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

Exploration of Gated Ligand Binding Recognizes an Allosteric Site for Blocking FABP4-Protein Interaction

Yan Li,[*,†,§] Xiang Li,[‡,§] and Zigang Dong[*,†]

[†]The Hormel Institute, University of Minnesota, Austin Minnesota 55912, United States

[‡]Department of Physiology and Pathophysiology, School of Basic Medical Sciences, Zhengzhou University, 450001 Zhengzhou, Henan, China

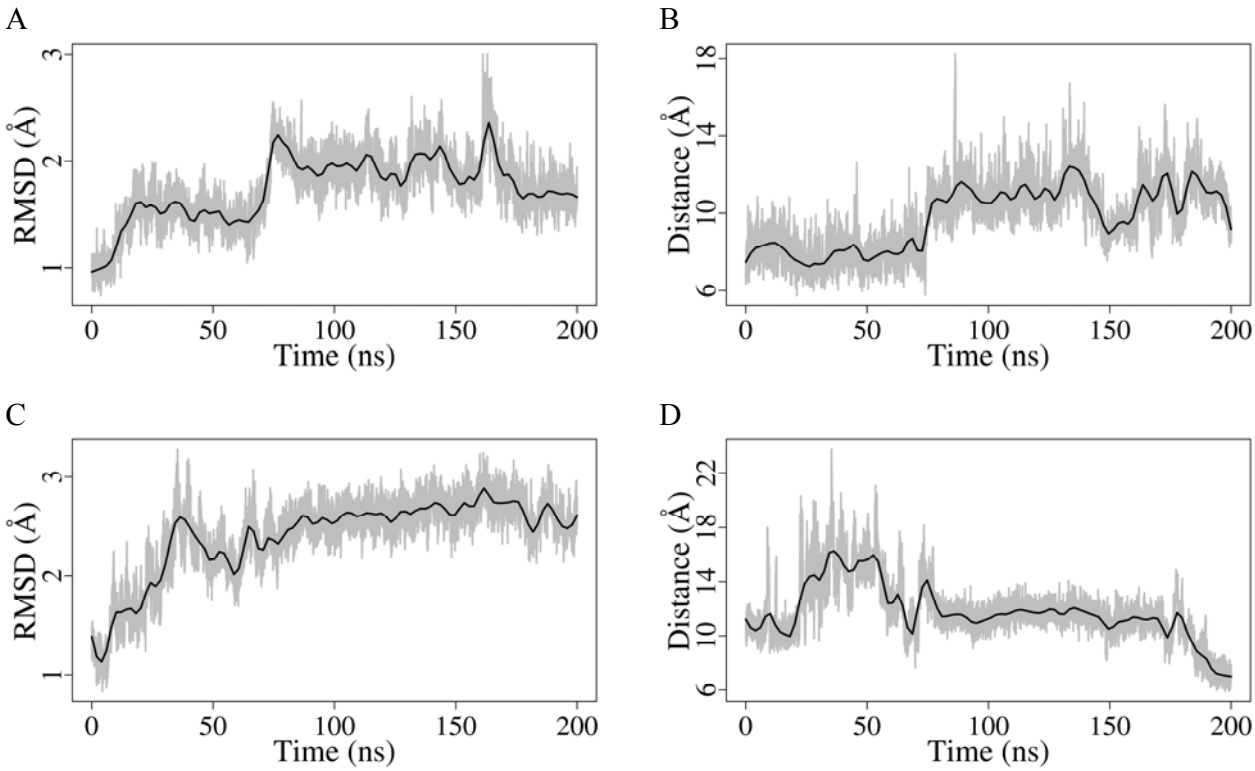

Figure S1. Time evolution of the backbone RMSD and the Thr29-Phe57 distance in two trajectories. (A) The backbone RMSD in Traj. 16. (B) The distance between Thr29 and Phe57 in Traj. 16. (C) The backbone RMSD in Traj. 178. (D) The distance between Thr29 and Phe57 in Traj. 178.

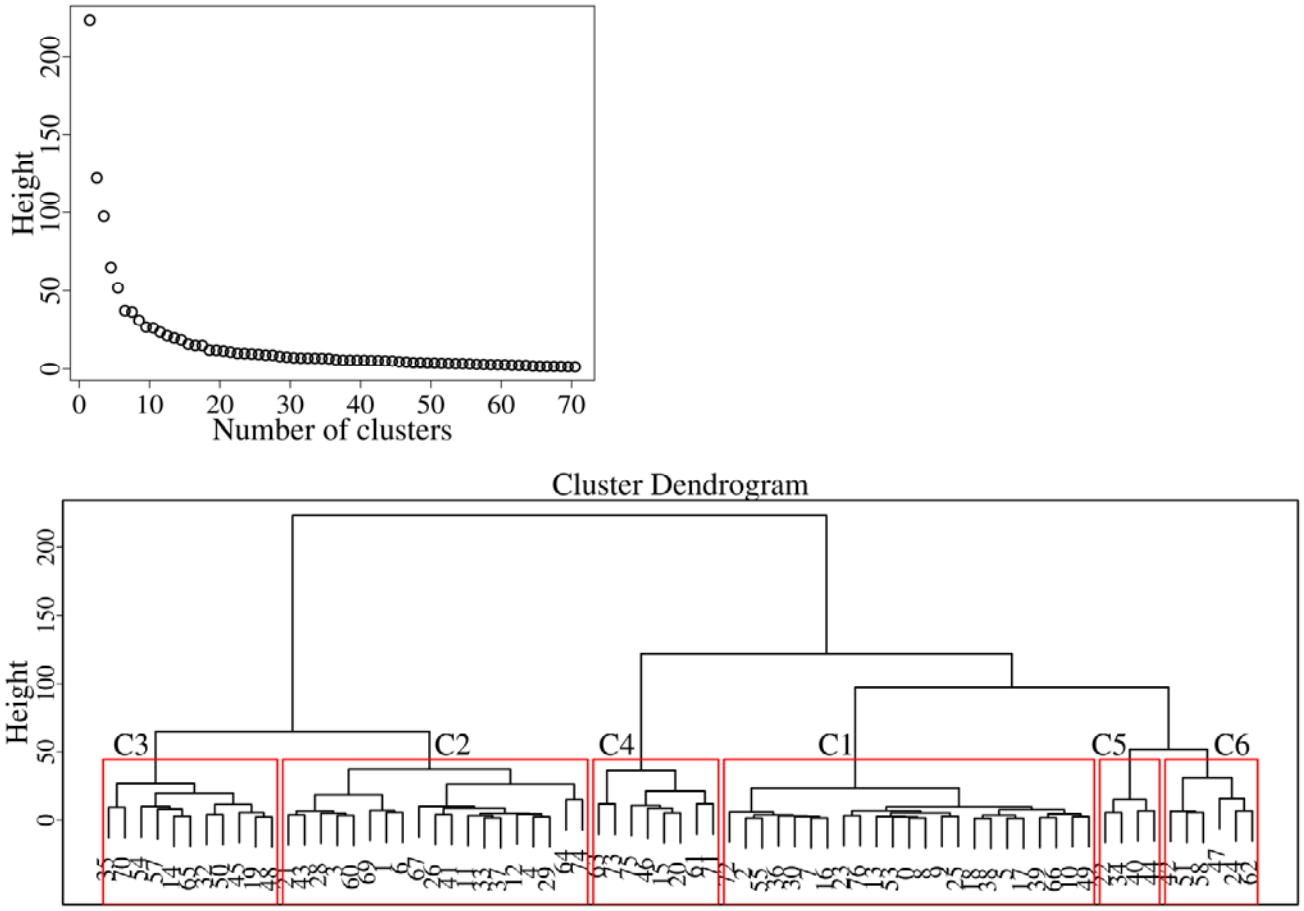

Figure S2. Hierarchical clustering of 71 classes obtained from Bayesian clustering of the dataset MD200. The macrostates are boxed by red line.

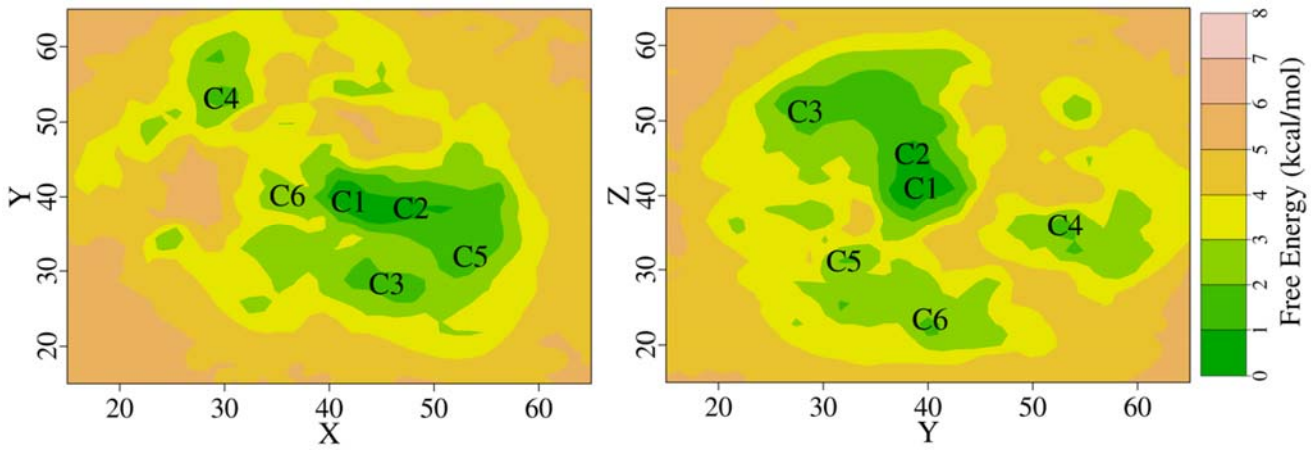

Figure S3. The 2D FES obtained from the dataset of MD200. Points sampled from MD simulations are assigned into 75×76 and 76×75 bins on the *xy* and *yz* plane with the bin area of 1.4×1.4 Å^2.

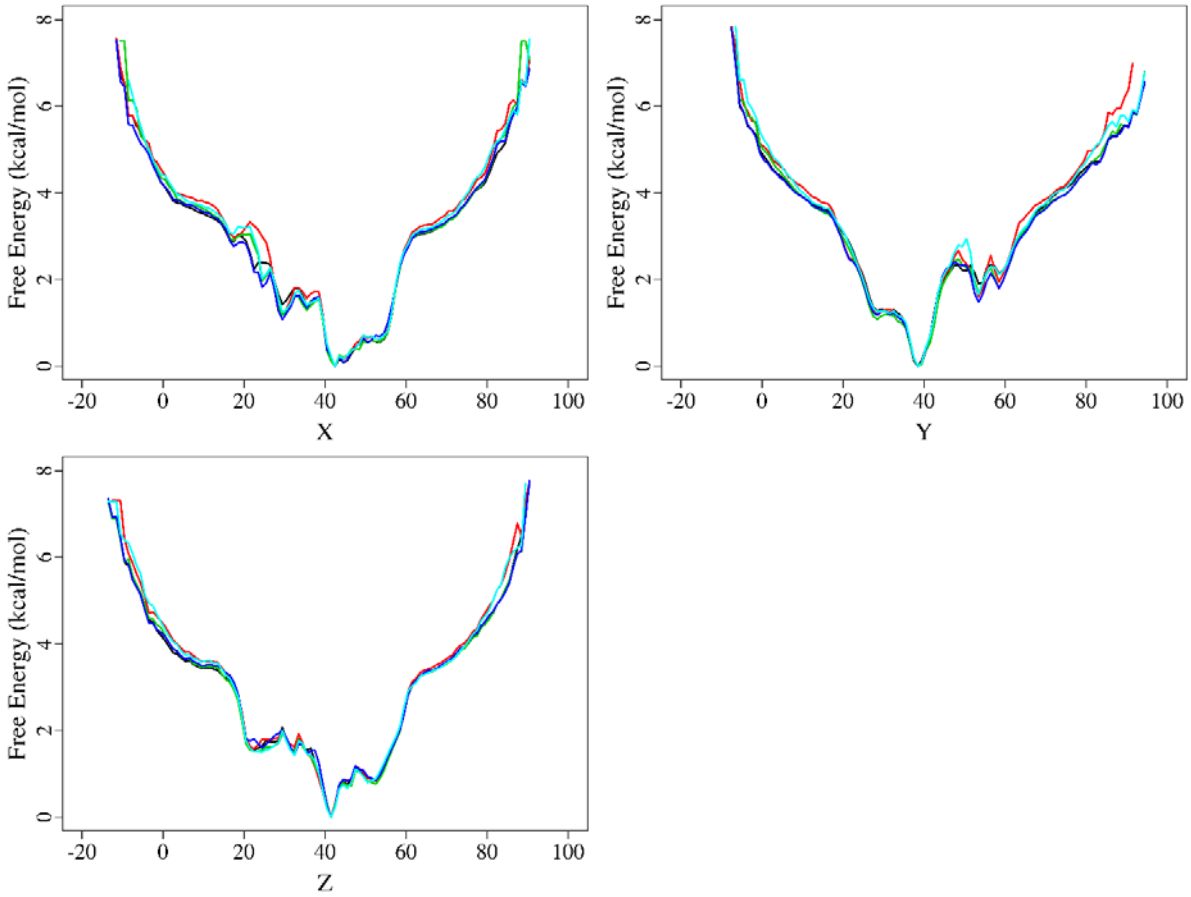

Figure S4. Convergence of free energy calculations in the dataset MD200. One-dimensional free energy surface on each coordinate of the atom OAQ, respectively, is calculated for 150 trajectories randomly chosen. It is repeated 5 times.

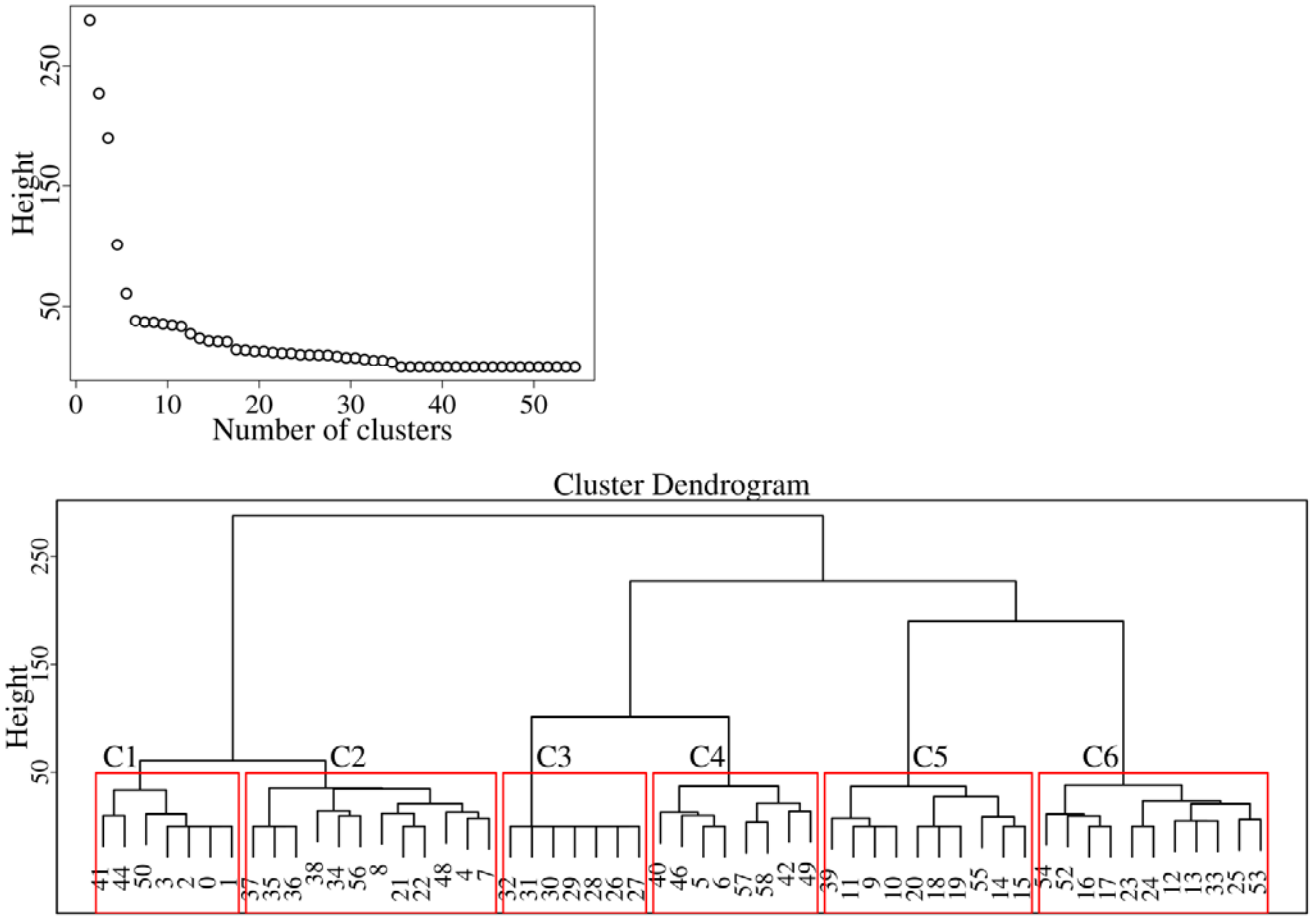

Figure S5. Hierarchical clustering of 55 classes obtained from Bayesian clustering of the dataset MD240. The macrostates are boxed by red line.

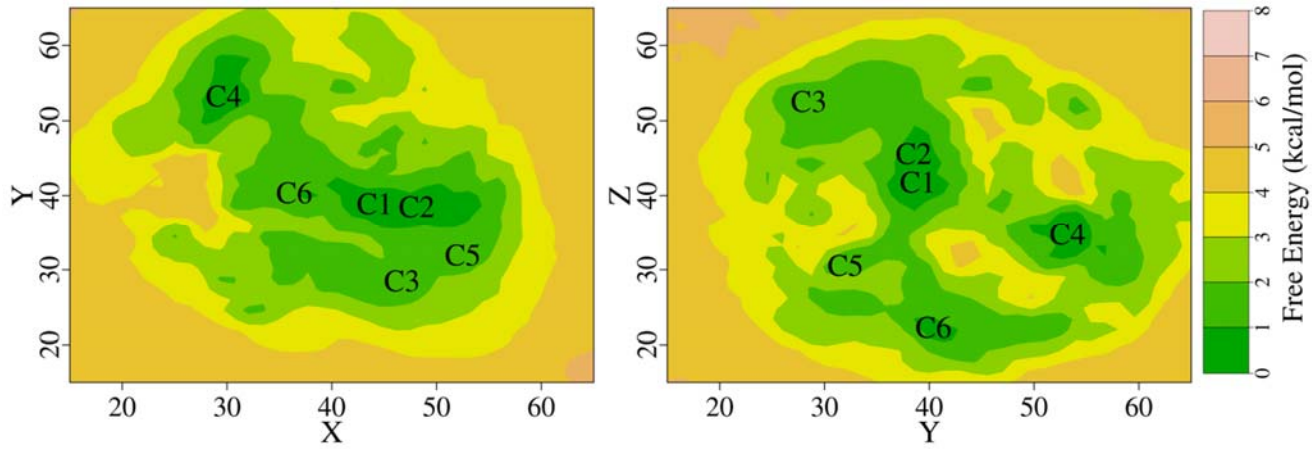

Figure S6. The 2D FES obtained from the dataset MD240. Points sampled from MD simulations are assigned into 75×79 and 79×80 bins on the *xy* and *yz* plane with the bin area of 1.4×1.4 $Å^2$.

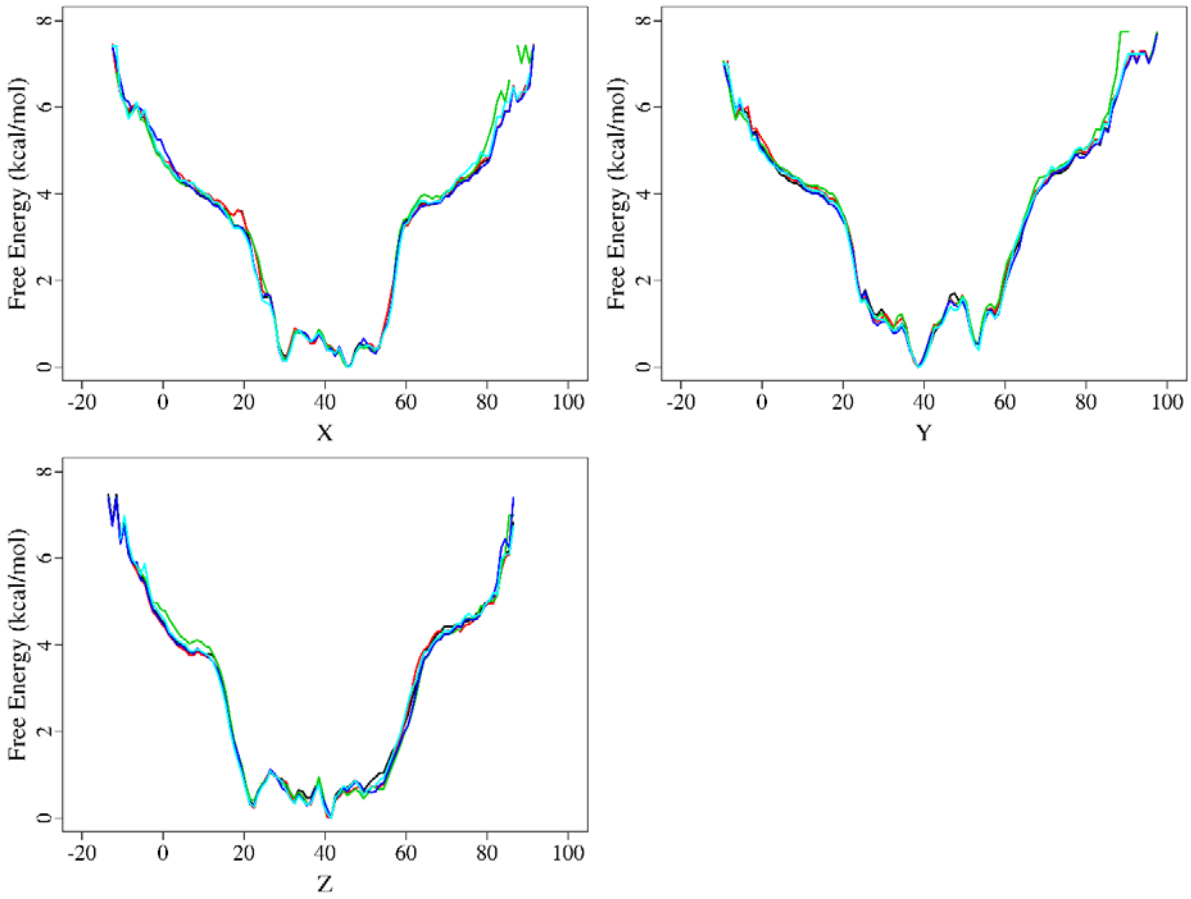

Figure S7. Convergence of free energy calculations in the dataset MD240. One-dimensional free energy surface on each coordinate of the atom OAQ, respectively, is calculated for 180 trajectories randomly chosen. It is repeated 5 time.

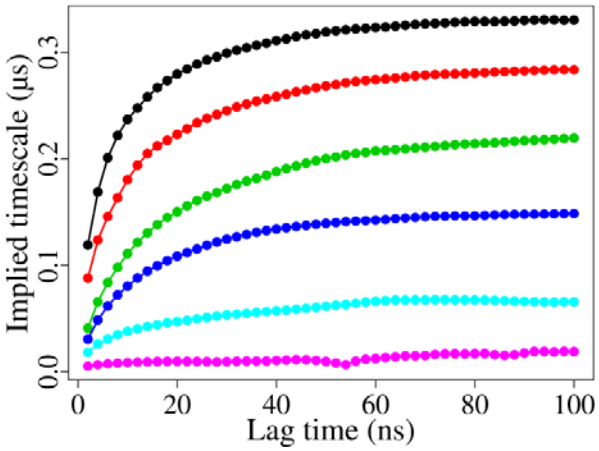

Figure S8. Evolution of the six slowest relaxation timescales as a function of the lag time. Convergence of the curves happens after 80 ns.

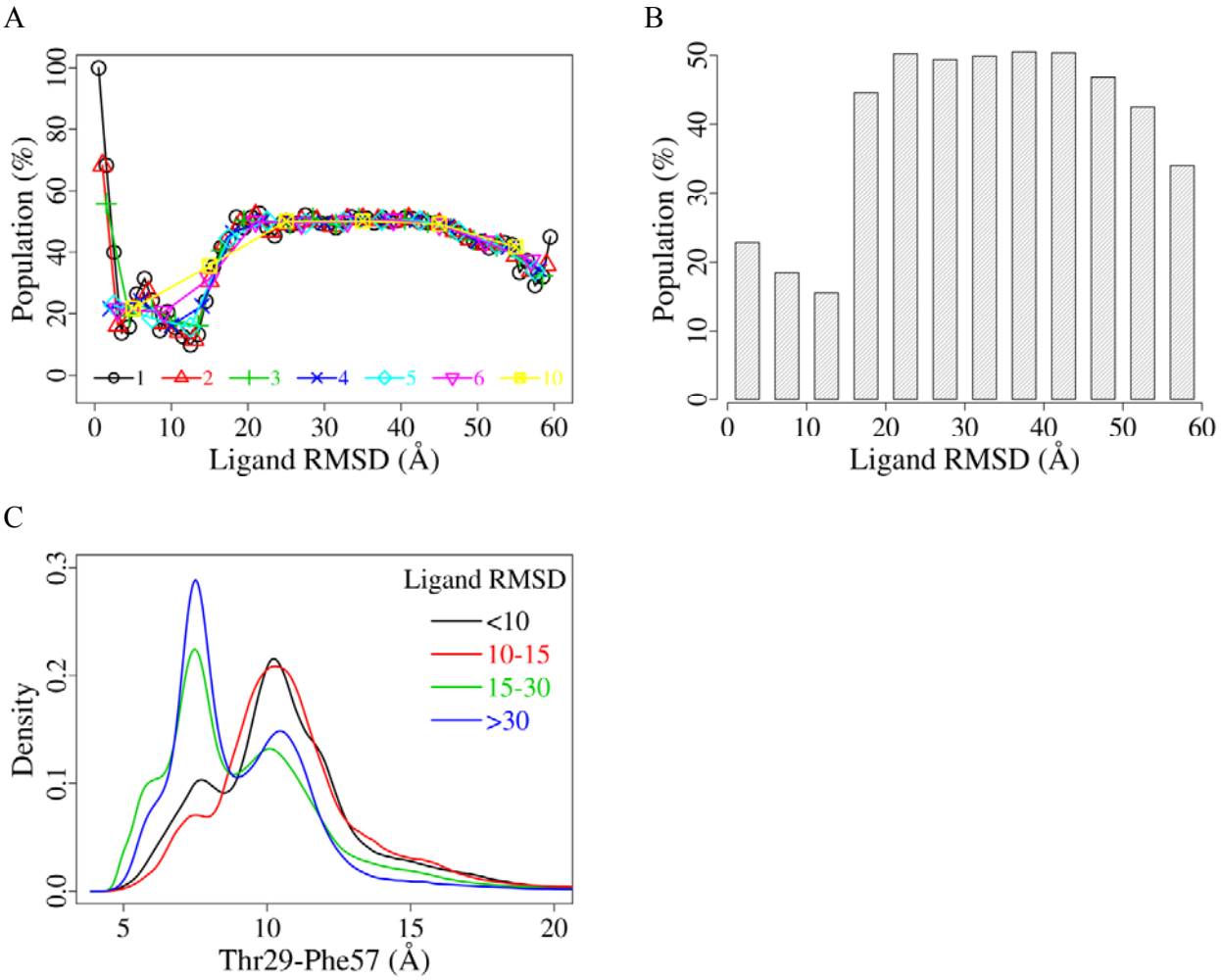

Figure S9. Distribution of FABP4 conformations as a function of the ligand RMSD for the dataset MD240. (A) Effect of the bin width on population distribution. The population of the closed form is calculated with a bin width of 1, 2, 3, 4, 5, 6 and 10 Å, respectively. As the bin width increases, the population at two ends decreases rapidly, suggesting inadequate sampling with a narrow bin (for example, 1 Å). With a wide bin, the variation of the population cannot be described very well. The bin width of 5 Å is determined to be optimal. (B) Population of the closed form with a bin width of 5 Å. (C) Distribution of FABP4 conformations at distinct ranges of the ligand RMSD. When the ligand RMSD is over 30 Å, the ligand is in the solvent. When the ligand RMSD is between 15 and 30 Å, the ligand is on the protein surface. When the ligand RMSD is between 10 and 15 Å, the ligand is around the portal. When the ligand RMSD is below 10 Å, the ligand is inside the protein. The results agree very well with the data shown in Figure 8 and Table 1.

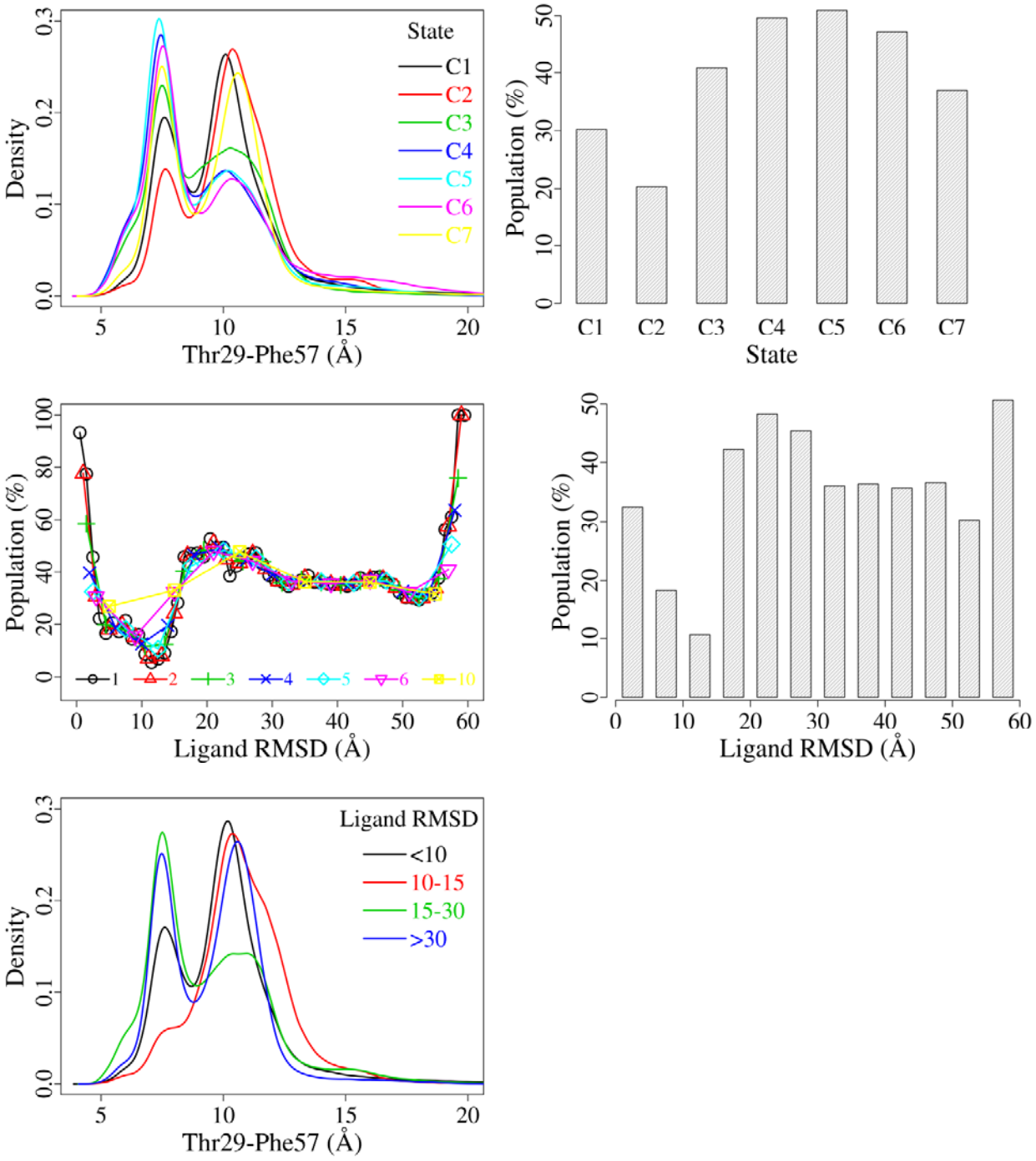

Figure S10. Distribution of FABP4 conformations as a function of the ligand RMSD for the dataset MD200. It is almost the same with Figure 8 and Figure S9.

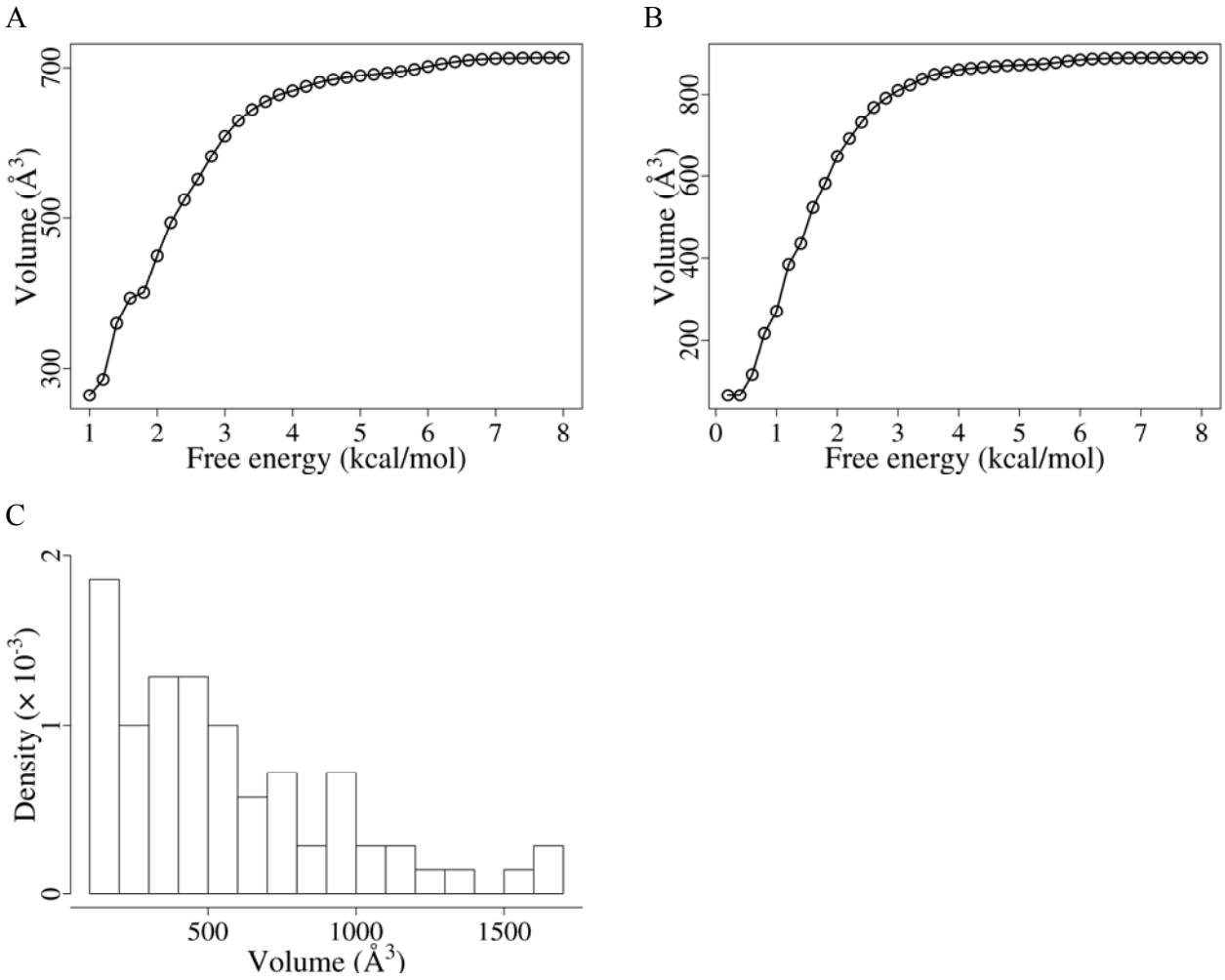

Figure S11. Calculation of the bound volume (V_b). (A) Sensitivity analysis of the 3D FES from the dataset MD200. (B) Sensitivity analysis of the 3D FES from the dataset MD240. (C) Histogram analysis of the volume of the binding cavity computed from 70 MD trajectories. For the FABP4-TGZ system, the bound volume should be equal to the volume of the binding cavity enclosed in the protein. The volume of the binding cavity in the crystal structure (PDB entry 2QM9) is computed to be 427.0 Å^3 using the software SiteMap implemented in Maestro v9.3. This value is achieved from the static structure, and cannot reflect the dynamics of the protein and ligand. Therefore, it is expected to be underestimated. Then we examine the covered space by the atom OAQ in MD simulations. For MD200, the ligand fluctuates inside the binding cavity in the first 70 MD simulations (Traj 1–71 except 16). The bound volume is estimated to be the space covered by the atom OAQ in the simulation, assuming that the space is cubic. The mean value of the volume with standard deviation from 70 trajectories is 561.48±379.42 Å^3. Thus, the bound volume obtained from the three-dimensional FES, the crystal structure and MD simulations is in the same order of magnitude.

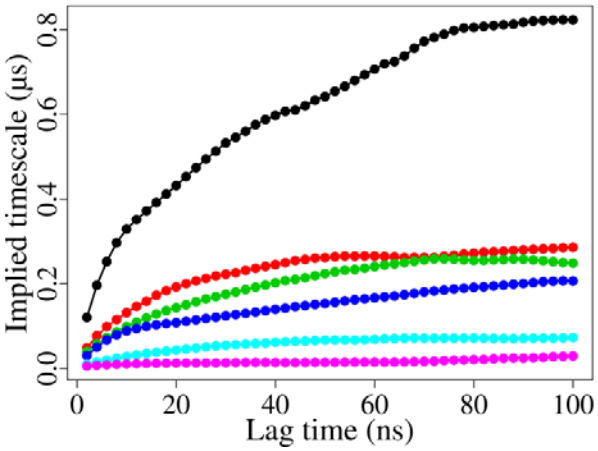

Figure S12. Evolution of the six slowest relaxation timescales as a function of the lag time for the dataset MD200. Convergence of the curves happens after 80 ns.

Table S1. The starting and final ligand RMSD (Å) in the 200 MD trajectories (MD200).

Traj	Start	Final	Traj	Start	Final	Traj	Start	Final	Traj	Start	Final	Traj	Start	Final
1	2.16	3.50	41	2.17	1.91	81	9.68	4.12	121	23.46	9.53	161	34.01	20.23
2	2.30	5.28	42	2.46	5.08	82	11.14	9.42	122	22.80	13.13	162	33.18	9.12
3	2.24	3.98	43	2.37	3.94	83	12.20	6.16	123	23.33	22.24	163	33.66	21.93
4	2.39	5.48	44	2.68	4.10	84	13.61	7.70	124	23.69	10.01	164	34.26	24.49
5	2.13	7.73	45	2.28	5.66	85	12.91	12.26	125	23.76	15.50	165	34.24	15.93
6	2.15	4.22	46	2.41	5.25	86	13.81	9.15	126	24.04	8.20	166	34.33	12.97
7	2.35	4.84	47	2.67	4.13	87	13.49	4.77	127	25.84	15.98	167	35.36	7.47
8	2.45	8.20	48	2.79	5.52	88	14.27	8.87	128	25.44	16.13	168	35.67	19.51
9	2.36	5.79	49	2.47	7.94	89	14.93	8.74	129	26.35	3.28	169	35.35	14.11
10	2.30	6.91	50	2.79	6.01	90	15.11	9.10	130	26.87	15.30	170	35.94	21.72
11	2.37	3.88	51	3.02	4.02	91	14.99	17.89	131	27.12	12.03	171	37.17	14.88
12	2.51	4.72	52	3.38	8.92	92	15.61	9.83	132	27.36	21.29	172	37.36	9.70
13	2.52	7.57	53	3.22	3.40	93	16.68	14.87	133	27.40	17.85	173	36.73	25.11
14	2.48	4.89	54	3.55	6.37	94	17.54	7.47	134	27.46	24.72	174	37.73	17.49
15	2.25	4.42	55	3.43	5.63	95	17.49	18.08	135	28.26	20.63	175	37.72	5.12
16	2.17	9.17	56	3.89	3.41	96	18.39	17.33	136	28.84	17.06	176	37.14	23.54
17	2.24	3.48	57	3.96	3.64	97	18.06	6.72	137	28.86	7.43	177	37.17	15.72
18	2.62	4.07	58	4.41	3.73	98	18.70	13.79	138	28.47	15.11	178	38.26	3.41
19	2.53	6.37	59	4.25	5.71	99	18.61	7.84	139	28.24	20.13	179	38.55	5.86
20	2.46	3.76	60	4.63	7.24	100	18.45	21.31	140	29.23	17.53	180	38.27	8.29
21	2.87	6.16	61	4.56	8.33	101	17.91	16.96	141	28.72	22.47	181	38.94	16.90
22	2.52	4.18	62	4.94	6.51	102	18.32	18.46	142	29.28	16.56	182	38.91	23.60
23	2.41	4.30	63	5.15	4.08	103	18.36	5.47	143	29.68	20.01	183	38.89	15.87
24	2.21	7.43	64	4.96	4.56	104	18.70	13.66	144	29.74	16.54	184	38.49	21.12
25	2.15	2.58	65	5.21	4.27	105	18.50	5.58	145	30.46	17.58	185	37.82	24.03
26	2.00	4.12	66	4.79	8.90	106	19.03	15.35	146	30.05	20.82	186	38.20	7.37
27	2.22	4.98	67	4.83	1.72	107	19.71	19.45	147	30.41	14.08	187	38.50	10.43
28	2.14	4.40	68	4.98	4.02	108	20.08	15.55	148	30.27	12.67	188	38.02	12.33
29	2.05	5.37	69	5.81	4.40	109	21.47	15.55	149	30.91	22.09	189	38.86	9.37
30	2.29	3.47	70	6.49	10.70	110	22.16	12.19	150	31.42	6.74	190	38.34	16.01
31	2.28	2.87	71	7.39	4.51	111	22.62	15.42	151	30.57	16.10	191	37.90	19.89
32	2.41	4.25	72	7.71	5.53	112	22.37	16.89	152	30.32	21.72	192	39.06	11.69
33	2.28	3.98	73	7.71	8.30	113	22.41	4.93	153	30.71	17.07	193	38.63	20.18
34	2.20	3.94	74	8.04	4.29	114	21.84	9.76	154	30.72	19.44	194	37.88	18.52
35	2.23	10.47	75	7.82	7.94	115	22.21	12.87	155	32.03	14.67	195	38.14	7.66
36	2.35	9.23	76	8.89	5.18	116	22.49	15.90	156	31.86	18.19	196	37.63	18.57
37	2.56	4.18	77	8.24	5.55	117	22.92	9.36	157	32.40	15.96	197	38.04	19.14
38	2.48	5.56	78	8.69	2.04	118	23.82	15.52	158	32.49	7.50	198	37.52	21.58
39	2.46	6.39	79	9.36	4.10	119	24.21	15.49	159	33.18	13.58	199	37.36	18.90
40	2.21	2.55	80	8.94	5.53	120	23.72	7.26	160	33.81	16.22	200	36.35	18.86

Table S2. Net flux between pairs of the seven states for ligand binding (from C7 to C1) computed with MSM and TPT.

	Net flux (ms^{-1})					
	All	Subset				
		1	2	3	mean	sd
C2→C1	12.19	11.07	10.83	11.98	11.29	0.61
C3→C1	1.50	1.53	1.52	1.54	1.53	0.01
C3→C2	5.58	4.86	4.29	5.37	4.84	0.54
C3→C5	1.49	1.39	1.76	1.28	1.47	0.25
C4→C1	0.08	0.08	0.10	0.09	0.09	0.01
C4→C2	0.22	0.16	0.27	0.14	0.19	0.07
C4→C3	3.02	2.88	2.57	3.22	2.89	0.33
C4→C6	0.15	0.10	0.07	0.19	0.12	0.06
C5→C1	0.32	0.35	0.27	0.30	0.31	0.04
C5→C2	1.24	1.23	1.29	1.08	1.20	0.11
C5→C4	0.60	0.50	0.88	0.52	0.63	0.22
C5→C6	0.11	0.16	0.13	0.15	0.14	0.02
C6→C1	0.10	0.13	0.06	0.10	0.10	0.04
C6→C2	0.28	0.24	0.25	0.34	0.28	0.06
C6→C3	0.37	0.30	0.43	0.48	0.40	0.09
C7→C1	1.42	1.80	1.25	1.18	1.41	0.34
C7→C2	4.87	4.58	4.74	5.05	4.79	0.24
C7→C3	5.17	4.59	4.57	4.48	4.55	0.06
C7→C4	2.87	2.73	2.12	3.13	2.66	0.51
C7→C5	0.77	0.85	0.81	0.78	0.81	0.04
C7→C6	0.49	0.41	0.55	0.58	0.51	0.09
total	42.82	39.94	38.76	41.97	40.22	1.62

Table S3. Percentage of net flux between pairs of the seven states for ligand binding (from C7 to C1) computed with MSM and TPT.

	Percentage (%)					
	All	Subset				
		1	2	3	mean	sd
C2→C1	28.47	27.7	27.94	28.54	28.06	0.43
C3→C1	3.49	3.83	3.93	3.66	3.81	0.13
C3→C2	13.04	12.17	11.06	12.78	12.00	0.88
C3→C5	3.47	3.47	4.54	3.04	3.68	0.77
C4→C1	0.18	0.21	0.26	0.22	0.23	0.03
C4→C2	0.50	0.41	0.69	0.33	0.48	0.19
C4→C3	7.06	7.22	6.62	7.68	7.17	0.53
C4→C6	0.35	0.24	0.18	0.46	0.30	0.15
C5→C1	0.74	0.88	0.71	0.72	0.77	0.09
C5→C2	2.88	3.08	3.32	2.58	2.99	0.38
C5→C4	1.39	1.24	2.28	1.23	1.58	0.60
C5→C6	0.25	0.40	0.32	0.35	0.36	0.04
C6→C1	0.22	0.33	0.16	0.24	0.25	0.09
C6→C2	0.66	0.60	0.65	0.81	0.69	0.11
C6→C3	0.86	0.75	1.11	1.13	1.00	0.22
C7→C1	3.31	4.51	3.23	2.80	3.52	0.89
C7→C2	11.38	11.45	12.22	12.03	11.9	0.40
C7→C3	12.08	11.5	11.80	10.68	11.33	0.58
C7→C4	6.70	6.83	5.48	7.46	6.59	1.02
C7→C5	1.80	2.13	2.09	1.85	2.02	0.15
C7→C6	1.15	1.03	1.41	1.37	1.27	0.21
total	100	100	100	100		

Table S4. MFPT to C1 from another state calculated with the transition probability matrix.

	MFPT (ns)					
	All	Subset				
		1	2	3	mean	sd
C2→C1	970.90	944.40	1079.60	1059.15	1027.72	72.88
C3→C1	1635.09	1509.88	1780.47	1727.48	1672.61	143.40
C4→C1	2173.26	2097.62	2320.50	2279.09	2232.40	118.55
C5→C1	1882.39	1775.76	1995.93	1981.28	1917.66	123.10
C6→C1	1991.48	1900.83	2113.95	2090.90	2035.23	116.96
C7→C1	1638.33	1537.44	1720.55	1733.04	1663.68	109.50

Table S5. MFPT from C7 to C1 calculated with the transition probability matrix from 120 MD trajectories.

	MFPT (ns)					
	All	Subset (90 trajectories randomly selected)				
		1	2	3	mean	sd
C7→C1	1278.85	1346.31	1226.23	1315.22	1295.92	62.32

Table S6. MFPT and rate constant for opening and closing of FABP4.

		All	Subset				
			1	2	3	mean	sd
MFPT (ns)	closed→open	175.32	175.13	176.46	173.46	175.02	1.50
	open→closed	309.78	315.07	307.07	304.80	308.98	5.39
Rate constant (ms^{-1})	k_{open}	5703.70	5710.10	5666.97	5764.98	5714.02	49.13
	k_{close}	3228.06	3173.89	3256.63	3280.88	3237.13	56.09

Table S7. Rate constant for ligand binding and unbinding.

	Rate constant (ms^{-1})					
	All	Subset				
		1	2	3	mean	sd
k_{on}	610.38	650.43	581.21	577.02	602.89	41.23
k_{off}	46.04	42.56	41.52	45.26	43.11	1.93

Table S8. Binding affinity of FABP4-TGZ.

	All	Subset				
		1	2	3	mean	sd
ΔG (kcal/mol)	-7.96	-7.85	-7.79	-7.71	-7.79	0.07

Table S9. The number of trajectories in three subsets of MD200. The trajectories are randomly selected from E1, E2, and E3, respectively.

	The number of trajectories			
	All	Subset		
		1	2	3
E1	71	35	50	64
E2	44	22	30	39
E3	85	43	60	77
total	200	100	140	180

Table S10. Net flux between pairs of the seven states for the transition from C7 to C1 computed with MSM and TPT for MD200.

	Net flux (ms^{-1})				Percentage (%)			
	All	Subset			All	Subset		
		1	2	3		1	2	3
C2→C1	18.91	24.51	21.22	15.82	34.94	34.47	34.43	34.41
C3→C1	1.29	0.16	1.22	0.83	0.22	1.99	1.81	2.34
C3→C2	3.53	3.68	2.13	3.18	5.25	3.46	6.92	6.42
C3→C6		0.01			0.01			
C4→C1	0.18	0.45	0.44	0.23	0.64	0.71	0.50	0.33
C4→C2	3.64	5.73	5.90	3.72	8.17	9.58	8.09	6.63
C4→C3	0.82	1.32	0.51	0.55	1.88	0.83	1.19	1.48
C5→C1	0.05	0.01	0.09	0.00	0.01	0.14	0.01	0.09
C5→C2	1.04	2.12	1.47	0.86	3.02	2.38	1.87	1.88
C5→C3	0.74	0.09	0.27	0.64	0.12	0.44	1.38	1.35
C5→C4	1.19	0.91	1.47	0.60	1.30	2.39	1.31	2.16
C6→C1	0.01	0.02	0.01	0.01	0.03	0.02	0.01	0.02
C6→C2	0.57	1.00	0.65	0.34	1.43	1.05	0.74	1.04
C6→C3	0.41		0.01	0.27		0.02	0.58	0.74
C6→C4	0.12	0.93	0.53	0.69	1.33	0.86	1.50	0.22
C6→C5	0.42	0.69	0.30	0.06	0.98	0.48	0.13	0.77
C7→C1	1.62	3.39	2.38	1.27	4.83	3.87	2.76	2.95
C7→C2	10.13	11.97	11.08	7.72	17.07	18.00	16.81	18.43
C7→C3	2.85	2.44	2.56	2.56	3.47	4.16	5.58	5.19
C7→C4	3.33	5.66	4.85	3.20	8.07	7.87	6.97	6.06
C7→C5	2.59	2.44	3.00	2.04	3.47	4.87	4.43	4.70
C7→C6	1.53	2.64	1.49	1.36	3.76	2.43	2.97	2.79
In total	54.97	70.15	61.57	45.94	100	100	100	100

Tables S11. MFPT to C1 from another state calculated with the transition probability matrix for MD200.

	MFPT (ns)			
	All	Subset		
		1	2	3
C2→C1	635.78	675.08	682.41	578.96
C3→C1	823.57	838.25	864.8	779.99
C4→C1	1497.99	1653.83	1812.74	1664.48
C5→C1	1465.4	1583.06	1730.76	1532.74
C6→C1	1576.64	1669.06	1784.09	1661.74
C7→C1	1069.98	1131.95	1217.02	1110.52