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

SUPPORTING INFORMATION

Structural view on the role of TRD loop in regulating DNMT3A activity: a molecular dynamics study

Hong Zhao[1,3], Jie Yu[1], Gaoqi Weng[1], Jiahui Yu[1], Ercheng Wang[1], Junbo Gao[1], Huanxiang Liu[2], Tingjun Hou[1], Zhe Wang*[1], Yu Kang*[1]

[1]*College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, China*

[2]*Faculty of Applied Science, Macao Polytechnic University, Macao, SAR, China*

[3]*Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare*

Corresponding authors

Zhe Wang

E-mail: wangzhehyd@zju.edu.cn

Yu Kang

E-mail: yukang@zju.edu.cn

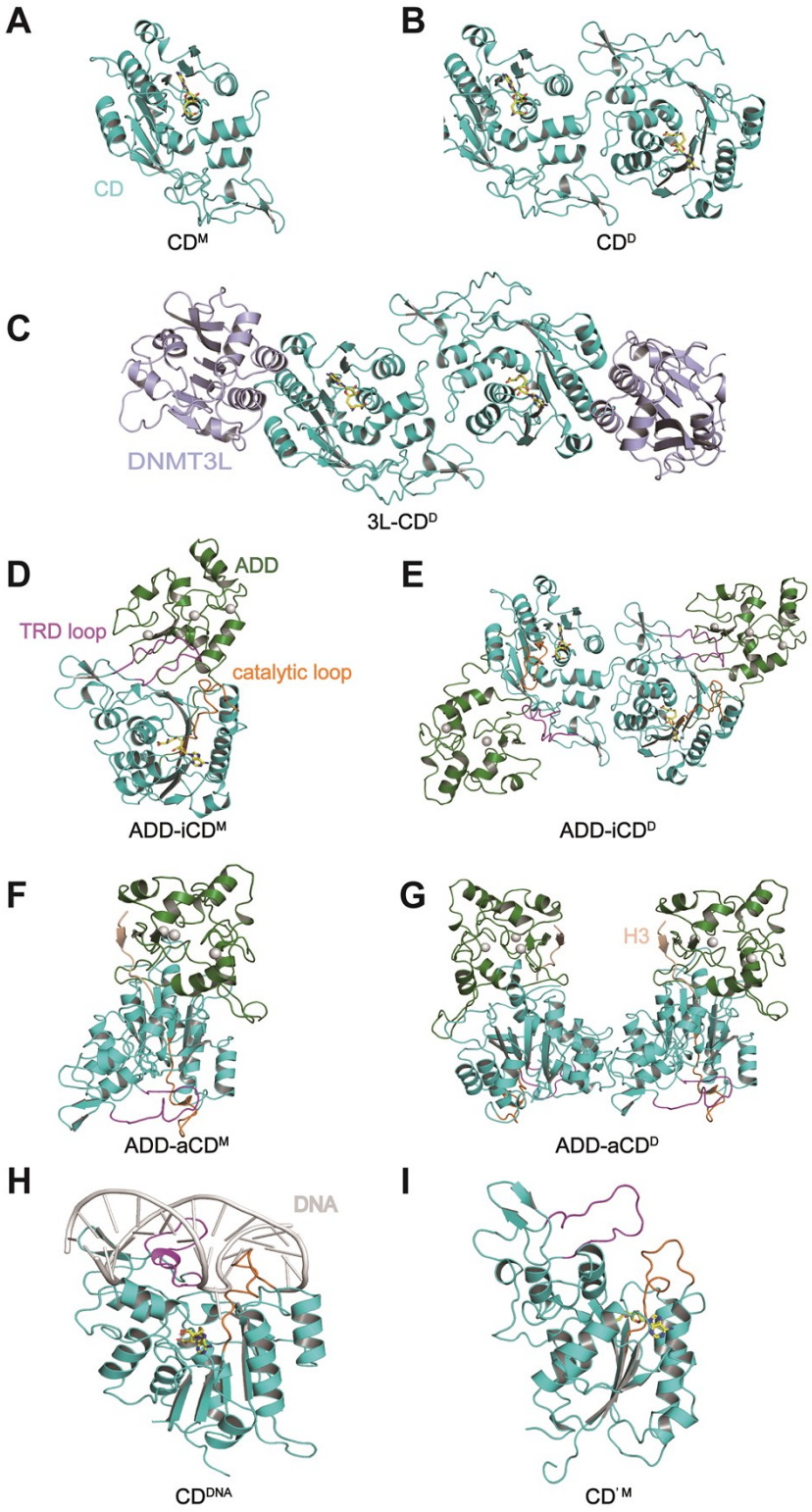

Figure S1. The initial structures of the systems CD^{M}, CD^{D}, 3L-CD^{D}, ADD-iCD^{M}, ADD-iCD^{D}, ADD-aCD^{M}, ADD-aCD^{D}, CD^{DNA} and CD'^{M} for the molecular dynamics simulations. AdoHcy is shown in stick.

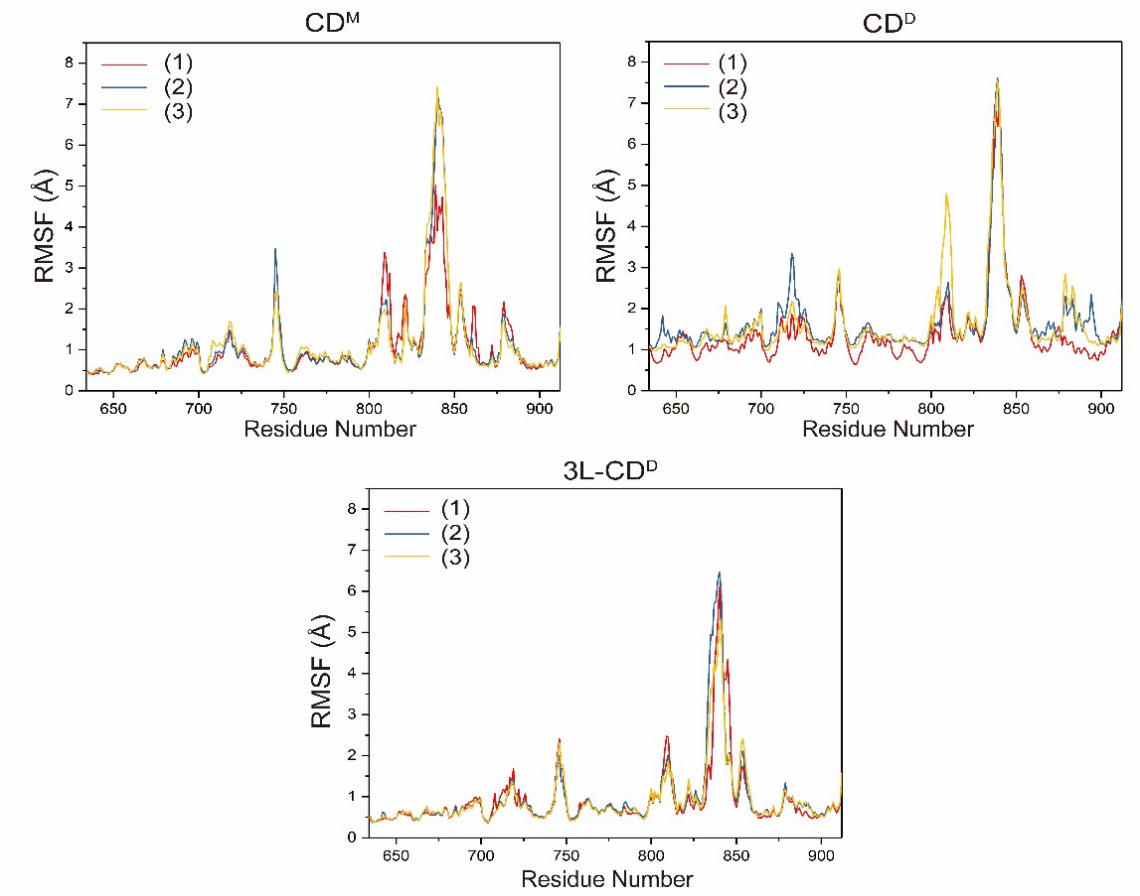

Figure S2. Residue-based RMSFs calculated from the three replicas of the CD monomer, homodimer, and heterotetramer systems.

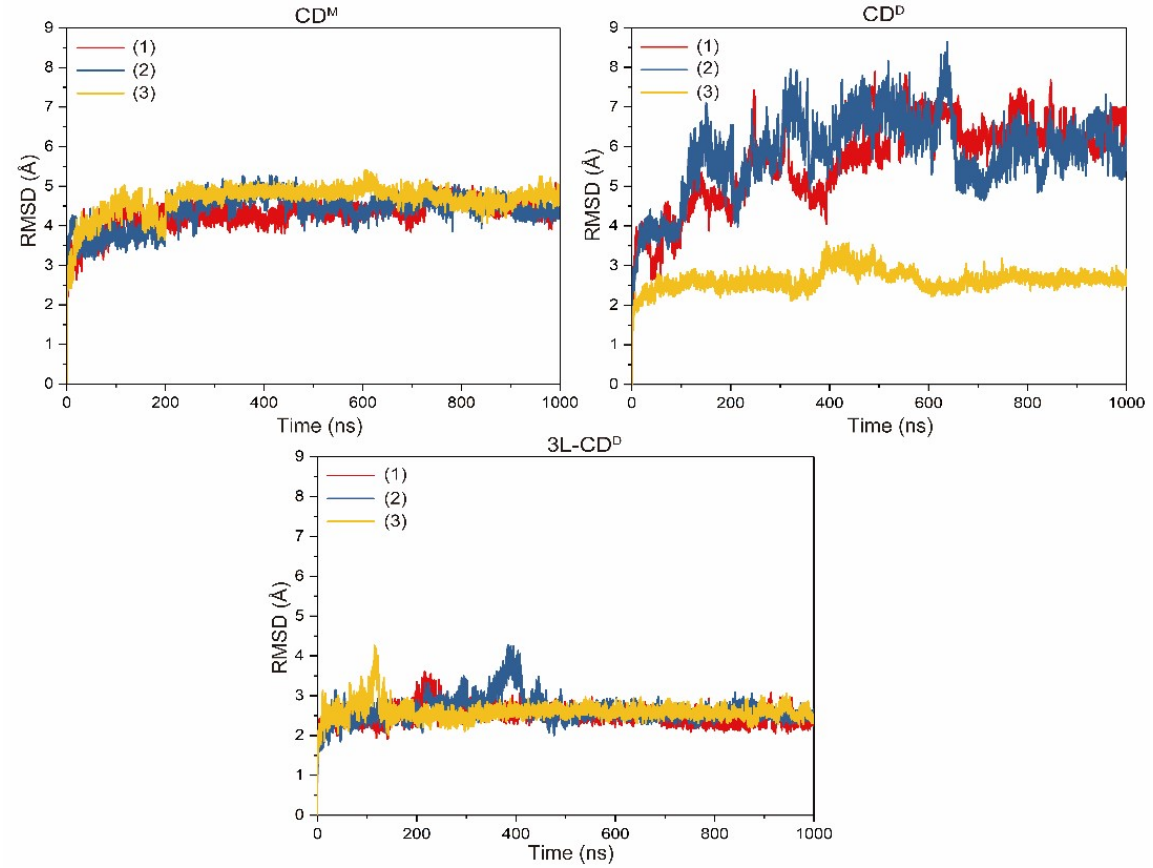

Figure S3. Three replicas of the RMSDs of the CD monomer, homodimer, and heterotetramer over the simulation time.

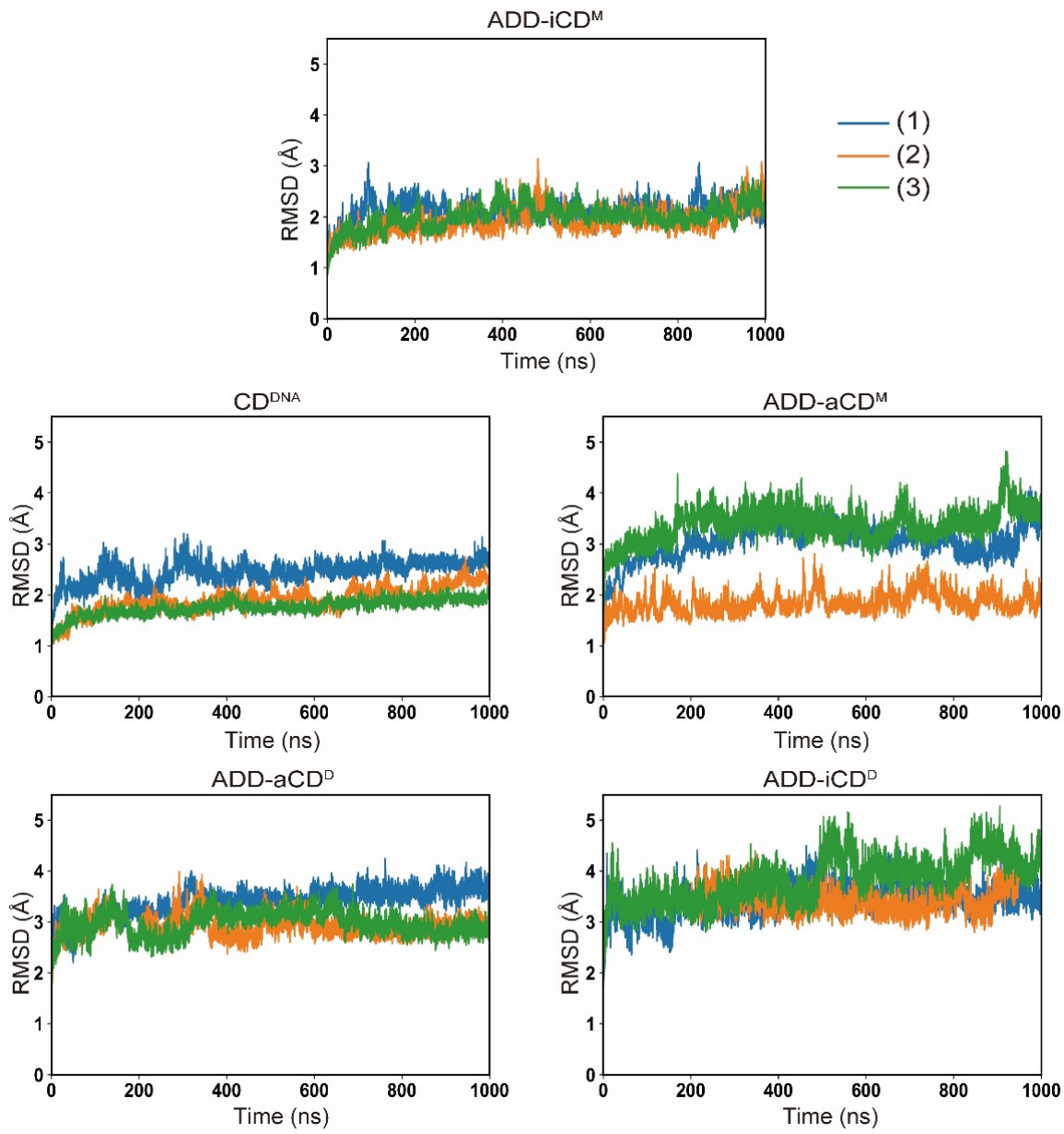

Figure S4. Three replicas of the RMSDs of the ADD-iCD^{M}, CD^{DNA}, ADD-aCD^{M}, ADD-aCD^{D}, ADD-iCD^{D} over the simulation time.

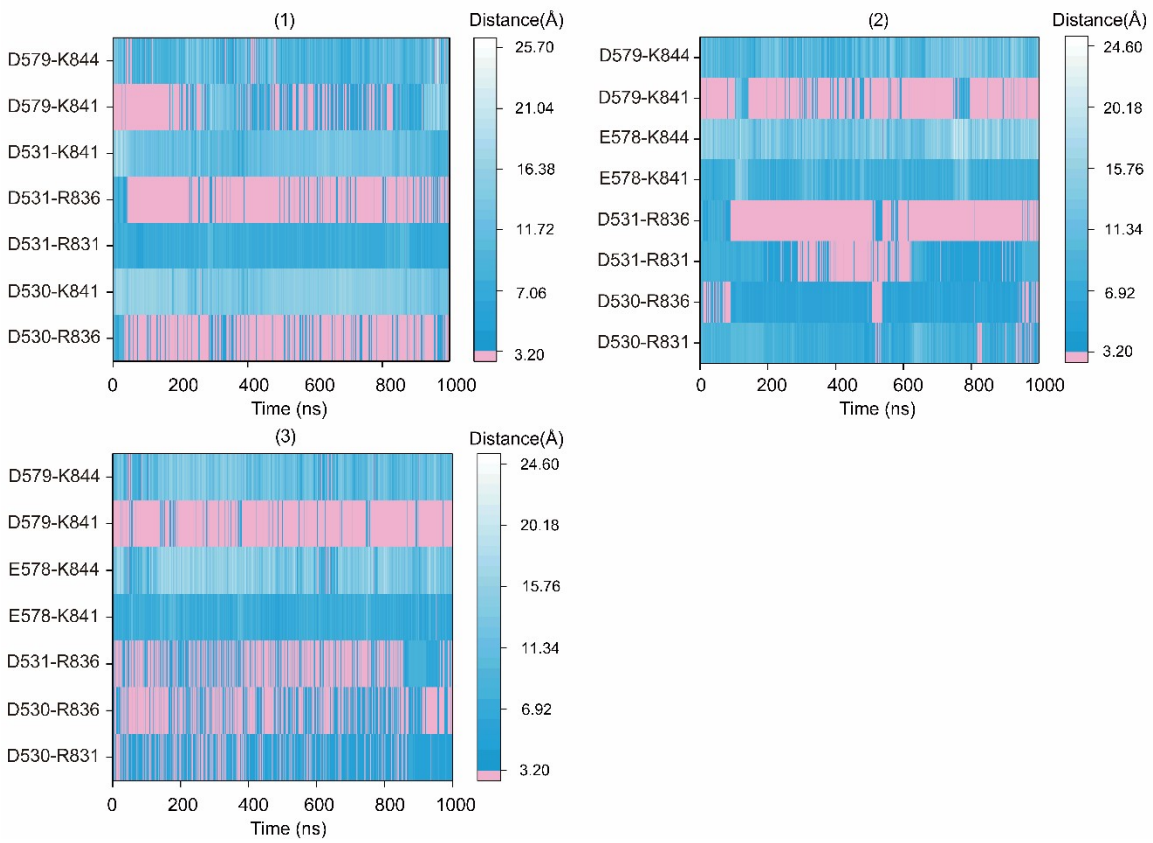

Figure S5. Salt bridges between the ADD domain and the TRD loop in ADD-iCD^{M} as a function of the simulation time (three replicas). The color bar represents the O-N distance between the acidic and basic side chain of the salt bridge pairs.

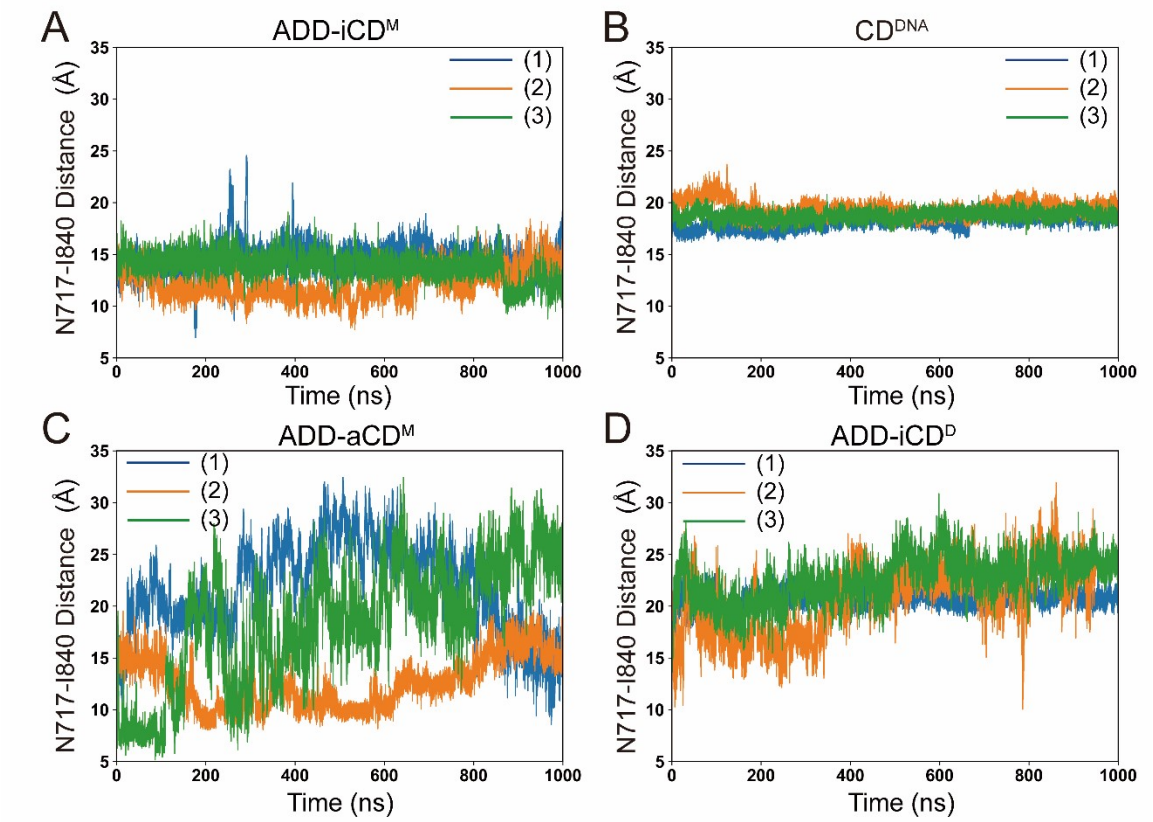

Figure S6. The center-of-mass distances (three replicas) between the I840 (TRD loop) and the N717 (catalytic loop) as a function of the simulation time in ADD-iCD^{M} (A), CD^{DNA} (B), ADD-aCD^{M} (C) and ADD-CD^{D} (D).

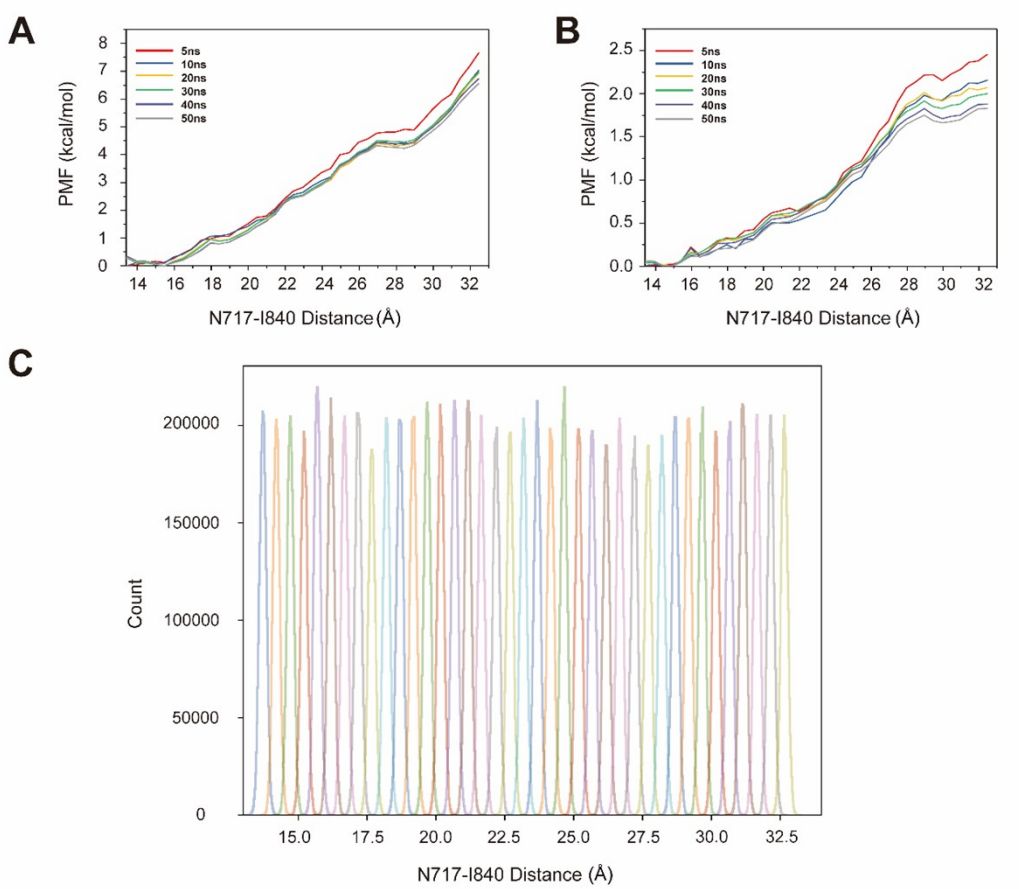

Figure S7. The convergence of the umbrella sampling for the systems (A) ADD-iCD^{M} and (B) CD'^{M}. (C) The histogram of the umbrella sampling for the system ADD-iCD^{M}.

Table S1. The densest clustered motion and max number of the cluster regions based on cluster-based wavelet feature extraction for CD^{M}, CD^{D}, and 3L-CD^{D}.

System	The densest clustered region	Total number of clusters
CD^{M} (1)	S807-D811	601
CD^{M} (2)	K783-H789	580
CD^{M} (3)	K693-Q696	634
CD^{D} (1)	R771-V912	333
CD^{D} (2)	A610-K812	47
CD^{D} (3)	A610-P718	295
3L-CD^{D} (1)	L723-H739	1111
3L-CD^{D} (2)	H789-G796	323
3L-CD^{D} (3)	P777-V850	1311

Table S2. H-bond occupancy (larger than 50% listed) in three replicas

System	Interface	(1)		(2)		(3)	
		H-bond[a]	Occupancy %	H-bond	Occupancy %	H-bond	Occupancy %
ADD-iCD^{M}	ADD-TRD loop	D531-R831 D531-R836	76.6 52.5	D531-R836[b] D531-R836[b]	69.8 56.4	D531-R836[c] D531-R836[c]	59.8 54.7
	TRD loop-TRD loop	K844-F848 Q842-Q842[d]	52. 5 50.5	K844-F848 K844-H847 I833-R836	77.7 59.2 50.2	I833-R836 K844-H847	93.2 56.4
ADD-iCD^{D}	ADD-TRD loop	K841-K577 D530-R831 V546-Q842	83.8 80.6 66.1	N552-R831	64.0	R544-Q842 E545-R836	75.7 52.5
	TRD loop-TRD loop	R836-S839 T835-N838 I833-R831 R831-T834 D845-F848 G843-Q846	93.6 87.6 79.5 63.0 58.6 54.0	R831-T834 K844-H847 Q846-R831 S839-K841 D845-F848	81.0 59.5 58.2 52.1 50.2	S839-Q842 S837-Q842 D845-F848 S839-K841 K844-H847	80.5 76.2 67.0 61.8 54.8
ADD-aCD^{M}	TRD loop-TRD loop	K844-H847 K844-F848	65.9 61.5	G843-H847 S837-I840	85.2 69.4	K844-F848 K844-H847	85.2 81.8
ADD-aCD^{D}		Q846-R831	85.9	Q846-R83	84.9	Q846-R831	71.2

CD^{DNA}	TRD loop-TRD loop	H847-Q842 Q842-R831 I840-Q842 G843-Q846 R836-S839 Q842-H847 R836-S839 S839-T834	97.8 95.6 94.0 79.6 73.8 72.5 68.1 63.1	S839-T834 R836-S839 R836-S839 G843-Q846 S837-I840 T834-R836 N838-G843 S839-Q842	94.2 91.7 83.0 82.1 79.3 67.8 53.7 52.8	N838-G843 S839-T834 S839-Q842 G843-Q846 R836-S839 Q842-D845 T834-R836 R836-S839 S837-I840	91.2 89.7 81.2 78.8 75.0 69.9 68.8 61.1 53.5
	TRD loop-DNA	G6-T835 G6-T835 G4-T832 G4-T832 T3-R831	66.6 66.1 60.6 57.4 55.1	G6-T835 G6-T835 T835-T11'[e] T11'-R836	83.7 83.1 61.0 51.3	G6-T835 G6-T835 T835- T11' G4-R831	98.7 98.3 59.4 54.0

[a]The hydrogen bond pairs are arranged in the order of hydrogen bond acceptor - hydrogen bond donor.

[b]The atoms OD2 and OD1 of D531 formed hydrogen bonds with atoms HH22 and HH12 of R836, respectively, with an occupancy of 69.8% and 56.4%.

[c]The atoms OD1 and OD2 of D531 formed hydrogen bonds with atoms HH22 of R836, respectively, with an occupancy of 59.8% and 54.7%.

[d]The atoms OE1 and H of Q842 formed hydrogen bonds with an occupancy of 50.5%.

[e]The base with ‘’ represents the single strand of DNA where the unmethylated cytosine is located.

Table S3. The number of the H-bonds of the catalytic loop with the surrounding residues and within itself in the ADD-CD^{M}, ADD-CD^{D} and CD^{DNA} systems.

System	(1)	(2)	(3)	Average
H-bonds between the catalytic loop and its surrounding residues				
ADD-aCD^{M}	8.3±1.1	7.5±1.6	6.4±1.6	7.4±1.7
ADD-iCD^{M}	6.6±1.1	6.3±1.3	7.2±1.6	6.7±1.4
CD^{DNA}	10.1±1.9	9.7±2.3	9.8±2.0	9.9±2.1
ADD-aCD^{D}	7.0±1.7	6.3±1.2	6.2±1.2	6.5±1.4
ADD-iCD^{D}	6.1±1.4	8.8±1.4	8.9±1.6	7.9±2.0
H-bonds within the catalytic loop itself				
ADD-aCD^{M}	8.8±1.7	7.0±1.5	7.5±1.9	7.8±1.9
ADD-iCD^{M}	7.0±1.4	7.1±1.4	7.2±1.4	7.1±1.4
CD^{DNA}	7.4±1.5	7.2±1.5	7.5±1.6	7.4±1.5
ADD-aCD^{D}	7.7±1.8	7.2±1.4	7.0±1.5	7.3±1.6
ADD-iCD^{D}	7.2±1.5	7.0±1.6	7.0±2.1	7.1±1.7

Table S4. The center-of-mass distances (Å) between the I840 (TRD loop) and the N717 (catalytic loop) as a function of the simulation time in ADD-iCD^{M}, CD^{DNA}, ADD-aCD^{M} and ADD-CD^{D}.

System	(1)	(2)	(3)	Average
ADD-iCD^{M}	14.6±1.4	12.2±1.4	13.8±1.2	13.5±1.7
CD^{DNA}	18.1±0.6	19.5±0.8	18.8±0.5	18.8±0.8
ADD-aCD^{M}	21.7±4.3	12.5±2.4	18.3±5.8	17.5±5.8
ADD-iCD^{D}	20.7±0.8	20.6±3.3	22.6±2.1	21.3±2.5