Towards the Rational Design of RsmE Small-RNA Binders: Insights from Molecular Dynamics Simulations

Agustín Ormazábal,[†] Juliana Palma,[*,†,‡] and Gustavo Pierdominici-Sottile[*,†,‡]

†*Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Sáenz Peña 352, Bernal B1876BXD, Argentina.*

‡*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Godoy Cruz 2290, CABA, Argentina.*

E-mail: juliana@unq.edu.ar; gsottile@unq.edu.ar

Supplementary Information

1 List of Figures

- Figure S1: Representative structure of the RsmE-(SL2-l*-SL2*) complex.
- Figure S2: Comparison of the PMFs obtained with WHAM, DHAM and BAR for selected processes.
- Figure S3: Comparison of the PMFs for selected processes calculated by dividing the entire dataset in three equal parts. The standard deviation of the PMF values at selected points was used to estimate the statistical uncertainty of the results.

- Figure S4: Symmetric Kullback-Leibler divergence test.
- Figure S5: Convergence of the QHA entropy estimations with respect to the number of individual trajectories added to the concatenated one.
- Figure S6: Correlation between the Free Energies of Binding (FEB) computed using Umbrella Sampling and MM/GBSA.
- Figure S7: Stem nucleotides interacting with RsmE residues in RsmE-$(SLX)_1$ and RsmE-$(SLX)_2$ complexes, with X = 0, 2, 3.

2 List of Tables

- Table S1: Free Energies of Binding estimated via the MM/GBSA approximation.
- Table S2: SL0–RsmE contacts in RsmE–$(SL0)_1$.
- Table S3: SL2–RsmE contacts in RsmE–$(SL2)_1$.
- Table S4: SL3–RsmE contacts in RsmE–$(SL3)_1$.
- Table S5: SL0–RsmE contacts in RsmE–$(SL0)_2$.
- Table S6: SL2–RsmE contacts in RsmE–$(SL2)_2$.
- Table S7: SL3–RsmE contacts in RsmE–$(SL3)_2$.

3 List of Movies

- Movie S1. Binding process of SL2 to the RsmE-$(SL2)_1$ complex.
- Movie S2. Binding process of SL3 to the RsmE-$(SL3)_1$ complex.
- Movie S3. Binding process of SL2-l-SL2 to RsmE.

4 Consistency and Statistical Error of the PMF Profiles

To quantify the statistical uncertainty of the PMFs, the complete US dataset was partitioned into three subsets of equal size. The BAR method was applied independently to each subset to obtain the corresponding PMF profiles. The standard deviation of the three PMF values at selected points was then taken as an estimate of the statistical uncertainty of the overall PMF.

The consistency between the resulting PMFs and the original raw US data was assessed by computing the biased probabilities for each US window using two alternative approaches and then quantifying the differences between them via the symmetric Kullback-Leibler divergence. We denote by $P_i(\chi)$ the biased probabilities directly obtained from the US simulations in window i, while $\mu_i(\chi)$ are the probabilities estimated from the unbiased distribution determined using WHAM by reintroducing the effect of the biasing potential. According to the symmetric Kullback-Leibler divergence criterion, the consistency between the two distributions is measured by,

$$S_i = \frac{1}{2}\, D(P_i, \mu_i) + \frac{1}{2}\, D(\mu_i, P_i)\,, \tag{1}$$

where,

$$D(f, g) = \sum_{k=1}^{N} f(\chi_k)\; ln\left(\frac{f(\chi_k)}{g(\chi_k)}\right). \tag{2}$$

In Eq. 2, N represents the number of bins employed in a discretized representation of the probability densities $f(\chi)$ and $g(\chi)$, while χ_k is the value of the reaction coordinate at the centre of bin k. The lower the value of S_i, the better agreement between the two distributions, and the more consistent the resulting unbiased probabilities with the original US data.

5 Data and Software Availability

5.1 Fortran codes

- **entropy.f**: Estimates the conformational entropy of a macromolecule using the QHA.
- **pmf.f**: Computes the Potential of Mean Force from US simulations using three algorithms: WHAM, DHAM, and BAR. It also performs the sKL consistency test.

5.2 RsmE-(BM) models

- **RsmE-BM_1.top**: Topology file for RsmE-(BM)_1.
- **RsmE-BM_1.crd**: Coordinates file for RsmE-(BM)_1.
- **RsmE-BM_2.top**: Topology file for RsmE-(BM)_2.
- **RsmE-BM_2.crd**: Coordinates file for RsmE-(BM)_2.

5.3 RsmE-(SL0) models

- **RsmE-SL0_1.top**: Topology file for RsmE-(SL0)_1.
- **RsmE-SL0_1.crd**: Coordinates file for RsmE-(SL0)_1.
- **RsmE-SL0_2.top**: Topology file for RsmE-(SL0)_2.
- **RsmE-SL0_2.crd**: Coordinates file for RsmE-(SL0)_2.

5.4 RsmE-(SL2) models

- **RsmE-SL2_1.top**: Topology file for RsmE-(SL2)_1.
- **RsmE-SL2_1.crd**: Coordinates file for RsmE-(SL2)_1.

- **RsmE-SL2_2.top**: Topology file for RsmE-(SL2)_2.
- **RsmE-SL2_2.crd**: Coordinates file for RsmE-(SL2)_2.

5.5 RsmE-(SL3) models

- **RsmE-SL3_1.top**: Topology file for RsmE-(SL3)_1.
- **RsmE-SL3_1.crd**: Coordinates file for RsmE-(SL3)_1.
- **RsmE-SL3_2.top**: Topology file for RsmE-(SL3)_2.
- **RsmE-SL3_2.crd**: Coordinates file for RsmE-(SL3)_2.

5.6 RsmE-(SL2lSL2) model

- **RsmE-SL2lSL2.top**: Topology file for RsmE-(SL2lSL2).
- **RsmE-SL2lSL2.crd**: Coordinates file for RsmE-(SL2lSL2).

5.7 SL2 model

- **SL2.top**: Topology file for SL2.
- **SL2.crd**: Coordinates file for SL2.

6 Figures

Figure S1: Representative structure of the RsmE-(SL2-l*-SL2*) complex.

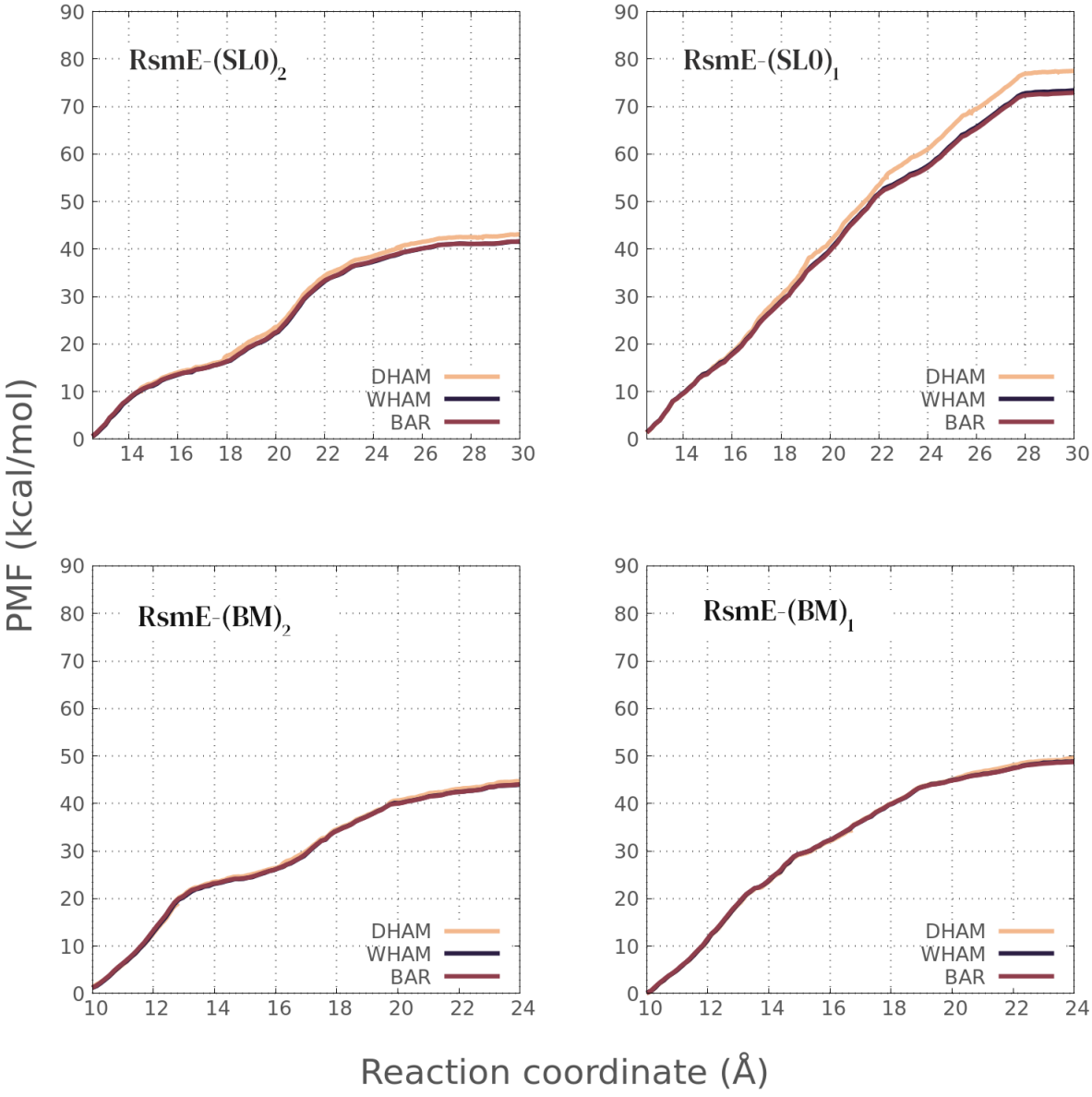

Figure S2: Comparison of the PMFs calculated using WHAM (blue), DHAM (orange), and BAR (red) for all models containing fragments of the *hcnA* transcript. In some cases, the results from different algorithms are so similar that they are visually indistinguishable. Comparable agreement was observed for the other systems analysed in this work.

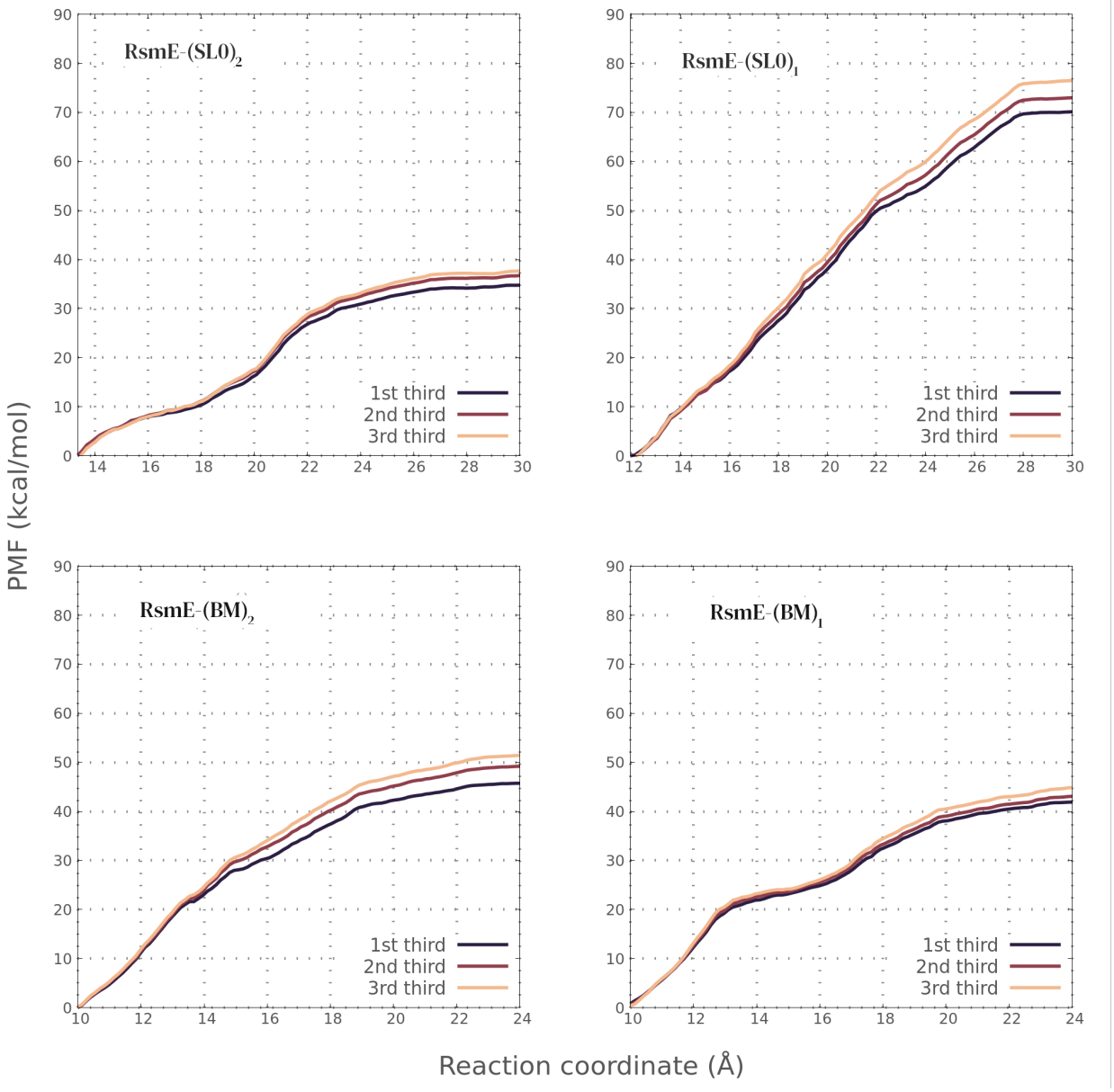

Figure S3: Comparison of the PMFs calculated with three alternative data sets obtained by dividing the whole set in three equal parts. The data in the plot correspond to the models containing fragments of the *hcnA* transcript. Comparable behaviour was observed for the other systems analysed in this work.

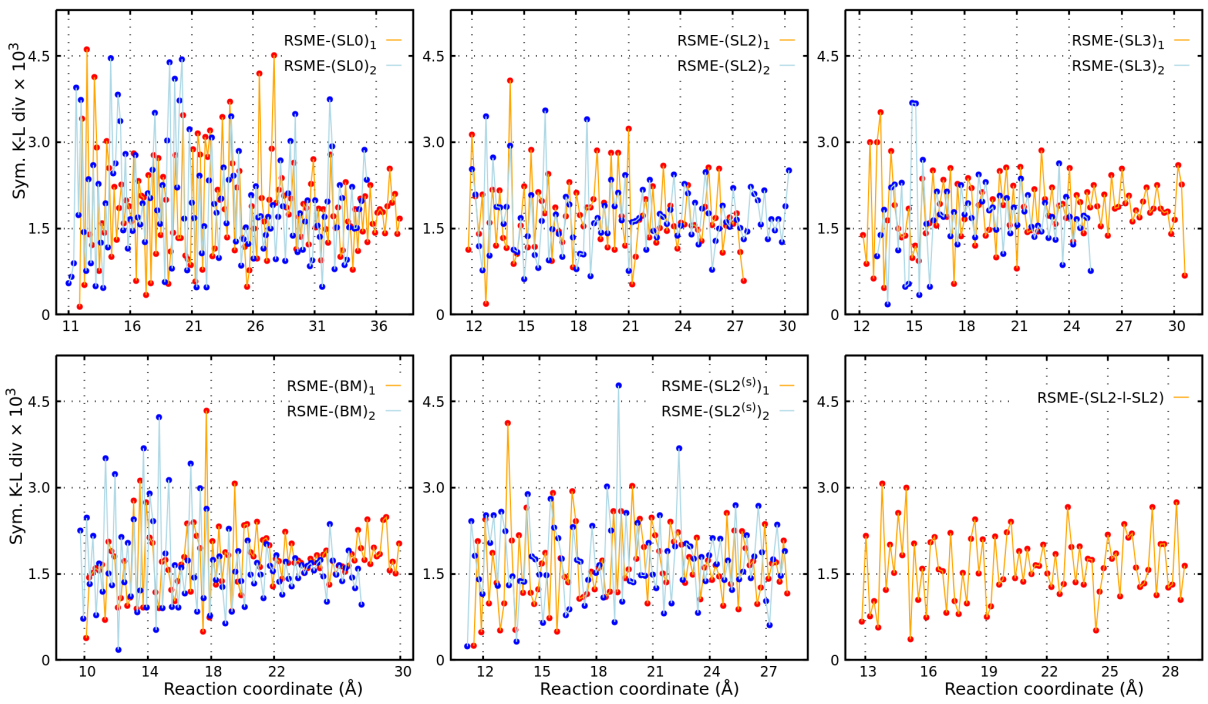

Figure S4: sKL-divergence (Eq. 1) as a function of the reaction coordinate χ, for all the PMFs presented in the main text.

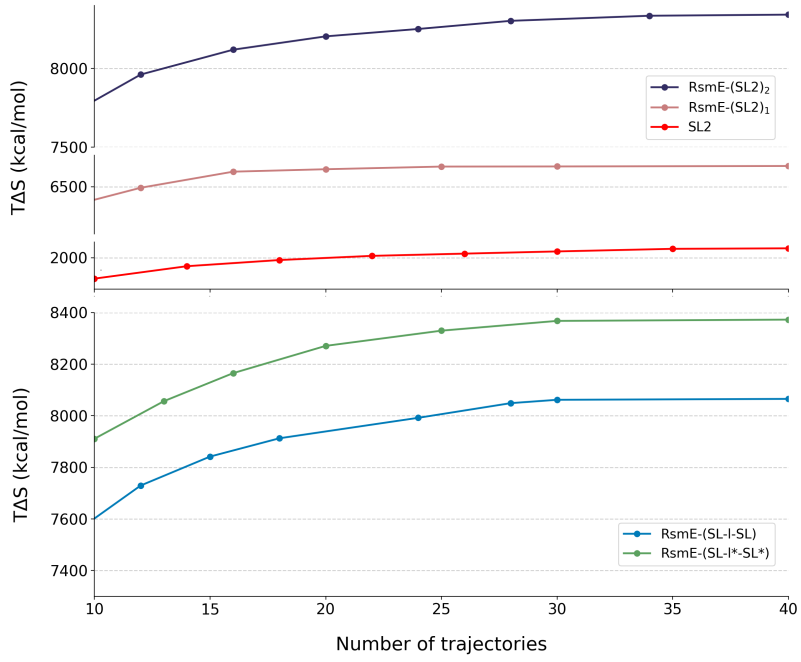

Figure S5: Convergence of the QHA-derived entropies as a function of the number of individual trajectories included in the concatenated dataset.

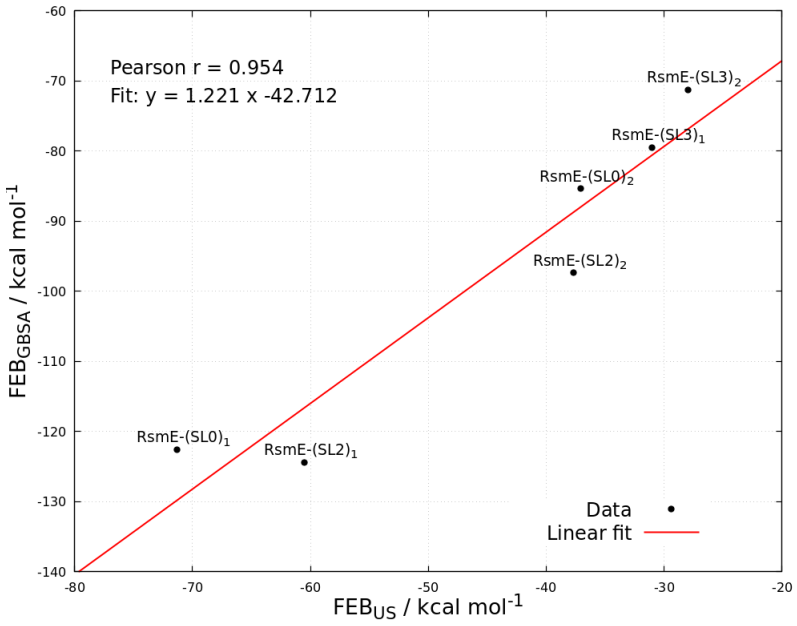

Figure S6: Correlation between the Free Energies of Binding (FEB) computed using Umbrella Sampling and MM/GBSA. Note that the MM/GBSA estimates do not include the unfavorable entropic contribution and therefore overestimate the binding strength. The Pearson correlation coefficient and the best linear fit between the two estimates are reported in the figure.

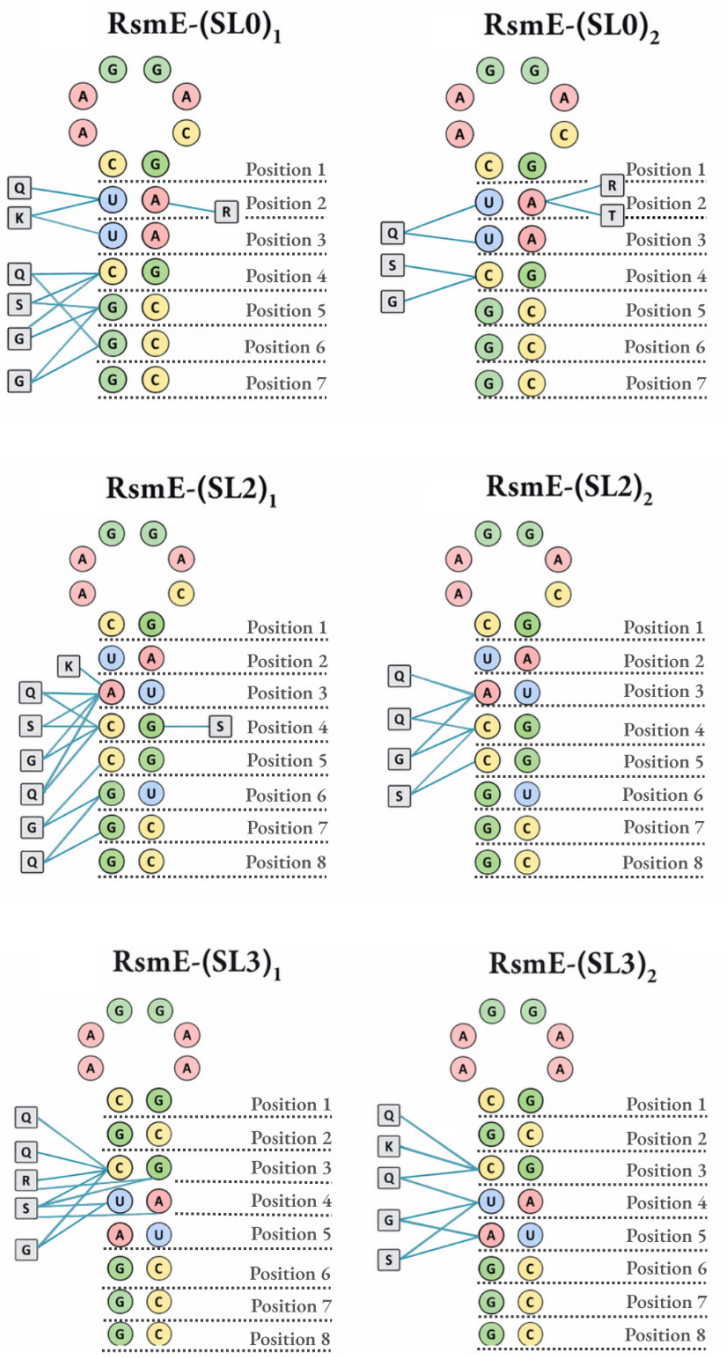

Figure S7: Stem nucleotides interacting with RsmE residues in RsmE-$(SLX)_1$ and RsmE-$(SLX)_2$ complexes, with X = 0, 2, 3.

7 Tables

Table S1: Free energies of binding (FEB) determined with the MM/GBSA approach for complexes RsmE-$(SLX)_1$ (FEB_1) and RsmE-$(SLX)_2$ (FEB_2). The reported uncertainties indicate the statistical error of the mean ($\sigma/\sqrt{(N)}$), where $N = 400$ is the number of samples used in the calculation. Note that, as the MM/GBSA computations do not include entropic effects they are better considered as estimations of the binding energies.

	FEB_1 (Kcal/mol)	FEB_2 (Kcal/mol)
SL0	-121.5±0.6	-85.4±0.6
SL2	-124.4±0.6	-97.4±0.5
SL3	-79.5±0.5	-71.3±0.9

Table S2: SL0–RsmE contacts in RsmE–$(SL0)_1$. (a) SL0 nucleotide ID and base-pair position relative to the loop-closing pair; (b) RsmE residue ID; (c) fraction of snapshots in which the contact is observed. Only contacts with a total fraction greater than 1% are shown. The subscript (A or B) on each RsmE residue indicates the dimer chain to which it belongs.

$SL0^{(a)}$	$RsmE^{(b)}$	Total $Fraction^{(c)}$
A15 - Position 2	$R31_B$	20.0975
C4 - Position 4	$S26_B$	16.0450
	$G27_B$	11.8750
	$Q28_B$	9.7575
G3 - Position 5	$G27_B$	11.6975
	$G27_A$	4.4550
	$S26_B$	3.3025
U6 - Position 2	$Q29_B$	10.7825
	$K7_A$	4.6500
G2 - Position 6	$Q28_A$	7.5225
	$G27_A$	3.2100
U5 - Position 3	$Q29_B$	5.0050

Table S3: SL2–RsmE contacts in RsmE–$(SL2)_1$. (a) SL2 nucleotide ID and base-pair position relative to the loop-closing pair; (b) RsmE residue ID; (c) fraction of snapshots in which the contact is observed. Only contacts with a total fraction greater than 1% are shown. The subscript (A or B) on each RsmE residue indicates the dimer chain to which it belongs.

SL2[(a)]	RsmE[(b)]	Total Fraction[(c)]
C5 - Position 4	$G27_B$	28.0750
	$S26_B$	17.6025
	$Q28_B$	6.0725
	$Q29_B$	1.8350
A6 - Position 3	$Q29_B$	26.6725
	$S26_B$	18.2775
	$Q28_B$	11.8775
	$G27_B$	3.7625
	$K7_A$	2.6600
C4 - Position 5	$G27_A$	4.5150
G2 - Position 7	$Q28_A$	4.1125
G18 - Position 4	$S26_B$	4.0225
G3 - Position 6	$Q28_A$	3.3775
	$G27_A$	1.8950

Table S4: SL3–RsmE contacts in RsmE–$(SL3)_1$. (a) SL3 nucleotide ID and base-pair position relative to the loop-closing pair; (b) RsmE residue ID; (c) fraction of snapshots in which the contact is observed. Only contacts with a total fraction greater than 1% are shown. The subscript (A or B) on each RsmE residue indicates the dimer chain to which it belongs.

SL3[(a)]	RsmE[(b)]	Total Fraction[(c)]
C6 - Position 3	$Q29_B$	27.3850
	$S26_B$	16.9975
	$Q28_B$	4.6275
	$G27_B$	1.8375
	$R31_B$	1.2550
U5 - Position 4	$G27_B$	12.5175
	$S26_B$	3.3075
G16 - Position 3	$R31_B$	5.4000
	$S26_B$	4.4000
A17 - Position 4	$S26_B$	5.0425

Table S5: SL0–RsmE contacts in RsmE–$(SL0)_2$. (a) SL0 nucleotide ID and base-pair position relative to the loop-closing pair; (b) RsmE residue ID; (c) fraction of snapshots in which the contact is observed. Only contacts with a total fraction greater than 1% are shown. The subscript (A or B) on each RsmE residue indicates the dimer chain to which it belongs.

SL0[(a)]	RsmE[(b)]	Total Fraction[(c)]
A15 - Position 2	$R31_A$	18.3975
	$T5_B$	1.9600
C4 - Position 4	$G27_A$	5.4750
	$S26_A$	3.2700
U6 - Position 2	$Q29_A$	3.8150
U5 - Position 3	$Q29_A$	1.2475

Table S6: SL2–RsmE contacts in RsmE–$(SL2)_2$. (a) SL2 nucleotide ID and base-pair position relative to the loop-closing pair; (b) RsmE residue ID; (c) fraction of snapshots in which the contact is observed. Only contacts with a total fraction greater than 1% are shown. The subscript (A or B) on each RsmE residue indicates the dimer chain to which it belongs.

SL2[(a)]	RsmE[(b)]	Total Fraction[(c)]
C5 - Position 4	$G27_A$	32.1075
	$Q28_A$	11.1000
	$S26_A$	5.8675
A6 - Position 3	$Q29_A$	18.5925
	$Q28_A$	15.7775
	$G27_A$	4.7400
C4 - Position 5	$S26_A$	1.1650

Table S7: SL3–RsmE contacts in RsmE–$(SL3)_2$. (a) SL3 nucleotide ID and base-pair position relative to the loop-closing pair; (b) RsmE residue ID; (c) fraction of snapshots in which the contact is observed. Only contacts with a total fraction greater than 1% are shown; (d) SL3 nucleotide ID in the alternative RNA construct. The subscript (A or B) on each RsmE residue indicates the dimer chain to which it belongs.

SL3[(a)]	RsmE[(b)]	Total Fraction[(c)]
A4 - Position 5	$G27_A$	16.9925
	$S26_A$	2.3125
U5 - Position 4	$G27_A$	12.5500
	$Q28_A$	9.6825
	$S26_A$	3.1525
C6 - Position 3	$Q29_A$	9.0600
	$K7_B$	6.2175
	$Q28_A$	2.7000
$SL3_a$[(a)]	$SL3_b$[(d)]	Total Fraction[(c)]
A4 - Position 5	G2 - Position 7	5.7175
	G3 - Position 6	4.2550
G3 - Position 6	G3 - Position 6	5.1300
U5 - Position 4	G2 - Position 7	2.6325