Supporting information for: Molecular Dynamics Simulations Reveal an Interplay between SHAPE Reagent Binding and RNA Flexibility

Vojtěch Mlýnský, Giovanni Bussi*

Scuola Internazionale Superiore di Studi Avanzati, SISSA, via Bonomea 265, 34136 Trieste, Italy

E-mail: vmlynsky@sissa.it,bussi@sissa.it

Contents

1 Structure preparation and simulation overview

The initial structure of RNA core from a signal recognition particle (SRP-RNA) was taken from the protein-RNA complex determined by X-ray crystallography at 1.8 Å resolution (PDB ID 1DUL)[S1]. The SRP-RNA structure (41 nucleotides in total) was used for four different molecular dynamics (MD) simulations, i.e., (i) one 500 ns-long classical MD simulation, (ii) one 1.8 μs-long enhanced sampling simulation with SRP-RNA alone, where we biased sugar-pucker transitions, and (iii) two similar enhanced sampling simulations (each 3.6 μs-long) with benzoyl cyanide (BzCN) SHAPE reagent, where all RNA atoms were kept frozen in one case. Additionally, we performed three enhanced sampling simulations involving SHAPE reagents and much smaller RNA systems, i.e., acgccGUAAggcgu (GNRA) tetraloop with BzCN (1.4 μs-long simulation) and two nucleotide analogs (3'-5'-cyclic-adenosine monophosphate (cAMP) and 3'-5'-cyclic-cytosine monophosphate (cCMP)) both with larger N-methylisatoic anhydride (NMIA) SHAPE reagent (each 500 ns-long). The initial structure of GNRA tetraloop was taken from protein:RNA complex (PDB ID 4AL6)[S2] and some nucleotides were mutated in order to analyze the same sequence as in SHAPE control datasets[S3].

2 Simulation protocols

All MD simulations were performed in GROMACS 4.6.7[S4] in combination with PLUMED 2.2[S5]. RNA was described using the standard Amber force field.[S6–S9] All simulated systems were neutralized by monovalent K^+ ions[S10] and immersed in a truncated dodecahedral box with an at least 10 Å thick layer of TIP3P water molecules all around the RNA solute. Bonds were constrained using the LINCS algorithm,[S11] allowing for a time step of 2 fs. The particle-mesh Ewald algorithm[S12] was used for long-range electrostatic interactions with a cutoff distance of 1 nm. Simulations were performed at temperature T = 298 K using a stochastic velocity rescale thermostat[S13]. Parameters for BzCN, NMIA, cAMP, and cCMP

molecules were derived according to the standard scheme by Cornell et al.[S6,S14]

We used two different setups for enhanced sampling simulations of SRP-RNA. Initially, bias-exchange simulations[S15] were used to enhance transitions of sugar puckers of SRP-RNA nucleotides. One sugar pucker was biased (as a pseudorotation dihedral Z_x) in each replica and all replicas were run simultaneously. Each replica was equilibrated for 1 ns and C2'-*endo*/C3'-*endo* sugar-pucker populations of SRP-RNA nucleotides were analyzed during a total simulation time of 1.8 μs (36 replicas $\times$50 ns per each replica). Subsequently, reactivities were estimated from the population of C2'-*endo* pucker of each nucleotide (Table S1).

Enhanced sampling simulations of SRP-RNA with BzCN reagent used the modified version of bias-exchange metadynamics that is in detail described elsewhere.[S16] We used the simplified version of that protocol, where (i) only one collective variable (CV), i.e., the distance between reactive carbon from BzCN (C*, see Figure 1 in the maintext) and O2' oxygen from 2'-OH group of particular nucleotide, was biased in each replica, and (ii) without the necessity to introduce additional bias potentials that were required to avoid the competition of different binding sites in the same replica as described by Cunha and Bussi in the original paper[S16]. Thus, one distance was biased at a time in each replica and all replicas were run simultaneously. However, replica specific restraints were used (see section Description of applied restraints and specific usage of reagents), making this setup slightly different from that of a standard bias-exchange simulation.[S15] Each replica was equilibrated with 10 ns-long MD run, where the BzCN reagent was allowed to move freely within the simulation box. Subsequently, we analyzed the relative binding rates of 36 nucleotides from SRP-RNA during a total simulation time of 3.6 μs (36 replicas $\times$100 ns). All states with the distance between BzCN(C*)...2'-OH(nucleotide) lower than 4.0 Å (binding cut-off) were considered as possible reactive conformations for particular nucleotide.

Simulations with smaller RNA motif (GNRA tetraloop) served for testing purposes, involved a smaller number of replicas (either 8 or 10 nucleotides were analyzed), and both BzCN and NMIA reagents were used (results not shown). In one specific simulation with

GNRA tetraloop and BzCN reagent, we applied the final simulation protocol (i.e., similar for SRP-RNA, 10 replicas $\times$140 ns = 1.4 μs in total) but immersed the system in a large box with at least 20 Å thick layer of TIP3P water molecules in order to estimate the absolute binding rates (see section Calculated reactivities and binding rates). Enhanced sampling simulations with nucleotide analogs (cAMP, cCMP) and NMIA reagent used slightly different setup. We performed a single 500 ns-long simulation, where we biased two CV's, i.e., (i) the distance between reactive carbon from NMIA and O2' oxygen from 2'-OH group, and (ii) χ dihedral angle of the nucleobase (O4'-C1'-N9-C8 and O4'-C1'-N1-C6 for cAMP and cCMP, respectively). The additional bias to χ dihedral in those simulations was applied in order to increase the number of *syn/anti* flips that might affect the sugar pucker of the ribose ring.[S17] GROMACS input files for all systems are available on github (https://github.com/srnas/shape-md).

3 Description of applied restraints and specific usage of reagents

We used four different restraints in SRP-RNA simulations, where two of them were specific for enhanced sampling simulations with SHAPE reagents. Firstly, we kept the BzCN reagent close to RNA structure (within the shell of 10 Å), which increased exchanges among replicas and allowed to observe multiple binding and unbinding events within each replica. The restraint is applied on the distance between C* and the O2' oxygen used for metadynamics in each replica. One of the simulations with GNRA tetraloop and BzCN reagent was done in much bigger box, where the reagent was allowed to spend a significant time away from any RNA interaction, i.e., to explore whole simulation box. The obtained results from that unrestrained simulation were then compared with restrained one, i.e., performed in smaller box, and the estimated reactivities were very similar. The unrestrained simulation revealed higher statistical error due to lower number of binding/unbinding events (data not shown). Similarly, estimated reactivities from testing simulations with GNRA tetraloop and larger

NMIA reagent showed higher statistical error than those with smaller BzCN reagent. In addition, NMIA was identified to have higher tendency to induce irreversible structural changes in RNA structure (see description of fourth restrain below). Hence, we opted to use the BzCN for the majority of MD simulations (except for the specific analysis of small nucleotide analogs) and keep the reagent close to RNA during entire simulation timescale. The restraint kept the BzCN close to the SRP-RNA system and we were able to benefit from the better convergency (see Figures S7, S11 and Table S1 for statistical errors) and computational efficiency by using much smaller water box (the solvated SRP-RNA system contained ∼50 000 atoms in total).

The second restraint kept the reagent away from the last base pair of SRP-RNA. As expected, the external molecule, i.e., SHAPE reagent, had a higher tendency to get stacked with the terminal nucleotides from RNA stem. Thus, the restraint prevented conformations where any heavy atom of BzCN reagent was situated closer than ∼8 Å from any heavy atom of nucleotides forming the terminal base pair. We note that both the terminal as well as the penultimate base pairs were excluded from the analysis of binding rates (SRP-RNA system contained 41 nucleotides and only 36 of them were analyzed). Unfortunately, we identified that the reactivity of third base pair from the end (G3-C39, from which only one nucleotide (G3) was considered for the reactivity analysis) was also affected because the reactive conformations, i.e., the reagent in close proximity from the particular 2'-OH group, were occurring not only from the specific replica biasing the particular distance but also from replicas of neighboring and from the opposite nucleotide. In other worlds, the reactivity of G3 nucleotide was affected by simulation settings and expected to be much lower for two reasons: (i) nucleotides from one of neighboring base pairs were excluded from the analysis and (ii) the opposite nucleotide (C39, see Figure 1 in the maintext) was not considered for the analysis as well as even number of analyzed nucleotides was required due to computational efficiency. Indeed, the reactivity of G3 was significantly hampered (see Figure S7) due to its 'unique' position and, ultimately, this nucleotide was excluded from subsequent calculation

of correlation coefficients.

Third and fourth restraints were applied in both classical and enhanced sampling simulations of SRP-RNA. Third restraint assured that the first and the last nucleotide, i.e., the terminal canonical C1-G41 base pair, preserved its Watson-Crick hydrogen bond contacts. A base-pair openings (breathing) was not allowed as it could potentially destabilize the overall SRP-RNA structure. Lastly and most importantly, we were forced to prevent any large conformation changes within whole SRP-RNA structure by including another restraint that allowed only local structural changes and fluctuations of RNA around its native conformation. Initial tests using GNRA tetraloop motif with both BzCN and NMIA SHAPE reagents revealed that the reagent binding had a tendency to induce irreversible structural changes of RNA, often resulting in formation of unfolded structures (data not shown). Even prolonged simulations did not reveal any refolding events, i.e. re-conformations back to native GNRA structure, probably, due to the limited timescale of our simulations. As the stability of unfolded structures in current force fields is questionable,[S18–S22] we applied the fourth restraint in a form of an RNA-dedicated metric (ϵRMSD)[S21,S23] arranging nucleobases close to the template (starting X-ray structure). We note, however, that the restraint may affect the reactivity (presumably decrease) of some, especially, unpaired nucleotides, which overall dynamics could be suppressed (Figure S7), ultimately decreasing the correlation against experimental data. The ϵRMSD was applied on all nucleotides from SRP-RNA except the one (U6, see Figure 1 in the maintext) that is part of closing U6-A36 base pair of the asymmetric loop. By omitting U6 from the ϵRMSD restraint, we aimed to allow some occasional base-pair openings as both nucleotides were probed as unpaired in the SHAPE experiment[S24]. A36 was however, still forced to keep its stacking interaction with neighboring nucleotides in order to prevent any further changes in the RNA fold. Notably, presence of reagent in close proximity of U6-A36 base pair was able to generate some occasional base-pair openings during enhanced sampling simulations (see Figure S12 for fluctuations of ϵRMSD and root-mean-square deviation (RMSD) during concatenated continuous trajectories). Notice that, for objectivity, the

ϵRMSD restraint was also applied during SRP-RNA simulations without SHAPE reagents, i.e., during classical MD as well as enhanced sampling simulations of SRP-RNA, where we studied sugar-pucker transitions. All restrains were applied by PLUMED and an example of PLUMED input file is available on github (https://github.com/srnas/shape-md).

4 Calculated reactivities and binding rates

The relative binding rates ($K_{a,rel}$) between reagent and each of 36 analyzed nucleotides from SRP-RNA were calculated using the equilibrium distributions recovered from the enhanced sampling simulations as described elsewhere.[S16] Subsequently, reactivity of particular nucleotide was derived as a relative free energy: $\Delta G^{calc}_{react,rel} = -R \times T \times \ln(K_{a,rel})$. Energies were then normalized against one specific nucleotide (C19) from the canonical C19-G26 base pair. The standard errors were estimated as leave-one-out jackknife approach from concatenated trajectories. We used four different numbers of blocks, i.e., 2, 4, 8, and 16, and considered the largest error obtained among them for each nucleotide.

Enhanced sampling simulations with smaller RNA motifs, i.e., cAMP, cCMP nucleotide analogs, and GNRA tetraloop, allowed us to use much larger solvent box and subsequently estimate absolute binding rates. The absolute binding rates (K_a) require to take into account the probability of the reagent to be found in the bulk (not being affected by any interaction with RNA). We considered that by calculating reagent occurrence in the region defined as a spherical shell with the width of 15 Å and located 25 Å away from the particular O2' atom (see ref.[S16] for the details).

The overall barrier of the SHAPE reaction ($\Delta G^{\ddagger}_{react}$) is related to the number of modifications (SHAPE adducts) per molecule (N_{mod}). $\Delta G^{\ddagger}_{react}$ could be estimated by using: $\triangle G^{\ddagger}_{react} = -R \times T \times \ln\left(\frac{N_{mod}}{N_{nt} \times t_{exp} \times c_{exp} \times K_a \times 10^{12}}\right)$, where N_{nt} is the nucleotide length of RNA molecule, t_{exp} duration of the experiment in seconds, c_{exp} the concentration of the reagent in the experiment, and absolute binding rate (K_a) with the pre-factor (10^{12}) originating from

the Eyring equation. Our calculated absolute binding rates for nucleotides from the GNRA tetraloop ranged from 0.2 M^{-1} to 6.4 M^{-1}. Thus, considering one modification per molecule ($N_{mod} = 1$) in the SHAPE experiment, ten analyzed nucleotides of our GNRA tetraloop ($N_{nt} = 10$), approximate duration of the SHAPE experiment in ten minutes ($t_{exp} = 600$ s), standard reagent concentration in experiments around hundreds of mM ($c_{exp} = 0.1$ M), and common calculated absolute binding rates ($K_a = 1$ M^{-1}), the rough estimation of $\Delta G^{\ddagger}_{react}$ would be around 20 kcal/mol.

5 Experimental reactivities

Experimentally measured SHAPE reactivities from SRP-RNA co-transcriptional experiment[S24] are freely available and were retrieved from RMDB database.[S25] We notice a small difference (one purine nucleotide) in sequences as our simulated SRP-RNA contained GAAA GNRA tetraloop, whereas the corresponding RNA chain probed by SHAPE experiment contained GGAA GNRA tetraloop. Results from 4 different SHAPE experiments are stored in the RMDB database, i.e., datasets labelled as:

- SRPECLI_BZCN_0001.rdat
- SRPECLI_BZCN_0002.rdat
- SRPECLI_BZCN_0003.rdat
- SRPECLI_BZCN_0004.rdat

Each dataset contains SHAPE reactivities for several RNA sequences ranging from 29 nucleotides to 124 nucleotides in length (as the sequence was prolonged during the co-transcription). We took averaged reactivities from last 16 longest sequences (see Figure S7 and Table S1 for statistical errors), where the motif of interest is supposed to be formed and not expected to undergo any structural change. The averaged reactivities of each nucleotide were then

derived as a free energy, i.e., $\Delta G^{exp}_{react} = -R \times T \times \ln(SHAPE_{react})$, and subsequently normalized against C19 nucleotide. All four datasets reveal comparable correlations against calculated reactivities (see Table S1). The dataset SRPECLI_BZCN_0001.rdat was used for plotting the correlation between experimental and calculated reactivities. Other experimental datasets are significantly correlated with the first one (R-values of 0.75, 0.73, and 0.93, for dataset 0002, 0003, and 0004, respectively). This indicates that any prediction with a R-values on the order of 0.75 or better would be practically undistinguishable from experimental datasets.

6 Fitting procedure involving sugar-pucker populations

In an attempt to improve the correlation between experiment and calculations, we *a posteriori* enhanced reactivity of those reactive conformations (reagent in the close proximity of particular 2'-OH group) having C2'-*endo* sugar pucker of the ribose ring. The relative binding rates were adjusted according to the equation: $K^{FIT}_{a,rel} = K^{3'-endo}_{a,rel} + \beta \cdot K^{2'-endo}_{a,rel}$, where $\beta = \exp\left(\frac{\lambda^{C2'-endo}}{RT}\right)$. Fits are expected to be in a range of few kcal/mol (Table S2) favoring C2'-*endo* sugar pucker in a free energy. The fitting procedure was done within MS EXCEL using 'what-if' analysis by the Solver.

7 Solvent accessibility of 2'-OH groups and other structural analysis

We explicitly tested a putative determinant for SHAPE reactivity the accessibility of 2'-OH groups from SRP-RNA nucleotides towards solvent and BzCN reagent computed by using different value of the probe radius. We compared the accessibility in the X-ray structure and the average accessibilities from a 500 ns-long MD simulation with experimental SHAPE reactivities. The calculated Pearson's linear R-values considering water molecule (1.4 Å

probe) and BzCN reagent (5.0 Å probe) show that accessibility is uncorrelated (analysis of X-ray structure) or even slightly anticorrelated (analysis of MD simulation) with experimental 2'-OH reactivity (Figures S1 and S2). Results are in agreement with larger structural datasets reported in the original studies.[S26–S28] We then considered the model proposed by Pinamonti and co-workers,[S29] where fluctuations of C2-C2 base atom distances as predicted by ENMs are used to infer SHAPE reactivities. The obtained correlation coefficient between theory and experiment (R = 0.48, Table S1) is in agreement with the results for similar systems reported in the original paper.[S29] By removing two outliers, i.e., highest reactive residues, the correlation could be increased up to 0.65, suggesting some potential of this computationally cheap and fast approach. We note however that using the ENM approach most nucleotides from SRP-RNA reveal similar fluctuations and are hardly distinguished among each other (Figure S3). A slightly better overall correlation (R = 0.61, Table S1) could be estimated by directly computing the fluctuations of C2-C2 base atom distances during a classical MD simulations (Figure S3), as it was also proposed.[S29]

We then investigated the transient C2'-*endo* sugar puckers as another possible determinant for SHAPE reactivity. The X-ray structure as well as the final structure of the 500 ns-long SRP-RNA MD simulation revealed just a single nucleotide among 36 analyzed with C2'-*endo* sugar pucker (A10 in X-ray and U6 in MD, respectively). Hence, we considered the populations of the C2'-*endo* sugar pucker during MD simulations as a possible structural factor for distinguishing among reactivity patterns. We observed that eight (including A10) among 36 nucleotides are sampling C2'-*endo* pucker during classical MD simulation and their population is correlated with experimental reactivity (Pearson's linear R-value of 0.67, Figure S4 and Table S1). The sugar-pucker sampling analysis is, however, able to distinguish less than 1/4 of the nucleotides from SRP-RNA since most of them never visited the C2'-*endo* conformation within the simulation timescale. Thus, we performed enhanced sampling simulations (see Simulation protocols for details), where all SRP-RNA nucleotides were allowed to visit C2'-*endo* sugar pucker. The obtained correlation (R = 0.53) was slightly

lower than the one from classical MD simulation (see Figure S4 for details).

8 Supporting Tables

EXP.	dataset_0001		dataset_0002		dataset_0003		dataset_0004		CALC.	BzCN_SRP-RNA_flex.		BzCN_SRP-RNA_fixed		BzCN_SRP-RNA_flex.	BzCN_SRP-RNA_fixed	SRP-RNA_flex._pucker-biased	SRP-RNA_flex._pucker	ENM_fluct.	MD_C2-C2_fluct.
nucl.	exp.	error	exp.	error	exp.	error	exp.	error		calc.	error	calc.	error	calc. + fitting	calc. + fitting	calc.	calc.	calc.	calc.
g3	*1.14E+00*	*5.28E-02*	*1.34E+00*	*5.82E-02*	*1.39E+00*	*3.83E-02*	*1.20E+00*	*8.29E-02*		*-7.68E-01*	*1.36E-01*	*-1.35E+00*	*2.10E-01*	*-8.04E-01*	*-1.36E+00*	*1.63E+00*	*-6.81E-01*	*1.08E-02*	*-1.84E-03*
u4	-1.13E-01	1.75E-01	2.36E-01	1.47E-01	4.99E-02	1.61E-01	1.63E-01	2.32E-01		2.71E-01	2.05E-01	-1.10E+00	2.33E-01	6.61E-01	-1.10E+00	1.87E+00	-6.36E-01	1.08E-02	1.03E-01
u5	1.95E+00	2.26E-02	2.10E+00	1.50E-02	2.25E+00	2.11E-02	2.26E+00	1.40E-02		1.40E+00	2.01E-01	-5.15E-01	1.82E-01	2.19E+00	-5.16E-01	1.98E+00	1.78E+00	1.24E-01	9.47E-01
u6	2.38E+00	1.94E-02	2.46E+00	2.39E-02	2.58E+00	2.09E-02	2.74E+00	1.39E-02		7.20E-01	1.76E-01	-4.83E-01	1.61E-01	3.18E+00	-4.84E-01	3.86E+00	3.50E+00	5.11E-01	2.34E+00
A7	2.29E+00	1.73E-02	2.49E+00	2.50E-02	2.61E+00	2.40E-02	2.71E+00	1.33E-02		2.54E-01	2.91E-01	-8.66E-01	1.57E-01	1.78E+00	-8.67E-01	1.66E+00	1.77E+00	4.05E-01	1.34E+00
C8	6.36E-01	8.96E-02	8.44E-02	1.37E-01	3.94E-02	1.67E-01	5.80E-01	8.71E-02		4.35E-01	2.89E-01	-3.28E-01	1.45E-01	1.76E+00	-3.29E-01	2.84E+00	1.89E+00	4.12E-01	8.18E-01
C9	2.28E+00	5.03E-02	1.53E+00	5.95E-02	1.39E+00	5.58E-02	1.96E+00	4.15E-02		2.13E+00	2.48E-01	3.51E-03	1.36E-01	4.89E+00	3.51E-03	3.29E+00	2.92E+00	1.11E+00	3.18E+00
A10	1.02E+00	4.24E-02	1.58E+00	2.98E-02	1.81E+00	2.57E-02	1.22E+00	5.60E-02		3.11E+00	1.76E-01	3.78E-01	1.32E-01	4.54E+00	3.29E+00	3.43E+00	1.55E+00	1.61E+00	2.30E+00
g11	2.53E-01	1.06E-01	7.22E-01	8.97E-02	3.66E-01	1.02E-01	-1.05E-01	2.42E-01		1.21E+00	1.49E-01	4.77E-01	1.97E-01	1.18E+00	4.79E-01	1.72E-01	-6.81E-01	3.56E-02	7.62E-02
g12	1.16E+00	5.98E-02	1.32E+00	3.31E-02	1.32E+00	7.71E-02	1.37E+00	5.67E-02		7.50E-01	1.91E-01	-6.16E-01	2.36E-01	7.15E-01	-9.55E-01	-6.95E-02	-6.81E-01	-3.60E-02	-4.46E-01
u13	-5.22E-02	1.70E-01	1.64E-01	1.11E-01	1.43E-01	1.09E-01	1.42E-02	1.16E-01		3.19E-01	1.78E-01	-9.54E-01	3.11E-01	2.84E-01	-6.17E-01	-2.37E-02	-6.81E-01	-5.85E-05	-1.76E-02
c14	1.47E+00	4.76E-02	3.28E-01	1.45E-01	3.94E-02	2.14E-01	1.17E+00	8.34E-02		1.68E+00	1.20E-01	-7.24E-01	2.59E-01	1.65E+00	-7.25E-01	-2.86E-02	-6.81E-01	1.46E-01	6.41E-02
a15	1.34E-01	7.12E-02	5.76E-01	5.83E-02	6.45E-01	4.73E-02	-4.52E-01	2.14E-01		4.28E-01	1.20E-01	-1.03E+00	1.36E-01	3.93E-01	-1.03E+00	-7.36E-02	-6.81E-01	2.51E-01	6.78E-01
g16	-9.72E-01	1.96E-01	-5.94E-01	1.70E-01	-4.18E-01	2.57E-01	-2.30E+00	3.31E-01		3.56E-01	1.20E-01	-2.20E-01	1.94E-01	3.20E-01	-2.21E-01	-7.27E-02	-6.81E-01	9.21E-02	3.15E-01
g17	6.09E-01	5.76E-02	8.42E-01	4.84E-02	8.79E-01	7.64E-02	9.71E-01	9.74E-02		2.25E-01	9.83E-02	-7.35E-01	1.68E-01	1.89E-01	-7.36E-01	2.03E+00	-6.81E-01	-1.37E-02	4.58E-02
u18	6.32E-03	1.16E-01	4.66E-01	5.33E-02	3.49E-01	1.31E-01	1.03E-01	9.32E-02		3.85E-01	1.08E-01	5.69E-01	1.39E-01	3.57E-01	5.71E-01	1.02E-02	-6.36E-01	-2.10E-02	1.59E-01
c19	0.00E+00	1.09E-01	0.00E+00	1.27E-01	0.00E+00	1.51E-01	0.00E+00	2.16E-01		0.00E+00	1.11E-01	0.00E+00	9.55E-02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
c20	9.41E-02	1.19E-01	-7.53E-01	3.40E-01	-4.15E-01	2.43E-01	-1.25E+00	4.77E-01		-6.58E-02	9.76E-02	-1.18E-02	1.60E-01	-1.02E-01	-1.18E-02	-4.38E-02	-6.15E-01	3.21E-02	2.11E-01
G21	1.92E-01	1.14E-01	1.53E-01	1.47E-01	2.98E-01	1.16E-01	-3.86E-02	2.24E-01		-1.00E+00	1.66E-01	1.28E-01	3.25E-01	-1.04E+00	1.29E-01	-5.47E-02	-6.81E-01	1.87E-01	9.74E-01
G/A22	2.17E+00	2.82E-02	2.33E+00	2.71E-02	2.51E+00	2.84E-02	2.56E+00	1.81E-02		1.39E-01	1.21E-01	-2.21E-01	1.72E-01	1.21E+00	-2.21E-01	8.74E-01	7.09E-01	3.45E-01	1.00E+00
A23	1.69E+00	3.74E-02	1.91E+00	2.38E-02	2.09E+00	3.19E-02	1.85E+00	4.84E-02		5.92E-01	1.73E-01	3.19E-01	1.63E-01	5.57E-01	3.20E-01	-7.37E-02	-6.81E-01	3.60E-01	3.98E-01
A24	1.25E+00	3.43E-02	1.65E+00	2.58E-02	1.88E+00	3.32E-02	1.41E+00	3.10E-02		4.44E-01	8.75E-02	2.39E-02	1.55E-01	4.09E-01	2.39E-02	-7.39E-02	-6.81E-01	2.33E-01	6.94E-01
g25	-3.73E-01	2.22E-01	-1.96E-01	1.34E-01	-3.15E-01	2.17E-01	-1.35E+00	4.10E-01		2.75E-01	1.27E-01	-5.44E-01	3.92E-01	2.40E-01	-5.44E-01	-7.37E-02	-6.81E-01	-2.35E-02	-2.05E-02
g26	-7.89E-02	2.87E-01	2.89E-01	1.17E-01	-5.14E-01	3.47E-01	-6.01E-01	6.00E-01		1.26E-01	1.21E-01	-2.78E-01	1.59E-01	9.04E-02	-2.78E-01	-3.15E-02	-6.81E-01	5.83E-03	2.15E-01
a27	2.00E-01	1.58E-01	1.19E-01	1.24E-01	-3.92E-02	3.00E-01	-8.23E-01	4.11E-01		1.47E-01	1.26E-01	1.28E-01	1.19E-01	1.11E-01	1.28E-01	9.42E-02	-6.81E-01	6.04E-02	1.35E-01
a28	4.35E-01	7.47E-02	8.24E-01	6.53E-02	9.02E-01	6.67E-02	7.41E-01	8.15E-02		3.58E-01	9.95E-02	3.02E-01	1.42E-01	3.23E-01	3.03E-01	1.16E+00	-6.81E-01	2.97E-02	4.68E-01
g29	1.46E-01	1.06E-01	-2.01E-02	2.48E-01	8.41E-02	1.70E-01	-4.08E-01	1.79E-01		8.29E-01	1.46E-01	1.13E-01	8.69E-02	7.94E-01	1.13E-01	6.54E-01	-6.81E-01	1.59E-01	-1.38E-01
c30	8.75E-01	9.18E-02	-7.65E-02	2.64E-01	-6.41E-02	3.27E-01	5.68E-01	1.18E-01		-1.64E-01	1.06E-01	6.71E-01	2.54E-01	-2.00E-01	6.74E-01	-7.39E-02	-6.81E-01	1.03E-01	2.79E-01
a31	1.07E-01	1.12E-01	1.14E-01	1.84E-01	2.03E-01	1.14E-01	-1.25E+00	4.40E-01		-1.41E-02	1.11E-01	-8.14E-01	1.99E-01	-4.98E-02	-8.15E-01	-7.39E-02	-6.81E-01	1.66E-01	5.23E-01
g32	5.37E-01	8.22E-02	9.31E-01	9.10E-02	7.06E-01	8.79E-02	6.38E-01	9.32E-02		4.75E-01	1.11E-01	-1.15E-01	2.80E-01	4.40E-01	-1.15E-01	-6.97E-02	-6.81E-01	1.25E-01	6.45E-01
c33	7.28E-02	1.12E-01	-5.25E-01	2.28E-01	-7.73E-01	2.29E-01	-9.15E-02	2.45E-01		4.17E-01	2.06E-01	-4.03E-02	1.04E-01	3.81E-01	-4.04E-02	-2.12E-02	-6.81E-01	2.38E-03	9.68E-02
c34	1.65E+00	9.42E-02	-1.02E+00	6.00E-01	-1.17E+00	3.72E-01	8.37E-01	1.86E-01		5.77E-01	3.49E-01	-3.67E-02	1.96E-01	5.42E-01	-3.68E-02	6.97E-01	-6.81E-01	8.91E-02	5.73E-01
A35	5.63E-01	7.46E-02	5.85E-01	8.33E-02	7.10E-01	1.02E-01	4.44E-01	1.25E-01		1.46E+00	1.32E-01	9.40E-02	1.86E-01	1.46E+00	9.42E-02	4.29E-01	-6.81E-01	1.26E-01	1.03E+00
a36	5.90E-01	6.74E-02	9.96E-01	6.73E-02	1.22E+00	5.26E-02	6.02E-01	8.82E-02		2.14E+00	1.79E-01	9.32E-02	2.89E-01	2.12E+00	9.34E-02	1.33E+00	-6.81E-01	5.81E-02	7.27E-01
g37	1.52E-01	1.03E-01	3.32E-01	1.20E-01	7.22E-01	9.85E-02	9.91E-02	1.99E-01		1.23E+00	2.68E-01	-4.17E-01	1.38E-01	1.20E+00	-4.18E-01	6.99E-02	-6.81E-01	8.10E-02	4.28E-01
g38	6.34E-02	2.06E-01	1.67E-01	7.27E-02	1.24E-01	1.06E-01	-9.57E-01	4.43E-01		8.77E-01	2.42E-01	1.73E-01	2.14E-01	8.42E-01	1.74E-01	1.25E-01	-6.81E-01	7.56E-02	-9.26E-02
									Correlations_Exp._vs_Calc.	dataset_0001		-0.04		0.59	0.01	0.53	0.67	0.48	0.61
									Pearson's linear R-values	dataset_0002		-0.09		0.55	0.04	0.50	0.60	0.46	0.55
										dataset_0003		-0.08		0.53	0.07	0.49	0.58	0.47	0.56
										dataset_0004		-0.05		0.55	0.02	0.56	0.62	0.42	0.54

Table S1: Normalized experimental and calculated reactivities for SRP-RNA nucleotides. Calculated reactivities were estimated from (i) enhanced sampling simulations involving BzCN reagent (involving both flexible and frozen RNA), (ii) sugar-pucker transitions during enhanced sampling simulations, (iii) sugar-pucker sampling analysis and fluctuations of C2-C2 distances during classical MD simulations, and (iv) fluctuations from elastic network models (ENMs). Both experimental and calculated reactivities are normalized to C19 nucleotide. Correlations (as Pearson's linear R-values) between calculated reactivities from different models and all four experimental datasets (SRPECLI_BZCN_0001-0004) are shown.

Table S2: The free-energy values corresponding to sugar-pucker fits that were applied to improve the correlation between computed and experimentally derived reactivities. Values correspond to reactivity predictions from the analysis of X-ray structure ($RNA_{constr.}$ + FIT_{pucker}), classical MD simulation ($RNA_{flex.}$ + FIT_{pucker}), enhanced sampling simulation biasing sugar-pucker flips ($RNA_{flex.(biased)}$ + FIT_{pucker}), and enhanced sampling simulations with BzCN reagent ($RNA_{constr.}$ + BzCN + FIT_{pucker} and $RNA_{flex.}$ + BzCN + FIT_{pucker}). Each number represents an estimate of the free-energy difference in the reactivity of C2'-*endo* and C3'-*endo* conformations.

	$RNA_{constr.}$ + FIT_{pucker}	$RNA_{flex.}$ + FIT_{pucker}	$RNA_{flex.(biased)}$ + FIT_{pucker}	$RNA_{constr.}$ + BzCN + FIT_{pucker}	$RNA_{flex.}$ + BzCN + FIT_{pucker}
β [kcal/mol]	6.9	4.6	4.2	0.5	2.9

9 Supporting Figures

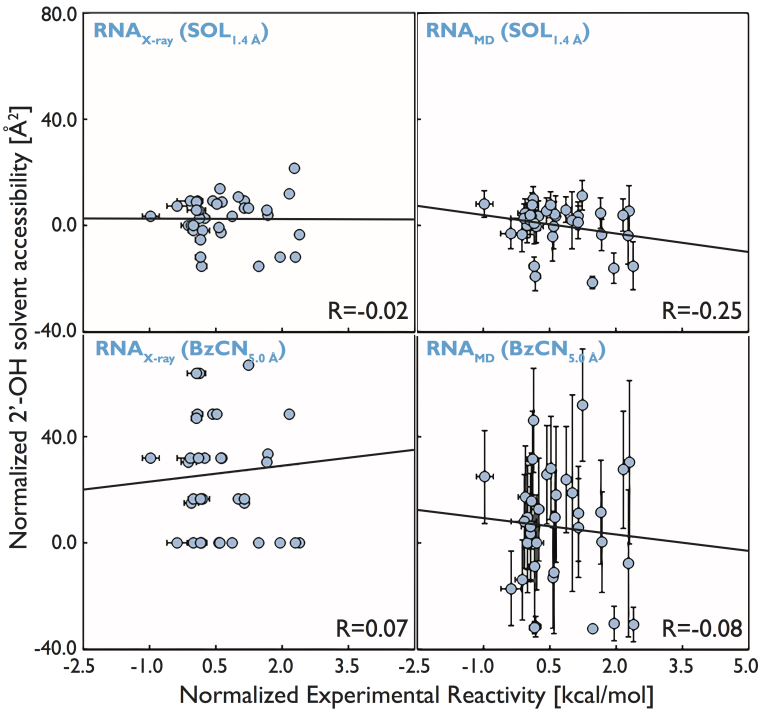

Figure S1: Correlations shown as Pearson's linear R-values between experimental SHAPE reactivities and solvent accessibilities of 2'-OH groups. Top and bottom panels correspond to the probe size of 1.4 Å (water, labelled as SOL) and 5.0 Å (BzCN SHAPE reagent), respectively. In the left side, panels reveal correlations for the X-ray structure. On the right side, averaged results from classical 500 ns-long MD simulation of SRP-RNA are shown. Considering the largely planar BzCN reagent, the better way to describe its accessibility should account its realistic shape, which was attempted in the simulation with rigid RNA (Figure S5 and Figure 2B in the maintext).

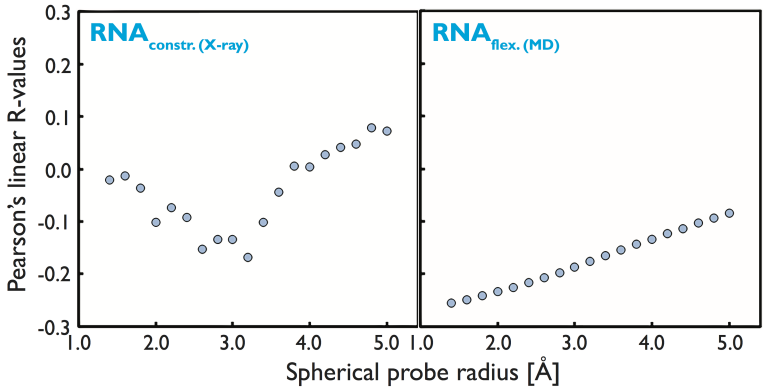

Figure S2: Changes in correlation between experimental SHAPE reactivity and solvent accessibility displayed as dependency on the probe size. Probes are ranging from water (˜1.4 Å) towards BzCN reagent (˜5.0 Å) with 0.2 Å steps. Left and right panels reveal results for the starting (X-ray) structure and averages from MD simulations of SRP-RNA.

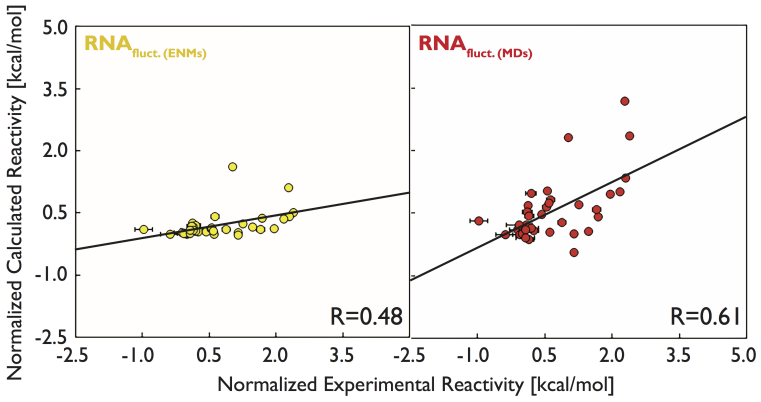

Figure S3: Correlations between experimental SHAPE reactivity and calculated reactivities estimated as fluctuations of C2-C2 distances from elastic network models (ENMs, left panel) and MD simulations (panel on the right) of SRP-RNA. MD fluctuations were calculated as changes in distances between C2 atoms from consecutive nucleobases.

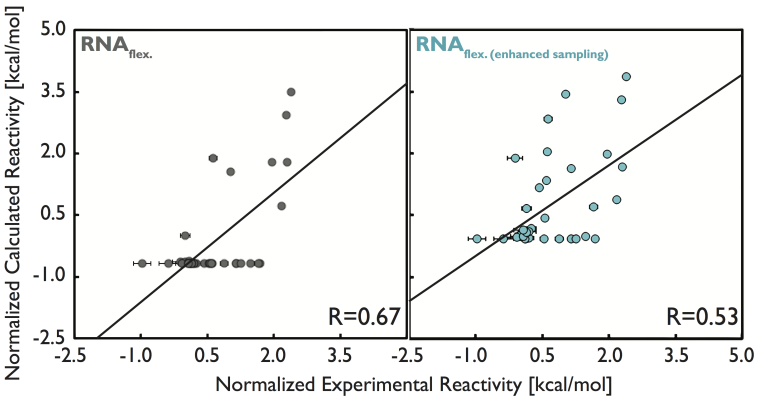

Figure S4: The effect of enhanced sampling on the estimation of SHAPE reactivity from structural analysis involving sugar-pucker populations of SRP-RNA nucleotides. Correlations between experimental SHAPE reactivity and calculated reactivities estimated from sugar-pucker populations during classical MD (left panel) and bias-exchange simulations (panel on the right, similar to panel C of Figure 2 in the main text). The normalized calculated reactivities are shifted from classical to enhanced sampling MD simulation due to the fact that C19 nucleotide, which was the one of 8 nucleotides successfully sampling C2'-*endo* sugar pucker during classical MD simulation, was used as a reference. 28 nucleotides are shown with negative normalized calculated reactivities because they did not visit C2'-*endo* state during classical MD simulation (panel on the left). Although results obtained with these two approaches are expected to be identical for infinitely long simulations, the enhanced sampling results are more statistically reliable for the simulated timescales.

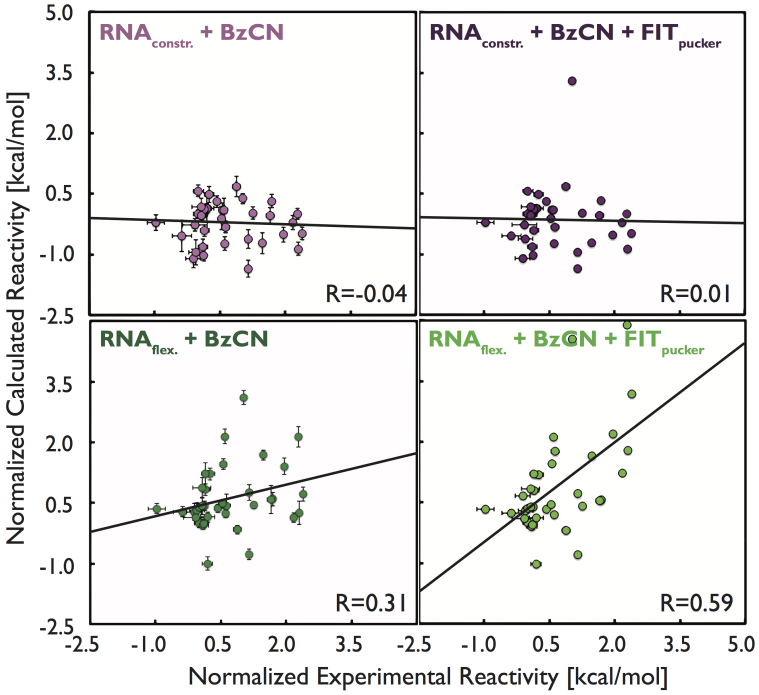

Figure S5: The effect of fitting procedure on the results from enhanced sampling simulations of SRP-RNA. Left panels reveal results from enhanced sampling simulations with BzCN reagent, where top-left panel shows results with frozen (X-ray) structure ($\text{RNA}_{constr.}$). The sampling analysis considering populations of C2'-*endo* sugar pucker (FIT_{pucker}) was applied *a posteriori* and improved correlations are shown as panels on the right.

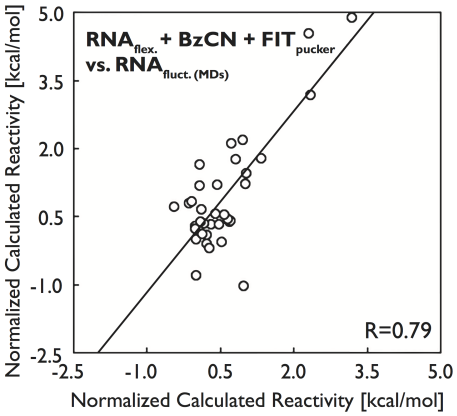

Figure S6: Correlations between results from enhanced sampling simulations and classical MD simulations of SRP-RNA. Reactivities from classical MD simulations were derived as changes in distances between C2 atoms from consecutive nucleobases. Results from bias-exchange like simulations contain corrections from population analysis of sugar puckers (FIT_{pucker}). Clearly, these two approaches provide results that are highly correlated.

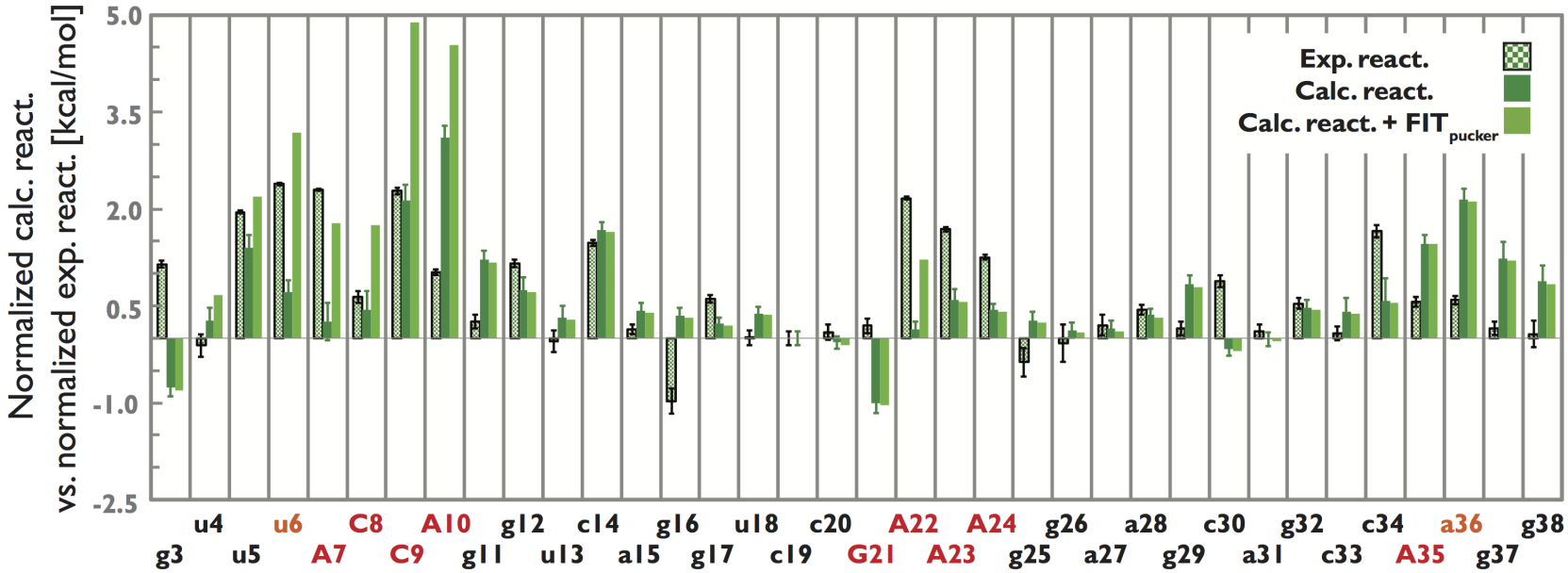

Figure S7: Normalized reactivities of all investigated nucleotides from SRP-RNA. Three columns for each nucleotide correspond to experimentally derived SHAPE reactivity (basketweave pattern with statistical errors derived from averaged reactivities among different sequences in length), calculated reactivities from bias-exchange like simulations with BzCN reagent (dark green with statistical errors highlighting the convergence of MD simulations), and same calculated reactivities corrected by the sugar-pucker population analysis (FIT_{pucker}, light green, errors not shown). All reactivities were normalized to C19 nucleotide. Capital letters mark nucleotides determined as unpaired in the X-ray structure[S1] and coloring scheme corresponds to Figure 1 in the maintext.

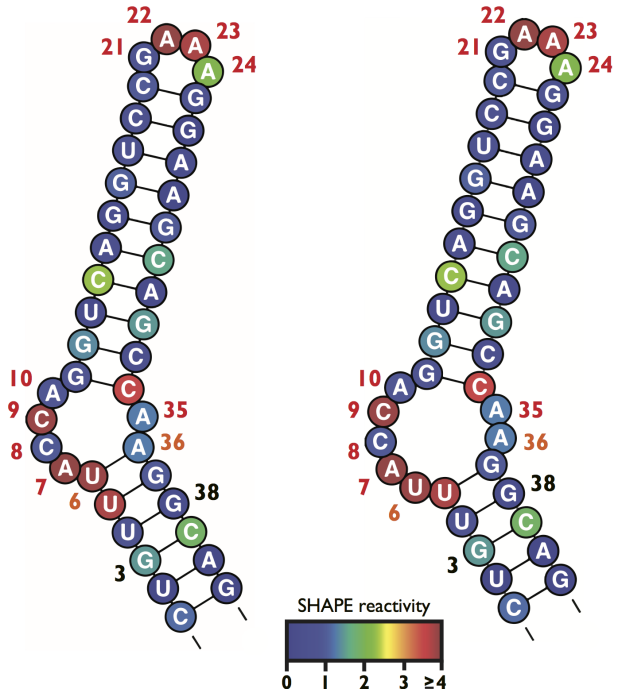

Figure S8: Different secondary structures determined either by X-ray crystallography[S1] (left panel) or probed by SHAPE experiment[S24] (panel on the right). Two structures of SRP-RNA differ in the asymmetric loop, where SHAPE analysis proposed opened U6-A36 base pair. Models were prepared by VARNA[S30] and coloring scheme for each nucleotide is based on averaged SHAPE reactivity.[S24]

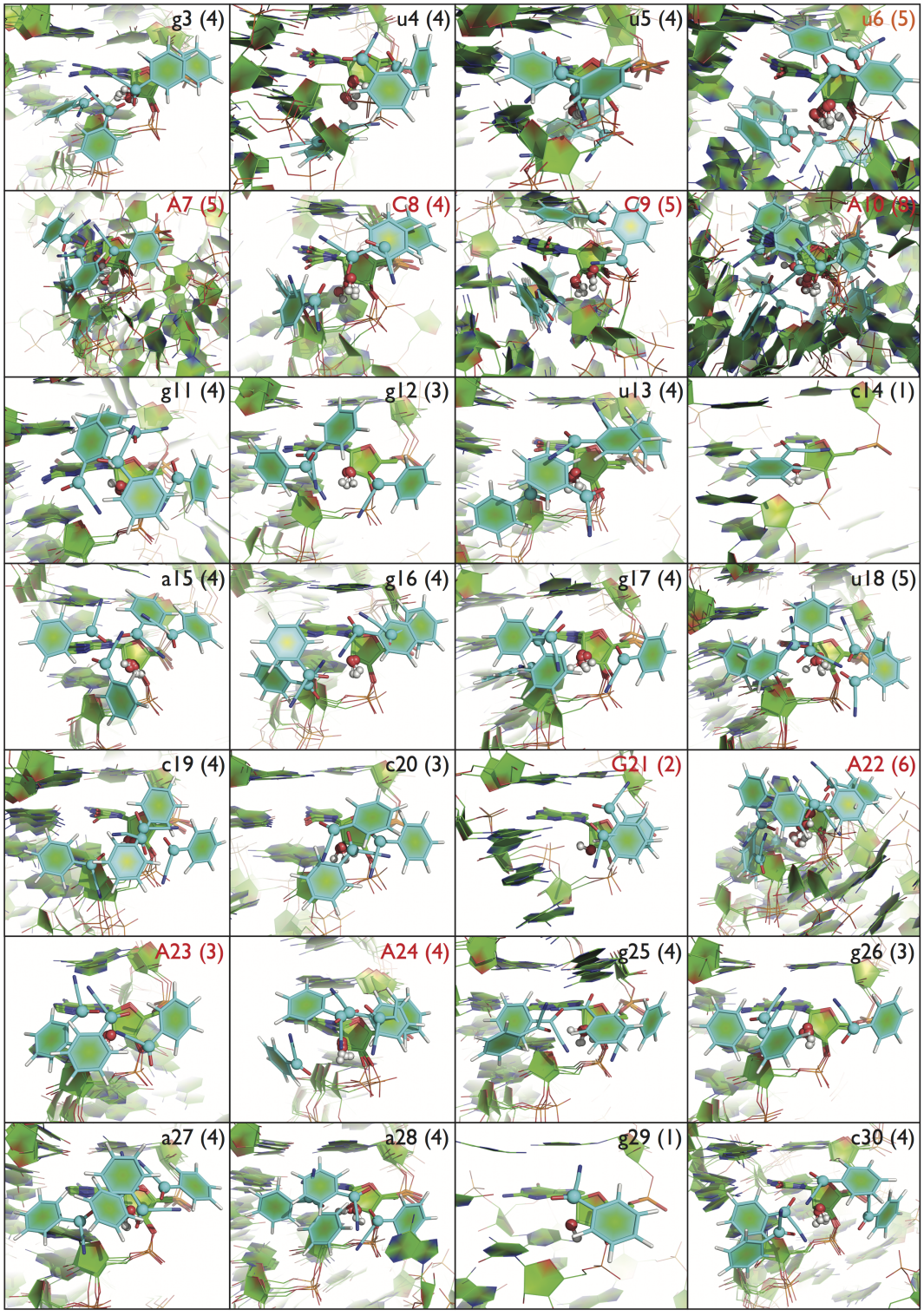

Figure S9: (part 1) See next page for the remaining part and the legend.

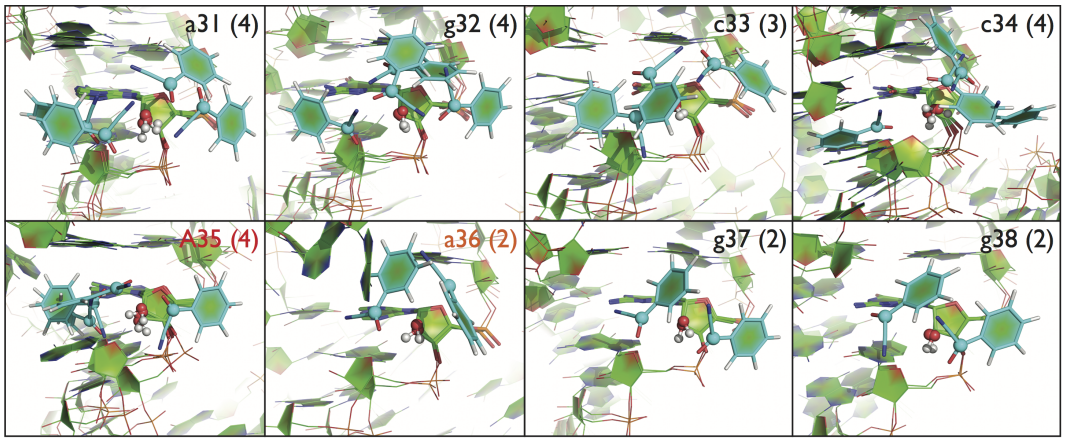

Figure S10: (part 2) Clustering of reactive geometries of BzCN reagent and SRP-RNA nucleotides. We used simple clustering scheme by GROMACS (gromos method, RMSD cutoff of 2.3 Å)[S4] in order to characterize potential reactive geometries between BzCN (in cyan with reactive carbon highlighted) and each one of 36 analyzed nucleotides. Representative snapshots of each cluster are fitted on particular nucleotide (shown in sticks), which 2'-OH group is highlighted as spheres. Water molecules, counter-ions and most of hydrogens (except those of BzCN and 2'-OH group) are not shown for clarity. Labelling and coloring scheme of each nucleotide is consistent with Figure 1 in the maintext and the number of different clusters for each nucleotide is displayed in a following bracket. Cartesian coordinates (in PDB format) of all clusters are available on github (https://github.com/srnas/shape-md).

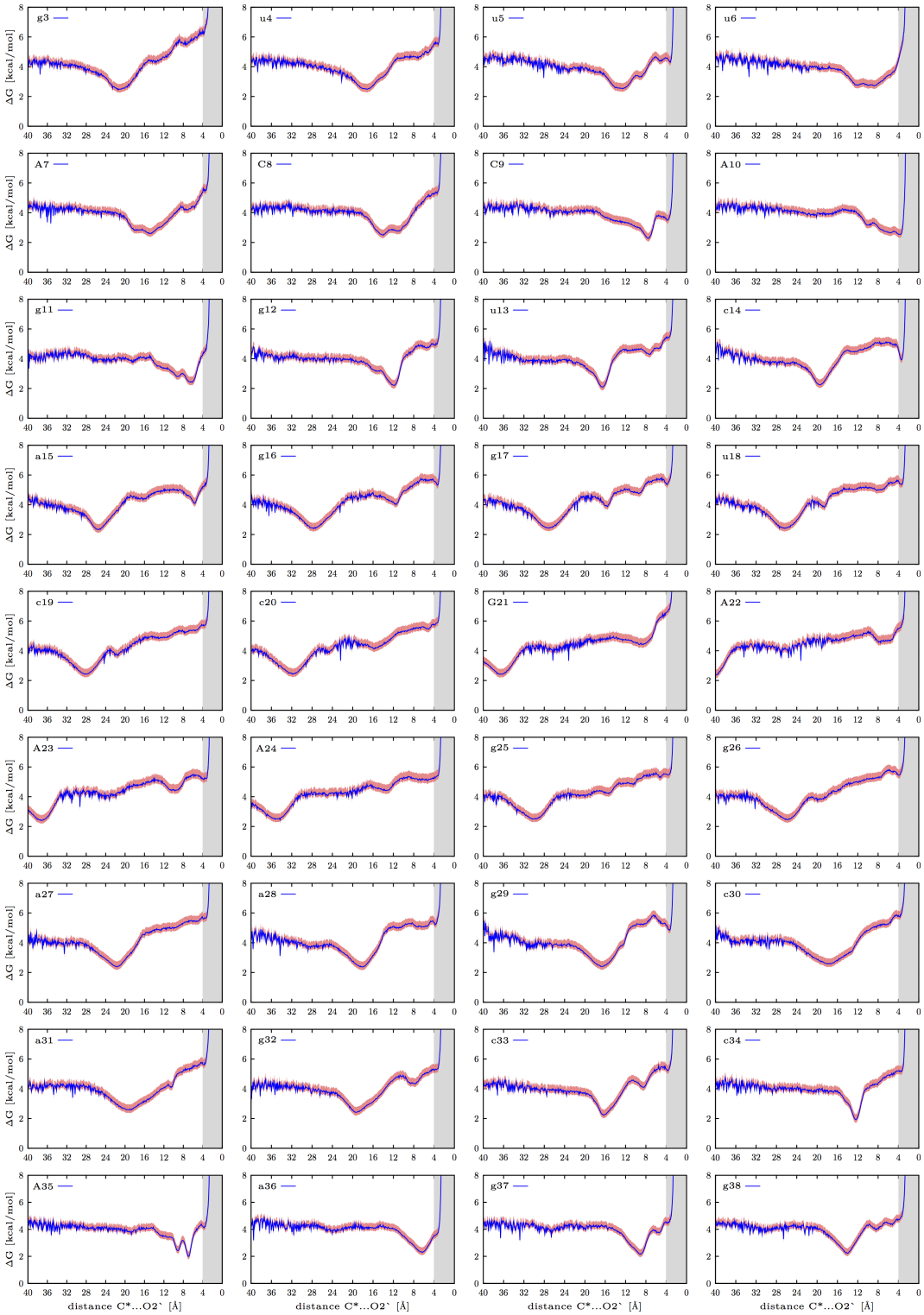

Figure S11: See next page for the legend.

Figure S11: Free-energy binding profiles of BzCN SHAPE reagent towards SRP-RNA nucleotides from enhanced sampling simulations with flexible RNA. Profiles (in blue) for 36 analyzed nucleotides are displayed as a function of distance between reactive carbon of BzCN (C*) and O2' oxygen of 2'-OH group from particular nucleotide ($-RT \ln P(distance)$). Statistical errors are displayed in red and are computed by averaging over four blocks. Shaded area corresponds to the distance lower than 4.0 Å (binding cut-off) that was used to estimate the reactivity of particular nucleotide. Note that the reagent was kept within ∼10 Å shell around RNA (see section Description of applied restraints and specific usage of reagents) and thus was interacting with at least one RNA residue in majority of snapshots. The minima (∼2 kcal/mol) at large distances for some nucleotides correspond to states where the reagent is bound in the asymmetric loop of SRP-RNA (near the close proximity of A10 nucleotide).

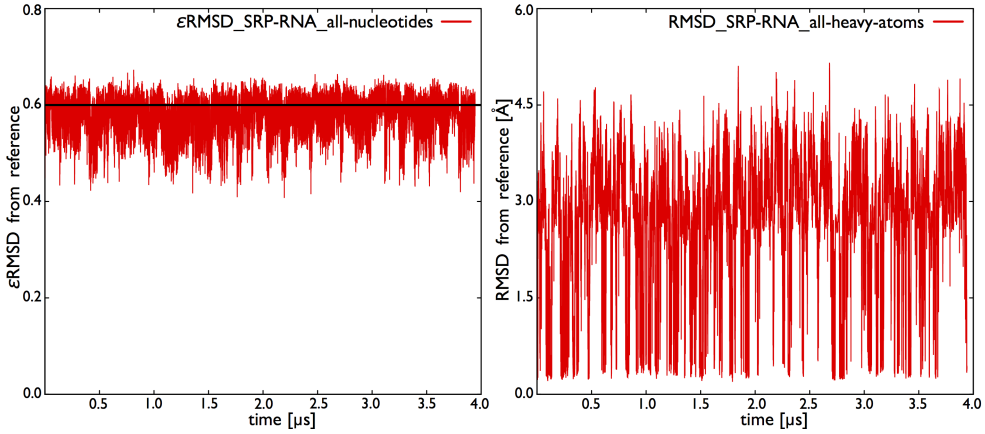

Figure S12: Fluctuations of ϵRMSD (left panel) and RMSD (panel on the right) during enhanced sampling simulations of SRP-RNA with BzCN reagent. Data were derived from concatenated continuous trajectories with the total simulation timescale almost 4 μs (36 replicas ×110 ns per replica). Notice that we used restraint (black line on the left panel) to prevent larger and irreversible structural changes of RNA upon BzCN binding (see section Description of applied restraints and specific usage of reagents).

References

(S1) Batey, R. T.; Rambo, R. P.; Lucast, L.; Rha, B.; Doudna, J. A. *Science* **2000**, *287*, 1232–1239.

(S2) Haurwitz, R. E.; Sternberg, S. H.; Doudna, J. A. *EMBO J.* **2012**, *31*, 2824–2832.

(S3) Hsiao, C.; Lenz, T. K.; Peters, J. K.; Fang, P.-Y.; Schneider, D. M.; Anderson, E. J.; Preeprem, T.; Bowman, J. C.; O'Neill, E. B.; et al., *Nucleic Acids Res.* **2013**, *41*, 3373–3385.

(S4) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; et al., *Bioinformatics* **2013**, *29*, 845–854.

(S5) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(S6) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(S7) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(S8) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.

(S9) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham III, T. E.; Jurečka, P. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.

(S10) Joung, I. S.; Cheatham III, T. E. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

(S11) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(S12) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(S13) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(S14) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.

(S15) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553–4559.

(S16) Cunha, R. A.; Bussi, G. *RNA* **2017**, *23*, 628–638.

(S17) Clay, M. C.; Ganser, L. R.; Merriman, D. K.; Al-Hashimi, H. M. *Nucleic Acids Res.* **2017**, *45*, e134.

(S18) Bergonzo, C.; Cheatham III, T. E. *J. Chem. Theory Comput.* **2015**, *11*, 3969–3972.

(S19) Condon, D. E.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Yildirim, I.; Turner, D. H. *J. Chem. Theory Comput.* **2015**, *11*, 2729–2742.

(S20) Kuhrová, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Šponer, J.; Otyepka, M.; Banáš, P. *J. Chem. Theory Comput.* **2016**, *12*, 4534–4548.

(S21) Bottaro, S.; Banáš, P.; Šponer, J.; Bussi, G. *J. Phys. Chem. Lett.* **2016**, *7*, 4032–4038.

(S22) Zgarbová, M.; Jurečka, P.; Banáš, P.; Havrila, M.; Šponer, J.; Otyepka, M. *J. Phys. Chem. B* **2017**, *121*, 2420–2433.

(S23) Bottaro, S.; Di Palma, F.; Bussi, G. *Nucleic Acids Res.* **2014**, *42*, 13306–13314.

(S24) Watters, K. E.; Strobel, E. J.; Angela, M. Y.; Lis, J. T.; Lucks, J. B. *Nat. Struct. Mol. Biol.* **2016**, *23*, 1124.

(S25) Cordero, P.; Lucks, J. B.; Das, R. *Bioinformatics* **2012**, *28*, 3006–3008.

(S26) Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. *J. Am. Chem. Soc.* **2005**, *127*, 4223–4231.

(S27) Gherghe, C. M.; Shajani, Z.; Wilkinson, K. A.; Varani, G.; Weeks, K. M. *J. Am. Chem. Soc.* **2008**, *130*, 12244–12245.

(S28) McGinnis, J. L.; Dunkle, J. A.; Cate, J. H.; Weeks, K. M. *J. Am. Chem. Soc.* **2012**, *134*, 6617–6624.

(S29) Pinamonti, G.; Bottaro, S.; Micheletti, C.; Bussi, G. *Nucleic Acids Res.* **2015**, *43*, 7260–7269.

(S30) Darty, K.; Denise, A.; Ponty, Y. *Bioinformatics* **2009**, *25*, 1974–1975.