Bowling Green State University

ScholarWorks@BGSU

11-2013

Isosteric And Nonisosteric Base Pairs In RNA Motifs: Molecular Dynamics And Bioinformatics Study Of The Sarcin Ricin Internal Loop

Marek Havrila

Kamila Réblová

Craig L. Zirbel

Neocles B. Leontis
Bowling Green State University, leontis@bgsu.edu

Jiří Šponer

Follow this and additional works at: https://scholarworks.bgsu.edu/chem_pub

Part of the Chemistry Commons

Isosteric and Nonisosteric Base Pairs in RNA Motifs: Molecular Dynamics and Bioinformatics Study of the Sarcin−Ricin Internal Loop

Marek Havrila,[†] Kamila Réblová,[‡] Craig L. Zirbel,[§,∥] Neocles B. Leontis,[§,∥] and Jiří Šponer*[,†,‡]

[†]Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

[‡]CEITEC—Central European Institute of Technology, Masaryk University, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic

[§]Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43403, United States

[∥]Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, Ohio 43403, United States

S *Supporting Information*

ABSTRACT: The sarcin−ricin RNA motif (SR motif) is one of the most prominent recurrent RNA building blocks that occurs in many different RNA contexts and folds autonomously, that is, in a context-independent manner. In this study, we combined bioinformatics analysis with explicit-solvent molecular dynamics (MD) simulations to better understand the relation between the RNA sequence and the evolutionary patterns of the SR motif. A SHAPE probing experiment was also performed to confirm the fidelity of the MD simulations. We identified 57 instances of the SR motif in a nonredundant subset of the RNA X-ray structure database and analyzed their base pairing, base−phosphate, and backbone−backbone interactions. We extracted sequences aligned to these instances from large rRNA alignments to determine the frequency of occurrence for different sequence variants. We then used a simple scoring scheme based on isostericity to suggest 10 sequence variants with a highly variable expected degree of compatibility with the SR motif 3D structure. We carried out MD simulations of SR motifs with these base substitutions. Nonisosteric base substitutions led to unstable structures, but so did isosteric substitutions which were unable to make key base−phosphate interactions. The MD technique explains why some potentially isosteric SR motifs are not realized during evolution. We also found that the inability to form stable cWW geometry is an important factor in the case of the first base pair of the flexible region of the SR motif. A comparison of structural, bioinformatics, SHAPE probing, and MD simulation data reveals that explicit solvent MD simulations neatly reflect the viability of different sequence variants of the SR motif. Thus, MD simulations can efficiently complement bioinformatics tools in studies of conservation patterns of RNA motifs and provide atomistic insight into the role of their different signature interactions.
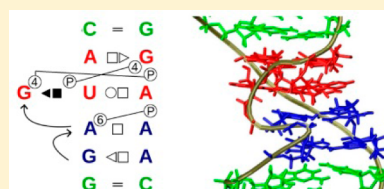
INTRODUCTION

Complex RNA 3D structures consist of short canonical double helices formed by Watson−Crick guanine−cytosine (GC), adenine−uracil (AU), and guanine−uracil (GU) base pairs alternating with noncanonical regions. The noncanonical regions include internal, hairpin, and multihelix junction loops. These loops often comprise conserved sequences of nucleotides and are highly structured.[1] Recurrent and conserved elements of primary and secondary structures forming precisely shaped 3D building blocks of RNA molecules are known as RNA motifs.[2−4]

The sarcin−ricin RNA motif (SR motif) (Figure 1) is a textbook example of an RNA building block that occurs in many different contexts and folds autonomously, that is, in a context-independent manner. The SR motif can be found in the core of an RNA or on its surface. SR motif instances are observed as internal loops within helices and as parts of complex multihelix junctions.[5,6] Both cases can be formed at the level of secondary structure (local motif formed by noncanonical base pairing of two strands) but also tertiary structure (composite motif involving more than two strands).

The diversity of locations and compositions of this motif suggests that it may serve a variety of functions.

The originally found SR motif is situated in helix 95 in domain VI of the large ribosomal subunit. It corresponds, together with an adjacent GAGA tetraloop, to one of the most conserved sequences of the ribosome and is referred to as the sarcin−ricin domain (SRD).[7] The name derives from the toxins sarcin and ricin that target the SRD of helix 95.[8−10] Sarcin cleaves the phosphodiester backbone at a specific position in the tetraloop, and ricin depurinates the adenine that is stacked at the top of the tetraloop.[8,11,12] Both the SR motif and the GAGA tetraloop are recognized by toxins.[13−15] The bulged G of this motif (Figure 1) is essential for specific recognition of SRD by sarcin.[14] Structural data reveal that sarcin recognizes the overall unique 3D conformation of the SR motif.[14,16,17]

The SRD forms part of the binding site for elongation factors EF-Tu and EF-G (in bacteria).[18,19] Disruption of the SRD by ribotoxins or deletion of SRD disables binding of elongation factors to the ribosome and causes its inactivation, although
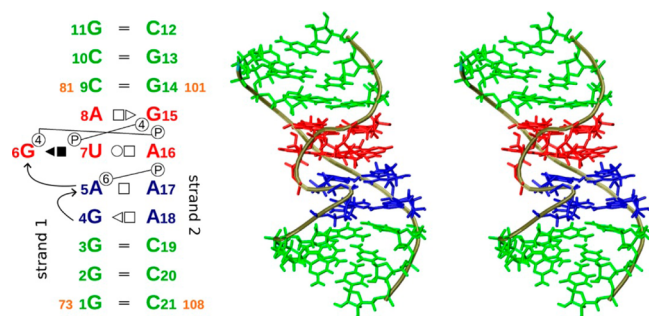
Figure 1. 2D and 3D (wall-eyed stereo) representations of a typical SR loop, depicting the structure that is used as the reference structure in this study (marked as G78-1S72 in the following, with GU wobble base pairs replaced by canonical GC pairs, see below). The G-bulge region is red, the flexible region is blue, and canonical regions are green. The red, green, and blue numbering of nucleotides is used in this study, while the original numbering (taken from the PDB file 1S72) is orange. Standard annotations of base pair (bp)[31] and base-phosphate (BPh)[34] interactions are used.

elongation factor-independent proteosynthesis is not affected.[20−22] The ability of the SRD to bind EF-G was confirmed by experiments with oligonucleotides.[14] According to structural data, the SRD is located close to the catalytic center of bound elongation factors.[21] Deletion, mutation, and disruption of SRD of the ribosomal Helix 95 are lethal.[22−24] There are several other universally conserved SR motifs in the ribosome.[5,6,25,26] It has been suggested that SR motifs serve to organize the structure of rRNA during folding, which is supported by the high occurrence of SR motifs in domains I and II of the large ribosomal subunit.[5]

Numerous atomic-resolution structures of SR motifs are available. Many occur in ribosomal X-ray structures,[5,6] while several others occur in riboswitches.[27−29] There are also high-resolution atomic structures of isolated SR motifs[16,17,30] alone or cocrystallized with ribotoxins.[16]

The evolution of folded RNA molecules is constrained by the principle of isostericity.[31,32] The sequence variability of structured RNA molecules can be understood by considering the geometry of the pairwise interactions (i.e., classification of base pairs) formed by the nucleotides in the motif. Viable base substitutions (mutations) at corresponding positions in different instances of the same motif almost always conserve the exact shape of its native signature interactions.[32] Although the basic principle of isostericity is now widely accepted, our knowledge of the full spectrum of sequence variants of different RNA motifs found in 3D structures and in sequence alignments is likely still incomplete. The sequence variability of known instances of a motif may be further constrained by exogenous (external) factors including distal tertiary interactions, interactions with a certain protein, or the need to avoid a certain folding pathway. Thus, it would be very useful to be able to predict likely sequence variants of known motifs using additional approaches.

It is well-established that violations of the principle of isostericity are rather rare, that is, nonisosteric substitutions are uncommon during evolution.[4−6] However, there are many sequences that should be able to form the SR motif (according to the base pair isostericity principle), but that were so far observed in neither structural nor sequence data.[5,6] Thus, some sterically viable substitutions are not realized during evolution. This can be caused either by exogenous factors that can

constrain sequence (see above) or by their intrinsic instability, that is, energy considerations.[33] To better understand which sequences are able to form SR motifs, we need to better understand why certain isosteric base combinations are not observed.

Some cases of unobserved sequences can be explained by conserved base−phosphate (BPh) interactions. BPh interactions were recently classified, extending the base pair classification.[34] While changes in base pairing are described by isostericity quite unambiguously, it is more difficult to describe the pattern of BPh interactions, because their formation is strongly dependent on the shape of the phosphodiester backbone.[34]

In the present work we investigate conservation patterns in SR motifs using computational approaches. We perform bioinformatics (using structural as well as sequence data) analyses of SR motif sequence variants followed by extended explicit solvent molecular dynamic (MD) simulations to evaluate the stabilities of different sequence variants of the SR motif. We analyze structures in the Protein Data Bank (PDB) database and evaluate base pairing and other interactions in these structures to get an initial idea about variability of the SR motif. Based on this survey, we chose 10 sequences with different base substitutions for MD simulations. The simulations are primarily used to investigate different SR sequence variants which are not seen in the 3D structure database, to explain why they are not realized in the course of the evolution. MD simulation is an established technique that captures spontaneous dynamic behavior of RNA motifs on the submicrosecond time scale, starting from the folded structure.[35,51] Although MD simulations do not directly reveal free energies, we can monitor structural instabilities of the SR motif caused by different base substitutions. It is likely that structures that develop significant instabilities on the submicrosecond time scale are less stable and thus unlikely to be able to form fully functional SR motifs. The simulations allowed us to better understand the relation between the sequence and tertiary structure of SR motifs and to propose a consensus sequence for viable SR motifs at the base pair level. Although the overall simulation performance was judged to be satisfactory and the simulation technique gives very useful results, we did notice some force field limitations which are briefly discussed in the Discussion.

■ METHODS

Description of SR Motif. As the reference structure, we use the SR motif from 5S rRNA of *Haloarcula marismortui* (*H. marismortui*) (PDB code 1S72; residues 73−83 and 99−108). This variant has good base pairing geometry, and furthermore its sequence has the highest occurrence in RNA X-ray structures.[52] Although the 1S72 crystal structure[52] is determined at 2.4 Å resolution, the structure of the selected SR motif very closely resembles the isolated SR motif determined at 1.1 Å resolution (PDB code, 483D[30]). In our study, the two GU wobble (cWW) base pairs in the vicinity of the SR motif (U82/G100 and G83/U99) were replaced by cWW (cis-Watson−Crick Watson−Crick, canonical) GC base pairs (Figure 1) to minimize end effects in simulations.

Numbering and Annotation. We investigate diverse sequence and structural variants of the SR motif. To unify the description of structures, we choose a common nucleotide numbering. The bulged G (strand 1) is numbered 6, and the A participating in the base triple (strand 2) is numbered 16

(Figure 1). To distinguish between crystal structures of SR motifs, we mark them by the original number of the bulged G and by the PDB code. Thus, the reference structure is marked as G78−1S72.

The backbone suites (set of backbone dihedrals from δ to $\delta+1$) are numbered according to the number of the residue containing the phosphorus atom.[53] SR motifs consist of canonical and noncanonical parts. The canonical part is modeled by cWW GC base pairs in the terminal regions. In the noncanonical part, we distinguish the G-bulge and flexible regions (Figure 1).[35] The noncanonical base pairs of this motif are shown in Figure 2.
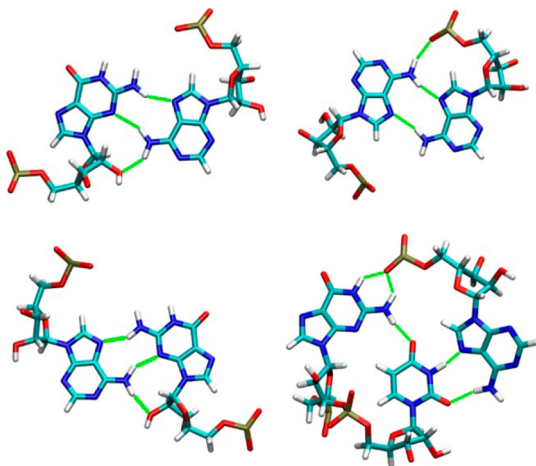

Figure 2. Noncanonical bp's present in the reference structure: tSH G4/A18 (upper left), tHH A5/A17 (upper right), tHS A8/G15 (lower left), and cSH/tWH G6/U7/A16 base triple (lower right). The H-bonds are colored green.

Base Pairing and H-Bonds. The G-bulge region in the reference structure consists of a tHS A8/G15 base pair (bp) and G6/U7/A16 base triple comprising cSH G6/U7 bp and tWH U7/A16 bp, which is quite nonplanar (Figure 2, Supporting Information, Figure S1). There is a characteristic cross-strand stacking of adenines (A8 and A16) and of the bulged G6 with A15 (Figure 3). The G-bulge region is also known as the GpUpA/GpA miniduplex and includes the G6-U7 dinucleotide platform.[54] The miniduplex has been studied in detail in refs 35, 54, and 55. Backbone suite 7 (GpU

platform) forms an edge participating in a 4BPh interaction (base phosphate interaction type 4)[34] by U7(O2P) that makes H-bonds with N1 and N2 of G15. Moreover, there is an intrabackbone U7(O2P)−G6(O2′) interaction.[54] Bulged G6 also forms a 4BPh G6/A16 interaction with backbone suite 16. Other interactions in the G-bulge region are base−sugar G6(O2′)−G15(O6) and sugar−sugar G15(O2′)−A16(O4′) interactions (Figure 3). The very complex system of 13 H-bonds and interstrand stackings make the G-bulge region very rigid[35,54,55] and means that the "bulged G" is actually very tightly integrated into the core of the motif.

The flexible region of the reference structure contains tSH G4/A18 and tHH A5/A17 bp's supplemented by a 6BPh A5/A17 contact. Note that the name *flexible region* likely originated from expectations in the older X-ray studies that the base pairing is weak and thus the region could fluctuate. Earlier MD simulations rather suggested that the whole SR motif is stiff.[35]

Backbone Conformation. A characteristic feature of SR motifs is the S-turn conformation of the backbone around the bulged G,[56] which reverses chain direction in the flexible region and then restores it in the G-bulge region (the double flip over). A5 is in the "3′-5′ like" arrangement that enables formation of the 6BPh A5/A17 contact. The common native backbone conformation of SR motifs is, according to Richardson's classification,[53] described as $G_31aG_45zA_54s$-$G_6\#aU_71aA_8$ (strand 1) and $G_{14}1aG_{15}1eA_{16}1aA_{17}1aA_{18}$ (strand 2) sequence of suites. The reference structure exhibits a few deviations which may be due to resolution; namely, there is unclassified suite instead of 5z in strand 1. The high-resolution structures show several backbone geometries in this region.

Searching, Structure Selection, and Preparation. We used webFR3D[26]—the online version of the FR3D[25]—program suite to search for structures of the SR motif. FR3D enables searching in the PDB database. The reference structure was used as a model structure. Both symbolic and geometric approaches were used to define search input. Details about searches are described in the Results section.

The reference structure was used as the starting structure for MD simulations. We introduced mutations into this structure to derive different variants of the SR motif. The mutations were introduced in three steps: (1) The PDB file was modified (residue name was changed, and atoms that are different in old and new nucleotides were removed). (2) The missing atoms were automatically added by the X-leap module of the Amber software package. (3) The geometries of mutated bp's were manually adjusted to remove severe steric clashes and to relieve other problems introduced by the automatic insertion of missing atoms in step 2.

Sequence Analysis. The following procedure was applied to get SRL sequence variants from RNA multiple sequence alignments. Bacterial small ribosomal subunit (SSU) and large subunit (LSU), archaeal LSU, and eukaryal SSU and LSU alignments were downloaded from the Silva webserver[57] on March 21, 2013. The GreenGenes[58] SSU alignment dated 2012 was downloaded from the GreenGenes download server. The bacterial LSU alignment from Stombaugh et al.[32] was also used. SR motif groups from the RNA 3D Motif Atlas release 1.0[59] were used to identify individual instances of the full SR motif appearing as an internal loop. Each SR motif group has an identification code (ID) in the RNA 3D Motif Atlas and can be viewed at a URL of the form http://rna.bgsu.edu/rna3dhub/motif/view/IL_49493.2 (a few other SR motif clusters are IL_85647.3, IL_95652.3, and IL_94744.1). Within each SR
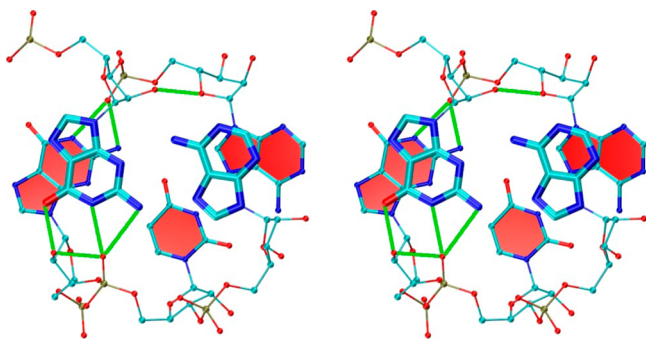

Figure 3. Wall-eyed stereo representation of the G-bulge region (i.e., the GpUpA/GpA miniduplex). The tHS A8/G15 bp is in the upper plane, and the G6/U7/A16 base triple is in the lower plane (filled red); H-bonds are green. BPh, sugar−sugar, sugar−phosphate, base−sugar, and the cross-strand purine stacking interactions are visible.

motif group there are instances with IDs in the form IL_1S72_103. The ID IL_1S72_103 corresponds to the reference example G78−1S72 used in this article. See the Supporting Information text for further explanation of the IL (internal loops) identifiers used in the RNA 3D Motif Atlas.

For each SRL instance from a ribosomal 3D structure, the full sequence of the 3D structure was aligned using the Needleman−Wunsch algorithm to each row of its corresponding alignment(s) to find the best matching entry; in every case, the exact sequence was found in the alignment with at most a handful of base substitutions. The nucleotides for each SRL instance were then mapped to columns of the alignment. Two primary extracts were made, the central 9 nucleotides, which correspond to positions 4−8 and 15−18 in Figure 1, and the central 7 nucleotides, which correspond to positions 5−8 and 15−17 in Figure 1. Only the 9 or 7 aligned columns were retained, after checking that the intervening columns are populated only by sporadic insertions and a small number of misalignments that would complicate the analysis but not change the overall results. The 9- or 7-nucleotide sequence variants thus obtained from each alignment were tallied and recorded.

For each of the 11 sequence variants studied in the article, the central 9 or central 7 nucleotides were extracted and searched in each sequence alignment extract(s). Supporting Information, Table S1, shows the alignments in which at least one match was found, the loop of the SRL instance that corresponds to that position in the alignment, the motif group ID where that instance appears in the RNA 3D Motif Atlas, and the corresponding designation of the structure used in this article. Numerical information in each line indicates the number of occurrences of the sequence in the corresponding columns of the alignment and the percentage of all sequence variants in those columns of that alignment. Thus, we present information on both the raw counts and the frequency of each sequence variant in each position of each alignment.

The conservation of nucleotides of different SR instances in LSU and SSU was extracted from sequence alignments, but also from a Comparative RNA Web Site[60] (Supporting Information, Table S2).

MD Simulation Protocol. The primary Amber nucleic acid Cornell et al. force field[61] was used with bsc0[62] and χ_{OL3}[63,64] corrections. Some systems were also simulated with older variants of the force field without the essential χ_{OL3} correction; however, the simulations were visibly less stable than those carried out with the bsc0χ_{OL3} variant. We used the TIP3P[65] water model and standard Amber Lennard−Jones parameters for Na$^+$ atoms (radius 1.868 Å and well depth 0.0028 kcal/mol). Discussion of the ion parameter choice can be found in our recent RNA simulation papers where many tests were presented with different ion parameters, ion concentrations, and water models, which led to the conclusion that the type of ion parameters and the ion concentration have a very small impact on the RNA dynamics.[36,66] Since the present study builds up on qualitative analysis of relative simulation behavior of different SR motif sequence variants, we assume the results are entirely unaffected by the choice of ion condition and water model.

Simulations were performed using the Amber suite of programs.[67] Files with topology and coordinates were generated by the X-leap module of AmberTools.[67] Using X-leap we added Na$^+$ ions to neutralize the system and an octahedral box of explicit water solvent with minimal distance of 10 Å between the border of the box and solute.

The standard equilibration and production inputs were used.[37,66] The assembled system was heated, minimized, and equilibrated in several steps before each MD run. Artificial force constants were applied to the atoms of solute during minimization, equilibration, and thermalization in order to fix their positions. This restraint was gradually reduced to zero, and the last step was 50 ps of free MD run.

Explicit solvent MD simulations were performed using the particle-mesh Ewald MD method.[68,69] The value of the nonbonded cutoff was 9 Å, and the integration time step was 2 fs. SHAKE[70] constraints were applied to all hydrogens to allow a longer time step. MD simulations were carried out with constant pressure boundary conditions and constant temperature of 300 K. The Berendsen weak-coupling algorithm[71] was used for temperature regulation.

SHAPE Probing. SR loop motif variants 0−10 (see below) were submitted to the EteRNA[72] synthesis pipeline[73] for SHAPE probing in June 2013. Sequences were designed as stem-loops consisting of a helix, a SRL variant, helix, a second SRL variant, helix, and hairpin. Each SRL variant was placed twice into each sequence, once near the terminal ends of the stem-loop and once next to the hairpin loop, cf. Supporting Information, Table S3. Each sequence was synthesized and then probed with the SHAPE technique.[74] SHAPE reactivities were measured and normalized at each position. These are reported in data file http://rmdb.stanford.edu/site_media/rdat_files/ETERNA_R73_0000/ETERNA_R73_0000.rdat; the relevant lines are annotated as data lines 895−915. Reactivity values are shown in Supporting Information, Figure S2, for each SRL variant, and standard errors are shown in Supporting Information, Figure S3.

Data Analysis. X-ray structures were analyzed with webFR3D and MolProbity[75−77] software. The trajectories were analyzed using Ptraj module of Amber and VMD.[78] VMD was also used for visualizing the trajectories. Graphs were plotted with Gnuplot 4.4.[79]

■ RESULTS

Searches. As a first step in understanding the relations between sequence and structure of the SR motif, we found existing X-ray structures of SR motifs and compared their characteristics. Five searches were performed using webFR3D.[25,26] webFR3D enables searching for RNA 3D structures using symbolic, geometric, or mixed approaches. We confined the search to a nonredundant set (release 0.58 at 4.0 Angstrom resolution) of RNA X-ray structures.[26]

First, we applied a symbolic approach. The following criteria can be specified as an input for the search based on the symbolic approach: the sequence of bases, base−base interactions, base−phosphate interactions, distances of nucleotides (in primary structure), and stacking of bases. The user can choose either an exhaustive description of the structure to find just a specified molecule or a less detailed description to find a wider range of matching structures. The critical step of searching using webFR3D is to define suitable searching criteria. Schematic illustrations of inputs of performed searches are shown in Figure 4.

In search 1 we specified the exact sequence and all base−base interactions of the noncanonical part of the reference structure. To exclude composite motifs, constraints on distances of nucleotides were imposed. The criteria in this search were really
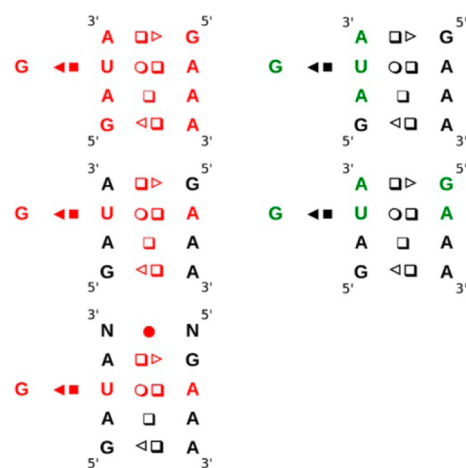
Figure 4. Scheme of webFR3D searches. Searches 1−3 are symbolic searches with constraints on nucleotide identity and basepairing; the constrained nucleotides and base pairs are indicated in red in the left panel (searches 1−3 from top to bottom). Query structures of geometric searches 4 (top) and 5 (bottom) are highlighted in green in the right panel. These diagrams do not indicate constraints on nucleotide order or connectivity; see the main text.

strict, and only eight structures were found (Supporting Information, Table S4), all being in perfect agreement with the definition of the SR motif.

Search 2 was similar to the first one, but the only sequence requirement corresponded to G6, U7, and A16 participating in the base triple. The base−base interactions and the continuity of strands remained identical as in search 1 (Supporting Information, Table S4). Twenty-four structures with different sequences but identical interactions between bases were found. All eight structures from search 1 were also found by search 2, and all structures found were local SR motifs.

The results of the first and second searches showed that the sequence of nucleotides at positions 4 and 18 is variable. In search 3 only the nucleotides G6, U7, and A16 were defined in sequence. Furthermore the base pairing in the G-bulge region and the geometry of 9/14 bp were specified. The nucleotides of the flexible region and the interactions between them were not specified. The third search identified 60 structures. Of these 14 were composite SR motifs, caused by the loose description of the flexible region. In comparison to search 2, this search found more structures, but a few structures found by search 2 were missed by search 3.

Our results illustrate that it is not simple to perform an unambiguous search using a purely symbolic approach. The main problem is that the user usually does not know the variability of sequence and interactions in the target motif.

We then performed geometric searches. The geometric approach searches for all 3D structures similar to the query structure defined by the user. The user-specified query structure should be highly characteristic for the searched motif.

In search 4 nucleotides 77−80 from the reference G78−1S72 structure were used as a query structure. These nucleotides include part of the S-turn, which is really a unique feature of the SR motif. All 56 structures found (Supporting Information, Table S4) were local SR motifs. This set of structures comprises all local SR motifs that were detected by symbolic searches. It seems that the geometric approach is the best choice for our purpose.

Finally, search 5, based on conserved nucleotides of the G-bulge region, was performed. The nucleotides of the G-bulge region were used as a query structure. Such criteria do not constrain variability of the flexible region and enable comprehensive searching for SR motifs in junction loops. This search identified 100 structures. It did not identify any new local SR motif, but some composite structures and geometrically similar structures (loop E motif) that are not relevant for our purposes were found.

SR Motif Structures. Several studies mapping the occurrence of the SR motif in RNA molecules have been published.[5,6] These studies describe sequence conservation and roles of SR motif in organizing RNA structures based on sequence data[5] or provide detailed descriptions of ribosomal SR motifs and their interactions with adjacent structural elements in atomic structures.[6] Our bioinformatics analysis provides the most complete overview of SR motifs carried out so far, with exhaustive structural analysis of all available structures supplemented by comparison of structural and sequence data.

Altogether 56 local SR motifs were identified by our searches (Supporting Information, Table S4). They were found in (i) large ribosomal subunits (38 cases), (ii) small ribosomal subunits (9 cases), and (iii) other molecules (10 cases). One structure (PDB code 4ERJ[29]) was published after our initial search and was added afterward, resulting in 57 structures. We suggest that these 57 SR motifs can be divided into four groups.

Group I (19 structures, colored purple in Supporting Information, Figure S4) represents structures that have all five characteristic noncanonical bp's flanked by canonical bp at each end, leading to the following consecutive bp interactions: cWW-tHS-tWH-cSH-tHH-tSH-cWW. Group II (13 structures, colored blue in Figure S4) includes structures with characteristic noncanonical bp's: tHS-tWH-cSH-tHH-tSH, but with noncanonical or wobble flanking bp or with another bulged base in the neighborhood. Group III (10 structures, colored yellow in Figure S4) contains some difference of base pairing in the noncanonical region. Nevertheless, all structures of the first three groups have the S-turn and the characteristic fold of the SR motif. Group IV represents structures that either do not have the S-turn or have S-turn, but there is some another variability in their fold (15 structures, colored red in Figure S4). While structures in groups I−III are internal loops, the fourth group contains mainly junction loop SR motifs.

There are 7 internal loop and 2 junction loop SR-motifs in the large ribosomal subunit of *H. marismortui* (Figure S4a, S4b). SR structures G213, G381, G464, G1370, and G2692 of 1S72 are universally conserved (based on sequence data) in all three domains of life, while the G175−1S72 SR motif is conserved across the archaea and bacteria (Supporting Information, Table S2). The 5S rRNA SR-motif (G78−1S72) is conserved across the eukaryota, but conservation in archaea is not so unambiguous (Supporting Information, Table S2). Only the G-bulge region of this motif has more than 90% conservation in archaea. Interactions of SR motifs with surrounding elements in large ribosomal subunit are described in Leontis et al.[6]

Each of the four nonredundant small ribosomal subunits deposited in the PDB database contains two SR motifs. The first (G890−1FJG) occurs as an internal loop and the second (G1347−1FJG) as a junction loop (Figure S4c). Both G890−1FJG and G1347−1FJG are universally conserved (Supporting Information, Table S2). Structure 1FJG contains an extra junction loop SR motif (G485−1FJG) not present in the other

Table 1. Geometries of bp's Found in Nonredundant X-ray Structures of SR Motifs[a]

position	base pair family	count	nucleotides	count
4/18	tSH	33 (33)	GA	12 (12)
			CC	9 (9)
			UC	7 (7)
			UA	2 (2)
			AA	2 (2)
			AC	1 (1)
	ntSH-wm	5 (5)	UC	3 (3)
			CC	2 (2)
	cWW	3 (2)	UU	1 (1)
			CG	1 (0)
			UG	1 (1)
	tWW	3 (0)	UU	3 (0)
	tSW	1 (1)	UG	1 (1)
	not present	11 (0)		11 (0)
	unpaired	1 (1)	UC	1 (1)
5/17	tHH	48 (35)	AA	48 (35)
	tHH-wm	5 (5)	AC	5 (5)
	tHW	1 (1)	AC	1 (1)
	tWH	1 (1)	AA	1 (1)
	unpaired	2 (0)	AU	2 (0)
6/7	cSH	57 (42)	GU	57 (42)
7/16	tWH	57 (42)	UA	57 (42)
8/15	tHS	57 (42)	AG	53 (40)
			CU	2 (0)
			AA	2 (2)

[a]Count means the number of instances of the SR motif comprising certain base pairs at a defined position. Numbers in parentheses show counts excluding the group IV. Note that the tWW U4/U18 bp occurred only in junction loop SR motifs. ntSH means near-tSH, and wm means water-mediated interaction. See Supporting Information, Figure S4.
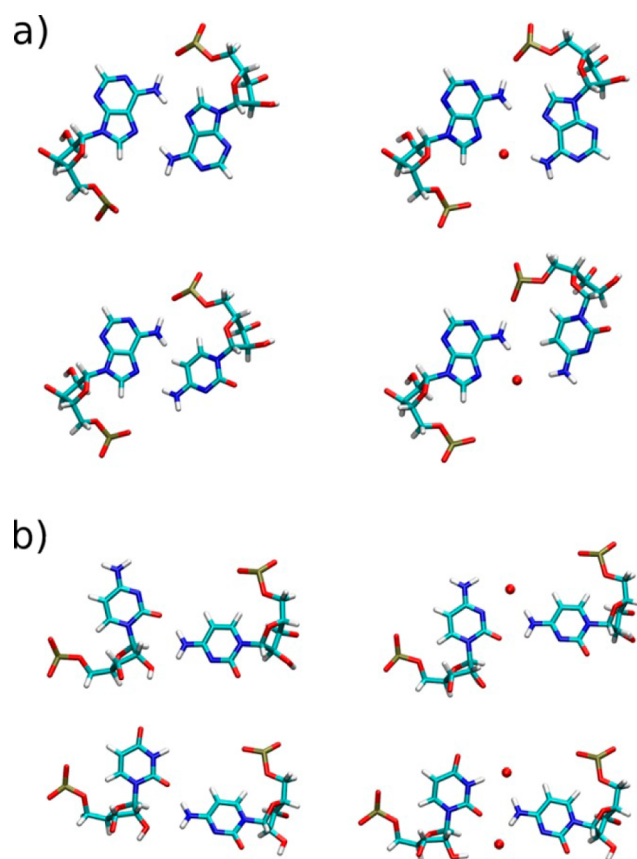

Figure 5. Examples of bp's observed in the crystal structures. (a) tHH and tHH-wm (the water molecule is marked by a red dot) geometries of A5/A17 and A5/C17 bp's. The 6BPh contact between A5(N6) and N17(O2P) is conserved. Note that tHH geometry of the A5/C17 bp (lower left) was observed neither in simulations nor in X-ray structures of SR motifs, and this geometry is taken from the bp catalog.[31] (b) tSH (left) and ntSH-wm (right) geometries of C4/C18 (top) and U4/C18 (bottom) bp's.

small subunit structures. SR motifs that were found in nonribosomal structures are shown in Figure S4d.

Base Pairing in SR Motifs. Table 1 contains a survey of bp's observed in X-ray structures of the SR motif. The core G6/U7/A16 base triple is present in all structures in the native cSH-tWH geometry. Thus, there is neither structural nor sequence variability in the triple. The tHS interaction of the 8/15 bp is also conserved in all structures. This bp is mostly formed by the G8/A15 combination (53 of 57 cases), so there is some sequence variability in the 8/15 interaction. The 5/17 bp is most often in tHH or tHH water-mediated (tHH-wm in the following) geometries (Figure 5a). At the 4/18 position there is often bp with tSH geometry, but also near-tSH-water mediated (ntSH-wm in the following) (Figure 5b), cWW and other geometries were detected. There is considerable variability in nucleotides forming the 4/18 bp. The SR motifs that are junction loops lack the 4/18 bp entirely.

Backbone Geometry. Several backbone conformations of the S-turn region were observed in the X-ray structures. The 1a-5z-4s-#a-1a sequence of suites is the most common in the SR motif,[53] and indeed 60% of structures match these criteria. There are some variations involving suites 6 (usually A_5sG_6) and 7 (usually $G_6\#aU_7$) in ~40% of structures. In several cases (for relevant examples see structures G182−2ZJR, G219−2ZJR, G1086−2XZM, and G47−4A1B) suite 7 exhibits strange geometry that does not fit any of the suites defined in Richardson et al.[53] This probably reflects resolution and

refinement limits, specifically of the sugar ring. It is indicated by an unusual value of the ε angle (around 145°)[53] and by inconvenient positions of H-bond donors and acceptors. Some of these geometries are close to &a or other A-form geometries, but with unrealistic distance between O2′ and O2P atoms (2.0−2.5 Å), which indicates refinement errors. Another group that does not match the expected geometries has β value ca. −80° in suite 6 (for relevant examples see structures G205−3UXR, G10−1JBR, and G588−1S72). Such geometries were not classified.[53] Note that many structures have X-ray resolution worse than 3 Å and backbone conformations of such structures are difficult to determine. Positions of bases and phosphates are more clearly visible in the electron densities.

BPh Interactions and Conservation in SR Motif. Nucleotides A5, G6, U7, A8, G15, and A16 are highly conserved in SR motifs. The BPh interactions[34] between G6/A16 (usually 4BPh with some 3BPh instances) and the A5/N17 6BPh interactions are universally conserved in structures. Nucleotides A5 and G6 are universally conserved in sequence, presumably in order to form the BPh interactions mentioned above. The third observed BPh interaction is G15/U7 4BPh or 3BPh and is present in more than 90% of structures. This interaction requires G at position 15. G15 is highly conserved in most of the ribosomal SR motifs (Supporting Information, Table S2).
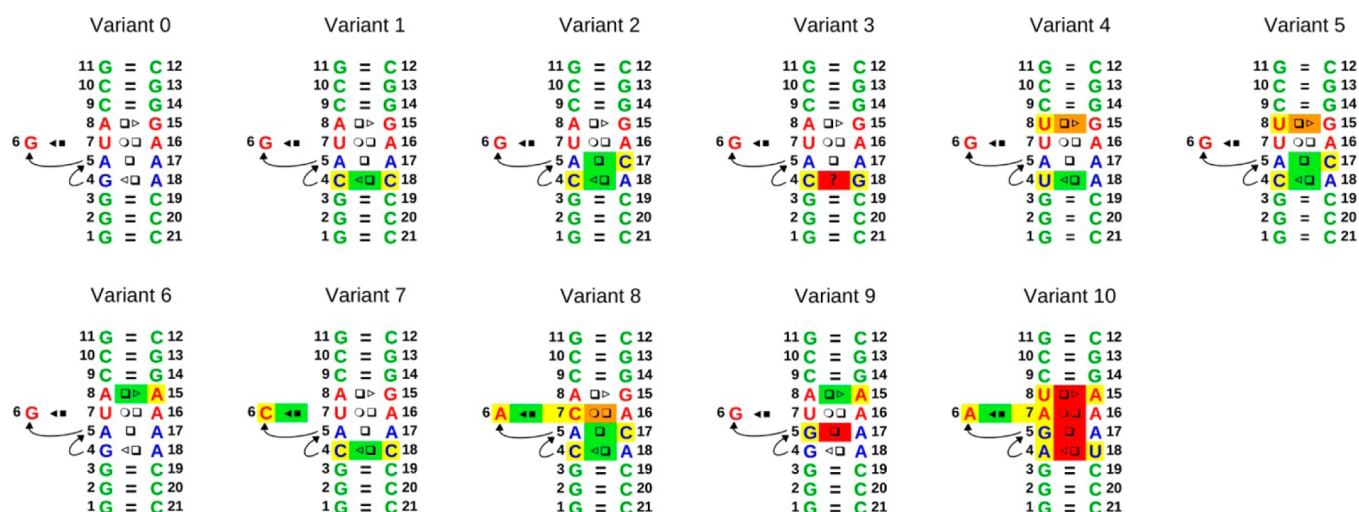
Figure 6. Annotated 2D structures of all simulated SR motifs (variant 0 is the reference structure). The G-bulge region is red, the flexible region is blue, and cWW GC base pairs are green. The yellow, orange, and red boxes indicate isosteric, near-isosteric, and not-isosteric substitutions, respectively. The structures are numbered from 0 to 10 according to decreasing score (considering base–base and base–phosphate interactions) that reflects the expected decreasing suitability of the system to form the SR motif (see the text). CG base combination is not able to form cSH base pair. We placed bases C4 and G18 near to required geometry and marked with ? in the variant 3.

Tertiary Interactions and Conservation in SR Motif. There are several universally conserved SR motifs in the ribosome. The bioinformatics data reveal that SR motifs at certain positions in the ribosome have different patterns of conserved nucleotides (Supporting Information, Table S2). These conservation patterns are imposed by tertiary interactions. Conservation of the U4/C18 bp in SR motif of helix 95 (right column in Figure S2b) can serve as an example. This SR motif interacts with L6 (bacteria) or L9 (archaea, eukaryota) ribosomal proteins. The sequence of protein interacting with C18 is not conserved. However, only C at position 18 is able to simultaneously serve as an H-bond acceptor to interact with the protein and an H-bond donor to interact with the nucleotide at SR position 4. Adenine lacks an H-bond acceptor to interact with the protein, uracil lacks an H-bond donor to interact with U4, and guanine cannot form an isosteric bp.

MD Analysis and SHAPE Probing of Substituted SR Motifs. Molecular simulations can be used to study the impact of base substitutions in nucleic acid structures based on known structures of the unsubstituted systems.[38,39] This application of the simulation technique is justified because the description of the base pairing and base stacking interactions belongs to the most reliable parts of the force field approximation.[39] In this study we investigated base substitutions in the G78−1S72 SR motif to understand the relation between sequence and structure of this motif (Figure 6). Further, we wanted to demonstrate that MD simulation can be used as a complement to bioinformatics and structural data to get insights into the role of base substitutions in known structures.

We simulated the reference (variant 0) system and 10 substituted structures (Figure 6). These were selected to include observed structures, unobserved structures supported by sequence analysis, structures that could be isosteric but are not supported by sequence analysis and nonisosteric cases. SR motif variants 0, 1, and 6 are observed in the X-ray database, and existence of variant 3 is supported by sequence data. The remaining seven sequences are either not found or are extremely rare in the sequence alignments, as shown by

Supporting Information, Table S1. These sequences are thus unlikely to form SR motifs.

How To Compare MD Simulations and SHAPE Data. All simulated sequences were probed using the SHAPE technique.[74] It should be noted that the comparison between simulations and SHAPE experiments is not always unambiguous. The simulations start with folded SR structure. The time scale of the simulations should allow to identify sequences which are unable to keep a stable SR motif. However, the simulations are not long enough to predict alternative folds of those variants that are not stable as SR motif, unless the rearrangement can be achieved just by a trivial movement with minimal energy barriers. On the other hand, the SHAPE data should reflect the equilibrium structural dynamics for all studied sequences which even may consist of a mixture of structures. However, the SHAPE technique provides only a very limited structural resolution. It basically maps flexibility of the sugar–phosphate backbone of the individual nucleotides. Thus, we expect that agreement between MD and SHAPE should be seen mainly for those sequences that do form SR motifs. For the other sequences, MD and SHAPE techniques should indicate instability and lack of formation of SR motif, respectively. However, no further agreement needs to be achieved as the simulation technique is unlikely to reach the equilibrium structure probed by the SHAPE experiment.

Initial Bioinformatics Ranking. The term isostericity, quantified by the IsoDiscrepancy Index,[32] considers the geometrical parameters of bp's. The IsoDiscrepancy Index is appropriate for comparison of one bp with another. To compare whole motifs consisting of several bp's and often comprising BPh interactions, a more comprehensive approach is necessary. We used an "in house" approximate scoring function (described in Supporting Information) that considers base pairing and base–phosphate interactions. The studied systems were numbered according to this scoring scheme from the (presumably) most likely variant 1 to the least likely variant 10. Note, however, that the scoring is only tentative, since the structure ranked as 6 is actually observed in structural data. As

shown below, the simulations considerably better reflect the suitability of a given sequence to form a viable SR motif.

All bp's were evaluated using Supplementary Table S8 from Stombaugh et al.,[32] which indicates occurrence frequencies of each bp in known RNA structural and sequence data (in the context of the given bp family) (Supporting Information, Table S5). The RMSD values, length of simulations, and occurrence of studied variants in known RNA structures and sequences are shown in Table 2. To exclude the contribution of unstable

Table 2. Occurrences of Studied Sequence Variants (All Non-Canonical Base Pairs Must Be Identical to Get Positive Occurrence) in Known RNA Sequences and Structures (Columns seq. and struct.), Column Length, and RMSD[a]

	occurrence in			RMSD	
	seq.	struct.	length (ns)	max	aver
reference	yes	yes	1000	2.2	0.95
Var. 1	yes	yes	500	2.16	1.05
Var. 2	no	no	500	3.42	1.56
Var. 3	yes	no	500	3.50	2.33
Var. 4	no	no	500	2.60	1.38
Var. 5	no	no	500	2.92	1.54
Var. 6	yes	yes	500	1.64	0.98
Var. 7	no	no	500	3.8	2.35
Var. 8	no	no	500	4.88	3.46
Var. 9	no	no	1000	2.67	1.28
Var. 10	no	no	100	3.56	1.96

[a]Column length gives the simulation length. In columns RMSD are RMSD values of noncanonical parts of the simulated motif with respect to the initial structure (first 200 ns of simulations are considered). Column max represents maximal RMSD values observed in simulation, and the column aver contains RMSD values averaged along the trajectory. Some additional simulations not summarized in the table are discussed in the text.

terminal GC bp's to the RMSD values, only the RMSD values of the noncanonical part of the structures were evaluated. RMSD developments can be found in Figure 7.

Common Simulation Behavior. Let us first describe results that apply to all simulations, are not related to the substitutions, and will not be mentioned further.

1. For the native structure, we see similar developments as reported in earlier shorter simulations.[35] In the G-bulge region, the BPh G15/U7, base−sugar G6(O2′)/G15(O6), and sugar− phosphate G6(O2′)/U7(O2P) contacts were disrupted at the beginning of the simulation. Instead, a new water-mediated contact between G15 and backbone suite 7 or 8 was formed (Figure 8). As discussed elsewhere,[55] this is a hitherto
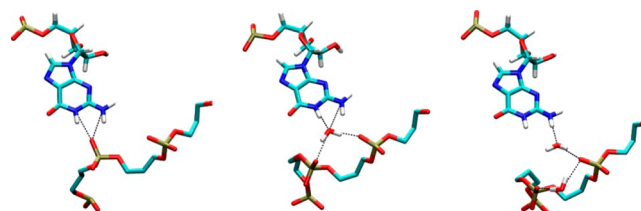

Figure 8. The 4BPh 15/7 interaction in the G-bulge region. Left, the X-ray structure; middle and right, the water-mediated substates dominant in simulations.

unresolved local force field problem. It does not prevent using the simulations to study base substitutions since the perturbation remains local on the present time scale, and it is affecting all simulated structures to a similar extent.

2. The changes in G-bulge region are connected with changes of backbone geometry of suite 7, which shows transition from family #a to family 6g according to the standard classification[53] (Supporting Information, Figure S5).

3. The native interactions and the backbone conformation described above were nevertheless often temporarily restored in the simulations. The population of the native arrangement of
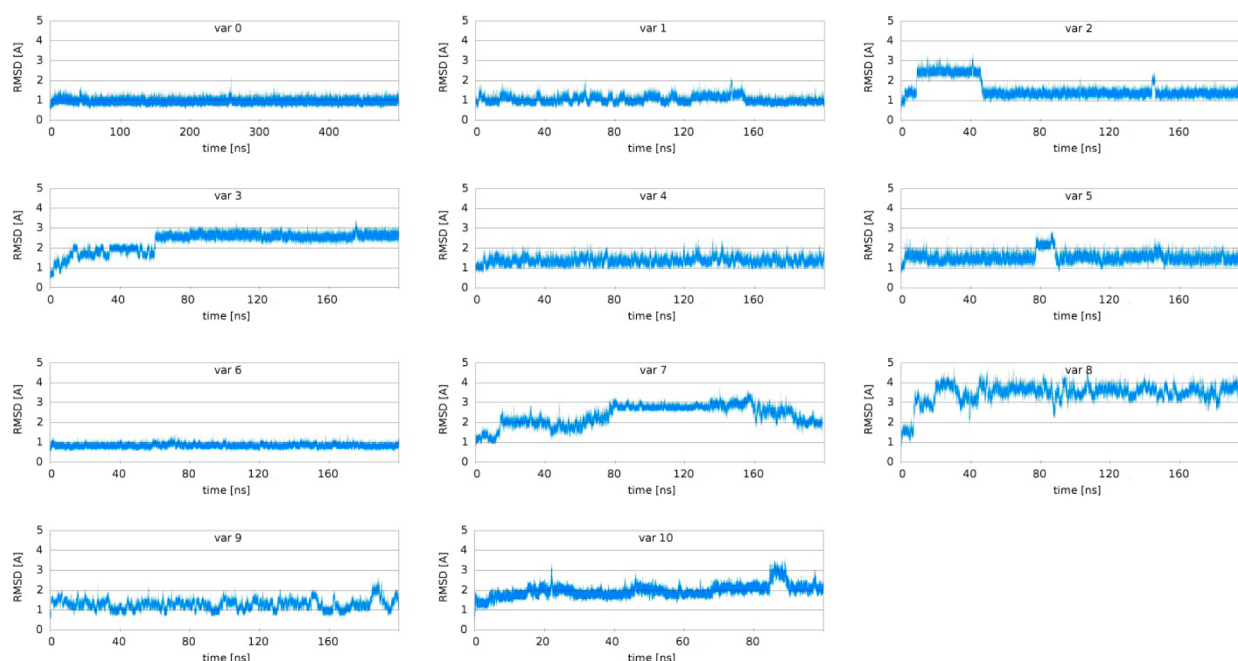

Figure 7. RMSD development of the noncanonical segments in the first 200 ns (500 for the reference structure) of the MD simulations.
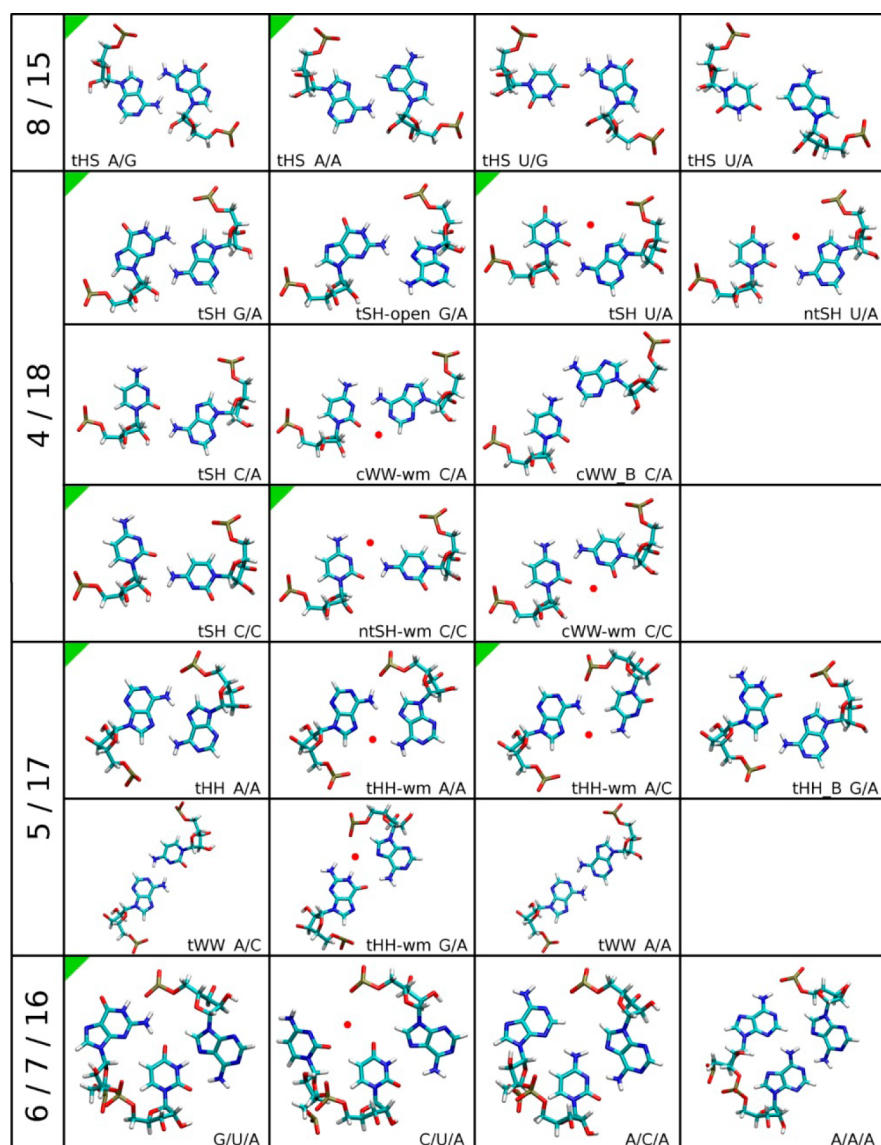
Figure 9. Survey of bp's observed in MD simulations. Bridging water molecules are marked by a dot. Protonated C/C and C/A bp's form cWW arrangement (not shown) which is, however, not consistent with the experimental structures. bp's observed in X-ray structures are marked by green triangles situated top left.

G-bulge region is approximately 10% which indicates that the force field imbalance is rather modest (1−2 kcal/mol free energy imbalance) and may be corrected by some future tuning of the force field.

4. A new G4(O2P)/A5(O2′) sugar−phosphate interaction was formed at the beginning of the simulations. This interaction can be observed in high resolution crystal structures of SR motif (PDB codes: 3DVZ, 483D, and 1Q96) though not in the reference structure.

5. The terminal cWW G11/C12 bp was frequently reversibly disrupted. This is, however, a rather common end effect in simulations which does not affect the behavior of the SR part of the structure.

6. Sequence variants 1, 2, 5, 7, and 8 contain C/C or C/A 4/18 bp, which tends to sample cWW geometries. Base pair classification suggests that, in most other RNA structures, cWW C/C and C/A bp's are protonated.[31] Since all simulations started with the 4/18 base pair in the tSH geometry, residue 18 was not protonated. Since fully paired cWW C/C and C/A bp's cannot be formed without protonation, water-mediated contacts were formed instead, and we call this arrangement cWW-wm base pairing. Similar geometries were observed in simulations for UC base pair.[35] Note that the experimental SRL structures do not indicate protonation of the 4/18 bp, as the observed cWW geometries are imperfect (Table 1). We have also carried out simulations with protonated cytosine (not shown), which immediately formed a firm cWW 4/18 bp. This was accompanied by visible fluctuations of the adjacent 5/17 base pair. This arrangement is not in agreement with the experimental SRL structures, indicating that protonation of C/C and C/A base pairs is unlikely to occur in the context of SR motif 4/18 bp.

Simulations of the Individual Structures. The simulations were initially run for 200 ns, which should be sufficient to see the basic effect of the base substitutions. We have observed the following behavior.

The Reference Structure (Variant 0). The sequence of the reference structure is abundant and can be also found in the
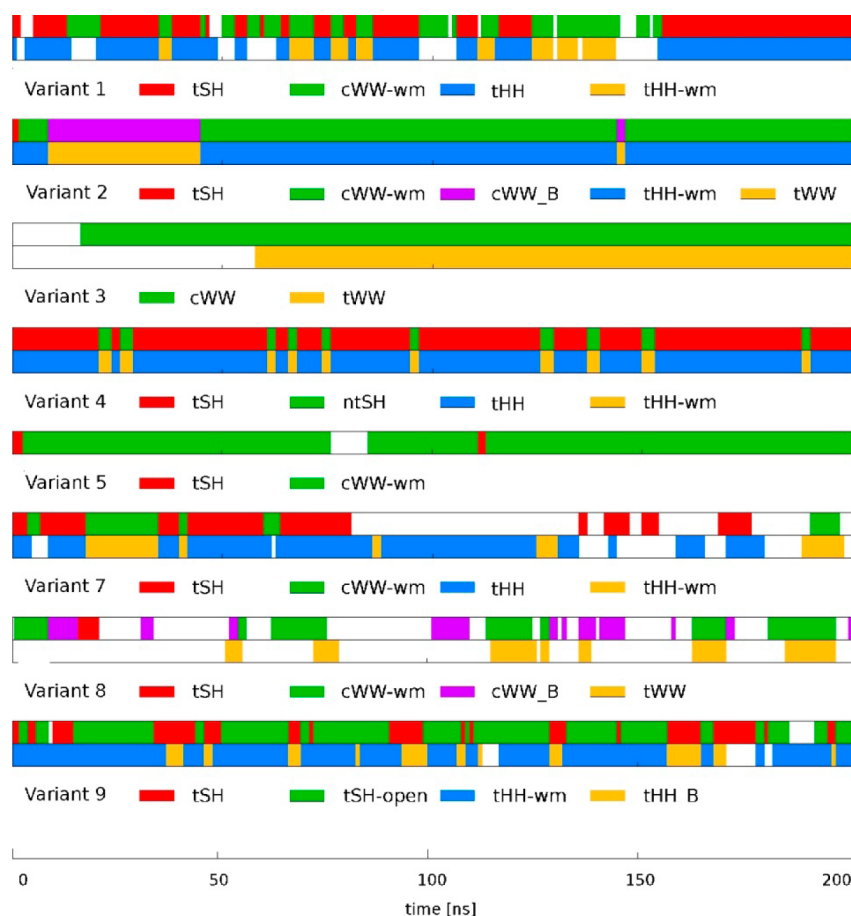
Figure 10. Development of 4/18 (1st row) and 5/17 (2nd row) bp geometries in 200 ns long MD simulations. Note that in the different variants the colors have different meanings because the bp's are formed by different nucleotides. White color indicates unpaired geometry (extended white areas) or very fast fluctuations between two geometries (small white areas). Variants 0 and 6 were stable in the simulations (no changes of the base pair pattern were observed), and variant 10 was so unstable that we are not able to visualize changes in the base pairing (see the main text). These variants are not included in the picture.

following 3D structures: G78−1S72, G77−3U5H, G242−3UXR, G24−3D0U, G1114−3U5F, G27−3DIR, G890−1FJG, and G1266−3UXR. All SR bp's were stable in the simulation. The 6BPh A5/A17 contact was stabilized at distance about 2.95 Å. Changes of backbone torsion angles in the S-turn region are captured in Supporting Information, Figure S5, and probably reflect local force field imbalance, as explained above.

Variant 0 has been shown by UV-monitored thermal melting to be very stable thermodynamically[80] and is a variant that is more prevalent in thermophiles as it contains two tHS AG pairs. It showed no significant SHAPE reactivity, indicating that it is rigid and very stable.

Variant 1. Variant 1 was constructed by the tSH G4/A18 → C4/C18 isosteric substitution. This variant is widespread based on sequence data and was observed in the following 3D structures: G175−1S72, G242−2QBG, G205−3UXR, G182−2ZJR, G47−4A1B, G84−4A1B, G10−1JBS, and G10−1Q96. All bp's in the G-bulge region were stable, but some fluctuations occurred in the flexible region. Particularly, the mutated C4/C18 bp fluctuated between tSH, ntSH-wm, and cWW-wm geometries (Figures 9 and 10). The ntSH-wm geometry is a transitional state between tSH and cWW-wm geometries. The adjacent A5/A17 bp fluctuates between tHH and tHH-wm geometries (Figures 9 and 10). Fluctuations of both base pairs were correlated. The most populated substate (4/18 tSH, 5/17 tHH) (Figure 10) is in agreement with experiment, while the

less populated (4/18 cWW-wm, 5/17 tHH-wm) was not observed in crystal structures. Note that the ntSH-wm and cWW-wm CC geometries are so similar that it is not possible to rigorously calculate their populations. Since the dominant populated state of the flexible region is the same as in the experiment and the other geometry is similar, MD suggests that the substitution does not destabilize the SR motif, in agreement with bioinformatics data.

Variant 1 showed no significant SHAPE reactivity, indicating that it is also rigid and stable.

Variant 2. Variant 2 includes isosteric tSH G4/A18 → C4/A18 and tHH A5/A17 → A5/C17 substitutions. The G-bulge region was stable, but we observed dynamics in the flexible region. Importantly, the mutated C4/A18 bp left the tSH geometry and fluctuated between cWW-wm and cWW_B geometries (Figures 9 and 10). The adjacent A5/C17 b̄p fluctuated between tWW and tHH-wm geometries (Figure 10). The 6BPh A5/C17 interaction was disrupted in the tWW geometry and fluctuated in the tHH-wm geometry. The changes in the flexible region were correlated. The considerable structural fluctuations in the flexible region indicate that the structure is rather unstable, in agreement with absence of this variant in sequence and structural data. The SHAPE data are consistent with the simulations. The SHAPE reactivity is increased compared to variants 0 and 1, albeit does not rule out formation of some folded structure with some features

resembling the SR motif. The loss of stable 5/17 interaction seen in simulations is nevertheless visible in the SHAPE data.

Variant 3. Variant 3 included tSH G4/A18 → C4/G18 substitution. In general, the tSH C/G bp configuration does not offer a viable base pairing, and indeed this base pair does not exist in RNA X-ray structures.[32] The cWW CG 4/18 base pair is observed in structure G485−1FJG. This, however, is one of the variants of SR visibly differing from perfect SR motif (group IV) as well as from the variant 3 (namely, structure G485−1FJG contains unusual 5/17 GC combination, Supporting Information, Figure S4c). We found few instances of potential variant 3 nucleotide combination in sequence data of SR motifs (Supporting Information, Table S1). However, in those cases, the suggested 3/19 bp flanking the SR motif has most often UU, UC, or CC base combinations, indicating that it takes further sequence changes outside the core of the SR motif to accommodate the potential single tSH G4/A18 → C4/G18 substitution. Thus, variant 3 is anticipated to be unstable in simulations as it has canonical 3/19 bp and no other substitution. Indeed, we observed changes in both the flexible and G-bulge regions. The C4/G18 bp changed to stable cWW bp, and this rearrangement affected the adjacent A5/A17 pair. The 6BPh A5/A17 interaction was disturbed, and the A5/A17 bp changed to tWW geometry (Figure 10). These nonisosteric rearrangements in the flexible region compromised the stability of the base triple, and all hydrogen bonds formed by G6 fluctuated considerably.

The SHAPE probing is consistent with the simulations and bioinformatics data. The SHAPE reactivity is evidently increased compared to variants 0 and 1, albeit it does not rule out formation of some folded structure with some features resembling the SR motif. Consistent with the formation of a stable cWW pair at positions 4 and 18 in the simulation, there is very little SHAPE reactivity at these positions. However, the neighboring positions 5, 6, 7, 16, and 17 show increased reactivity, consistent with the simulation revealing perturbation of the motif induced by the cWW 4/18 pairing. Thus we suggest that the SR motif formed by sequence variant 3 could be occasionally stabilized by specific tertiary context in large RNA molecules consistent with the sequence data and requires a noncanonical bp at position 3/19.

Variant 4. Variant 4 included isosteric tSH G4/A18 → U4/A18 and near-isosteric tHS A8/G15 → U8/G15 substitutions. We observed changes in the flexible and G-bulge regions. The mutated U4/A18 bp oscillated between tSH and near-tSH (ntSH in the following) geometries (Figures 9 and 10). In the ntSH geometry the U4(O2)/A18(N6) H-bond was maintained, while the U4(O2′)/A18(N6) distance was stretched to 4.5 Å. The adjacent A5/A17 bp oscillated between tHH and tHH-wm geometries (Figures 9 and 10). The changes in the flexible region were correlated (Figure 10). The G6/U7/A16 base triple was affected by the adjacent U8/G15 bp. In contrast to the tHS A/G bp, where 3 H-bonds are present, there is only one H-bond in the mutated tHS U/G bp. The U8/A15 bp had a notable propeller twist that allowed formation of the U8(O2)/A16(N6) contact replacing the U7(O2)/A16(N6) interaction. These changes occurred repeatedly during the simulation. The G15(O2′)/A16(O4′) sugar−sugar interaction fluctuated. Although the instabilities observed in G-bulge region are rather small, we need to consider that the simulation time scale is short and the structure can develop larger perturbations on longer time scales (see below). SHAPE data for this and some other variants are commented on later.

Variant 5. Variant 5 included isosteric tSH G4/A18 → C4/A18, tHH A5/A17 → A5/C17, and near-isosteric tHS A8/G15 → U8/G15 substitutions. We observed changes in the flexible and the G-bulge regions. The mutated C4/A18 bp was mostly in cWW-wm geometry (Figures 9 and 10). The adjacent A5/C17 bp adopted a tHH-wm geometry. As for variant 4, the G6/U7/A16 base triple was affected by the adjacent U8/G15 bp, which again propeller-twisted, forming the U8(O2)/A16(N6) contact instead of U7(O2)/A16(N6). We evidenced large reversible fluctuations in both the G-bulge and flexible regions. For a 10 ns long time period, the G6/U7/A16 base triple disrupted, while a non-native U8/G15/A16 base triple was formed. That results in perturbation of all bp's in the flexible region. Simulation of variant 5 confirmed that the tHS A8/G15 → U8/G15 mutation could cause sizable structural changes as was proposed for variant 4. It also explains why the near-isosteric U8/G15 bp is not observed in sequence data. The absence of two H-bonds in the 8/15 bp and poor stacking of U8 with adjacent bases result in the instability of the G-bulge region (variants 4, 5, and 10 below).

The native 4BPh G15/U7, base−sugar G6(O2′)/G15(O6), and sugar−sugar G15(O2′)/A16(O4′) contacts which disappeared at the beginning of the simulation were surprisingly restored in the perturbed geometry. As noted above, the loss of these interactions occurs in essentially all simulations and is rather due to force field imbalance than due to the substitutions. Restoration of these signature interactions in variant 5 simulation is likely due to incidental compensation of errors, where perturbation and loosening of the adjacent parts of the SR motif due to the substitution reduce the force field imbalance in the 4BPh interaction region.

Variant 6. Variant 6 contains the isosteric tHS A8/G15 → A8/A15 substitution. Existence of this variant is supported by sequences, and it has been observed twice in 3D structures (G890−2AW7 and G168−1NBS). Although the substitution in the 8/15 bp prevented formation of the 4BPh 15/7 contact, all SR bp's were stable. The 6BPh in A5/A17 and the newly formed sugar−phosphate G4(O2P)/A5(O2′) contacts fluctuated. The stable trajectory of this variant is in accord with experimental data.

Thermodynamic studies[80] show that this single substitution, which converts the strongly H-bonded AG tHS base pair to a weaker AA tHS pair, significantly lowers the melting temperature of the motif compared to variant 0, more than is expected just from the change in H-bonding in the tHS pair itself. The large change in thermal stability is consistent with loss of the crucial base−phosphate interaction that connects position 15 with positions 6 and 7 of the base triple and weakens the tHS 8/15 base pair. This is corroborated by SHAPE analysis, which shows increased reactivity at positions 6, 7, and 16, even though those bases are not mutated. Nevertheless, the SHAPE data of variant 6 keep some resemblance of the native SR motif SHAPE reactivity pattern. Overall, the data suggest that variant 6 is capable to form SR motif; however, it is thermodynamically less stable than the variant 0. The SR motifs formed by sequence variant 6 could be stabilized by specific tertiary context in large RNA molecules.

Variant 7. Variant 7 included isosteric tSH G4/A18 → C4/C18 and cSH G6/U7 → C6/U7 substitutions. The cSH C/U bp has only one H-bond mediated by a water molecule (C6(O2)/U7(O4)). The substitution in the 6/7 bp prevented formation of the 4BPh 6/16 contact. Changes in the flexible and G-bulge regions were observed. Particularly, the mutated

cSH C6/U7 bp was disrupted at the beginning of the simulation. The tWH U7/A16 bp oscillated between tWH and tSH geometries. In the flexible region the C4/C18 bp oscillated between tSH, cWW-wm, and unpaired geometries (Figures 9 and 10). The adjacent A5/A17 bp fluctuated between tHH and tHH-wm geometries (Figure 10). The 6BPh A5/A17 and the newly formed sugar−phosphate G4(O2P)/A5(O2′) contacts fluctuated considerably. Loss of the most characteristic cSH C6/U7 feature of the SR motif shows that it is not a viable SR motif variant.

Variant 8. Variant 8 included isosteric tSH G4/A18 → C4/A18, tHH A5/A17 → A5/C17, cSH G6/U7 → A6/C7, and near-isosteric tWH U7/A16 → C7/A16 substitutions. The substitution in 6/7 bp prevented formation of the 4BPh 6/16 contact. The mutated C4/A18 bp oscillated between tSH, cWW-wm, cWW_B, and unpaired geometries (Figures 9 and 10). The adjacent A5/C17 bp oscillated between tWW and unpaired geometries (Figure 10). The A6/C7 bp was disrupted at the beginning of the simulation. C7 repeatedly paired with G15 and formed a C7/G15 cWW bp. As indicated also by RMSD (Figure 7) the sequence of variant 8 appears to be unable to form a correctly folded SR motif.

Variant 9. Variant 9 included nonisosteric tHH A5/A17 → G5/A17 and isosteric tHS A8/G15 → A8/A15 substitutions. The substitution in the 8/15 bp prevented formation of the 4BPh 15/7 interaction and the substitution in the 5/17 bp prevented formation of the 6BPh 5/17 interaction. bp's in the G-bulge region were stable, but we observed changes in the flexible region. The mutated G5/A17 bp oscillated between tHH-wm and tHH_B geometries (Figures 9 and 10) and the adjacent G4/A18 bp oscillated between tSH and tSH-open geometry (Figure 10). Although the G5/A17 bp is not isosteric and considerable fluctuations in its geometry were observed, 200 ns was not sufficient to disrupt the flexible region (Figure 7).

Variant 10. Variant 10 included nonisosteric tSH G4/A18 → A4/U18, tHH A5/A17 → G5/A17, tWH U7/A16 → A7/A16, tHS A8/G15 → U8/A15, and isosteric cSH G6/U7 → A6/A7 substitutions. The substitutions prevented the formation of 4BPh 6/16, 4BPh 15/7, and 6BPh 5/17 contacts. The mutated tSH A4/U18 bp and tHS U8/A15 bp changed to cWW geometry at the beginning of the simulation. Other significant changes occurred subsequently. Particularly the A16 changed position and formed tHS A16/U8 bp. The A17 formed contact with A7. These changes were reversible and occurred several times. Finally the cWW A4/U18 bp was disrupted, and U18 left its position. The A17 relocated and stacked with the U18 out of the duplex. Variant 10 thus lost all characteristics of SR motif in a very short time period.

Prolongation of Simulations. All simulations, except of variant 10 that lost all important structural features of the SR motif in 100 ns, were prolonged to 500 ns. The majority of systems (structures 0, 1, 2, 3, 6, and 8) showed identical behavior as in the first 200 ns of simulations. RMSD developments can be found in the Supporting Information, Figure S6.

Prolongation of simulations 7 and 9 revealed moderate additional changes supporting the expectation that systems 7 and 9 will gradually lose the characteristic interactions of the SR motif on longer time scales and therefore are not able to form the SR motif. In variant 7 the tWW geometry of A5/A17 was sampled to a certain extent. Bp's 4/18 and 5/17 of variant 9 changed to cWW and tWW geometries which may (see below)

later on result in disruption of the whole motif. Thus we prolonged simulation of variant 9 to 1 μs, which resulted in complete disruption of its structure.

Major perturbations occurred in prolonged simulations of variants 4 and 5, further propagating instabilities sensed in the first 200 ns period. The flexible region formed by a combination of nucleotides not viable according to the sequence data was not stable in the simulations. Structures 4 and 5 were disrupted in several consecutive steps which took dozens of nanoseconds each: (i) the 4/18 bp changed geometry to cWW; (ii) the 5/17 bp was disrupted, and the 6BPh 5/17 interaction was lost; (iii) the other BPh interactions were lost, while the 5/17 bp changed to tWW, and (iv) all noncanonical bp's changed to cis or trans WW geometry, the second strand changed backbone geometry to A-form, and the bulged G lost all interactions. No such changes occurred in variant 2 lacking the A8/G15 bp mutation despite having the same substitutions in the flexible region as variant 5. This suggests the stabilizing effect of the A8/G15 bp influencing even the nonadjacent bp's.

The transition of multiple bp's to WW geometry resulted from instability of the 4/18 bp. The backbone conformation and stacking interactions push the 4/18 bp to the cWW geometry. Therefore, nucleotide combinations supporting stable tSH geometry in the 4/18 bp position and its inability to form the cWW geometry support the SR motif. The simulations suggest that nucleotide combinations with good potential to form stable cWW bp are not suitable for the 4/18 bp position.

SHAPE Probing. The flexibility of the RNA backbone can be monitored at single nucleotide resolution using the SHAPE technique.[74] Nucleotides that are involved in either base−base or base−phosphate interactions exhibit low SHAPE reactivities (indicating low flexibility of backbone), while nucleotides not constrained by mentioned interactions are more flexible and exhibit higher values of SHAPE reactivity. Thus the stability of the system with a known structure can be monitored by the SHAPE technique.

All studied variants of SR motif were probed with SHAPE technique. Variants that are according to sequence data viable (variants: 0, 1, 6, and potentially also variant 3, see above) are supposed to form the characteristic fold of the SR motif in the SHAPE probing experiment, and therefore the SHAPE data can be directly compared to MD simulation. Variants 0 and 1 have very low SHAPE reactivities (Supporting Information, Figure S2) in agreement with high stabilities of these variants in MD simulations. Variants 3 and 6 have somewhat increased SHAPE reactivities that are, however, still consistent with potential formation of the SR motif and in agreement with the simulation data. As discussed in detail above, we assume that variant 6 forms SR motif, while for variant 3 we suggest that the 4/18 CG pair adopts cWW geometry, and in the absence of an additional compensatory substitution the structure does not form fully a precisely shaped SR motif.

Comparison of SHAPE data with MD simulations is much more complicated for the variants that are not supported by sequence data to form SR motif and which also progressively deviate from the SR structure in the simulations. While MD simulations are too short to allow substantial refolding and simulated systems remain in the "SR-like" fold, their global fold of internal loops in SHAPE experiment could be different. Therefore, the comparison of SHAPE reactivities (flexibilities of backbone) with MD data is not straightforward. However, the

SR motif was shown to be rigid (structural data, MD simulations, SHAPE probing of variant 0, 1), and high values of SHAPE reactivity are inconsistent with the correctly folded SR motif. Strong SHAPE reactivities (Supporting Information, Figure S2) suggest that variants 4, 5, 7, 8, 9, and 10 do not form precisely shaped SR motif, and it is in good agreement with MD data.

Simulations with the Older Force Field. We have carried some simulations with older variants of the Amber RNA force field. They confirmed that the χ_{OL3} correction is essential to achieve stable RNA simulations.[36,63,64,66] Without applying this correction, the force field imbalances would disturb the simulated structures before we could get enough data to assess the base substitutions. More details can be found in the Supporting Information.

■ DISCUSSION

Bioinformatics Data. We identified 57 individual examples of SR motifs in the available set of nonredundant RNA 3D structures, which is the most complete survey of SR motifs currently available. Base pairing, base−phosphate interactions, and backbone geometries of all structures have been evaluated. Afterward, we extracted sequences aligned to these instances from large rRNA alignments in order to see the degree of support for different sequence variants.

It is not trivial to unambiguously interpret the bioinformatics data since, as we mentioned above, SR motifs at certain positions in the ribosome have different patterns of conserved nucleotides due to tertiary interactions in which they participate. Thus, there are general trends that apply to different instances of ribosomal SR motifs to different extent. For more details see the Supporting Information, Table S2.

Bioinformatics analysis confirmed (i) high conservation of nucleotides forming base triple, (ii) high conservation of geometries of 5/17, 6/7, 7/16, and 8/15 base pairs, (iii) high conservation of all base−phosphate interactions in SR motifs, and (iv) some variability in base pairing and in sequence of the flexible region. Conservation of sequence in the G-bulge region and its variability in the flexible region have been described before,[5,6] but the present study is the most complete so far.

Based on bioinformatics and structural analyses we suggest the following four classes of SR motifs: (i) perfect SR motifs with all their native features flanked by canonical bp, (ii) perfect SR motifs which, however, are flanked by additional non-canonical bp, (iii) SR motifs with some modification of the base pairing pattern, and (iv) SR-like structures with some unusual (atypical) fold, often lacking the bottom bp from the flexible region and the S-turn (cf. Supporting Information, Figure 4 for a full survey). The later motifs occur in junction regions.

MD Simulations Can Predict the Ability of Different Sequence Variants To Form a Precisely Folded SR Motif. The simulations do not allow one to straightforwardly analyze the stability of the studied systems in a thermodynamic sense. However, we suggest that increased fluctuations and deviations of the simulated structures from the native geometries can be used as an indirect indicator of the stability of the simulated systems. Fast large-scale perturbations definitely allow one to conclude that a given sequence is unlikely to form a stable SR motif.

Some 0.5 μs scale MD simulations of SR motifs with different sequences but starting from the native structure were performed to examine their dynamic properties. To reach the basic conclusions, 200 ns simulations would be sufficient. In

some cases, 500 ns data were useful for clarification. Note that unlimited prolongation of the simulations would not necessarily bring more reliable results since even the native SR motif is not perfectly described by the presently available force field variants.[55] Thus, very long simulations may start to accumulate some force field imbalances. Nevertheless, simulations in the range of 200−500 ns very well respond to base substitutions in SR motifs and may be used to complement bioinformatics analyses.

Results of MD simulations are in very good agreement with occurrence of the observed structures and with the SHAPE experiment. All three sequences that were observed to form the SR motif in X-ray structures (variants 0, 1, and 6) are stable in simulations, and all variants that are unstable in simulations exhibited strong SHAPE reactivities. Actually, the MD analysis is more reliable than our initial ranking based on an approximate bioinformatics scoring function which would rank one of the existing variants as sixth, that is, not capable to form the SR motif.

G6/U7/A16 Base Triple Is Absolutely Conserved. The most conserved element of the SR motif is the G6/U7/A16 base triple. It consists of two bp's which share nucleotide U7. The G6/U7/A16 cSH + tWH bp's combination occurs in all 57 observed occurrences (Table 1), and these nucleotides are also highly conserved in all universally conserved ribosomal SR motifs in sequences (Supporting Information, Table S2).

The coupling between the G6/U7 and the U7/A16 bp's reduces the possible base combinations that are able to form isosteric structures. Another limitation of base combinations results from the inability of adenine, cytosine, and uracil to form the 4BPh interaction[34] that is characteristic for the G-bulge region. Absolute conservation of nucleotides in the base triple can be explained considering both constraints. Although there could be isosteric or near-isosteric substitutions at the bp level, none of them would provide the 4BPh interaction[34] between nucleotides 6 and 16.

The native G/U/A base triple was stable in simulations except cases where the whole structure was perturbed due to other substitutions (variants 4 and 5).

We performed simulations of three systems with a substitution in the G6/U7/A16 base triple. All caused instabilities. Isosteric G6 → C6 substitution in variant 7 disrupted the cSH C6/U7 bp, although the perturbation did not propagate further on the present simulation time scale. Near-isosteric G6/U7/A16 → A6/C7/A16 substitution in variant 8 also disrupted the A6/C7 bp. This slightly affected the adjacent A8/G15 bp, while the flexible region seemed to be unaffected. It behaved similarly to variant 2 where the same substitutions were introduced in the flexible region while having the native triple. Adenine at position 6 could in principle form a 6BPh interaction with A16, but due to the overall geometrical requirements this interaction was not formed. G6/U7/A16 → A6/A7/A16 substitution in variant 10 is nonisosteric at the bp level. Despite the fact that it enables formation of the 6BPh interaction, it leads to decisive deterioration of the whole structure.

A8/G15 bp Conservation Rules. The tHS A8/G15 base pair is highly conserved; however, C/U and A/A tHS bp (two occurrences of each) 8/15 combinations are occasionally observed in X-ray structures. Opening of this base pair has been studied by ^{1}H-NMR exchange measurements on a millisecond time scale.[81] The A/G base combination is conserved in the sequence of all universally conserved internal

loop SR motifs of large ribosomal subunit (Supporting Information, Table S2). All three variants observed in X-ray structures are isosteric at the bp level, but only the A/G bp enables formation of the G15/U7 4BPh interaction. There are combinations of nucleotides that can form isosteric bp's, but there is no isosteric base combination that would enable formation of the 4BPh 15/7 interaction. This interaction nevertheless appears to be dispensable.

We performed simulations with three different mutations in this bp. Note that the assessment of simulations targeting this bp is complicated by the inability of the force field to stabilize the 4BPh 15/7 interaction even with the native A/G bp, probably due to imbalance in the torsional part of the force field.[55] Despite this, the native tHS A8/G15 bp was absolutely stable in all simulations; that is, the simulation finds an alternative backbone geometry without the native BPh interaction.

Variant 6 with the isosteric tHS A/G → A/A substitution showed the same behavior as the reference structure, despite abolishing the 4BPh interaction. It is fair to admit that, since the force field inaccuracy results in the loss of this interaction even with the native tHS A/G bp, the simulations might be less sensitive to fully visualize the effect of the 8/15 substitution. Nevertheless, the very stable A8/A15 simulation is consistent with occurrence of this variant in the available structures. Thermodynamics studies[80] as well as our SHAPE data show that the A/G → A/A substitution reduces the stability of the SR motif.

The near-isosteric tHS A8/G15 → U8/G15 substitution (variants 4 and 5) is not supported by experimental data and is also not supported by simulations. The WC edges of uracils 7 and 8 are oriented similarly to the Hoogsteen edge of A16 and thus mutually compete for interaction with A16. Indeed, poor stabilization of U8 at its position (only single H-bond) and location near to U7 resulted in perturbation of the G-bulge region. The decreased stability of the G-bulge region resulted in bigger structural changes in prolonged simulations of variants 4 and 5. In the case of the nonisosteric A8/G15 → U8/A15 substitution (variant 10), this bp changed geometry to cWW not compatible with the SR motif. In summary, the native tHS A8/G15 bp comprising three "internal" H-bonds and another two H-bonds with sugar of G6 and backbone of U7 is by far the most stable bp at this position. Substitution of A8/G15 by less stable bp's that are not fully isosteric leads to instability of the G-bulge region, showing good correspondence between the experimental data and simulations.

5/17 bp Conservation Rules. This bp consists of conserved adenine at position 5 and adenine or cytosine at position 17. Dominant in observed structures (∼85%) is tHH A/A followed by water-mediated tHH-wm A/C (∼10%) (Table 1). SR motifs in junction loops with complicated structures (1S72-G381 and 1S72-G464) have tHH A/A. This observation is supported by sequence data (Supporting Information, Table S2). The junction SR motifs are those lacking the adjacent 4/18 bp. Thus, geometry of the 5/17 bp is solely tHH for A5/A17 combination and preferably tHH-wm in case of A5/C17. These combinations are isosteric, and the 6BPh 5/17 interaction is present in both cases. AU bp that was also observed in X-ray structures is not isosteric with tHH AA bp but is geometrically very similar to tHH-wm A/C bp; thus its occurrence is not surprising. There is no other isosteric bp preserving the 6BPh 5/17 interaction,[34] yet there is a single instance of a GC combination in the X-ray structures (Table 1)

which, however, occurs in rather unusual G485−1FJG SR loop structure (Supporting Information, Figure S4c).

An important finding of the simulations not apparent from the bioinformatics data is profound coupling between structural dynamics of the 5/17 and 4/18 bp's (Figure 10). Normally, the A5/A17 bp sampled tHH and tHH-wm geometries. tWW A5/A17 geometry that is very different from tHH was sampled in variants 3, 4, and 7 where the adjacent 4/18 pair rearranged to a stable cWW bp. The A5/C17 base combination sampled tHH-wm, tWW, and unpaired geometries. The 6BPh 5/17 interaction was disrupted in tWW geometry. However, note that in variants with A5/C17 also the 4/18 bp was mutated to C4/A18 (variants 2, 5, and 8).

The nonisosteric tHH G5/A17 bp (variants 9 and 10) does not preserve the BPh 5/17 interaction. tHH-wm (similar to cWH G/A) and tHH_B geometries (Figure 9) were sampled. These geometries of base pairing are not similar and transition between them requires large movement of the bases. This is indicated by the RMSD graph of variant 9 (Figure 7) where lower regions correspond to the tHH-wm geometry of the 5/17 bp. The fact that the stability of the flexible region comprising the G5/A17 bp is lower in comparison with the A5/A17 or A5/C17 bp's corresponds to experimental data that do not suggest viability of the G5/A17 bp.

4/18 bp Should Not Support the cWW Arrangement. The sequence and geometry of the last noncanonical bp of SR motif is quite variable. In structural data nine different 4/18 combinations of bases were observed (Table 1). The most frequent geometries are tSH and ntSH-wm, but also instances of cWW and tWW were observed, demonstrating that this bp is indeed flexible (Table 1). Some base combinations were observed in two different geometries in X-ray structures. In addition, a closer inspection of the structures shows that nucleotides 4 and 18 are typically not paired in junction loop SR motifs, which also lack the S-turn. Further, all three observed instances of tWW U4/U18 bp are in structures near junctions, where no helix is formed under the SR motif. So, the tWW U4/U18 combination appears rather as deviation from optimal SR motif. Finally, two of the three instances of cWW geometry are seen in structures with another bulged base below the flexible region (G485−1FJG and G300−3G78). Moreover, sequence data showed that cWW 4/18 bp is viable only in structures with additional flanking noncanonical bp's. The structure G75−1UN6 occurring in a complex of 5S rRNA with part of TFIIIA protein contains cWW UU bp and is an exception, because it has cWW 4/18 bp combined with canonical 3/19 bp. However, this sequence variant is not supported by 5S rRNA sequence data, so its significance is unclear. Considering all of these facts we suggest that despite the variability of the 4/18 base pairing the tSH and ntSH-wm are those geometries of 4/18 bp that intrinsically support local SR motifs. The other geometries and sequence variants should be considered as exceptions and are associated with some additional structural features not present in perfect SR motifs.

Several variants with isosteric mutations in 4/18 bp were simulated. Only G4/A18 (variant 0 and 6) bp remained in tSH geometry. The C4/C18 bp in X-ray structures has variable geometries, and indeed these were sampled in MD simulation of variants 1 and 7. Simulation behavior of G/A and C/C 4/18 bp's thus nicely agrees with sequence and structural data. The isosteric G4/A18 → C4/A18 and A5/A17 → A5/C17 double substitutions were introduced in variants 2, 5, and 8. This combination decreased stability of the flexible region, and the

14315

dx.doi.org/10.1021/jp408530w | J. Phys. Chem. B 2013, 117, 14302−14319

tWW A5/C17 and cWW_B C4/A18 geometries were sampled (Figure 9). Nonisosteric C4/G18 (variant 3) and A4/U18 (variant 10) bp's adopt canonical cWW geometry not compatible with SR motif, which further disrupts the adjacent 5/17 bp confirming that these substitutions are not viable. Note that sequence data give some support for the variant 3 to be potentially able to form the SR motif (Supporting Information, Table S1), but closer inspection of sequences showed the presence of noncanonical 3/19 bp in such cases (data not shown).

As discussed above, despite the fact that CA and AA combinations in position 4/18 could potentially form protonated cWW base pairs, such base pairing does not appear to be supported by the overall context of SR motifs.

SR Motif Consensus Model. The simulations revealed an important aspect of the 4/18 base pair that was not apparent from the bioinformatics and structural data. The simulations show that the 4/18 bp tends to adopt cWW or cWW like geometry in simulations which competes with the tSH geometry. The population of the cWW geometries depends on a combination of bases forming the 4/18 bp and on stability of the adjacent 5/17 bp. A stable A5/A17 bp supports the tSH geometry of the 4/18 pair. While in case of the 5/17 and 8/15 bp's and the 6/7/16 base triple, it was obvious that the BPh interactions constrain possible sequences; it was not straightforward to find what constrains the 4/18 bp besides isostericity. MD simulations clarified that the inability to form stable cWW geometry is an important factor in case of the 4/18 bp. The ability of bp's to adopt cWW (4/18 bp) and tWW (5/17bp) geometries correlates with perturbation of the SR motif. Indeed, structural data reveal only such 4/18 base combinations that are unable to form stable cWW bp (or have only small propensity to form this geometry). The most populated 4/18 combinations are not able to form cWW bp without protonization (AC, CC) or are very stable in tSH geometry (GA). Considering dynamical behavior of the simulated variants supplemented by sequence and structural data, we propose the consensus sequence of the internal loop SR motif (Figure 11). The most stable variant of SR motif is
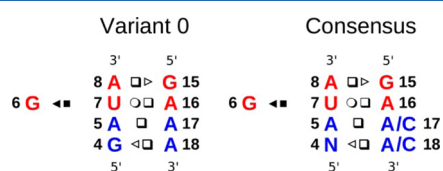

Figure 11. Consensus sequence of the SR motif based on the present data. The most stable variant of SR motif is left and the suggested consensus SR motif right. N means any nucleotide that is able to form a tSH bp with the nucleotide at position 18 but is unable to form a stable cWW geometry.

unambiguously variant 0. Reasons for conservation of the G-bulge region (A8/G15 + G6/U7/A16) are described above. While variants with different sequence of the flexible region are intrinsically less stable, they are often useful to form tertiary interactions. Thus, there is some sequence variability in this region. There can be either A or C at positions 17 and 18. A at position 5 is required to form the 6BPh interaction. Finally, nucleotide 4 must be able to form a stable tSH bp and should not be able to form a competitive cWW geometry.

CONCLUSIONS

We identified 57 instances of the SR motif in a nonredundant subset of the RNA 3D structures and divided them into four groups (Supporting Information, Figure S4).

We extracted sequences aligned to these instances from large rRNA alignments to see the degree of support for different sequence variants. We used a simple scoring scheme based on isostericity to prepare 10 sequence variants of the SR motif with a highly variable degree of expected compatibility with the perfect SR motif (Figure 6).

MD simulations were applied to assess the structural stability of these SR motif variants. The MD simulations show stable trajectories for SR variants that correspond to existing SR motifs. In contrast, nonisosteric base substitutions lead to unstable structures, but so did isosteric substitutions which are unable to make base−phosphate interactions observed in 3D structures. The outcome of the 0.5 μs MD simulation runs thus very well agrees with the structural and sequence data.

The results of SHAPE probing that was performed for all studied SR motif variants agreed with MD data. The simulation technique reflects properties of the different SR motif variants better than the simple bioinformatics scoring function that we used for the initial scoring.

MD simulation is an established theoretical technique capable to address many issues of structural dynamics and physical chemistry of RNA molecules.[35−51] Obviously, the method has some limitations stemming from the affordable simulation time scales[82,83] and force field approximations.[64,84,85] Our present study nevertheless suggests that the MD simulation technique can very efficiently assess the impact of base substitutions in known RNA motifs. Therefore, the MD simulation technique may be used as a robust complement to bioinformatics techniques analyzing noncanonical RNA building blocks and their conservation patterns.[4−6,86−88] Note that, although the SR motif is a small molecule, it has quite a complicated structure which is challenging for accurate force field description. Still, regarding the assessment of different base substitutions, the MD results are unambiguous. The χ_{OL3}[63,64] force field correction is essential to achieve stable RNA simulations. Without applying this correction, the force field imbalances would deteriorate the simulated structures before we could assess the effects of the base substitutions.

Based on the simulations, we identify coupling between structural dynamics of the two base pairs of the flexible region and suggest that the inability of the 4/18 base pair to form stable cWW arrangement is an important hitherto unknown sequence constraint for proper folding of the SR motif. We suggest a refined consensus scheme for the SR motif (Figure 11).

ASSOCIATED CONTENT

Ⓢ Supporting Information

Description of initial scoring scheme for sequence variants; summary of simulations of selected variants done with older force fields; description of SR group identifiers; tables and figures. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: sponer@ncbr.muni.cz. Tel.: +420 541 517 133. Fax: +420 541 212 179.

Notes

■ REFERENCES

(1) Moazed, D.; Stern, S.; Noller, H. F. Rapid Chemical Probing of Conformation in 16 S Ribosomal RNA and 30 S Ribosomal Subunits Using Primer Extension. *J. Mol. Biol.* **1986**, *187*, 399−416.

(2) Michel, F.; Westhof, E. Modelling of the Three-Dimensional Architecture of Group I Catalytic Introns Based on Comparative Sequence Analysis. *J. Mol. Biol.* **1990**, *216*, 585−610.

(3) Leontis, N. B.; Westhof, E. Analysis of RNA Motifs. *Curr. Opin. Struct. Biol.* **2003**, *13*, 300−308.

(4) Lescoute, A.; Leontis, N. B.; Massire, C.; Westhof, E. Recurrent Structural RNA Motifs, Isostericity Matrices and Sequence Alignments. *Nucleic Acids Res.* **2005**, *33*, 2395−2409.

(5) Leontis, N. B.; Westhof, E. A Common Motif Organizes the Structure of Multi-Helix Loops in 16 S and 23 S Ribosomal RNAs. *J. Mol. Biol.* **1998**, *283*, 571−583.

(6) Leontis, N.; Stombaugh, J.; Westhof, E. Motif Prediction in Ribosomal RNAs Lessons and Prospects for Automated Motif Prediction in Homologous RNA Molecules. *Biochemie* **2002**, *84*, 961−973.

(7) Gutell, R. R.; Schnare, M. N.; Gray, M. W. A Compilation of Large Subunit (23S- and 23S-Like) Ribosomal RNA Structures. *Nucleic Acids Res.* **1992**, *20*, 2095−2109.

(8) Endo, Y.; Mitsui, K.; Motizuki, M.; Tsurugi, K. The Mechanism of Action of Ricin and Related Toxic Lectins on Eukaryotic Ribosomes. The Site and the Characteristics of the Modification in 28 S Ribosomal RNA Caused by the Toxins. *J. Biol. Chem.* **1987**, *262*, 5908−5912.

(9) Qin, S.; Zhou, H.-X. Dissection of the High Rate Constant for the Binding of a Ribotoxin to the Ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6974−6979.

(10) Lacadena, J.; Álvarez-García, E.; Carreras-Sangrà, N.; Herrero-Galán, E.; Alegre-Cebollada, J.; García-Ortega, L.; Oñaderra, M.; Gavilanes, J. G.; Martínez Del Pozo, Á. Fungal Ribotoxins: Molecular Dissection of a Family of Natural Killers. *FEMS Microbiol. Rev.* **2007**, *31*, 212−237.

(11) Schindler, D. G.; Davies, J. E. Specific Cleavage of Ribosomal RNA Caused by Alpha Sarcin. *Nucleic Acids Res.* **1977**, *4*, 1097−1110.

(12) Endo, Y.; Tsurugi, K. RNA N-Glycosidase Activity of Ricin A-Chain. Mechanism of Action of the Toxic Lectin Ricin on Eukaryotic Ribosomes. *J. Biol. Chem.* **1987**, *262*, 8128−8130.

(13) Glück, A.; Endo, Y.; Wool, I. G. Ribosomal RNA Identity Elements for Ricin A-Chain Recognition and Catalysis: Analysis with Tetraloop Mutants. *J. Mol. Biol.* **1992**, *226*, 411−424.

(14) Munishkin, A.; Wool, I. G. The Ribosome-In-Pieces: Binding of Elongation Factor EF-G to Oligoribonucleotides that Mimic the Sarcin/Ricin and Thiostrepton Domains of 23S Ribosomal RNA. *Proc. Natl. Acad. Sci.* **1997**, *94*, 12280−12284.

(15) Pérez-Cañadillas, J. M.; Santoro, J.; Campos-Olivas, R.; Lacadena, J.; Martínez Del Pozo, A.; Gavilanes, J. G.; Rico, M.; Bruix, M. The Highly Refined Solution Structure of the Cytotoxic Ribonuclease A-Sarcin Reveals the Structural Requirements for Substrate Recognition and Ribonucleolytic Activity. *J. Mol. Biol.* **2000**, *299*, 1061−1073.

(16) Yang, X.; Gérczei, T.; Glover, L.; Correll, C. C. Crystal Structures of Restrictocin−Inhibitor Complexes with Implications for

RNA Recognition and Base Flipping. *Nat. Struct. Biol.* **2001**, *8*, 968−973.

(17) Correll, C. C.; Beneken, J.; Plantinga, M. J.; Lubbers, M.; Chan, Y.-L. The Common and the Distinctive Features of the Bulged-G Motif Based on a 1.04 Å Resolution RNA Structure. *Nucleic Acids Res.* **2003**, *31*, 6806−6818.

(18) Hausner, T.-P.; Atmadja, J.; Nierhaus, K. H. Evidence That the G2661 Region of 23S Rrna Is Located at the Ribosomal Binding Sites of Both Elongation Factors. *Biochemie* **1987**, *69*, 911−923.

(19) Moazed, D.; Robertson, J. M.; Noller, H. F. Interaction of Elongation Factors EF-G and EF-Tu with a Conserved Loop in 23S RNA. *Nature* **1988**, *334*, 362−364.

(20) Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. The Kink-Turn: A New RNA Secondary Structure Motif. *EMBO J.* **2001**, *20*, 4214−4221.

(21) Chan, Y.-L.; Correll, C. C.; Wool, I. G. The Location and the Significance of a Cross-Link between the Sarcin/Ricin Domain of Ribosomal RNA and the Elongation Factor-G. *J. Mol. Biol.* **2004**, *337*, 263−272.

(22) Lancaster, L.; Lambert, N. J.; Maklan, E. J.; Horan, L. H.; Noller, H. F. The Sarcin−Ricin Loop of 23S Rrna Is Essential for Assembly of the Functional Core of the 50S Ribosomal Subunit. *RNA* **2008**, *14*, 1999−2012.

(23) Macbeth, M. R.; Wool, I. G. The Phenotype of Mutations of G2655 in the Sarcin/Ricin Domain of 23 S Ribosomal RNA. *J. Mol. Biol.* **1999**, *285*, 965−975.

(24) Chan, Y.-L.; Sitikov, A. S.; Wool, I. G. The Phenotype of Mutations of the Base-Pair C2658·G2663 That Closes the Tetraloop in the Sarcin/Ricin Domain of Escherichia Coli 23 S Ribosomal RNA. *J. Mol. Biol.* **2000**, *298*, 795−805.

(25) Sarver, M.; Zirbel, C. L.; Stombaugh, J.; Mokdad, A.; Leontis, N. B. FR3D: Finding Local and Composite Recurrent Structural Motifs in RNA 3D Structures. *J. Math. Biol.* **2008**, *56*, 215−252.

(26) Petrov, A. I.; Zirbel, C. L.; Leontis, N. B. Webfr3d—A Server for Finding, Aligning and Analyzing Recurrent RNA 3D Motifs. *Nucleic Acids Res.* **2011**, *39*, W50−W55.

(27) Garst, A. D.; Héroux, A.; Rambo, R. P.; Batey, R. T. Crystal Structure of the Lysine Riboswitch Regulatory Mrna Element. *J. Biol. Chem.* **2008**, *283*, 22347−22351.

(28) Serganov, A.; Huang, L.; Patel, D. J. Structural Insights into Amino Acid Binding and Gene Control by a Lysine Riboswitch. *Nature* **2008**, *455*, 1263−1267.

(29) Garst, A. D.; Porter, E. B.; Batey, R. T. Insights into the Regulatory Landscape of the Lysine Riboswitch. *J. Mol. Biol.* **2012**, *423*, 17−33.

(30) Correll, C. C.; Wool, I. G.; Munishkin, A. The Two Faces of the Escherichia Coli 23 S Rrna Sarcin/Ricin Domain: The Structure at 1.11 Å Resolution. *J. Mol. Biol.* **1999**, *292*, 275−287.

(31) Leontis, N. B.; Stombaugh, J.; Westhof, E. The Non-Watson-Crick Base Pairs and Their Associated Isostericity Matrices. *Nucleic Acids Res.* **2002**, *30*, 3497−3531.

(32) Stombaugh, J.; Zirbel, C. L.; Westhof, E.; Leontis, N. B. Frequency and Isostericity of RNA Base Pairs. *Nucleic Acids Res.* **2009**, *37*, 2294−2312.

(33) Šponer, J.; Šponer, J. E.; Petrov, A. I.; Leontis, N. B. Quantum Chemical Studies of Nucleic Acids: Can We Construct a Bridge to the RNA Structural Biology and Bioinformatics Communities? *J. Phys. Chem. B* **2010**, *114*, 15723−15741.

(34) Zirbel, C. L.; Sponer, J. E.; Sponer, J.; Stombaugh, J.; Leontis, N. B. Classification and Energetics of the Base-Phosphate Interactions in RNA. *Nucleic Acids Res.* **2009**, *37*, 4898−4918.

(35) Spackova, N.; Sponer, J. Molecular Dynamics Simulations of Sarcin-Ricin Rrna Motif. *Nucleic Acids Res.* **2006**, *34*, 697−708.

(36) Sklenovský, P.; Florová, P.; Banáš, P.; Réblová, K.; Lankaš, F.; Otyepka, M.; Šponer, J. Understanding RNA Flexibility Using Explicit Solvent Simulations: The Ribosomal and Group I Intron Reverse Kink-Turn Motifs. *J. Chem. Theory Comput.* **2011**, *7*, 2963−2980.

(37) Réblová, K.; Fadrná, E.; Sarzynska, J.; Kulinski, T.; Kulhánek, P.; Ennifar, E.; Koča, J.; Šponer, J. Conformations of Flanking Bases in

HIV-1 RNA DIS Kissing Complexes Studied by Molecular Dynamics. *Biophys. J.* **2007**, *93*, 3932−3949.

(38) Réblová, K.; Špačková, N.; Štefl, R.; Csaszar, K.; Koča, J.; Leontis, N. B.; Šponer, J. Non-Watson-Crick Basepairing and Hydration in RNA Motifs: Molecular Dynamics of 5S Rrna Loop E. *Biophys. J.* **2003**, *84*, 3564−3582.

(39) Ditzler, M.; Otyepka, M.; Sponer, J.; Walter, N. Molecular Dynamics and Quantum Mechanics of RNA: Conformational and Chemical Change We Can Believe In. *Acc. Chem. Res.* **2010**, *43*, 40−47.

(40) Orozco, M.; Noy, A.; Pérez, A. Recent Advances in the Study of Nucleic Acid Flexibility by Molecular Dynamics. *Curr. Opin. Struct. Biol.* **2008**, *18*, 185−193.

(41) Sanbonmatsu, K. Y. Computational Studies of Molecular Machines: The Ribosome. *Curr. Opin. Struct. Biol.* **2012**, *22*, 168−174.

(42) Goh, G. B.; Knight, J. L.; Brooks, C. L. Ph-Dependent Dynamics of Complex RNA Macromolecules. *J. Chem. Theory Comput.* **2013**, *9*, 935−943.

(43) Wolf, A.; Baumann, S.; Arndt, H.-D.; Kirschner, K. N. Influence of Thiostrepton Binding on the Ribosomal Gtpase Associated Region Characterized by Molecular Dynamics Simulation. *Bioorg. Med. Chem.* **2012**, *20*, 7194−7205.

(44) Jung, S.; Schlick, T. Candidate RNA Structures for Domain 3 of the Foot-And-Mouth-Disease Virus Internal Ribosome Entry Site. *Nucleic Acids Res.* **2013**, *41*, 1483−1495.

(45) Romanowska, J.; Mccammon, J. A.; Trylska, J. Understanding the Origins of Bacterial Resistance to Aminoglycosides through Molecular Dynamics Mutational Study of the Ribosomal A-Site. *PLoS Comput. Biol.* **2011**, *7*, E1002099.

(46) Do, T. N.; Carloni, P.; Varani, G.; Bussi, G. RNA/Peptide Binding Driven by Electrostatics—Insight from Bidirectional Pulling Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 1720−1730.

(47) Caulfield, T.; Devkota, B. Motion of Transfer RNA from the A/T State into the A-Site Using Docking and Simulations. *Proteins: Struct., Funct., Bioinf.* **2012**, *80*, 2489−2500.

(48) White, K. H.; Orzechowski, M.; Fourmy, D.; Visscher, K. Mechanical Unfolding of the Beet Western Yellow Virus −1 Frameshift Signal. *J. Am. Chem. Soc.* **2011**, *133*, 9775−9782.

(49) Veeraraghavan, N.; Ganguly, A.; Chen, J.-H.; Bevilacqua, P. C.; Hammes-Schiffer, S.; Golden, B. L. Metal Binding Motif in the Active Site of the HDV Ribozyme Binds Divalent and Monovalent Ions. *Biochemistry* **2011**, *50*, 2672−2682.

(50) Sethaphong, L.; Singh, A.; Marlowe, A. E.; Yingling, Y. G. The Sequence of HIV-1 TAR RNA Helix Controls Cationic Distribution. *J. Phys. Chem. C* **2010**, *114*, 5506−5512.

(51) Lee, T.-S.; Giambaşu, G. M.; Harris, M. E.; York, D. M. Characterization of the Structure and Dynamics of the HDV Ribozyme in Different Stages along the Reaction Path. *J. Phys. Chem. Lett.* **2011**, *2*, 2538−2543.

(52) Klein, D. J.; Moore, P. B.; Steitz, T. A. The Roles of Ribosomal Proteins in the Structure Assembly, and Evolution of the Large Ribosomal Subunit. *J. Mol. Biol.* **2004**, *340*, 141−177.

(53) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; et al. RNA Backbone: Consensus All-Angle Conformers and Modular String Nomenclature (An RNA Ontology Consortium Contribution). *RNA* **2008**, *14*, 465−481.

(54) Lu, X.-J.; Olson, W. K.; Bussemaker, H. J. The RNA Backbone Plays a Crucial Role in Mediating the Intrinsic Stability of the Gpu Dinucleotide Platform and the Gpupa/Gpa Miniduplex. *Nucleic Acids Res.* **2010**, *38*, 4868−4876.

(55) Mládek, A.; Šponer, J. E.; Kulhánek, P.; Lu, X.-J.; Olson, W. K.; Šponer, J. Understanding the Sequence Preference of Recurrent RNA Building Blocks Using Quantum Chemistry: The Intrastrand RNA Dinucleotide Platform. *J. Chem. Theory Comput.* **2011**, *8*, 335−347.

(56) Duarte, C. M.; Wadley, L. M.; Pyle, A. M. RNA Structure Comparison, Motif Search and Discovery Using a Reduced Representation of RNA Conformational Space. *Nucleic Acids Res.* **2003**, *31*, 4755−4761.

(57) Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F. O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res.* **2012**, *41*, D590−D596.

(58) Desantis, T. Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E. L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G. L. Greengenes, A Chimera-Checked 16S Rrna Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72*, 5069−5072.

(59) Petrov, A. I.; Zirbel, C. L.; Leontis, N. B. Automated Classification of RNA 3D Motifs and the RNA 3D Motif Atlas. *RNA* **2013**, DOI: 10.1261/rna.039438.113.

(60) Cannone, J. J.; Subramanian, S.; Schnare, M. N.; Collett, J. R.; D'Souza, L. M.; Du, Y.; Feng, B.; Lin, N.; Madabusi, L. V.; Müller, K. M.; et al. The Comparative RNA Web (CRW) Site: An Online Database of Comparative Sequence and Structure Information for Ribosomal, Intron, and Other Rnas. *BMC Bioinf.* **2002**, *3*, 2.

(61) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(62) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.

(63) Zgarbová, M.; Otyepka, M.; Sponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E., 3rd; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.

(64) Banáš, P.; Hollas, D.; Zgarbová, M.; Jurečka, P.; Orozco, M.; Cheatham, T. E.; Šponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6*, 3836−3849.

(65) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(66) Beššeová, I.; Banáš, P.; Kührová, P.; Košinová, P.; Otyepka, M.; Šponer, J. Simulations of A-RNA Duplexes. The Effect of Sequence, Solute Force Field, Water Model, and Salt Concentration. *J. Phys. Chem. B* **2012**, *116*, 9899−9916.

(67) Case, D.; Darden, T.; Cheatham, T. E., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R.; Zhang, W.; et al. *AMBER 10*; University of California: San Francisco, CA, 2008.

(68) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(69) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(70) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, 327−341.

(71) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(72) Eterna Home Page; http://Eterna.Cmu.Edu/Web/ (accessed July 9, 2013).

(73) Das, R. Stanford University, Stanford, CA. *Personal Communication*, 2013.

(74) Deigan, K. E.; Li, T. W.; Mathews, D. H.; Weeks, K. M. Accurate SHAPE-Directed RNA Structure Determination. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 97−102.

(75) Davis, I. W.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. Molprobity: Structure Validation and All-Atom Contact Analysis for Nucleic Acids and Their Complexes. *Nucleic Acids Res.* **2004**, *32*, W615−W619.

(76) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B.; Snoeyink, J.; Richardson, J. S.; et al. Molprobity: All-Atom Contacts and Structure Validation for Proteins and Nucleic Acids. *Nucleic Acids Res.* **2007**, *35*, W375−W383.

(77) Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. Molprobity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D: Biol. Crystallogr.* **2010**, *66*, 12−21.

(78) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (33−38), 27−28.

(79) Williams, T.; Kelley, C. *Gnuplot 4.4: An Interactive Plotting Program*, 2010; http://www.gnuplot.info/.

(80) Khisamutdinov, E.; Sweeney, B.; Leontis, N. Context-Sensitivity of Isosteric Substitutions of Non-Watson-Crick Basepairs in Recurrent RNA 3D Motifs. 2013, in preparation.

(81) Chen, C.; Jiang, L.; Michalczyk, R.; Russu, I. M. Structural Energetics and Base-Pair Opening Dynamics in Sarcin-Ricin Domain RNA. *Biochemistry* **2006**, *45*, 13606−13613.

(82) Kührová, P.; Banáš, P.; Best, R. B.; Šponer, J.; Otyepka, M. Computer Folding of RNA Tetraloops? Are We There Yet? *J. Chem. Theory Comput.* **2013**, *9*, 2115−2125.

(83) Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Reliable Oligonucleotide Conformational Ensemble Generation in Explicit Solvent for Force Field Assessment Using Reservoir Replica Exchange Molecular Dynamics Simulations. *J. Phys. Chem. B* **2013**, *117*, 4014−4027.

(84) Banáš, P.; Sklenovský, P.; Wedekind, J. E.; Šponer, J.; Otyepka, M. Molecular Mechanism of Preq1 Riboswitch Action: A Molecular Dynamics Study. *J. Phys. Chem. B* **2012**, *116*, 12721−12734.

(85) Faustino, I.; Pérez, A.; Orozco, M. Toward a Consensus View of Duplex RNA Flexibility. *Biophys. J.* **2010**, *99*, 1876−1885.

(86) Bida, J. P.; Das, R. Squaring Theory with Practice in RNA Design. *Curr. Opin. Struct. Biol.* **2012**, *22*, 457−466.

(87) Sripakdeevong, P.; Beauchamp, K.; Das, R. Why Can't We Predict RNA Structure at Atomic Resolution? In *RNA 3D Structure Analysis And Prediction*; Leontis, N., Westhof, E., Eds.; Nucleic Acids And Molecular Biology; Springer: Berlin, 2012; Vol. 27, pp 43−65.

(88) Zhong, C.; Zhang, S. Clustering RNA Structural Motifs in Ribosomal Rnas Using Secondary Structural Alignment. *Nucleic Acids Res.* **2011**, *40*, 1307−1317.