Supplemental Material:

Predicting Structure and Stability for RNA Complexes with Intermolecular Loop-loop Base Pairing

Song CAO, Xiaojun XU and Shi-Jie CHEN
Department of Physics and Department of Biochemistry
University of Missouri, Columbia, MO 65211

[1] Virtual bond representation and loop entropy calculation

Due to the rotameric nature of RNA backbone torsional angles, the original six-bond nucleotide can be reduced to a two-bond system, where each bond is a virtual bond (Fig.1). In the Vfold model we compute the conformational entropy by enumerating the virtual bond conformations (Fig. 1). In the current version of the Vfold model, the virtual bonds are configured on a diamond lattice with three equiprobable torsional angles (60°, 180°, 300°) and fixed bond length of 3.9 $\mathring{A}$ and bond angle of 109.47^o.

An advantage of the model for the loop entropy calculation is the ability to account for chain connectivity, excluded volume effect and completeness of the conformational ensemble. With the Vfold model, we have computed the entropy parameters for several RNA motifs, such as hairpin loops, internal loops, bulge loops, H-type pseudoknots and hairpin-hairpin kissing motifs:

1. Hairpin, internal and bulge loops, in Ref [1];
2. Pseudoknots without inter-helix junction, in Ref [2];
3. Pseudoknots with inter-helix junction, in Ref [3];
4. Hairpin-hairpin kissing, in Ref [4].

[2] Partition function

Based on a recursive algorithm[1−4], we compute the partition function for all the possible secondary and pseudo-knotted structures from the sequence. For each given structure, the helix free energy is computed from the empirical thermodynamic parameters of the base stacks (Turner rules) and the loop free energies are calculated from the Vfold model.

For the loop treatment, the Vfold model distinguishes itself from other models by accounting for the intraloop mismatched base stacking interactions. The formation of intraloop stacks can cause significant constraint in loop conformation and a resultant reduction of conformational entropy. For given intraloop stacks, We use the Turner rules to determine the free energy of the mismatched base stacks and use the Vfold model to compute the loop entropy. Depending on the loop sequence, such Vfold-predicted loop entropy parameters could be different from the experimentally measured loop entropy parameters, which is the Boltzmann average over all the possible intraloop base stacks. The partition function of a loop (see Fig. S1) is calculated as the sum over all the possible intra-loop mismatched base stacks. Such Vfold-predicted loop free energy can be temperature and sequence dependent.

From the partition function $Q(T)$, we can calculate the heat capacity C(T) melting curves: $C(T) = \frac{\partial}{\partial T}[k_B T^2 \frac{\partial}{\partial T} \ln Q(T)]$. Furthermore, from the conditional partition function $Q(i, j, T)$ for the ensemble of conformations with base pair (i, j), we compute the probability $P_{ij}(T)$ for the formation of the (i, j) pair: $P_{ij}(T) = Q(i, j, T)/Q(T)$. From the distribution of the base-pairing probability, we can deduce the stable structures at temperature T.

The computational time scales with the chain length N as $O(N^6)$ and the memory scales as $O(N^2)$.

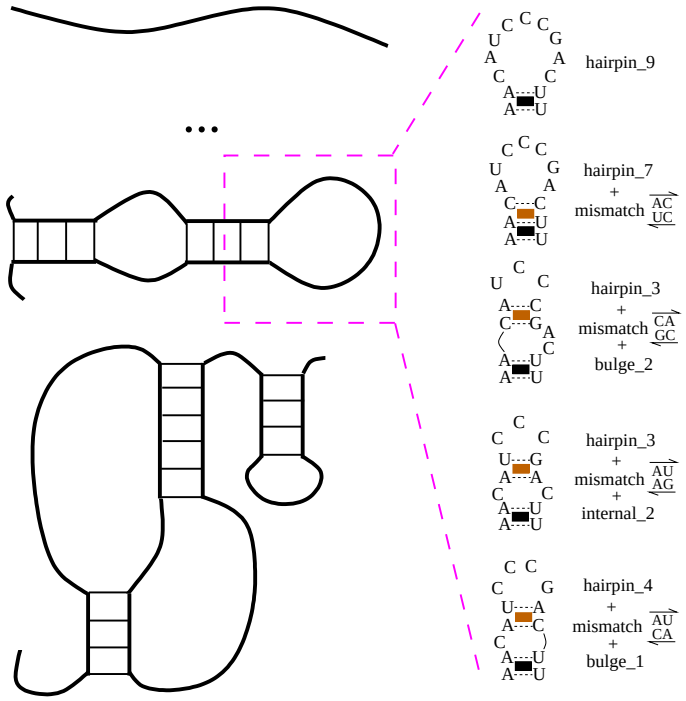

Figure S1: In the Vfold model, we enumerate all the possible structures (see the left panel) that contain secondary/pseudoknot structures using a recursive algorithm. Helix free energies are calculated from the Turner's thermodynamic parameters and loop free energies are computed from the partition function over all the possible arrangements of the intra-loop mismatched base stacks for a given loop sequence. Here, we show an example of a hairpin of size 9 nts closed by a A-U base pair. The ensemble of the loop conformations contains 5 different arrangements of mismatched base stacks within the loop, as shown in the right panel.

[3] Additional experimental tests for the Vfold model

We calculated the stabilities for three small RNA systems[5] using the Vfold model and compare the results with the experimental data as well as the RNAfold predictions. The stacking energy parameters used in Vfold is from the Turner rule with 1M NaCl, while the experimental data is for the 150 mM NaCl and 10 mM Na_2HPO_4. Another possible source for the inaccuracy of the models may come from the intraloop noncanonical interactions (beyond mismatched base stacks). A more accurate model should consider these effects.

Overall, RNAfold and Vfold give close results for the stabilities of the bistable RNAs. For sequence 4, the energy difference between the two (bistable) states is 1.73 k_BT (from the experimental result and from the RNAfold), while Vfold predicts 0.41 k_BT. The slight difference between Vfold and RNAfold for the three RNAs may come from the different energy parameters for the loops.

Table S1: Comparison of the stabilities predicted by the Vfold with the experimental results and the RNAfold results for three bistable RNAs [5]. Sequence 3, 4 and 5 denote the RNAs (no. 3, 4 and 5) studied in Ref [5]. The ratios n:m in the table are the relative populations of the bistable states as determined from the experiment, the RNAfold model and the Vfold model, respectively.

Sequence	Exp. ratio	RNAfold	Vfold
3	50:50	8:92	22:78
4	85:15	85:15	60:40
5	45:55	21:79	21:79

Table S2: The size-dependent (sequence-independent) entropy parameters ($\Delta S/k_B$, which is dimensionless) for hairpin, bulge and internal loops derived from the Vfold model.

Size	Hairpin	Bulge	Internal
1	-	-5.87	-
2	-	-6.64	-8.74
3	-7.28	-7.47	-9.19
4	-7.35	-7.81	-9.14
5	-7.90	-8.25	-9.27
6	-8.09	-8.51	-9.34
7	-8.41	-8.78	-9.45
8	-8.61	-9.01	-9.55
9	-8.82	-9.21	-9.66

Table S3: Using bulge loop as a simple test for the approximation that we have used in the entropy calculation for a multi-branched kissing loop: replacing the helix terminal base pair in the loop with a single nucleotide. With this approximation, an n-nt bulge loop turns into an $(n+1)$-nt hairpin loop. In the table, $\Delta\Delta S/k_B$ denotes the difference between the entropies of the bulge and of the hairpin loops. $|\Delta\Delta S/\Delta S_{\rm bulge}|$ is the relative error. In our calculation, we use the approximation only for multibranched kissing loops, for which the approximation is necessary for the entropy calculation.. For simple hairpin, bulge and internal loops, we use the rigorous entropy values without using the approximation.

Bulge	Hairpin	$\lvert\Delta\Delta S/k_B\rvert$	$\lvert\Delta\Delta S/\Delta S_{\rm bulge}\rvert$
2	3	0.64	0.096
3	4	0.12	0.016
4	5	0.09	0.011
5	6	0.16	0.019
6	7	0.10	0.011
7	8	0.17	0.019
8	9	0.19	0.021

Table S4: A simple test similar to the one shown in Table S3. Here we use internal loop instead of bulge loop as the test case. With the appoximation, an internal loop of size n is converted into a hairpin loop of size of n+1.

Internal	Hairpin	$\lvert\Delta\Delta S/k_B\rvert$	$\lvert\Delta\Delta S/\Delta S_{\rm internal}\rvert$
2	3	1.46	0.167
3	4	1.84	0.200
4	5	1.24	0.135
5	6	1.18	0.127
6	7	0.93	0.099
7	8	0.84	0.088
8	9	0.73	0.076

Table S5: A simple test for the effect of helix migration on loop entropy. The entropy difference $\Delta\Delta S/k_B$ between a bulge loop and an internal loops of the same size. The entropy parameters $\Delta S/k_B$ are from Table S2. The third column shows the relative error.

Size	$\lvert\Delta\Delta S/k_B\rvert$	$\lvert\Delta\Delta S/\Delta S_{\mathrm{internal}}\rvert$
2	2.10	0.240
3	1.72	0.187
4	1.33	0.145
5	1.02	0.110
6	0.83	0.088
7	0.67	0.070
8	0.54	0.056

Table S6: A simple test for the effect of helix migration on loop entropy, specifically, for the effect of relative position (l_1, l_2) between the helices in an internal loop with size l_1 + l_2 = 6. $\Delta\Delta S/k_B$ is the entropy difference between the (l_1, l_2) internal loop and the entropy parameter (in Table S2) for a 6-nt internal loop averaged over all the possible (l_1, l_2) values.

l_1	l_2	$\Delta S/k_B$	$\lvert\Delta\Delta S/k_B\rvert$	$\lvert\Delta\Delta S/\Delta S_{\mathrm{average}}\rvert$
1	5	-9.51	0.17	0.018
2	4	-9.33	0.01	0.001
3	3	-9.24	0.10	0.011
4	2	-9.21	0.13	0.014
5	1	-9.39	0.05	0.005
Average		-9.34	0.09	0.010

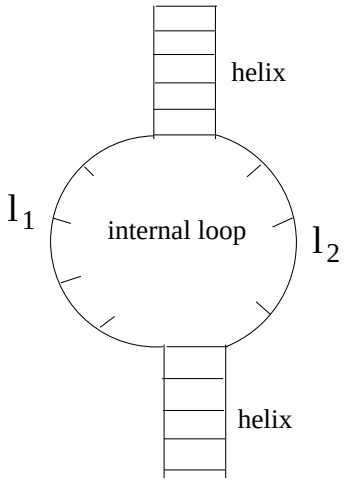

Figure S2: A schematic diagram for an internal loop of size (l_1+l_2).

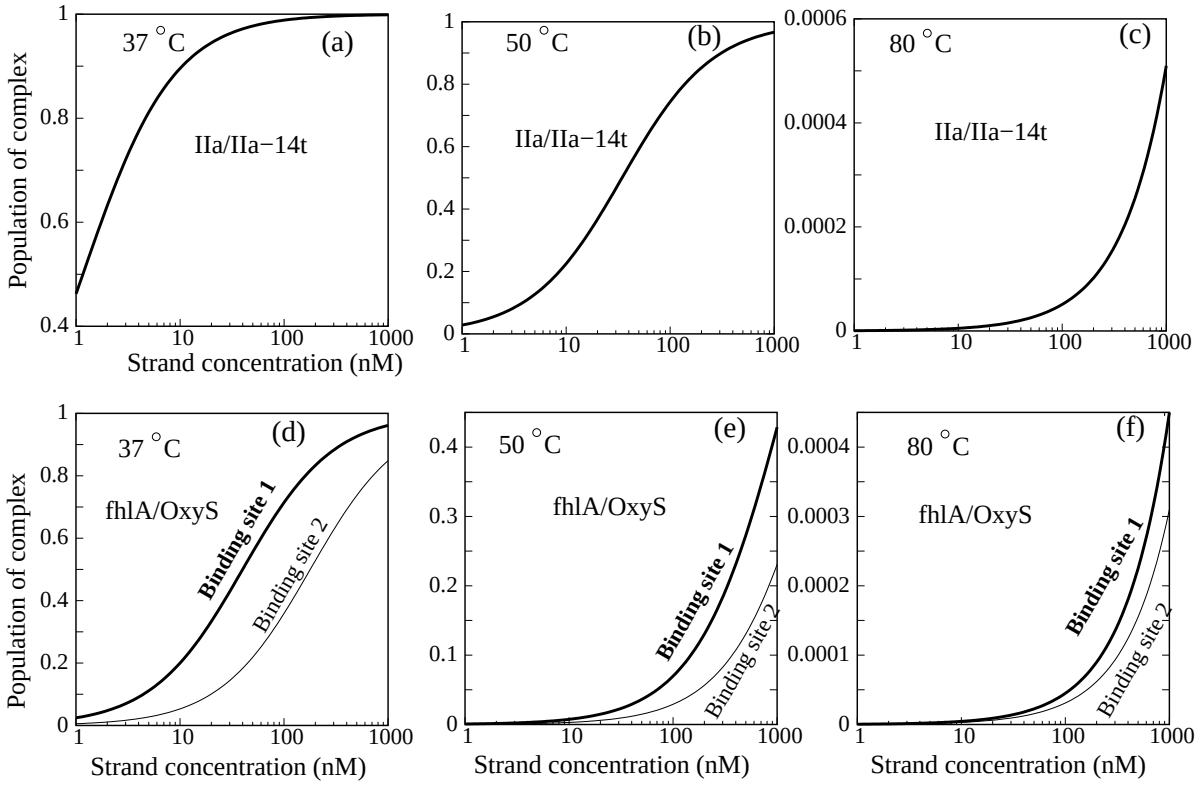

Figure S3: The predicted fractional population of the RNA/RNA complex at the different temperatures. (a), (b) and (c) are the results for the IIa/IIa-14t complex and (d), (e) and (f) are for the fhlA/OxyS complex. The RNA strand concentration ranges from 1 nM to 1000 nM in the calculation.

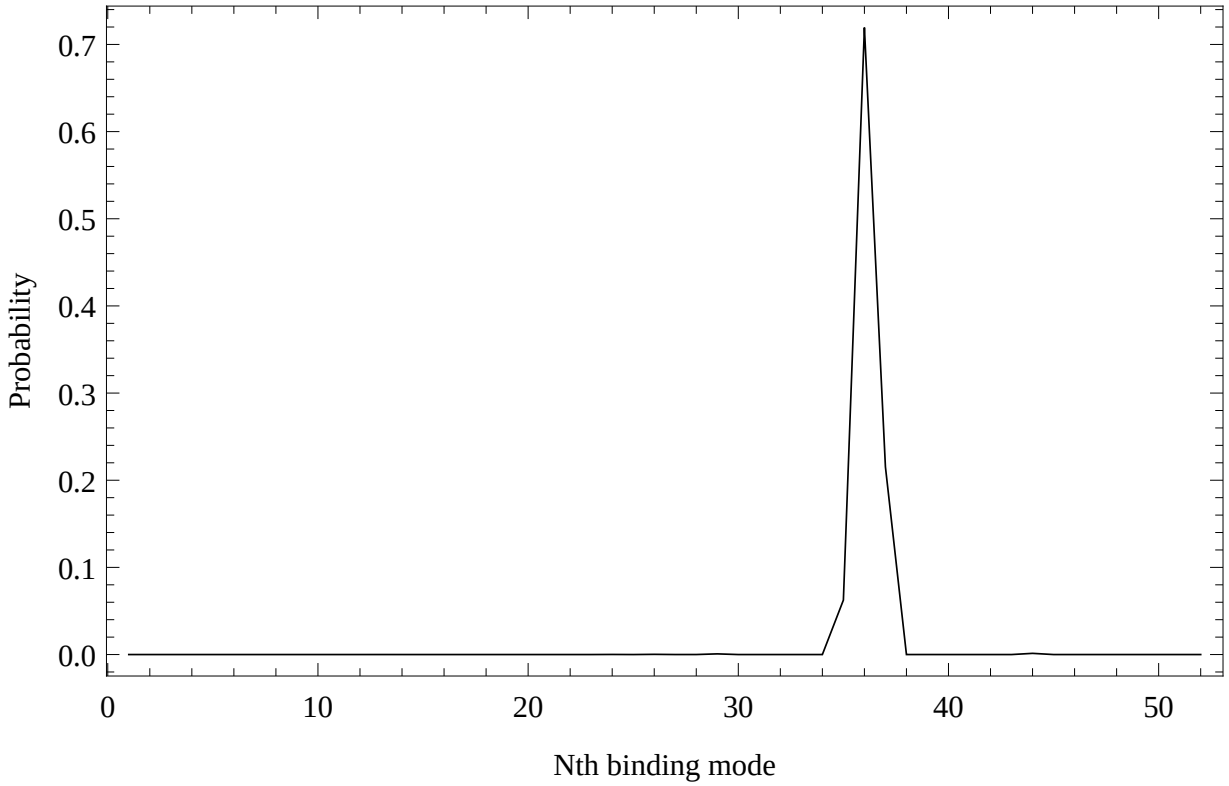

Figure S4: The probability distribution of the 52 binding modes identified by the model for the IIa/IIa-14t complex. The figure shows that there exists a single overwhelmingly dominant binding mode (binding site). This allows us to use the two-step process for structure prediction, namely, to identify the most probable binding site first, then predict the structure of the RNA-RNA complex with the dominant binding site.

References

[1] Cao, S. & Chen, S.-J. (2005). Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA*, **11**: 1884-1897.

[2] Cao, S. & Chen, S.-J. (2006). Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res.*, **34**: 2634-2652.

[3] Cao, S. & Chen, S.-J. (2009). Predicting structures and stabilities for H-type pseudoknots with inter-helix loop. *RNA*, **15**: 696-706.

[4] Cao, S. & Chen, S.-J. (2011). Structure and stability of RNA/RNA kissing complex: with application to HIV dimerization initiation signal. *RNA*, **17**: 2130-2143.

[5] Hobartner C. & Micura R. (2003). Bistable secondary structures of small RNAs and their structural probing by comparative imino proton NMR spectroscopy. *J. Mol. Biol.*, **325**: 421-431.