Supplementary Materials

Predicting Cotranscriptional Folding Kinetics For Riboswitch

Ting-ting SUN[1,2‡], Chenhan ZHAO[2‡] and Shi-Jie CHEN[2*]

[1]Department of Physics, Zhejiang University of Science and Technology, Hangzhou 310023, P.R. China

[2]Department of Physics, Department of Biochemistry, and University of Missouri Informatics Institute, University of Missouri, Columbia, MO 65211

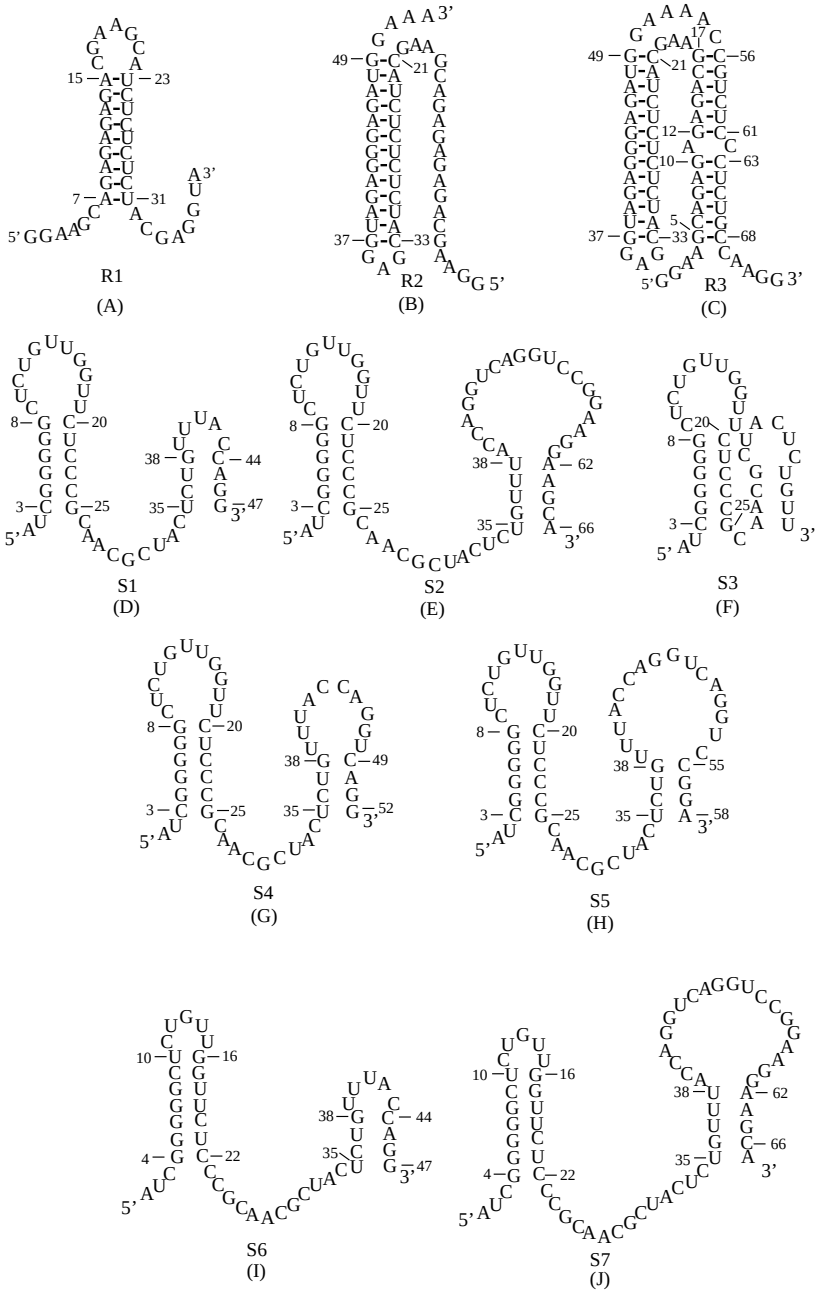

Figure S1: Identified kinetically important conformations for the 73-nt "reverse" RNA and SRP RNA. These conformations are determined from the complete conformational ensemble predictions. (A), (B) and (C) are secondary structures of R1, R2 and R3, which are predicted from the populational kinetics for the "reverse" sequence. (D),(F),(G),(F),(I) and (J) are the secondary structures of S1, S2, S3, S4, S5, S6 and S7, respectively. S1-S7 are predicted from the populational kinetics for the 90-nt SRP RNA.

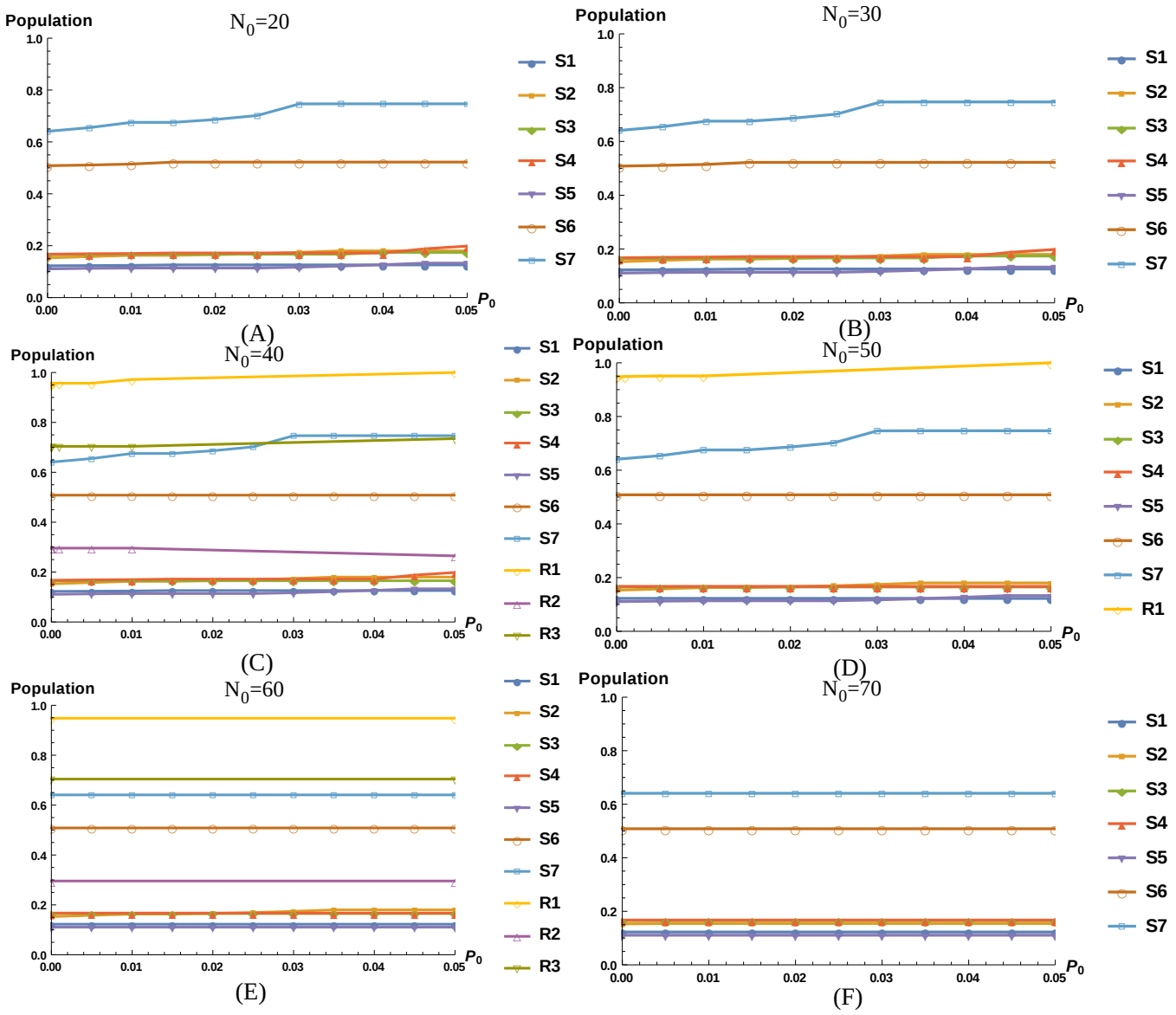

Figure S2: Test of the CRKR method for a 73-nt "reverse" RNA sequence[1] and 90-nt SRP RNA.[2] We show the predicted populations of the kinetically important states, namely, the states emerging with non-negligible populations, in the different step of cotranscriptional folding. The same kinetically important conformations are identified from both the CRKR method and the original complete conformation ensemble-based method.[3] The figure shows the results with the different population resampling-off parameter P_0 and the different minimum chain length N_0 for which CRKR is applied. We find that the population resampling approach gives nearly the same results as the complete conformational ensemble[3] for $N_0 = 30, 40, 50$ and P_0 below 1%, and for $N_0 = 60$, P_0 below 5%. For the 102-nt *pbuE* riboswitch and 117-nt *E.Coli.* SRP RNA, we set $N_0 = 50$ and P_0=0.5%.

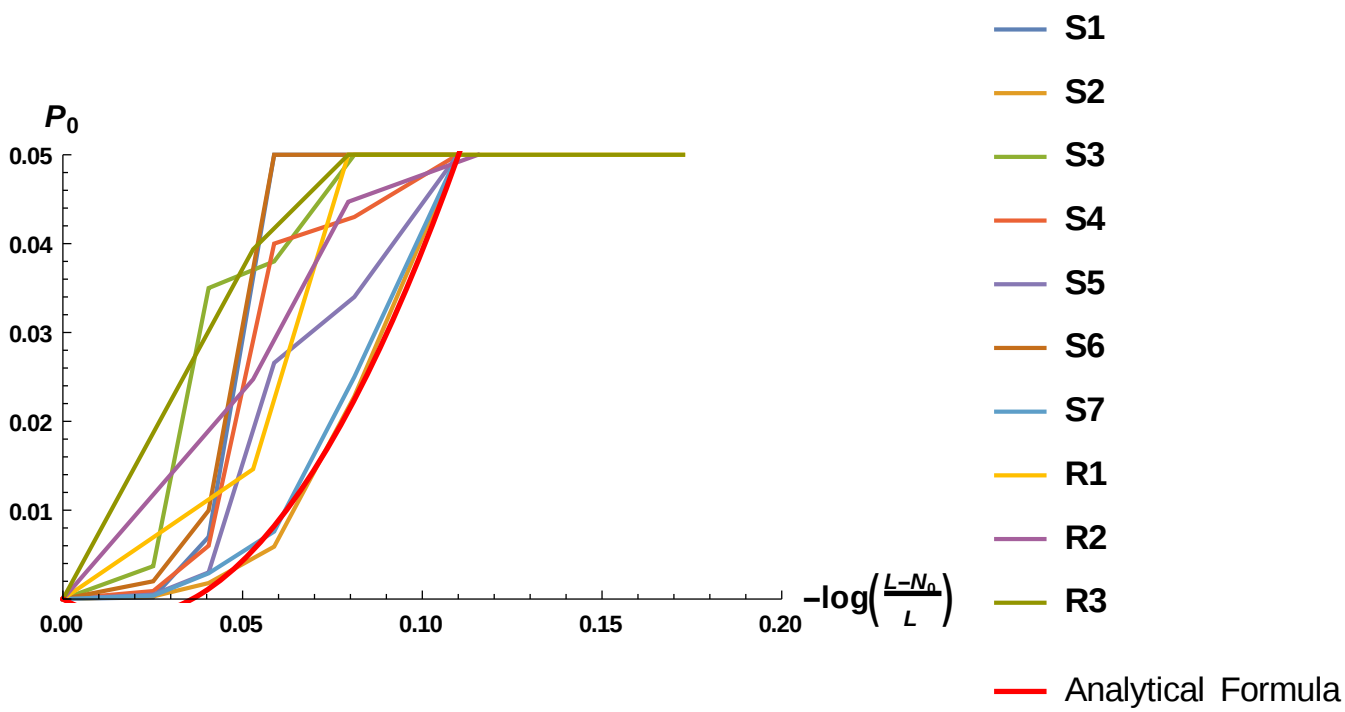

Figure S3: Parameter set for CRKR approach. In this figure, each curve indicates the minimum P_0 determined from the populational kinetics test at the different N_0s for each kinetically important conformation. From the test curves, we fitted a red curve which can give parameter sets for other sequences.

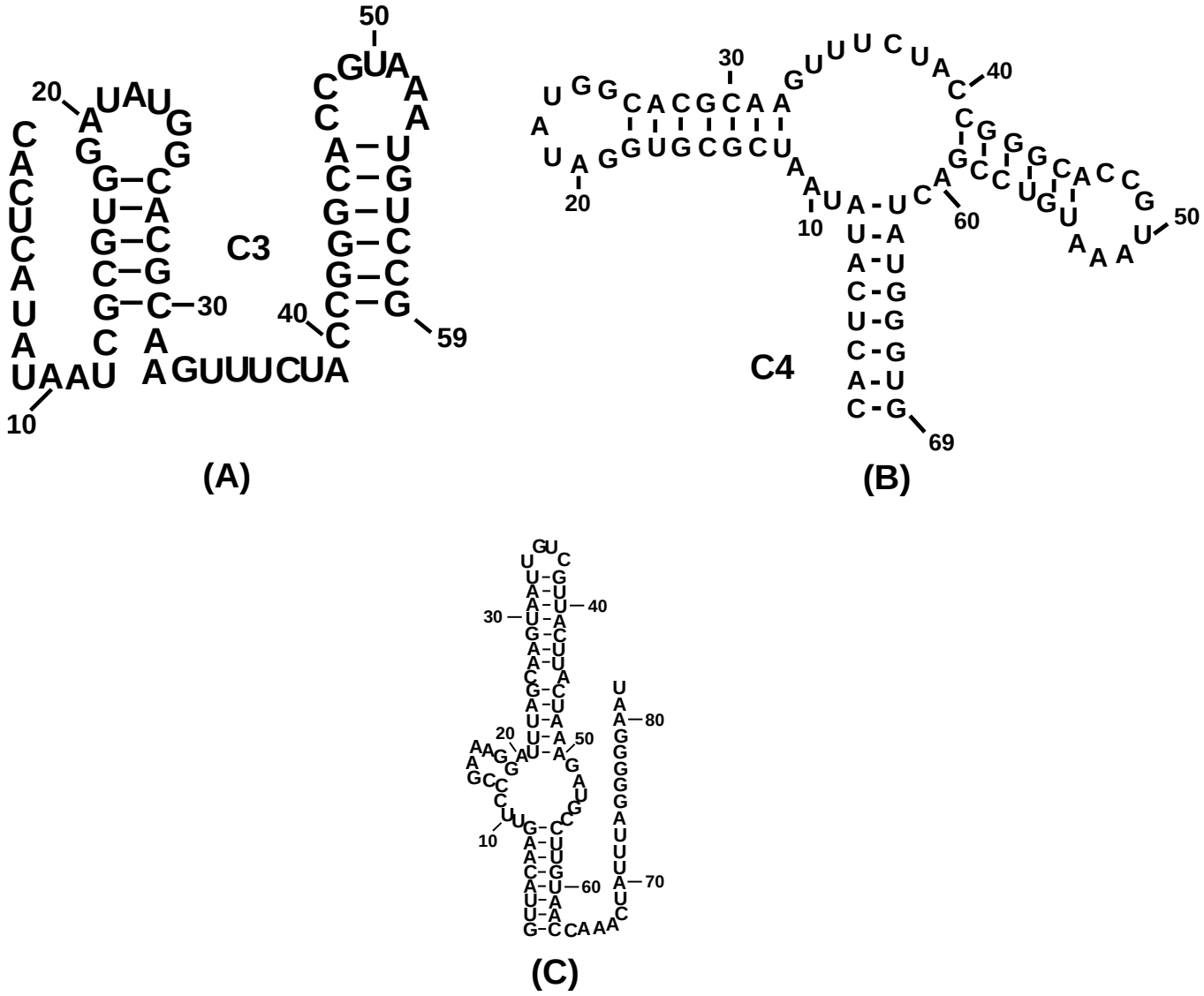

Figure S4: Kinetically important conformations used in the test of the CRKR method. (A) and (B) are the secondary structures of the intermediate state C3 and the final structure C4 for the 69-nt *xpt-pbuX* riboswitch, respectively.[4] (C) shows the secondary structure of the 82-nt *Enterococcus faecalis* SAM-III (S_{mk}) riboswitch.[5]

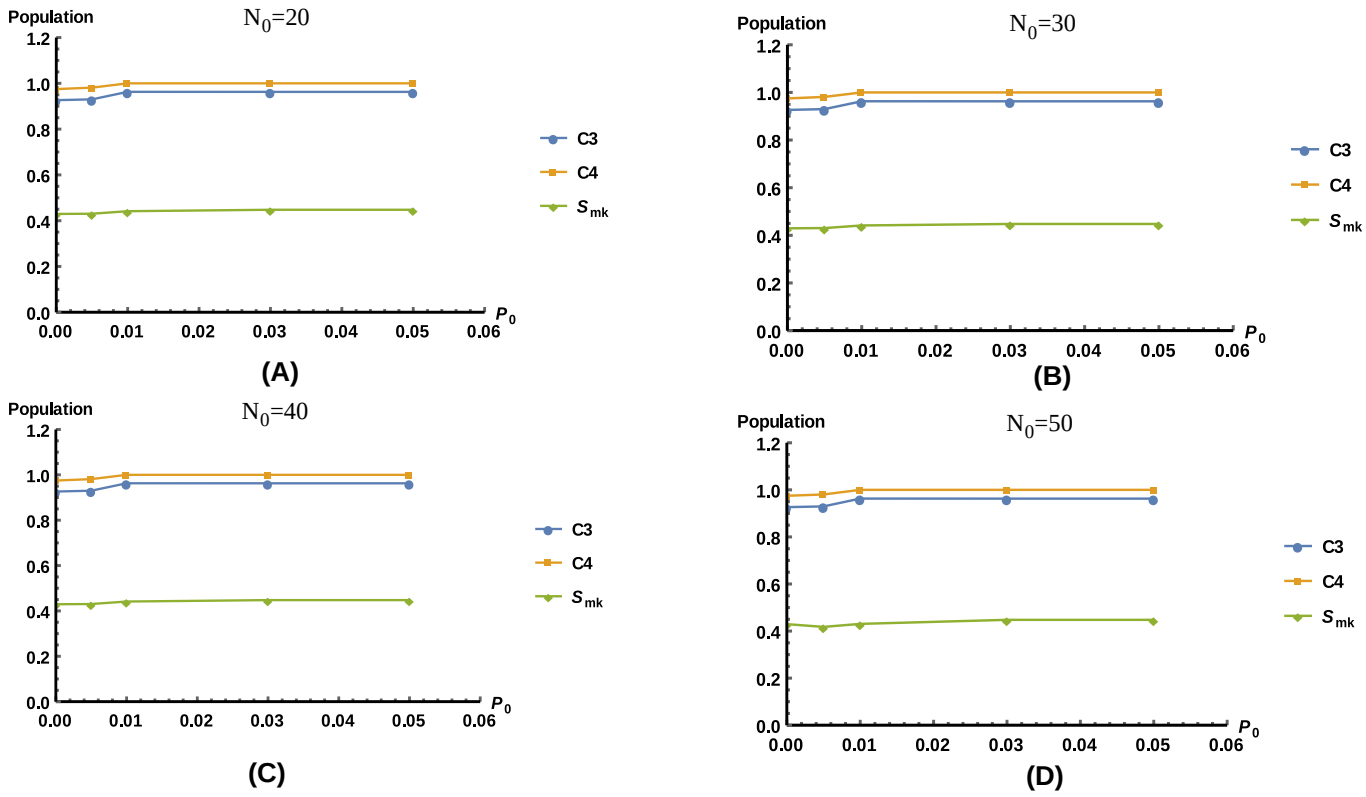

Figure S5: Test of the CRKR method using the 69-nt *xpt-pbuX* riboswitch and the 82-nt *Enterococcus faecalis* SAM-III (S_{mk}) riboswitch. (A), (B), (C), and (D) show the comparisons between the populations calculated from the CRKR method and that from the complete conformational ensemble. We apply the CRKR method to the two riboswitches with different N_0 and P_0 values. The plot shows the populations of the intermediate state C3 at step 59, the final structure C4 at step 69, and the structure S_{mk} at step 82, respectively. As shown in (A), (B), (C), and (D), the results from the CRKR method are consistent with those from the complete ensemble (N_0=0, P_0=0). The CRKR method reduces the computational time and the number of states. For example, with the CRKR method, the number of states for the *xpt-pbuX* riboswitch at step 59 is reduced from 140 to 4 with N_0=20, P_0=0.01. The computational time for *Enterococcus faecalis* SAM-III (S_{mk}) riboswitch is reduced from 6 hours to 4 minutes with N_0=50, P_0=0.01.

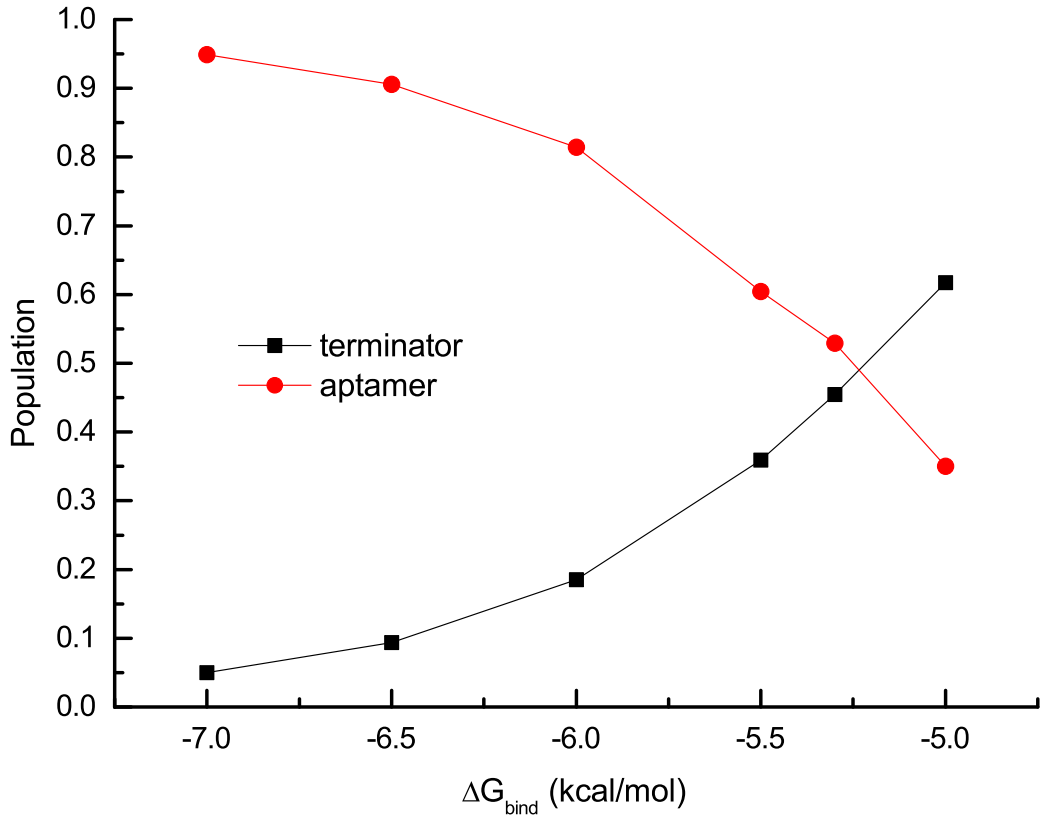

Figure S6: Population of the aptamer and the terminator as a function of the ligand binding-induced additional stability ΔG_{bind} for the aptamer with no pulling force. Force spectroscopy experiment shows that the termination level is decreased to 50% in the presence of adenine. We choose the free energy parameter ΔG_{bind} to be -5.3 kcal/mol, which is close to the experimentally determined value.

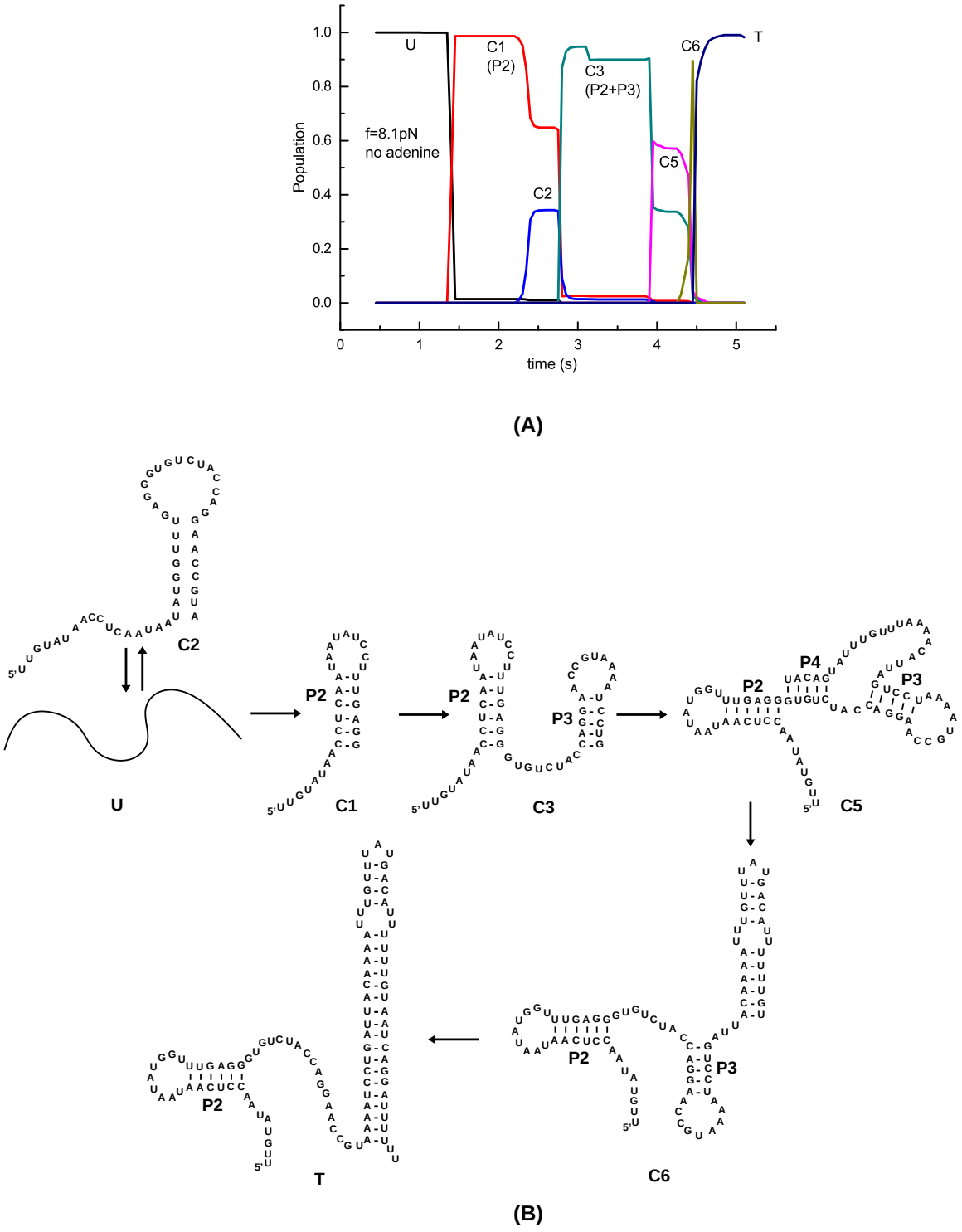

Figure S7: (A) The population kinetics for the cotranscriptional folding of the *pbuE* riboswitch under 8.1 pN pulling force and transcription speed 20 nt/s without the presence of adenine. There are 6 kinetically important states identified during the folding process, U (Black), C1 (Red), C2 (Blue), C3 (Dark Cyan), C4 (Magenta), and T (Dark Yellow). (B) The folding pathway inferred from the populational kinetics.

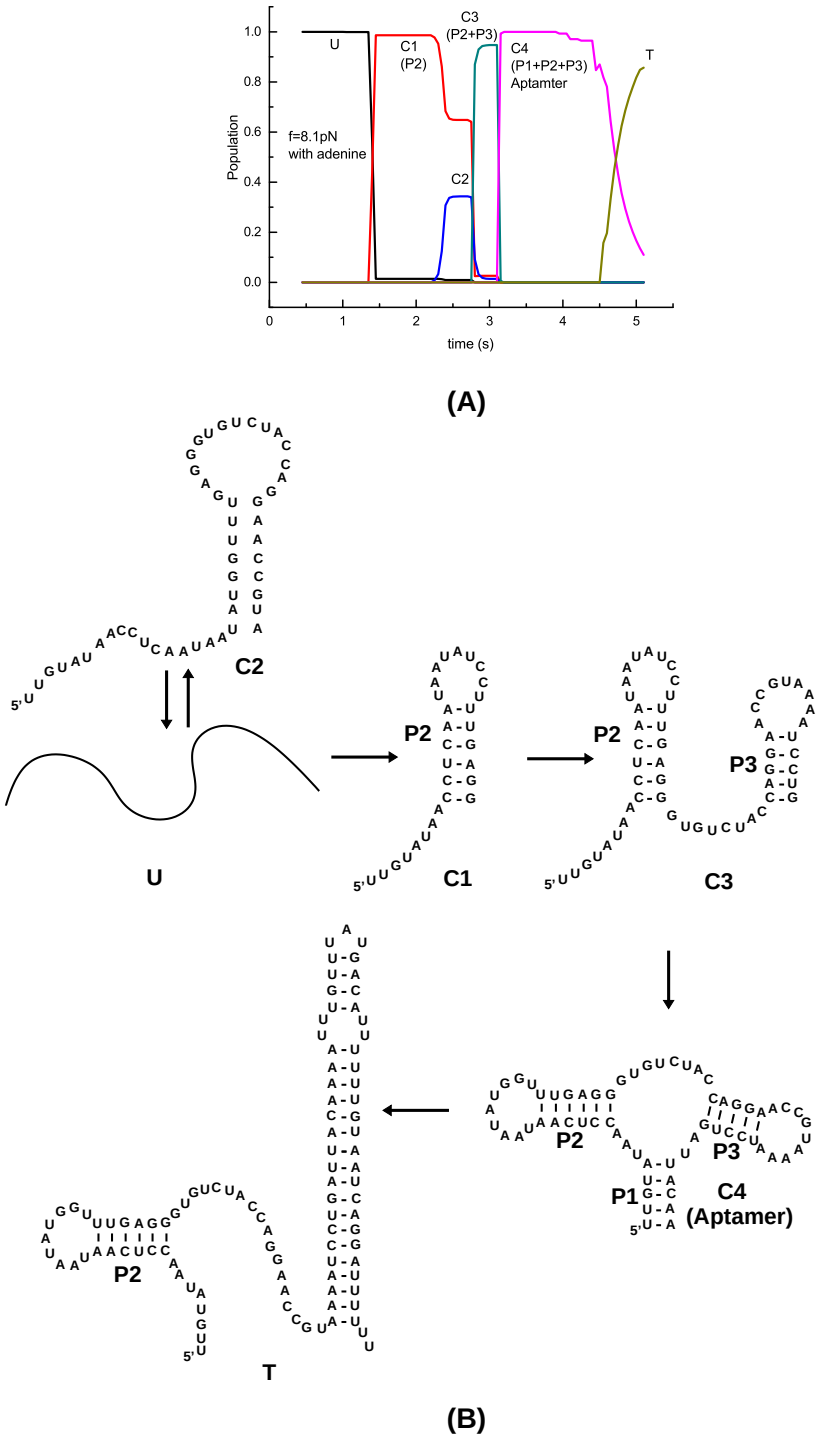

Figure S8: (A) The population kinetics for the cotranscriptional folding of the *pbuE* riboswitch under 8.1 pN pulling force and transcription speed 20 nt/s with the presence of adenine. There are also 6 kinetically important states identified during the folding process, U (Black), C1 (Red), C2 (Blue), C3 (Dark Cyan), C4 (Magenta), and T (Dark Yellow). (B) The folding pathway inferred from the populational kinetics.

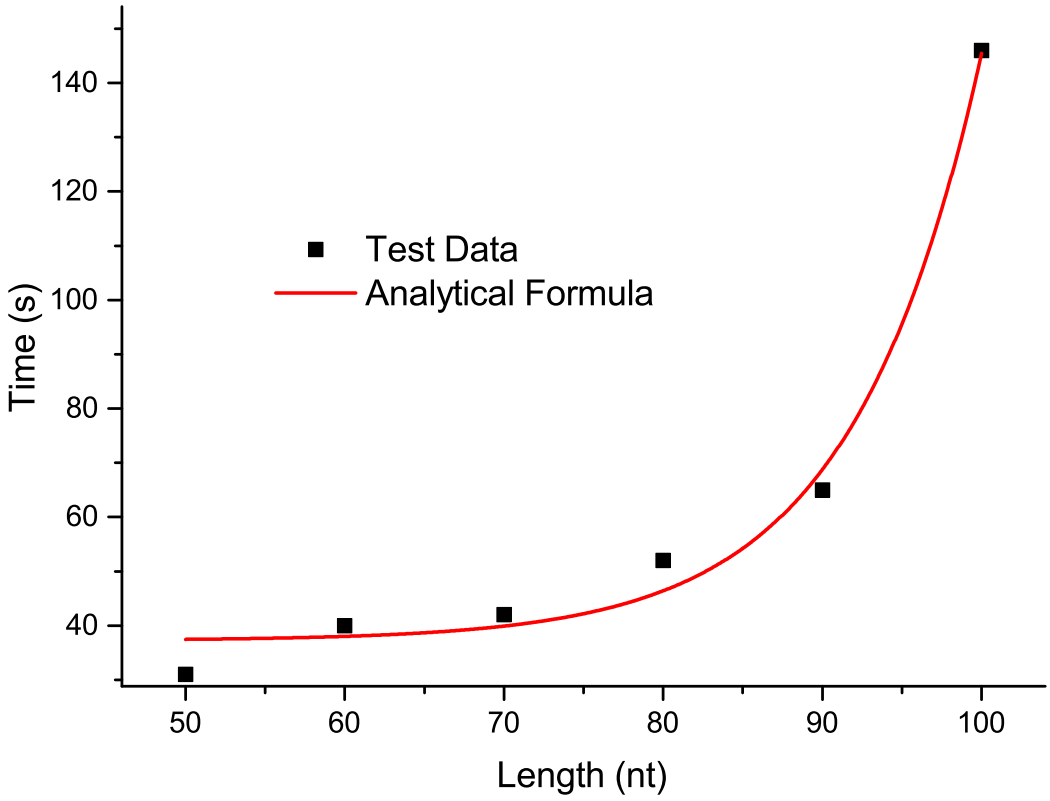

Figure S9: The computational time as a function of the chain length with the CRKR method. Red dots in the figure show the required computer time in order to make reliable predictions for the cotranscriptional folding kinetics for the give sequence length. The blue curve represents the fitted analytical formula. For (red) test data, we calculate the cotranscriptional folding for the different length L of the *E. Coli.* SRP RNA and for the different (N_0, P_0) parameters, we choose the one that can give the minimum time t required for an accurate prediction. For each length, we compare the population of the most populated state in the final step calculated from the complete ensemble (N_0=0, P_0=0) with the one calculated from the CRKR method. If the predicted population of a state is in the range of $[1 \pm (\frac{N_0}{L})^4]\, P_{\text{exact}}$, where P_{exact} is the predicted population based on the complete conformational ensemble, we consider the result as a successful prediction. For L=50 nt, P_0=1% and N_0=20 nt gives the minimum time t=31 s. For L=60 nt, P_0=5% and N_0=30 nt gives the minimum t=40 s. For L=70 nt, P_0=0.5% and N_0=30 nt gives the minimum t=42 s. For L=80 nt, P_0=0.5% and N_0=30 nt gives the minimum t=52 s. For L=90 nt, P_0=3% and N_0=40 nt gives the minimum t=65 s. For L=100 nt, P_0=1% and N_0=40 nt gives the minimum t=146 s. The analytical formula $t = 37.2429 + 4.72 * 10^{-4} * e^{(0.1234*L)}$ (sec) is fitted according to the test data. All calculations are performed on a PC with Intel(R) Xeon(R) CPU E5-1650 v3 CPU and 64 GB RAM.

Table S1: Computer (in seconds) on a PC with Intel(R) Xeon(R) CPU E5-1650 v3 CPU and 64 GB RAM for the different sequence length L of the *E. Coli.* SRP RNA with the different P_0 and N_0 parameters.

N_0 \ P_0	0.5%	1%	3%	5%	10%	20%	30%
L=50 nt							
20 nt	32	31	32	31	32	31	32
30 nt	33	32	32	31	32	32	32
40 nt	37	37	37	36	37	37	37
L=60 nt							
20 nt	41	41	40	41	39	39	39
30 nt	41	42	41	40	41	40	41
40 nt	46	46	46	45	45	45	44
50 nt	262	262	260	262	263	269	268
L=70 nt							
20 nt	44	47	43	41	42	41	41
30 nt	42	43	42	43	43	42	42
40 nt	49	49	50	48	47	48	47
50 nt	282	280	274	275	275	275	275
60 nt	1977	1979	1983	1977	1974	1988	1988
L=80 nt							
20 nt	58	52	50	49	50	49	49
30 nt	52	52	50	49	51	52	48
40 nt	57	57	56	54	54	51	52
50 nt	290	288	282	283	284	283	285
60 nt	2008	1997	1996	1995	1995	1994	1994
70 nt	10970	10974	10968	10962	10960	10973	10961
L=90 nt							
20 nt	75	67	60	58	55	54	57
30 nt	77	70	61	59	60	56	56
40 nt	82	71	65	62	62	60	60
50 nt	315	304	297	290	291	291	288
60 nt	2019	2015	2007	1997	2003	2002	2003
70 nt	10988	10962	10964	10947	10960	10969	10959
80 nt	51442	52969	52910	52943	52804	525000	52474
L=100 nt							
20 nt	199	141	80	73	71	64	65
30 nt	199	142	80	73	72	65	66
40 nt	203	146	85	77	76	702	68
50 nt	438	379	314	307	304	301	298
60 nt	2140	2087	2030	2024	2018	2006	2008
70 nt	11107	11047	10973	10973	10909	10911	10882
90 nt	205520	214443	216114	215683	214963	214377	216603

Conformational sampling and 3D structure prediction

For a given 2D structure, we first extract the motifs such as helices, hairpin loops, bulge loops from the 2D structure. For each non-helix motif we search our database VfoldMTF (http://rna.physics.missouri.edu/vfoldMTF/index.html) for the best template to identify the appropriate template structures. We then assemble the templates to construct the 3D scaffold. For motifs without available templates, we use a coarse-grained (CG) RNA model[6] where the Pyrimidines/purines are represented by 4/5 CG beads, respectively, and a knowledge-based force field to simulate the 3D structure. The simulation is carried out with Langevin dynamics in the modified LAMMPS packages.[7] The time step for the simulation is 0.5 fs, and each structure is simulated for 2000000 steps. We cluster the structures with an RMSD cut-off of 3 Å and choose the cluster center whose cluster has the largest population. Finally, we refine the 3D structure with 10000000 steps of energy minimization.

References

[1] Xayaphoummine, A.; Viasnoff, V.; Harlepp, S.; Isambert, H. Encoding Folding Paths of RNA Switches. *Nucleic Acids Res.* **2007,** *35*, 614-622.

[2] Batey, R. T.; Rambo, R. P.; Lucast, L.; Rha, B.; Doudna, J. A. Crystal Structure of the Ribonucleoprotein Core of the Signal Recognition Particle. *Science.* **2000,** *287*, 1232-1239.

[3] Zhao, P. N.; Zhang, W. B.; Chen, S. J. Cotranscriptional Folding Kinetics of Ribonucleic Acid Secondary Structure. *J. Chem. Phys.* **2011,** *135*, 245101.

[4] Gong, S.; Wang, Y.; Wang, Z.; Zhang, W. Co-Transcriptional Folding and Regulation Mechanisms of Riboswitches. *Molecules.* **2017,** *22*, 1169.

[5] Priyakumar, U. D. Atomistic details of the ligand discrimination mechanism of S MK/SAM-III riboswitch. *J. Phys. Chem. B.* **2010,** *114*, 9920-9925.

[6] Zhang, D.; Chen, S. J. IsRNA: An Iterative Simulated Reference State Approach to Modeling Correlated Interactions in RNA Folding. *J Chem Theory Comput.* **2018,** *14*, 22302239.

[7] Plimpton, S. Fast Parallel Algorithms for Short-range Molecular Dynamics. *J. Comput. Phys.* **1995,** *117*, 1-19.