Fucini, P. (2008) Translational regulation via L11: molecular switches on the ribosome turned on and off by thiostrepton and micrococcin. *Mol. Cell*, **30**, 26–38.
53. Krasilnikov, A.S., Yang, X., Pan, T. and Mondragon, A. (2003) Crystal structure of the specificity domain of ribonuclease P. *Nature*, **421**, 760–764.
54. Serganov, A., Huang, L. and Patel, D.J. (2008) Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature*, **455**, 1263–1267.
55. Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S. and Ban, N. (2011) Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science*, **334**, 941–948.
56. Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G. and Yusupov, M. (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, **334**, 1524–1529.
57. Lu, D., Searles, M.A. and Klug, A. (2003) Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature*, **426**, 96–100.
58. Correll, C.C., Beneken, J., Plantinga, M.J., Lubbers, M. and Chan, Y.L. (2003) The common and the distinctive features of the bulged-G motif based on a 1.04 Å resolution RNA structure. *Nucleic Acids Res.*, **31**, 6806–6818.
59. Leontis, N.B., Stombaugh, J. and Westhof, E. (2002) The non-Watson–Crick base pairs and their associated isosterity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.