Supplementary Information:

Understanding the kinetic mechanism of RNA single base pair formation

Xiaojun XU[1], Tao YU[1,2], Shi-Jie CHEN[1‡]

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

[2]Department of Physics, Jianghan University, Wuhan, Hubei 430056, China

[1] Initial 3-bp folded structures

The initial coordinates for the 3-bp helix are determined from the experimentally determined structure (PDB ID: 1c0o), which is a hairpin loop with a 4-bp helix. The hairpin structure is embedded in a TIP3 water box with the water shell of 12 Å, using the *Solvate* plugin in VMD. Sodium and chloride ions are used to neutralize the system and to keep 1M sodium concentration of the system. Three minimization steps are used to minimize the whole RNA system: (1) fix the atoms in RNA, minimize the added water and ion molecules for 1000 steps; (2) fix the heavy atoms of RNA, minimize the whole system for 1000 steps; (3) minimize the whole system without any constraint for 1000 steps. We keep the temperature at 310 K by coupling the system to a Langevin heat bath. We retain the first three base pairs and deleted the rest 8 nucleotides to build the initial folded structures of the 3-bp helix.

[2] Molecular dynamics simulations

For each initial (unfolded) structure, generated by the rotations of four torsional angles shown in Fig. 1B in blue, we rebuild the water box with the water shell of 12 Å. We add sodium and chloride ions to the solution to neutralize the system and to maintain 1M NaCl concentration. It should be noted that the total number of water molecules and ions may be slightly different for each solvated system, because different unfolded structures have different positions of the G1 base, resulting in slightly different dimensions of water boxes.

We use a cutoff of 12 Å switching function starting at 10 Å for the van der Waals interaction. An integration time step of 2 fs is employed with the rigidBonds "on" option. We keep the temperature at 310 K by coupling the system to a Langevin heat bath. Long-range electrostatic interactions are evaluated using the smooth particle mesh Ewald (PME) method with a grid spacing of 1 Å. After a 500-step minimization for the whole system, we perform molecular dynamics simulations for all the atoms of the first (G1) nucleotide as well as all the water molecules and the ions.

[3] Details of extracting clusters

We use two RMSD values to represent the structural distances between two snapshots of MD trajectories: (1) RMSD_{base} based on 11 atoms (N9, C4, N3, C2, N2, N1, C6, O6, C5, N7, C8) in the G1 base, (2) $\text{RMSD}_{backbone}$ based on 6 atoms (C5', C4', C3', C2', C1', O4') in the G1 backbone, to denote the base and the backbone conformations, respectively, of the G1 nucleotide. If the RMSDs with respect to one cluster satisfies the condition: $\text{RMSD}_{base} < 1.2$ Å and $\text{RMSD}_{backbone} < 0.5$ Å at a time window (T_s, T_e), the time duration (residence time) $T_{duration}$ = T_e - T_s is calculated for the cluster. The extracted clusters are ranked by the average residence time. Top 50 clusters are used for the kinetic studies (See Tables S1 and S2).

[4] Master equation

A Master equation (ME) is a set of first-order differential equations describing the time evolution of the population of each one of a discrete set of states with regard to a continuous time variable t, $dp_i/dt = \sum_{j=1}^{\Omega}[k_{j\to i}p_j - k_{i\to j}p_i]$, where, $k_{j\to i}$ and $k_{i\to j}$ are the rate constants for the respective transitions, p_i is the population of the i^{th} conformation ($i = 1, \cdots, \Omega$, where Ω is the total number of chain conformations). The above ME has an equivalent matrix form: $d\mathbf{P}/dt = \mathbf{M} \cdot \mathbf{P}$, where $\mathbf{M}$ is the rate matrix defined as $M_{ij} = k_{j\to i}$ for $i \neq j$, and $M_{ii} = -\sum_{i\neq j} k_{i\to j}$. $\mathbf{P}$ is the fractional populational vector col(p_1, p_2, $\cdots$, p_Ω). By solving the eigenvalues λ_m and eigenvectors $\mathbf{n}_m$ of the rate matrix $\mathbf{M}$,the solution of the ME $\mathbf{P}(t) = \sum_{m=1}^{\Omega} C_m \mathbf{n}_m e^{\lambda_m t}$ gives a rigorous and exact relaxation kinetics of the system for a given initial folding condition at $t = 0$. Here, C_m is the coefficient that is dependent on the initial condition.

The eigenvalue spectrum gives the rates of the kinetic modes of the system. The eigenvectors give the basic modes of the kinetic process and are intrinsically related to the energy landscape. In fact, from the eigenvectors we can obtain the rate-limiting steps of the kinetics[1, 2, 3]. However, The ME solution can only give ensemble-averaged macroscopic kinetics and cannot give detailed information about the microscopic pathways.

[5] Force field dependence.

Following the same protocol as described above, we also rum simulations using the CHARMM27 force field (see Fig. S12). The overall kinetics is similar to the predictions from the AMBERff99 force field (Fig.3). In both force fields, the process can be described by the 4-cluster kinetics. The main difference between them is that AMBER predicts much faster than that from the CHARMM force field. For example the population of the trapped cluster reaches the peak value at $t \approx 10^{-8}$s by AMBER (blue in Fig.3) and 10^{-7} s by CHARMM (blue in Fig. S12), respectively. Such one-order difference in magnitude of the kinetics time scale may be explained by the different treatment of the RNA strand terminals. The absence of the terminal (G1) phosphate group in the simulations with AMBER causes a lighter mass of the G1 as well as the different backbone charge, resulting in the faster movement towards the folded cluster. Another difference is that the predicted kinetics with CHARMM lacks the fast pre-equilibration between the different unfolded clusters (Fig.3). This pre-equilibration of the unfolded intermediates suggests that the AMBER-predicted local energy landscape around the unfolded cluster may be more bumpy (with several basins) than that predicted by CHARMM, which has only one dominant unfolded basin. Furthermore, the role of the intermediate conformational cluster (the brown lines in Figs.3 and S12) are slightly different. CHARMM predicts a much large population for the intermediate cluster than AMBER.

Both force fields predict the same five kinetically dominant structures (black, brown, blue, orange and red ones shown in Fig. S12), corresponding to the unfolded, intermediate, trapped and folded clusters. As shown in Fig. S12, except for the unfolded cluster, which is relatively flexible due to weak base stacking/pairing interaction, the other four clusters have similar backbone orientations and base positions. The result suggests that both force fields provide similar base pairing and base stacking interactions and hence similar roles for these clusters in the overall folding kinetics. CHARMM predicts two other stable clusters (purple and dark green ones in Fig. S12) which are absent in the AMBER-predicted kinetics. These two clusters are less folded compared with the other four clusters (brown, blue, orange and red), thus we concluded that, both force fields can properly treat near-native (kinetically important) clusters but handle the non-native (unfolded) clusters slightly differently.

In summary, we find that the AMBER and CHARMM force fields gave similar overall energy landscapes, especially near the folded state. Both force fields predict similar kinetic intermediates. The only difference is that the local energy landscape around the unfolded states given by CHARMM is less bumpy than that given by AMBER. The AMBER force field resulted in several kinetic basins with comparable stabilities. Because the pre-equilibration between these (unfolded) conformations is much faster than the transitions to the intermediate conformational cluster and to the trapped cluster, the difference in the unfolded cluster may not alter the overall folding kinetics. We conclude that both force fields can properly treat the interplay between the base pairing and base stacking interactions, especially for the folded and partially folded clusters.

[6] Kinetic Monte Carlo simulation.

The KMC simulation gives an ensemble of simulated sequences (trajectories) for inter-cluster transitions. The average over all the sampled trajectories for the population of a cluster i at time t (as represented by the centroid conformation) gives the populational kinetics $p_i(t)$ of the cluster. The trajectories can also provide the information about the transition states. For a two-state reaction, starting from the transition state, the probability of committing to either state would be 50% [4]. Due to the high free energy, a transition state has relatively low population and hence could be missed in our conformational cluster network. As a result, a transition state is unlikely a centroid structure of a cluster. However, if we start from a cluster that is sufficiently close to the actual transition state, the probabilities for going to the reactant and the product states may be close to 50%. In our calculation, we start from each conformation and ran 50000 KMC trials. By calculating the probability of falling to the folded/unfolded states, we identify the clusters (centroid structures) that may be close to the transition state.

Table S1: Top 50 clusters ranked by the average residence time T^{ave} (in ns), extracted from MD simulations with the AMBER force field.

id	T^{ave}	id	T^{ave}	id	T^{ave}	id	T^{ave}	id	T^{ave}
M_1	-	M_2	1.47893	M_3	0.43596	M_4	0.3804	M_5	0.17694
M_6	0.06516	M_7	0.04629	M_8	0.04113	M_9	0.0376	M_{10}	0.03107
M_{11}	0.0289	M_{12}	0.02518	M_{13}	0.02272	M_{14}	0.02237	M_{15}	0.02154
M_{16}	0.0213	M_{17}	0.02127	M_{18}	0.02051	M_{19}	0.02003	M_{20}	0.0196
M_{21}	0.01882	M_{22}	0.01877	M_{23}	0.01862	M_{24}	0.01859	M_{25}	0.01854
M_{26}	0.01811	M_{27}	0.01774	M_{28}	0.01757	M_{29}	0.01737	M_{30}	0.0173
M_{31}	0.0169	M_{32}	0.01671	M_{33}	0.0166	M_{34}	0.02237	M_{35}	0.01772
M_{36}	0.01646	M_{37}	0.01626	M_{38}	0.01614	M_{39}	0.01675	M_{40}	0.01605
M_{41}	0.01569	M_{42}	0.01568	M_{43}	0.01566	M_{44}	0.01563	M_{45}	0.01542
M_{46}	0.0154	M_{47}	0.01524	M_{48}	0.01516	M_{49}	0.01516	M_{50}	0.01509

Table S2: Top 50 clusters ranked by the average residence time T^{ave} (in ns), extracted from MD simulations with the CHARMM force field.

id	T^{ave}	id	T^{ave}	id	T^{ave}	id	T^{ave}	id	T^{ave}
M_1	-	M_2	0.30671	M_3	0.23692	M_4	0.19779	M_5	0.1828
M_6	0.05143	M_7	0.046	M_8	0.045	M_9	0.04456	M_{10}	0.04373
M_{11}	0.04215	M_{12}	0.0407	M_{13}	0.04	M_{14}	0.04	M_{15}	0.03972
M_{16}	0.03714	M_{17}	0.03541	M_{18}	0.035	M_{19}	0.035	M_{20}	0.03467
M_{21}	0.03435	M_{22}	0.033	M_{23}	0.03267	M_{24}	0.03143	M_{25}	0.03139
M_{26}	0.03	M_{27}	0.03	M_{28}	0.03	M_{29}	0.03	M_{30}	0.03
M_{31}	0.03	M_{32}	0.02876	M_{33}	0.02875	M_{34}	0.02835	M_{35}	0.02749
M_{36}	0.027	M_{37}	0.02662	M_{38}	0.0265	M_{39}	0.02539	M_{40}	0.02519
M_{41}	0.02509	M_{42}	0.025	M_{43}	0.025	M_{44}	0.025	M_{45}	0.025
M_{46}	0.025	M_{47}	0.025	M_{48}	0.02463	M_{49}	0.02458	M_{50}	0.02438

Table S3: Averaged values (over randomly selected 1000 structures in each cluster) of the order parameters for the 50 clusters. The clusters are extracted from MD simulations with the AMBER force field. The units of the distances and interactions (van der Waals, electrostatic, and total energy) are Å and kcal/mol, respectively. The table is continued in Table S4.

id	d_{12}	d_{15}	d_{16}	$Elec_{12}$	VdW_{12}	E_{12}	$Elec_{15}$	VdW_{15}	E_{15}	$Elec_{16}$	VdW_{16}	E_{16}
M_1	4.0	8.3	5.6	-34.72	-12.70	-47.42	-2.91	-0.31	-3.22	-26.80	0.15	-26.65
M_2	5.0	9.8	6.4	-49.17	-9.77	-58.94	-1.13	-0.13	-1.26	-10.03	-2.49	-12.53
M_3	4.2	8.1	5.7	-47.75	-9.55	-57.31	-2.68	-0.34	-3.02	-21.24	-1.83	-23.08
M_4	4.7	9.0	6.6	-47.84	-10.57	-58.42	1.21	-0.18	1.03	2.78	-1.51	1.27
M_5	4.4	8.6	5.6	-48.13	-6.86	-54.99	-2.45	-0.22	-2.68	-18.66	-2.15	-20.82
M_6	5.4	10.2	6.7	-53.44	-7.23	-60.67	0.91	-0.06	0.85	7.54	-1.41	6.12
M_7	6.6	10.2	7.9	-59.57	-3.26	-62.83	-0.53	-0.05	-0.59	-1.16	-0.35	-1.51
M_8	6.8	10.5	8.8	-56.81	-5.45	-62.26	0.79	-0.10	0.68	0.90	-0.35	0.55
M_9	6.7	9.9	8.5	-62.95	-4.54	-67.49	0.98	-0.19	0.79	1.35	-0.57	0.78
M_{10}	6.1	11.0	7.6	-49.94	-8.14	-58.09	1.47	-0.04	1.43	8.28	-0.93	7.35
M_{11}	10.1	15.7	13.0	-53.42	-2.74	-56.17	-0.28	-0.00	-0.28	-1.04	-0.01	-1.06
M_{12}	9.7	14.2	11.2	-56.93	-1.44	-58.38	-0.01	-0.01	-0.02	0.52	-0.09	0.42
M_{13}	7.2	11.0	9.2	-46.19	-5.08	-51.27	-0.73	-0.04	-0.77	-1.23	-0.23	-1.47
M_{14}	9.4	13.4	10.8	-54.68	-1.67	-56.35	0.10	-0.01	0.09	0.66	-0.10	0.55
M_{15}	7.1	11.1	8.0	-50.04	-3.88	-53.93	0.21	-0.03	0.17	2.78	-0.49	2.28
M_{16}	7.5	11.7	8.3	-44.84	-4.93	-49.77	0.18	-0.02	0.16	1.12	-0.35	0.76
M_{17}	6.6	9.8	8.8	-55.45	-3.69	-59.15	0.46	-0.10	0.36	-0.29	-0.23	-0.52
M_{18}	7.3	12.2	8.7	-46.70	-5.46	-52.16	0.02	-0.02	0.00	0.95	-0.46	0.49
M_{19}	10.6	15.3	12.3	-55.81	-1.87	-57.68	0.13	-0.00	0.12	0.34	-0.08	0.26
M_{20}	7.2	10.7	8.2	-48.50	-4.61	-53.11	-0.40	-0.03	-0.44	-0.45	-0.35	-0.80
M_{21}	11.7	17.4	15.2	-45.47	-4.19	-49.67	-0.08	-0.00	-0.08	0.02	-0.00	0.01
M_{22}	9.6	14.1	11.0	-56.94	-1.41	-58.36	-0.39	-0.00	-0.40	-0.66	-0.11	-0.77
M_{23}	14.0	19.5	16.9	-52.42	-0.40	-52.83	-0.09	0.00	-0.09	-0.04	-0.00	-0.05
M_{24}	6.5	10.7	7.2	-46.87	-6.60	-53.48	0.88	-0.04	0.84	6.40	-0.94	5.45
M_{25}	8.6	12.5	9.4	-56.08	-1.29	-57.37	0.16	-0.01	0.14	-0.11	-0.24	-0.36

Table S4: Table S3 continued.

id	d_{12}	d_{15}	d_{16}	$Elec_{12}$	VdW_{12}	E_{12}	$Elec_{15}$	VdW_{15}	E_{15}	$Elec_{16}$	VdW_{16}	E_{16}
M_{26}	7.1	11.6	8.2	-48.21	-5.43	-53.65	-0.54	-0.03	-0.58	-2.17	-0.53	-2.71
M_{27}	8.8	13.2	9.8	-46.80	-3.52	-50.33	0.13	-0.00	0.13	1.52	-0.11	1.41
M_{28}	6.7	10.1	8.0	-55.77	-4.77	-60.55	0.18	-0.13	0.05	-0.17	-0.88	-1.05
M_{29}	9.5	14.0	10.8	-56.68	-1.36	-58.05	0.18	-0.00	0.17	0.11	-0.10	0.01
M_{30}	13.8	19.2	16.9	-47.60	-2.10	-49.70	-0.05	0.00	-0.05	0.22	-0.00	0.22
M_{31}	6.5	10.8	7.2	-45.93	-7.03	-52.96	1.05	-0.04	1.01	7.31	-1.10	6.21
M_{32}	11.2	15.6	13.0	-56.63	-1.47	-58.11	-0.05	-0.00	-0.06	-0.57	-0.07	-0.65
M_{33}	6.6	10.2	8.3	-51.20	-5.96	-57.16	0.20	-0.12	0.08	0.61	-0.70	-0.09
M_{34}	5.6	10.7	7.5	-40.57	-10.39	-50.97	-0.58	-0.07	-0.65	-3.70	-1.49	-5.19
M_{35}	12.2	17.9	15.3	-49.82	-1.48	-51.30	-0.07	-0.00	-0.07	-0.01	-0.00	-0.02
M_{36}	9.0	12.4	10.6	-59.05	-2.34	-61.39	0.50	-0.02	0.48	0.54	-0.15	0.39
M_{37}	10.2	14.7	11.5	-47.72	-3.62	-51.35	0.13	-0.00	0.13	0.08	-0.07	0.01
M_{38}	8.8	12.5	10.0	-50.32	-4.31	-54.63	-0.16	-0.01	-0.18	0.01	-0.10	-0.08
M_{39}	6.8	10.5	7.4	-43.99	-5.66	-49.65	-0.49	-0.04	-0.53	-0.75	-0.49	-1.25
M_{40}	11.1	15.5	13.4	-50.06	-1.92	-51.99	0.04	-0.00	0.03	0.24	-0.00	0.24
M_{41}	9.3	14.3	10.8	-48.14	-3.66	-51.80	0.17	-0.00	0.17	1.67	-0.09	1.57
M_{42}	7.7	11.2	8.6	-48.05	-4.60	-52.66	-0.49	-0.03	-0.53	-0.64	-0.27	-0.91
M_{43}	6.7	11.5	8.0	-45.31	-7.36	-52.67	-0.90	-0.04	-0.95	-3.95	-0.84	-4.79
M_{44}	12.6	18.4	16.0	-44.27	-3.39	-47.67	-0.10	-0.00	-0.10	0.55	-0.00	0.54
M_{45}	8.3	12.0	10.0	-44.47	-5.08	-49.55	-0.93	-0.02	-0.96	-1.23	-0.15	-1.38
M_{46}	11.0	15.1	12.6	-57.25	-1.34	-58.60	-0.08	-0.00	-0.09	-0.46	-0.07	-0.53
M_{47}	11.6	16.2	13.3	-49.56	-3.67	-53.23	0.09	-0.00	0.08	-0.21	-0.06	-0.28
M_{48}	7.4	12.0	8.6	-47.17	-5.87	-53.04	-0.73	-0.03	-0.76	-3.33	-0.48	-3.81
M_{49}	13.7	19.4	17.4	-44.19	-2.86	-47.06	-0.09	-0.00	-0.09	0.18	-0.00	0.17
M_{50}	10.5	16.1	13.3	-51.72	-1.60	-53.33	-0.12	-0.00	-0.12	-0.44	-0.01	-0.45

Table S5: Averaged values (over randomly selected 1000 structures in each cluster) of the order parameters for the 50 clusters. The clusters are extracted from MD simulations with the CHARMM force field. The units of distances and interactions (van der Waals, electrostatic, and total energy) are Å, and kcal/mol, respectively. The table is continued in Table S6.

id	d_{12}	d_{15}	d_{16}	$Elec_{12}$	VdW_{12}	E_{12}	$Elec_{15}$	VdW_{15}	E_{15}	$Elec_{16}$	VdW_{16}	E_{16}
M_1	4.1	8.3	5.64	22.24	-11.95	10.29	-3.88	-0.33	-4.22	-25.53	-0.53	-26.07
M_2	4.7	9.5	6.17	34.99	-10.94	24.05	-2.49	-0.23	-2.72	-7.35	-2.31	-9.66
M_3	4.2	8.2	5.68	1.38	-10.46	-9.08	-3.03	-0.34	-3.37	-20.98	-1.27	-22.25
M_4	4.7	8.9	6.60	7.20	-9.49	-2.29	1.92	-0.20	1.72	6.30	-1.46	4.84
M_5	4.5	9.3	6.01	3.27	-10.92	-7.65	-2.00	-0.21	-2.21	-10.20	-0.78	-10.98
M_6	7.1	10.6	8.56	-8.45	-2.38	-10.84	-0.27	-0.05	-0.32	2.80	-0.27	2.52
M_7	10.4	15.9	12.90	21.45	-6.58	14.87	0.08	-0.04	0.05	3.85	-0.23	3.61
M_8	13.8	18.8	17.44	2.95	-0.23	2.72	0.00	-0.00	0.00	0.37	-0.00	0.37
M_9	5.5	9.2	8.30	27.14	-8.52	18.62	-0.53	-0.18	-0.71	-1.40	-0.36	-1.76
M_{10}	4.9	8.6	6.33	-8.34	-6.17	-14.51	2.67	-0.27	2.39	9.10	-1.75	7.35
M_{11}	6.6	9.9	8.68	-4.31	-4.01	-8.32	1.40	-0.10	1.30	4.19	-0.27	3.92
M_{12}	7.0	12.4	9.44	9.42	-7.41	2.00	-0.93	-0.02	-0.95	-5.26	-0.40	-5.67
M_{13}	9.2	14.5	11.26	6.06	- 3.91	2.15	-0.01	-0.00	-0.01	1.11	-0.04	1.06
M_{14}	12.4	16.9	14.67	11.34	-2.35	8.99	-0.05	-0.00	-0.05	0.40	-0.01	0.39
M_{15}	8.1	11.8	9.12	-4.29	-2.84	-7.13	-0.01	-0.04	-0.04	1.16	-0.31	0.85
M_{16}	8.1	11.5	10.23	-0.89	-3.01	-3.90	1.25	-0.06	1.19	6.66	-0.24	6.42
M_{17}	6.7	11.2	7.61	3.44	-4.28	-0.84	0.62	-0.02	0.60	7.49	-0.57	6.92
M_{18}	9.2	14.2	10.77	17.82	-5.93	11.89	0.38	-0.05	0.32	4.42	-0.50	3.92
M_{19}	12.2	16.8	15.09	5.30	-2.98	2.33	0.42	-0.02	0.40	4.50	-0.23	4.28
M_{20}	7.2	12.1	8.96	15.07	-6.41	8.66	-0.30	-0.02	-0.32	-0.97	-0.28	-1.25
M_{21}	6.4	10.1	8.45	-2.46	-5.89	-8.35	0.20	-0.15	0.04	1.69	-0.79	0.91
M_{22}	7.6	11.3	8.88	-6.80	-2.04	-8.84	-0.20	-0.04	-0.24	1.58	-0.26	1.31
M_{23}	5.4	9.7	6.25	2.17	-7.07	-4.90	-2.14	-0.19	-2.33	-5.88	-1.92	-7.80
M_{24}	7.3	12.1	9.12	16.68	-6.12	10.56	-0.36	-0.02	-0.38	-1.77	-0.22	-2.00
M_{25}	5.9	10.8	7.43	28.26	-9.78	18.49	1.19	-0.08	1.10	9.89	-0.99	8.91

Table S6: Table S5 continued.

id	d_{12}	d_{15}	d_{16}	$Elec_{12}$	VdW_{12}	E_{12}	$Elec_{15}$	VdW_{15}	E_{15}	$Elec_{16}$	VdW_{16}	E_{16}
M_{26}	11.4	16.7	13.3	3.03	-1.94	1.09	-0.06	-0.00	-0.06	-0.06	-0.01	-0.06
M_{27}	14.4	20.0	17.8	-6.49	0.11	-6.38	-0.10	-0.00	-0.10	0.61	-0.00	0.61
M_{28}	13.4	19.1	16.5	-1.55	0.13	-1.42	-0.09	-0.00	-0.09	0.46	-0.00	0.46
M_{29}	9.0	12.6	11.2	2.78	-2.65	0.13	0.51	-0.02	0.49	5.39	-0.17	5.22
M_{30}	12.1	17.2	15.0	9.72	-5.73	3.98	0.67	-0.04	0.63	5.91	-0.36	5.55
M_{31}	5.2	9.6	6.2	28.39	-8.65	19.74	-1.86	-0.19	-2.05	-0.93	-2.29	-3.22
M_{32}	6.4	9.9	8.0	-0.26	-5.87	-6.13	-0.27	-0.15	-0.43	0.83	-1.05	-0.23
M_{33}	13.6	19.3	17.2	0.33	-0.82	-0.49	-0.09	-0.00	-0.09	0.52	-0.00	0.52
M_{34}	7.1	12.6	9.7	18.82	-10.96	7.86	-1.15	-0.02	-1.17	-0.61	-0.23	-0.84
M_{35}	4.8	8.5	6.8	5.35	-7.19	-1.84	1.82	-0.31	1.51	0.77	-1.06	-0.29
M_{36}	9.0	14.3	11.1	6.31	-4.38	1.93	-0.05	-0.00	-0.05	1.01	-0.05	0.97
M_{37}	8.1	11.9	10.2	3.84	-4.59	-0.75	-0.62	-0.03	-0.65	-0.59	-0.15	-0.74
M_{38}	9.5	15.1	12.3	20.56	-6.50	14.07	0.11	-0.01	0.11	4.73	-0.07	4.66
M_{39}	12.0	17.2	14.1	8.78	-3.07	5.72	-0.08	-0.00	-0.08	0.45	-0.00	0.45
M_{40}	6.7	10.0	8.4	-8.81	-4.28	-13.09	1.39	-0.20	1.20	6.09	-0.60	5.49
M_{41}	10.8	15.6	14.0	2.21	-4.74	-2.54	0.12	-0.00	0.12	-0.36	-0.03	-0.39
M_{42}	9.7	15.2	12.1	9.20	-4.29	4.91	-0.15	-0.00	-0.15	0.74	-0.03	0.71
M_{43}	12.6	17.1	15.3	0.90	-1.49	-0.59	0.23	-0.01	0.22	3.22	-0.12	3.10
M_{44}	14.2	19.6	17.8	7.13	-3.18	3.95	-0.08	0.00	-0.08	0.70	-0.00	0.70
M_{45}	14.4	19.9	17.9	1.95	-1.74	0.21	-0.07	-0.00	-0.07	1.51	-0.01	1.50
M_{46}	10.2	14.6	11.6	3.60	-1.72	1.89	-0.09	-0.00	-0.09	-0.06	-0.03	-0.09
M_{47}	10.3	15.9	13.2	-0.47	-3.87	-4.34	0.35	-0.01	0.34	5.34	-0.11	5.23
M_{48}	7.0	10.8	7.6	15.33	-5.19	10.13	-0.09	-0.05	-0.14	0.62	-0.49	0.13
M_{49}	7.4	11.2	9.9	-1.38	-5.17	-6.55	0.89	-0.09	0.80	1.97	-0.28	1.69
M_{50}	12.8	18.5	16.1	1.30	-1.00	0.29	-0.07	-0.00	-0.07	1.08	-0.01	1.07

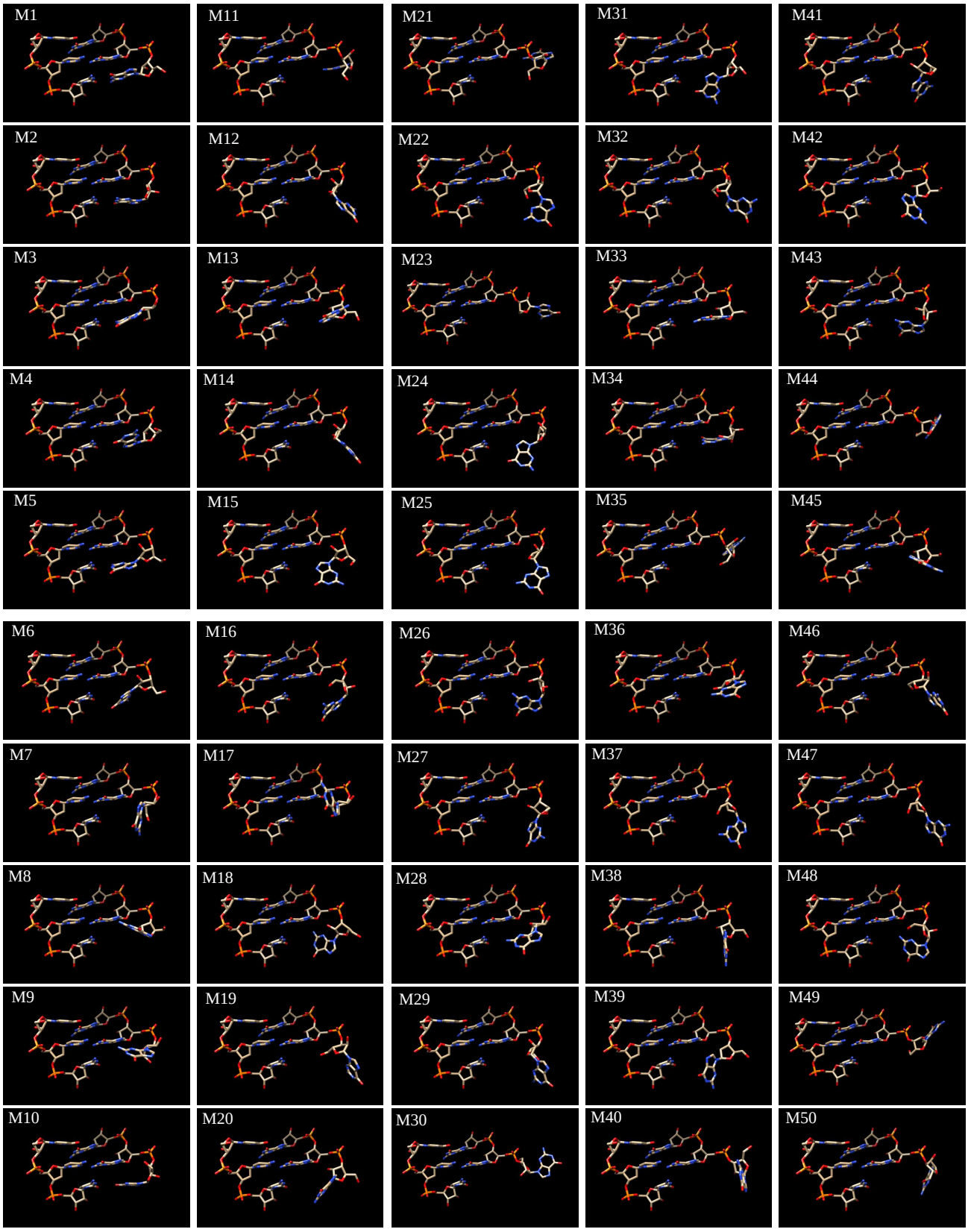

Figure S1: Structures of the 50 clusters extracted from MD simulations with AMBER force field.

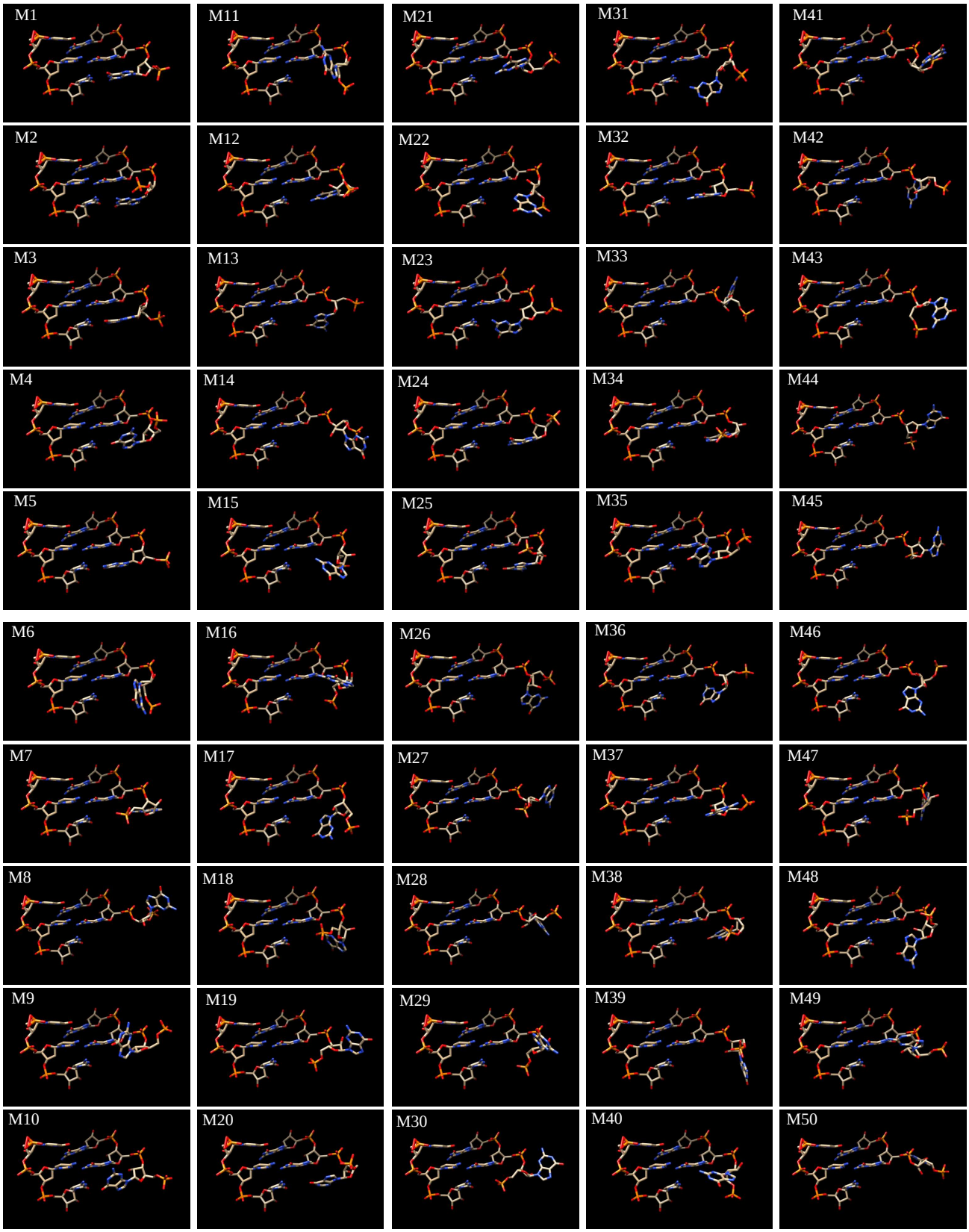

Figure S2: Structures of the 50 clusters extracted from MD simulations with CHARMM force field.

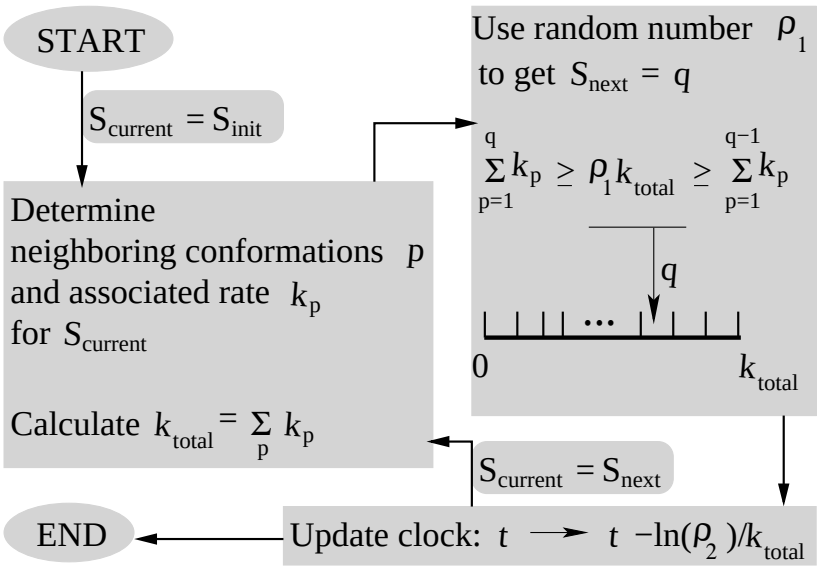

Figure S3: Flowchart illustrating the steps in the kinetic Monte Carlo algorithm. The total out_of_state (exit) rate k_{total} is partitioned into p individual rates k_p for the transitions from the current state to its neighbors. (See details in the published paper: Xu, Chen, 2012, J Am Chem Soc. 134, 12499-12507 and references therein.)

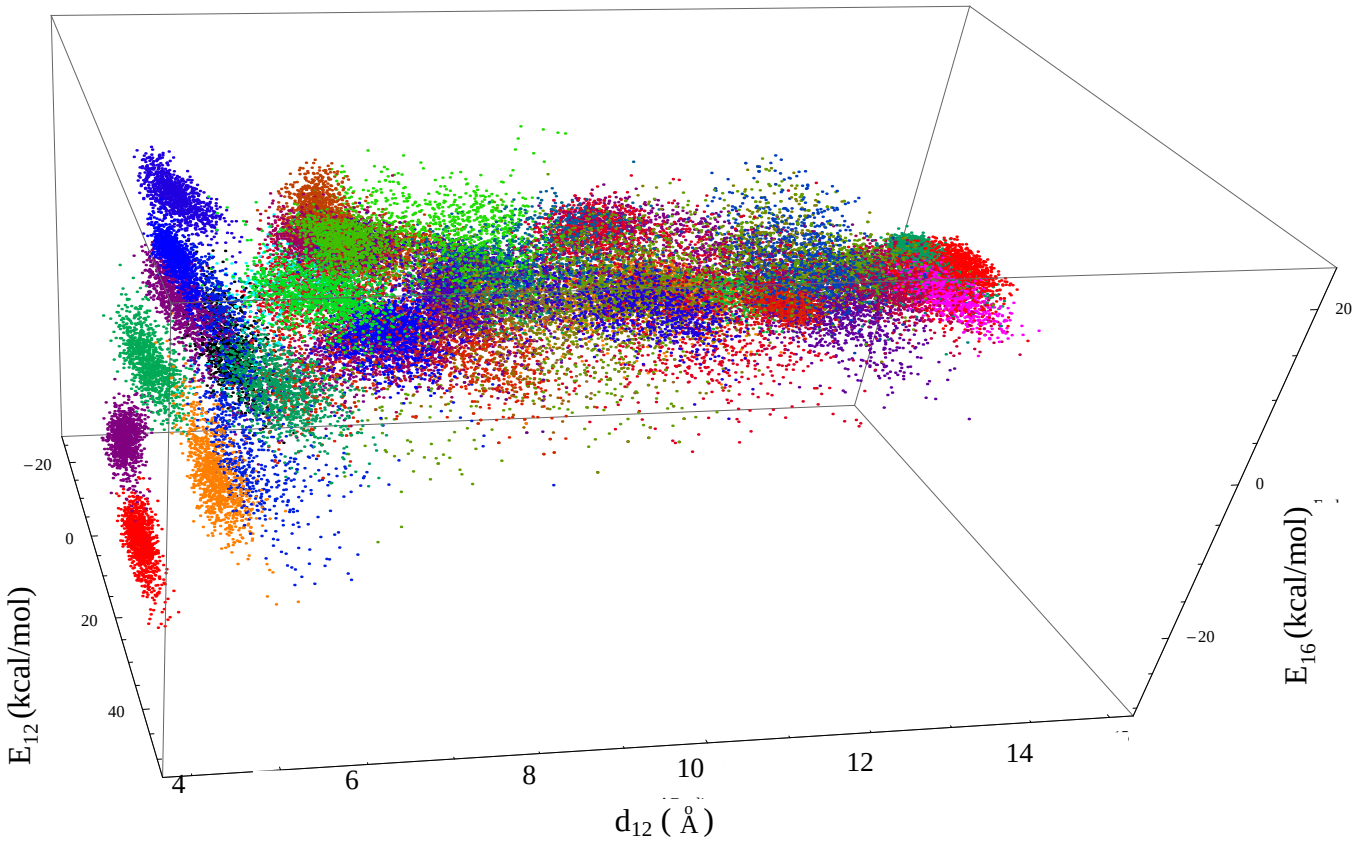

Figure S4: The order parameters of the 50 extracted clusters and their nearby (with small RMSD values) structures using the CHARMM force field. d_{12} is the center of geometry distance between all the base heavy atoms of nucleotides G1 and G2. E_{12} and E_{16} are the non-bonded interactions (VdW and electrostatic energies measured by VMD) between nucleotides G1 and G2 and between G1 and C6, respectively, reflecting the base stacking and pairing interactions.

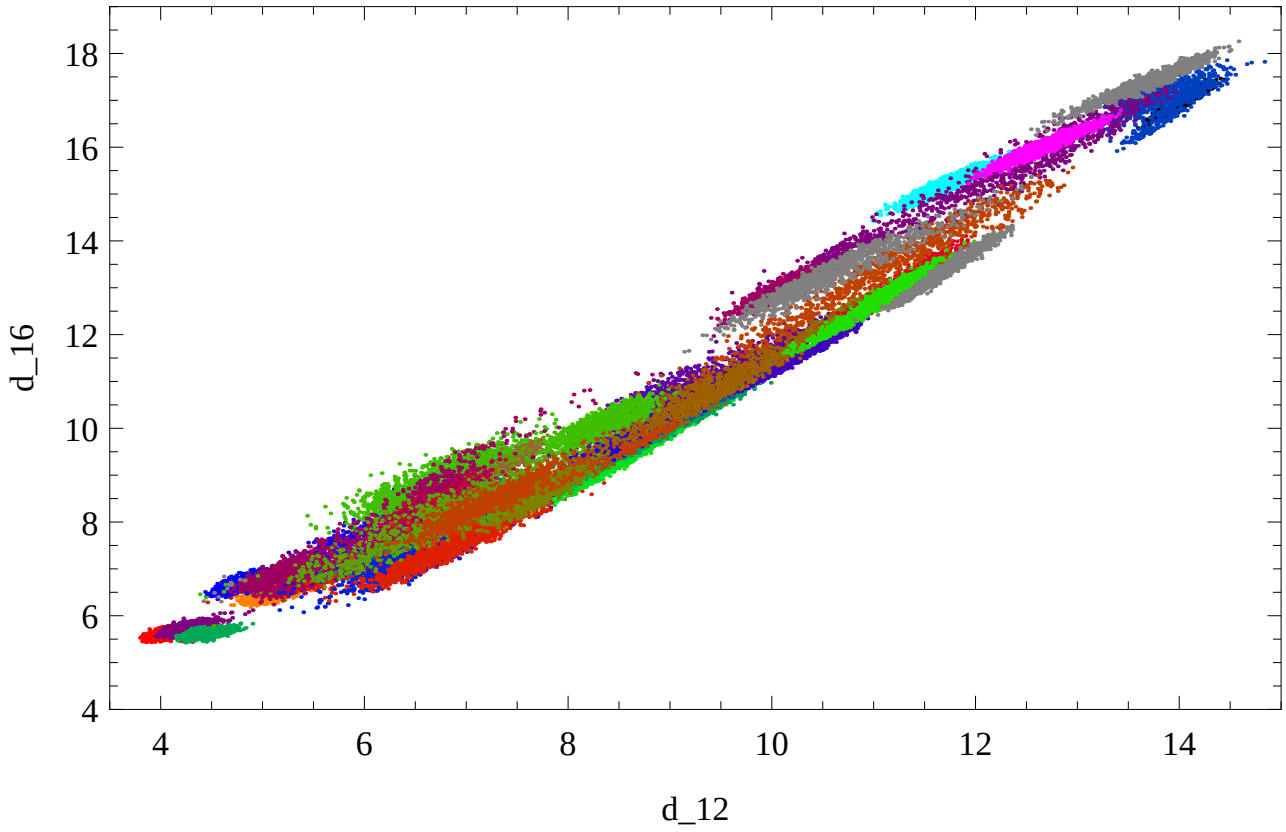

Figure S5: The distribution of the 50 extracted clusters and their nearby (with small RMSD values) structures using the AMBER force field as the function of d_{12} (the center of geometry distance between all the base heavy atoms of nucleotides G1 and G2) and d_{16} (the center of geometry distance between all the base heavy atoms of nucleotides G1 and C6).

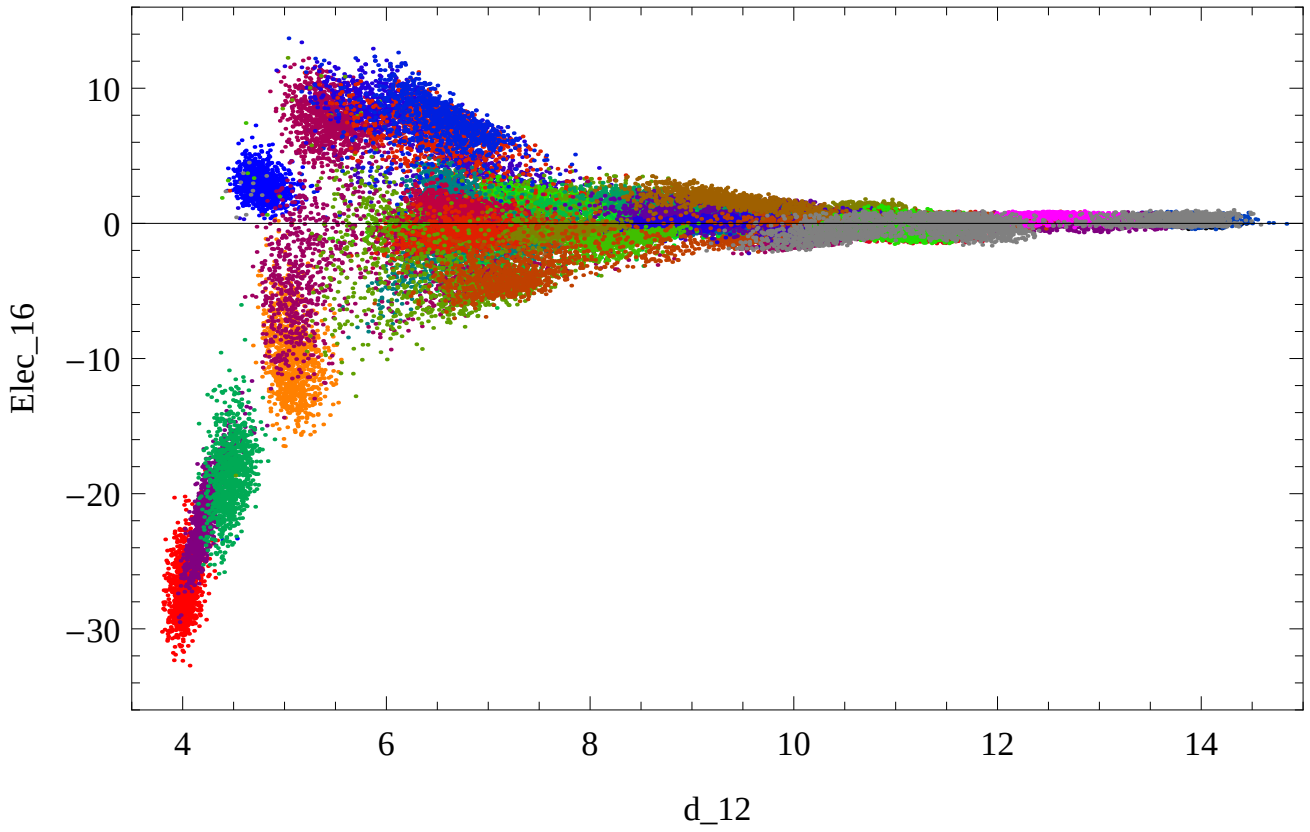

Figure S6: The distribution of the 50 extracted clusters and their nearby (with small RMSD values) structures using the AMBER force field as the function of d_{12} (the center of geometry distance between all the base heavy atoms of nucleotides G1 and G2) and $Elec_{16}$ (the electrostatic interactions between the nucleotides G1 and C6).

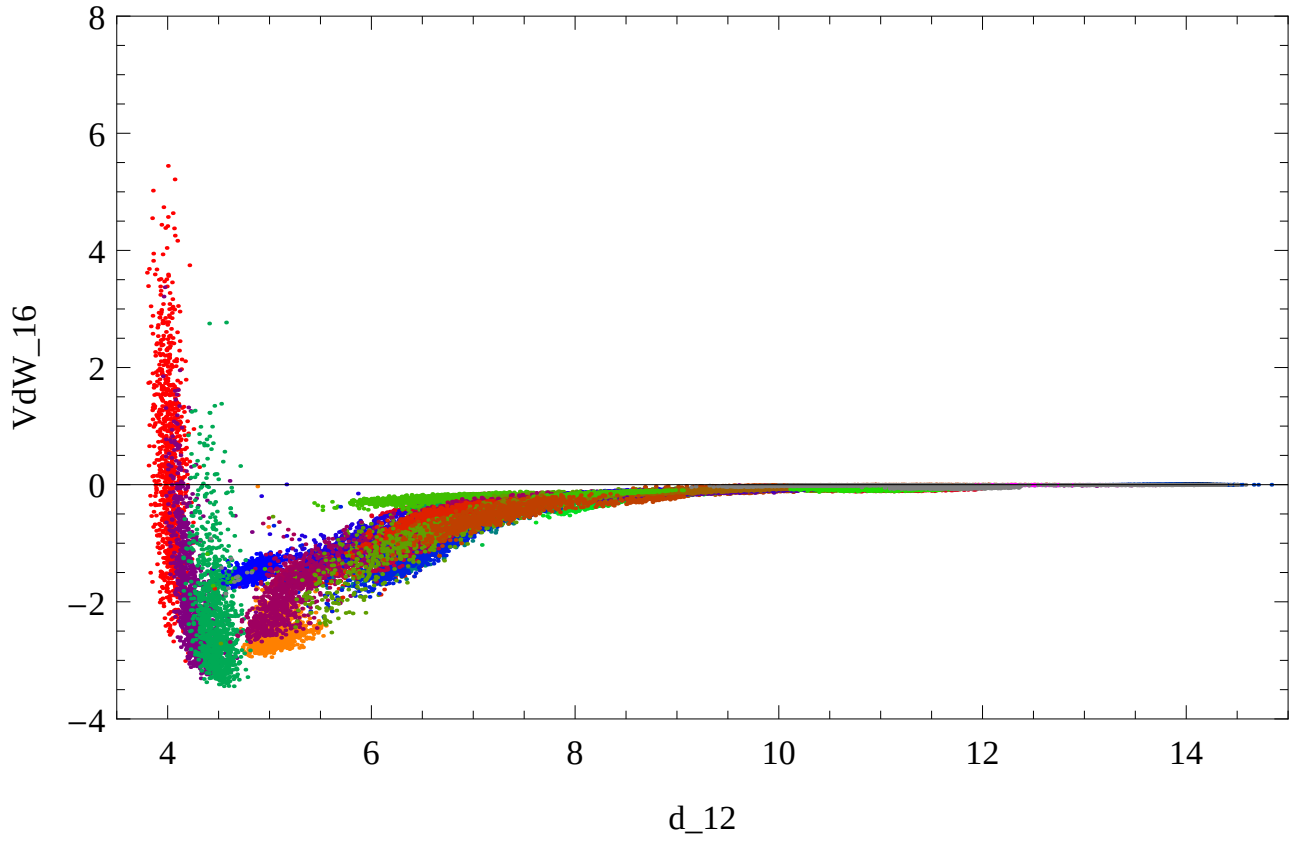

Figure S7: The distribution of the 50 extracted clusters and their nearby (with small RMSD values) structures using the AMBER force field as the function of d_{12} (the center of geometry distance between all the base heavy atoms of nucleotides G1 and G2) and VdW_{16} (the van der Waals interactions between the nucleotides G1 and C6).

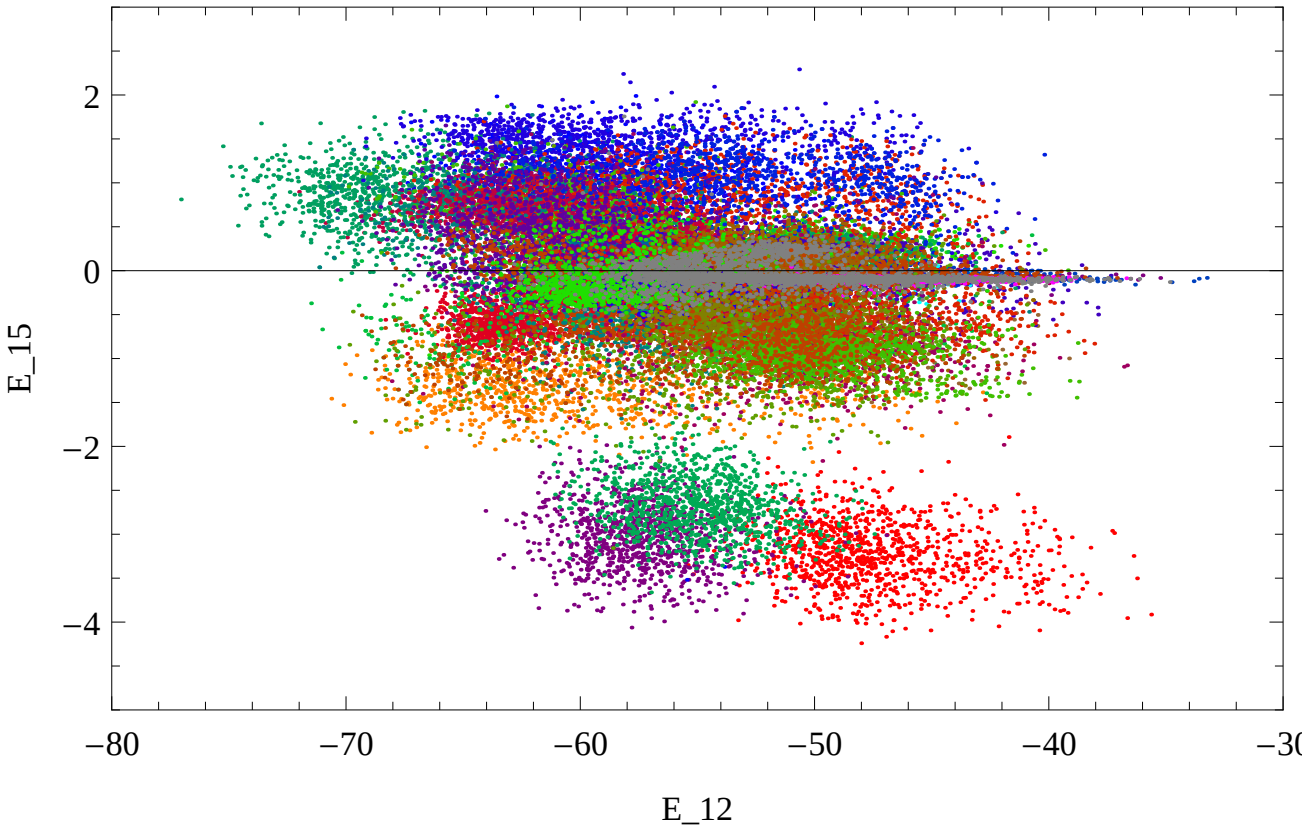

Figure S8: The distribution of the 50 extracted clusters and their nearby (with small RMSD values) structures using the AMBER force field as the function of E_{12} (the total non-bond interactions between the nucleotides G1 and G2) and E_{15} (the total non-bond interactions between the nucleotides G1 and C5).

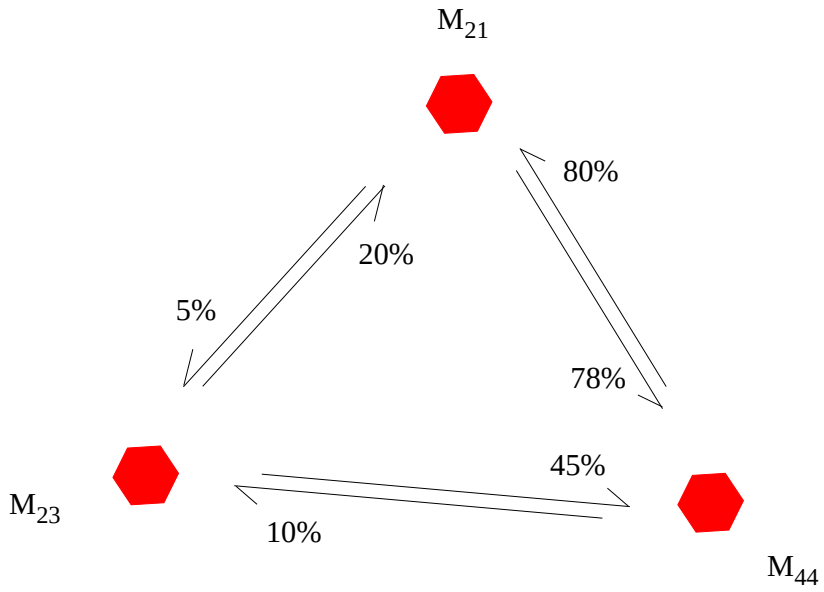

Figure S9: Distribution of transitions between clusters M_{21}, M_{23}, and M_{44} obtained from the MD simulations with AMBER force field. The transitions out of one cluster are dominated by the transitions to the other two clusters (for example, 20% and 45% transitions from cluster M_{23} to the cluster M_{21} and M_{44}, respectively), indicates that, there is probably a fast transition between the three clusters.

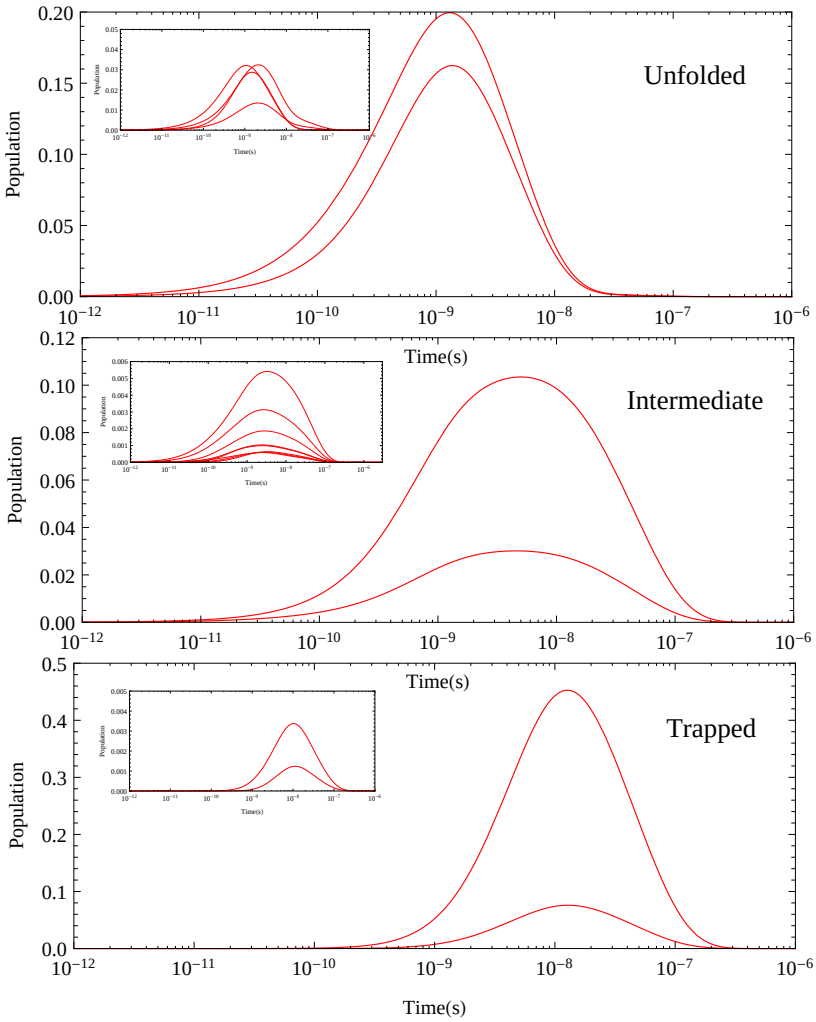

Figure S10: Time dependence of the populations of the clusters, with significant maximum values, for the unfolded, intermediate, and folded clusters.

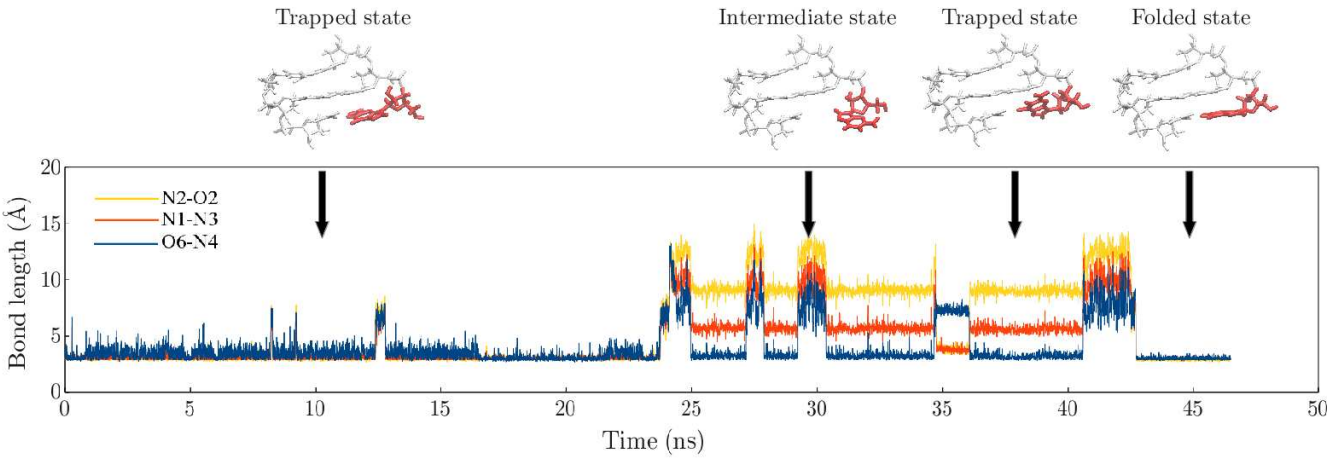

Figure S11: A MD trajectory for single base pair formation process. There are four representative clusters in RNA folding process: A trapped state with nonnative hydrogen bonds; an intermediate state with no hydrogen bonds between atoms O6, N1, N2, N4, N3, and O2; a trapped state with only one O6-N4 hydrogen bond; the folded state with three stable hydrogen bonds. In the last trapped state from 36 ns to 41 ns, a large number of Na^+ ions accumulate around the 6 hydrogen bonding atoms.

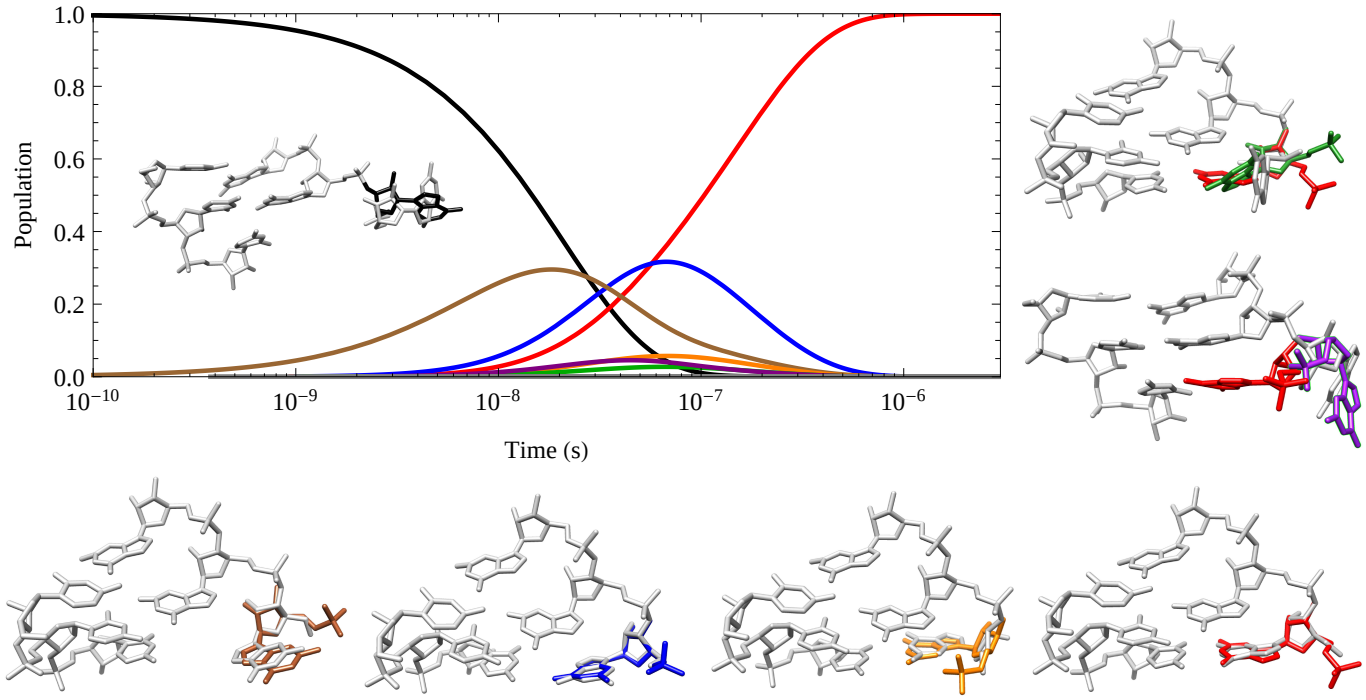

Figure S12: Long-time single-nucleotide folding kinetics starting from the unfolded cluster (shown in the inset with the G1 base shown in black) predicted by the master equation method based on the MD simulations with the CHARMM force field. Structures of the clusters with significant populations in the folding kinetics are shown with same color code as the lines shown in the folding kinetics. The gray G1 bases in each structures are the corresponding closest structures extracted from the MD simulations using the AMBER force field. Especially, the ones corresponding to the black, brown, blue, orange, and red are exactly the ones in same color with significant populations in the folding kinetics shown in Fig.3.

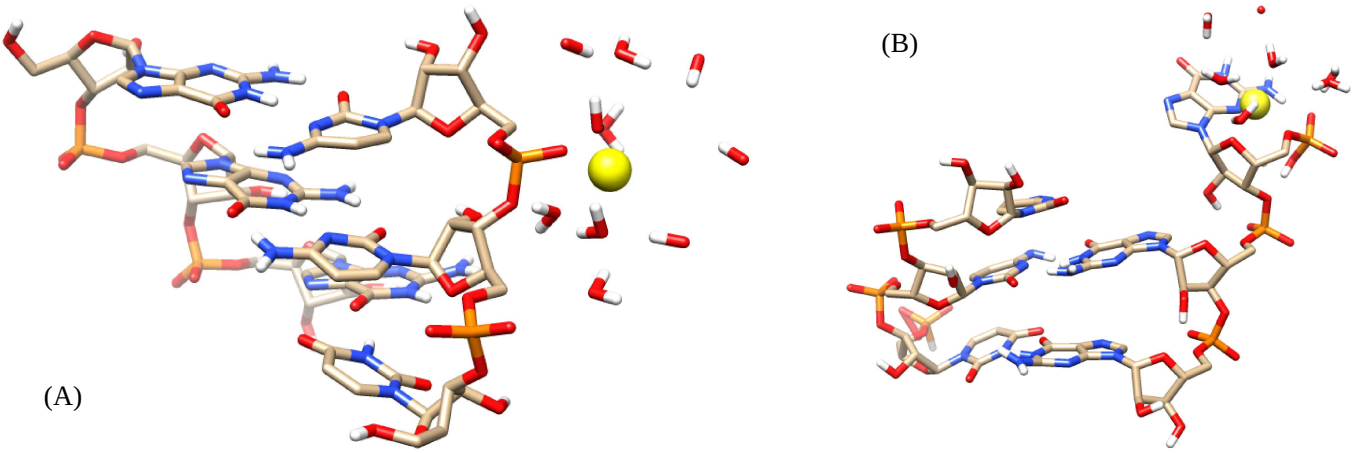

Figure S13: Dehydration of the Na^+ ion bound to the surface of RNA: (a) Na^+ ion binding to the phosphate groups of RNA. (b) Na^+ binding to both the phosphate and the base of the first G nucleotide (G1).

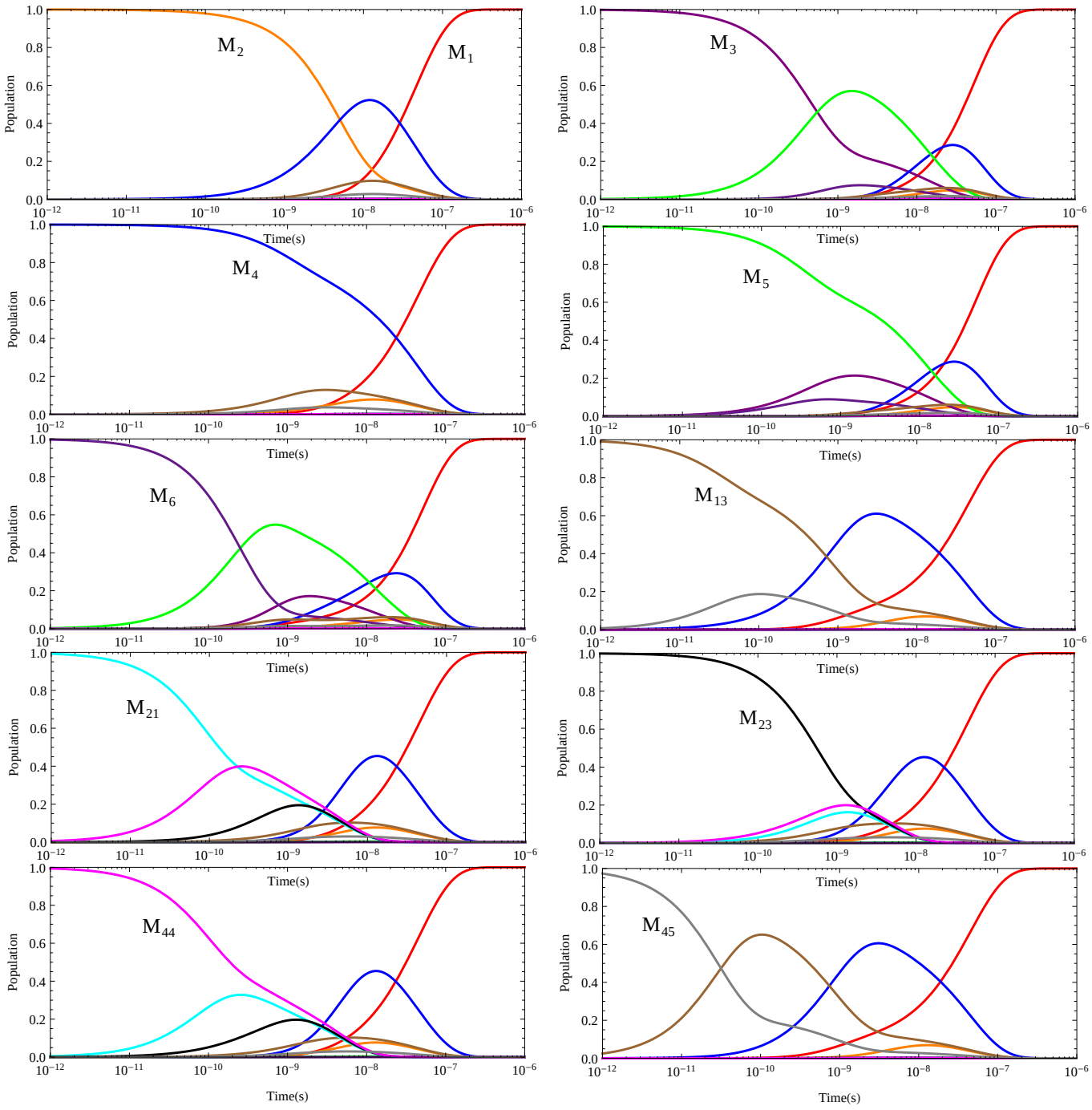

Figure S14: ME-predicted long-time single-nucleotide folding kinetics starting from the different conformational clusters based on AMBER MD simulations. M_2, M_3, M_4, M_5, and M_6 are in the trapped state. M_{13}, and M_{45} are in the intermediated state. M_{21}, M_{23}, and M_{44} are in the unfolded state.

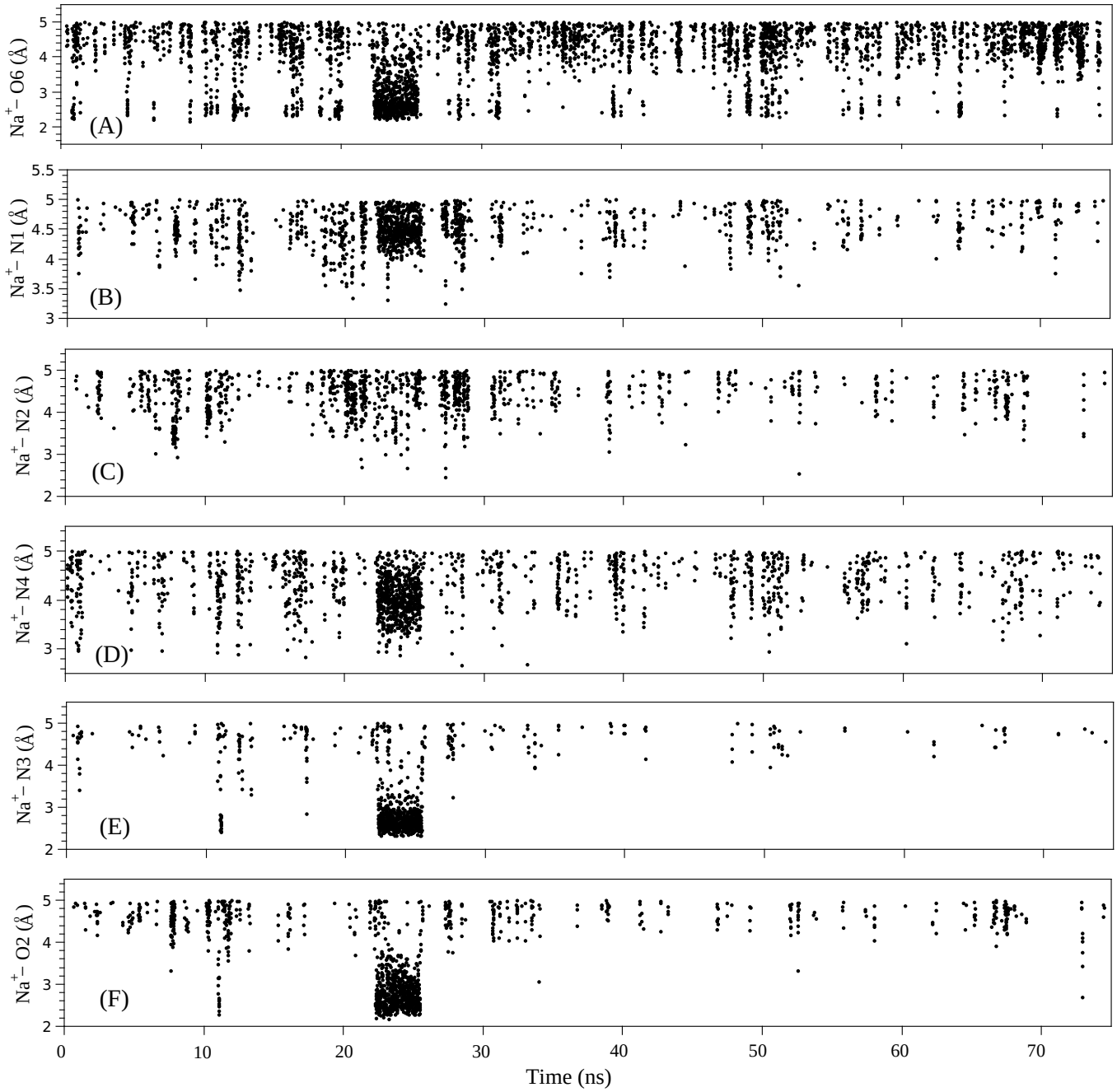

Figure S15: Na^+ ion distribution within a cutoff distance 5 Å from the six hydrogen bonding atoms: (A) O6, (B) N1, (C) N2, (D) N4, (E) N3, and (F) O2.

References

[1] Zhang W, Chen S-J. (2002). RNA hairpin-folding kinetics *Proc Natl Acad Sci USA*, **99**, 1931-1936.

[2] Zhang W, Chen S-J. (2003). Master equation approach to finding the rate-limiting steps in biopolymer folding. *J Chem Phys*, **118**, 3413-3420.

[3] Zhang W, Chen S-J. (2003). Analyzing the biopolymer folding rates and pathways using kinetic cluster method. *J Chem Phys*, **119**, 8716-8729.

[4] Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich ES. (1998). On the transition coordinate for protein folding. *J Chem Phys*, **108**, 334-350.