Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

(Supporting Information)

Predicting the aptamer SYL3C-EpCAM complex's structure with the Martini-based simulation protocol

Xu Shang,[†] Zhen Guan,[†] Shuai Zhang,[†,‡] Lulin Shi,[†,‡] and Haihang You[*,†]

†*State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China*

‡*School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, 100190, China*

E-mail: youhaihang@ict.ac.cn

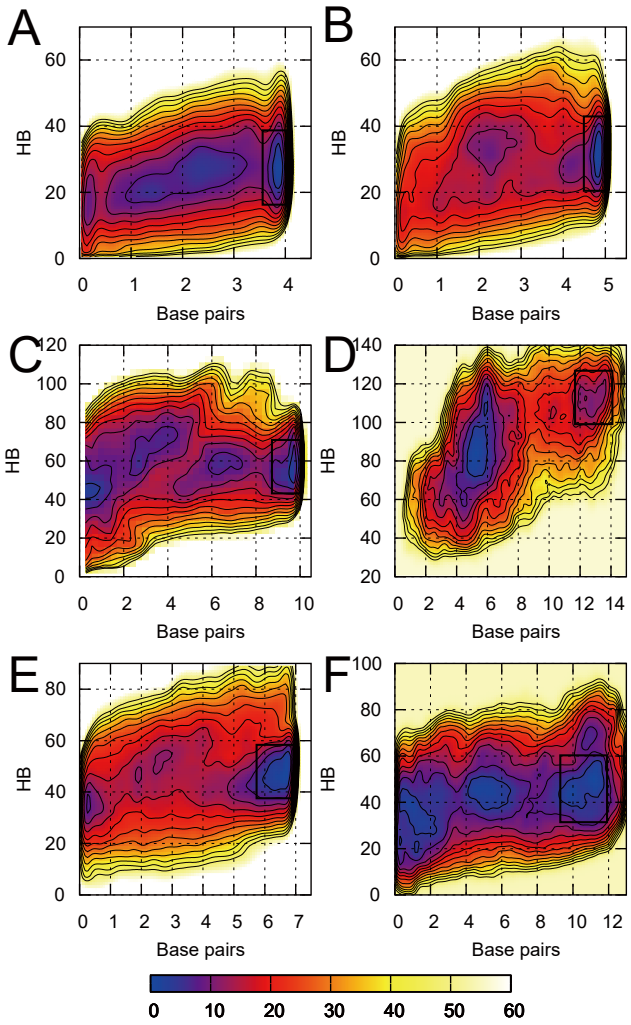

Figure S1: Free energy as a function of the intrachain base pairs and hydrogen bonds (HB) for ssDNAs with PDB IDs: 1BJH (A), 1LA8 (B) , 2L5K (C), 6U82 (D), 1XUE (E) and 1EN1 (F), respectively. In 6U82 both aptamer and protein are contained in this PDB. Here we only used the ssDNA aptamer for structure prediction's verification.

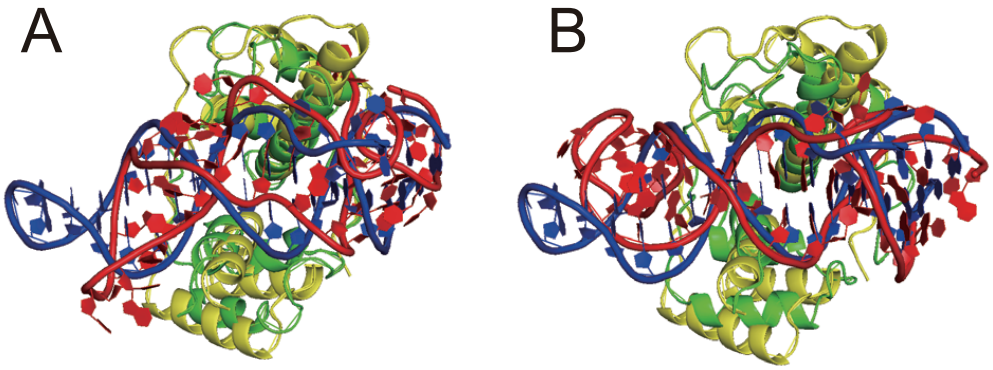

Figure S2: Structure prediction of the DUX4-aptamer complexes from simulations starting from experimental structure modeling aptamer (A) and the prediction structure modeling aptamer (B) coupled with the protein.

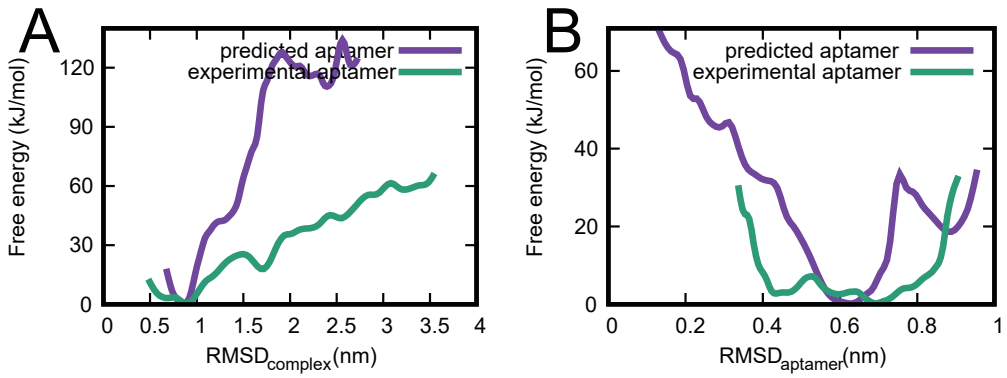

Figure S3: FES as a function of the aptamer-DUX4 complex's RMSD (A) and of the aptamer's RMSD (B) to the corresponding native structures. The FES calculated from simulations starting from the prediction modeling aptamer and experimental modeling aptamer coupled with the protein are colored in purple and green. The Martini aptamer was modeled with a initial structural difference of RMSD 0.545 nm to the native structure. The complex prediction starting from this aptamer structure shows narrower free energy basin of RMSD, but the value at the free energy minima doesn't change much comparing with the result from experimental measurement modeling aptamer. Thus the initial modeling could doesn't influence the bound state much if the modeling structure is reliable enough.

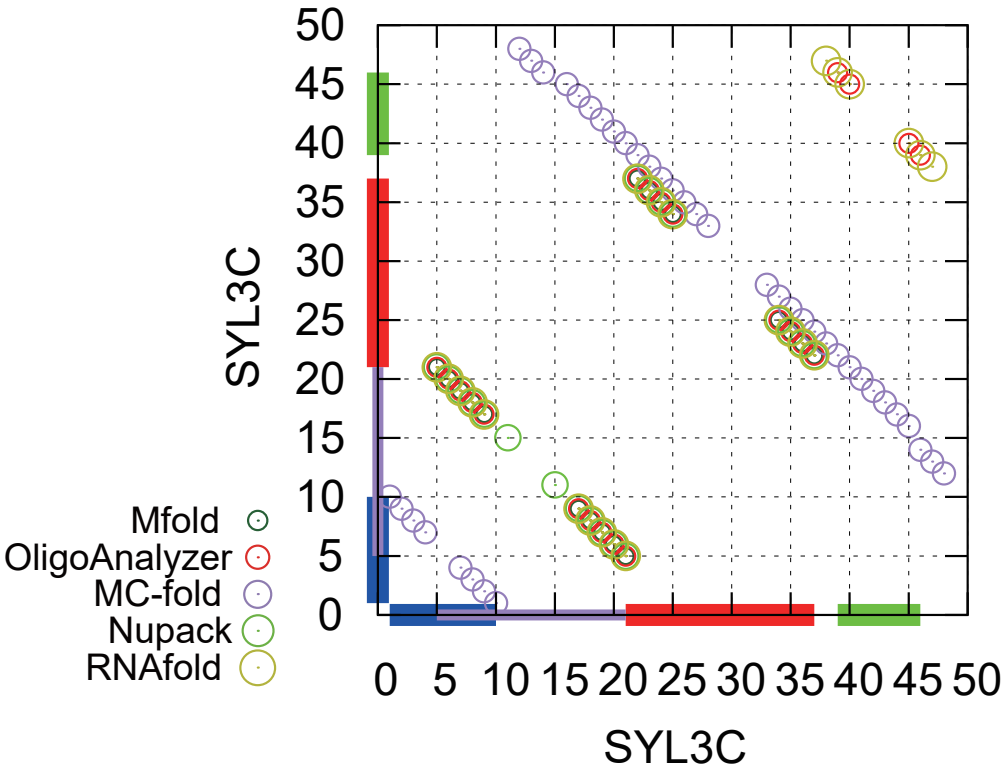

Figure S4: Secondary structure generated by all the selected prediction tools of Mfold (dark green), OligoAnalyzer3.0 (red), MC-fold (purple), Nupack (green) and RNAfold (yellow). The four hairpin loop regions are divided by blue (HP1), purple (HP2), red (HP3) and green (HP4) bars.

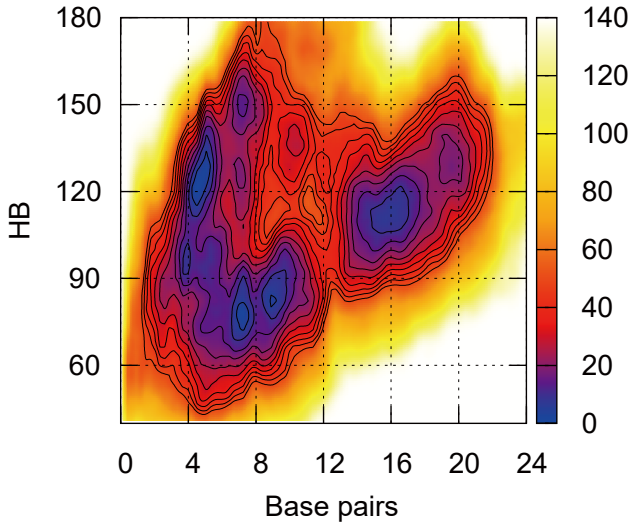

Figure S5: Free energy as a function of the intrachain base pairs (BP) and hydrogen bonds (HB) for the aptamer SYL3C.

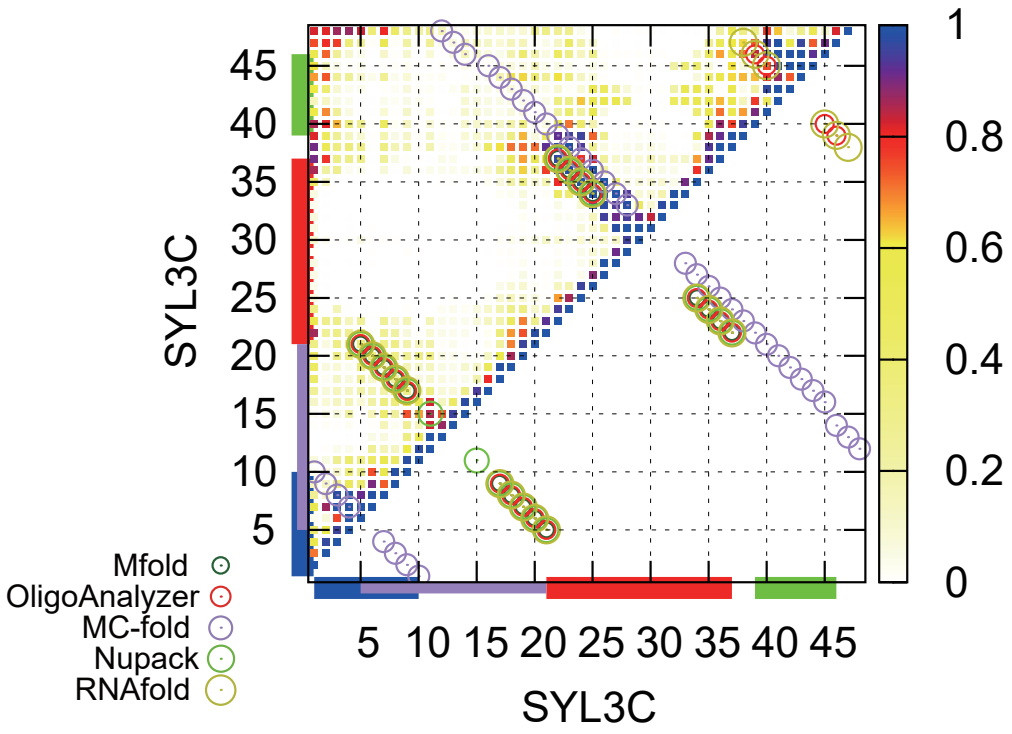

Figure S6: Contact map in the state 2 of the aptamer SYL3C's folding FES. Secondary structure generated by all the selected prediction tools of Mfold (dark green), OligoAnalyzer3.0 (red), MC-fold (purple), Nupack (green) and RNAfold (yellow). The four hairpin loop regions are divided by blue (HP1), purple (HP2), red (HP3) and green (HP4) bars.

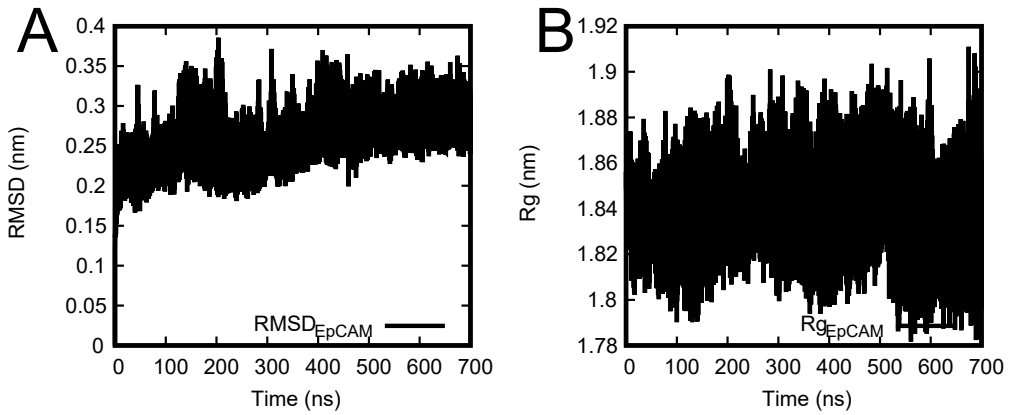

Figure S7: The RMSD (A) and the Rg (B) trajectories for the EpCAM's structural variation in the Martini-based MetaD simulations during the binding process.

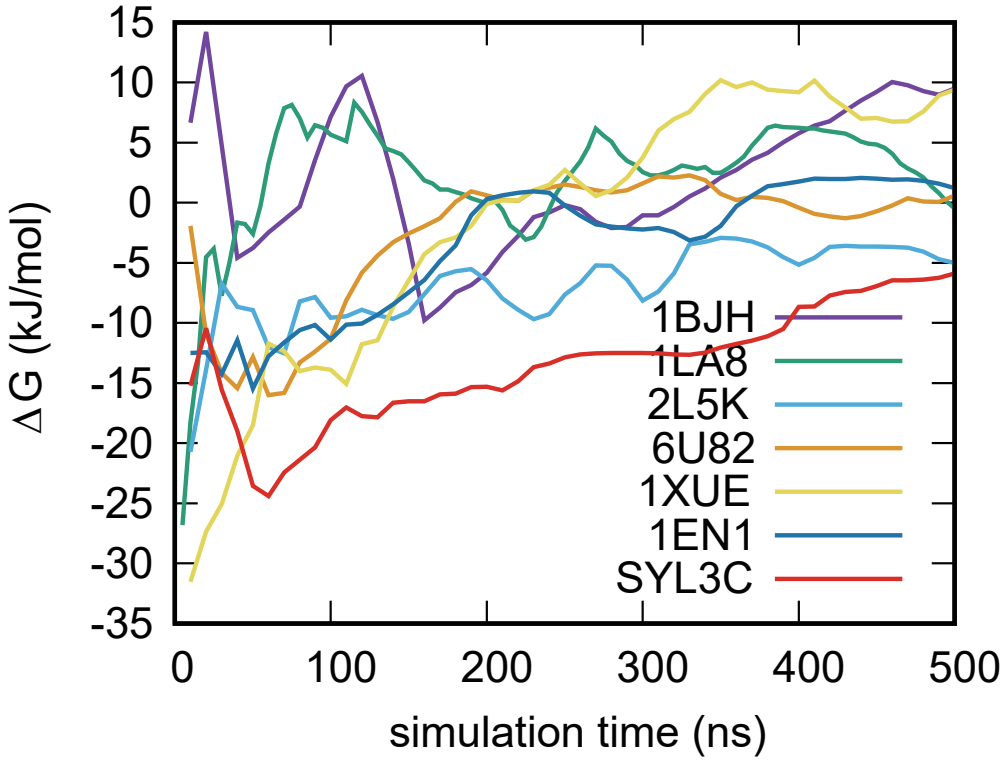

Figure S8: Convergence verification of the metadynamics for the ssDNAs (aptamers) folding using the free energy difference between states ($\Delta G = F_{state1} - F_{state2} = k_b T * log(P_{state2}/P_{state1})$ as a function of time. The evolution of the CV of the $BP_{aptamer}$ projection was calculated. States were divided by $BP_{aptamer}$ at 2 (1BJH), 3 (1LA8), 4 (2L5K),6 (6U82), 4 (1XUE), 6 (1EN1), and 12 (aptamer SYL3C), respectively.

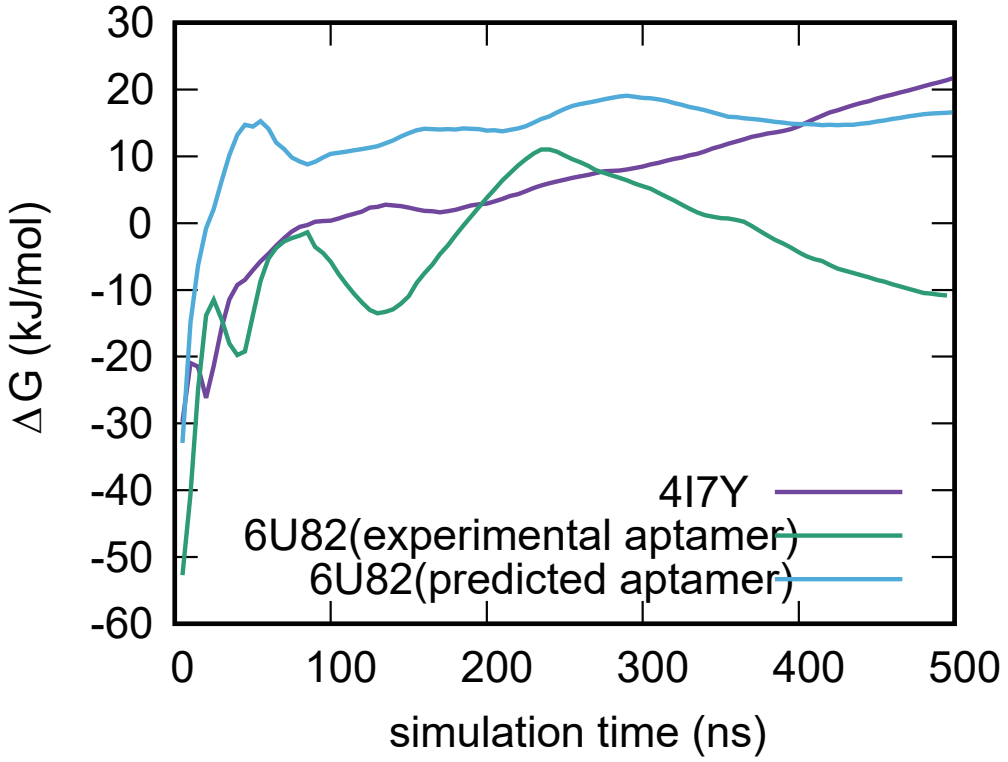

Figure S9: Convergence verification of the metadynamics for the aptamer-EpCAM binding using the free energy difference between states (ΔG) as a function of time. The evolution of the CV of the D_{COM} projection was calculated. States were divided by D_{COM} at 2.5 nm for 4I7Y and 2 nm for two 6U82 binding tests.

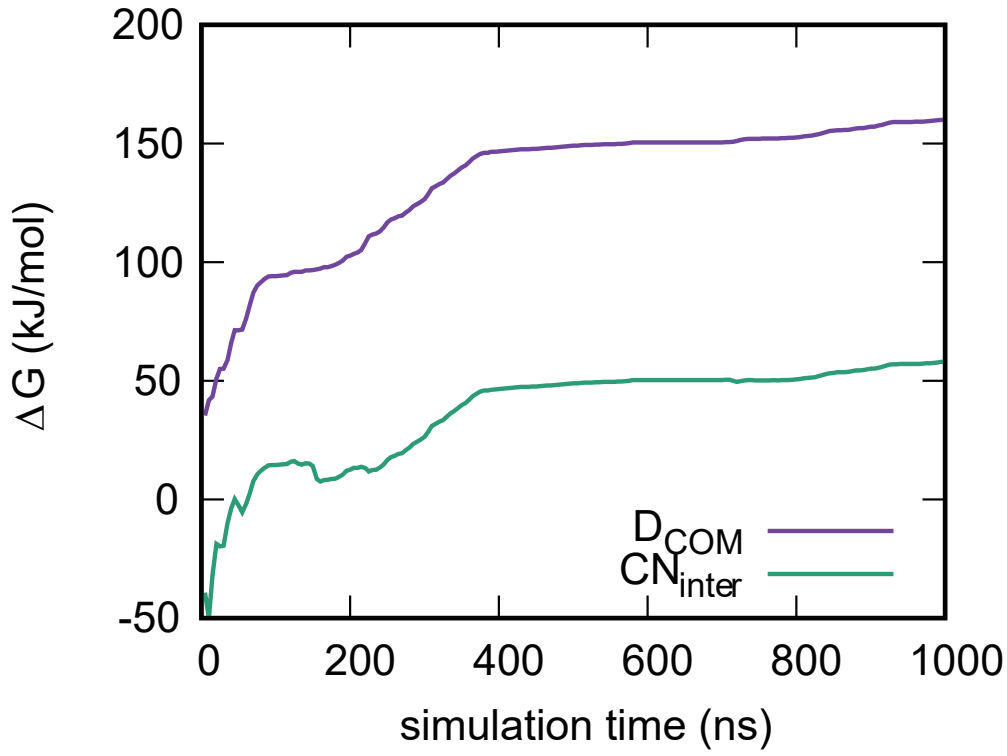

Figure S10: Convergence verification of the metadynamics for the aptamer-EpCAM binding using the free energy difference between states (ΔG) as a function of time. The evolution of both CVs of the D_{COM} and CN_{inter} projections were calculated. States were divided by D_{COM} at 5 nm and by CN_{inter} at 150.

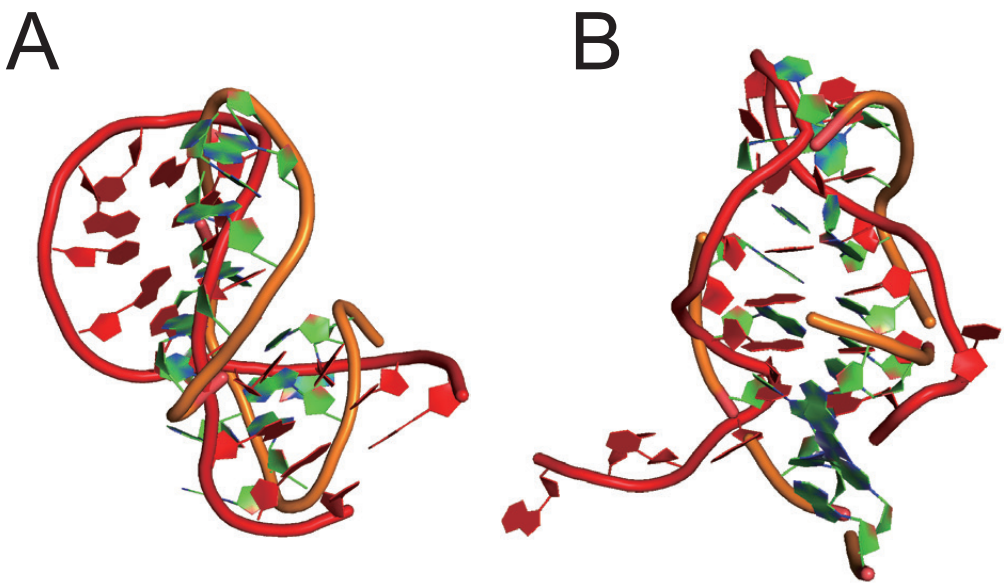

Figure S11: The 3D structure predictions of 1XUE (A) and the 1EN1 (B) by software iFoldRNA. Native structures are aligned and colored in red.

Table S1: Secondary structure predictions by Martini-based simulation and prediction tools.

ssDNA	secondary structure (dot-bracket)
1BJH	GTACAAAGTAC ((((...))))
1LA8	CGCGGTGTCCGCG (((((...)))))
2L5K	CAGTTGATCCTTTGGATACCCTG ((((((((((...))))))))))
6U82(DNA)	GCTAATCTAATCAACCGCAGGTTGATTAGCCCATTAGC ((((((((((((((...))))))))))...))))))
1XUE	GTGGAATGCAATGGAAC (((((((...)))))))
1EN1	GTCCCTGTTCGGGCGCCA (.(((.....))))....