Supporting information

A probabilistic network model for structural transitions in biomolecules

Michael Habeck[1,2,*], Thach Nguyen[2]

February 12, 2018

Running times and computational resources

We ran the 94 morphing simulations on a high throughput cluster (HPC)[1] where each node is a Ivy-Bridge Intel E5-2670 v2 2.5GHz processor with 64 GB memory. We simulated 5000 Gibbs sampling iterations for each morphing task. To sample a conformational transition, much fewer Gibbs sampling iterations ($\sim$ 200 iterations) are required (see Fig. 3 in manuscript). The running times, memory usage, initial and final RMSD are summarized in Supplementary Table S1.

Transition	Size (AA)	CPU time (s)	Avg memory (MB)	Initial RMSD (Å)	Final RMSD (Å)
1ysy_A $\rightarrow$ 2ahm_D	71	652.87	54.51	7.73	1.01
2ahm_D $\rightarrow$ 1ysy_A	71	1243.74	69.78	7.73	0.89
1szv_A $\rightarrow$ 1vet_B	91	838.47	71.3	6.31	1.02
1vet_B $\rightarrow$ 1szv_A	91	1784.59	96.32	6.31	0.79
1l5e_B $\rightarrow$ 1l5b_A	101	682.32	60.78	6.52	0.52
1l5b_A $\rightarrow$ 1l5e_B	101	808.26	63.72	6.52	0.55
1wrp_R $\rightarrow$ 3wrp_A	108	706.67	61.33	1.98	0.46
3wrp_A $\rightarrow$ 1wrp_R	108	760.08	62.74	1.98	0.53

[1] https://www.gwdg.de/application-services/high-performance-computing

1xfr_A → 2fjy_A	123	832.1	73.57	5.27	0.71
2fjy_A → 1xfr_A	123	1775.8	96.86	5.27	0.9
1e7x_A → 1dzs_B	129	806.11	68.37	3.4	0.69
1dzs_B → 1e7x_A	129	1064.5	68.9	3.4	0.6
1cfd_A → 1cfc_A	148	2237.67	120.81	5.21	0.33
1cfc_A → 1cfd_A	148	2520.83	120.79	5.21	0.34
1hd2_A → 1oc3_C	158	1086.58	86.92	8.56	0.76
1oc3_C → 1hd2_A	158	1134.79	86.47	8.56	0.73
2gja_B → 1rfl_A	162	2512.23	152.04	8.73	1.03
1rfl_A → 2gja_B	162	1103.06	92.05	8.73	1.39
1r3e_A → 1ze1_D	169	1823.95	154.31	1.53	0.92
1ze1_D → 1r3e_A	169	2062.04	133.78	1.53	0.96
1ybj_A → 1dk0_B	173	1123.31	93.68	5.64	0.99
1dk0_B → 1ybj_A	173	2435.46	125.75	5.64	0.88
1aje_A → 1ees_A	174	2799.43	182.67	6.75	1.24
1ees_A → 1aje_A	174	2545.91	144.22	6.75	1.51
1cbu_B → 1c9k_B	180	1165.33	89.65	3.11	0.62
1c9k_B → 1cbu_B	180	1107.19	91.32	3.11	0.55
1ex6_A → 1ex7_A	186	1232.52	96.29	3.64	0.42
1ex7_A → 1ex6_A	186	1388.97	96.85	3.64	0.43
1s2h_A → 1go4_D	190	1315.2	105.79	4.93	1.02
1go4_D → 1s2h_A	190	3587.34	179.48	4.93	0.93
1bcc_E → 2bcc_E	196	1263.48	98.75	7.45	0.63
2bcc_E → 1bcc_E	196	1227.62	98.75	7.45	0.73
2rh5_A → 2rgx_A	202	1349.35	104.29	5.85	0.55
2rgx_A → 2rh5_A	202	1259.65	104.38	5.85	0.54
4ake_A → 1ake_B	214	1527.12	109.95	7.14	0.56
1ake_B → 4ake_A	214	1349.17	106.12	7.14	0.63
1ggg_A → 1wdn_A	220	1624.13	112.31	5.34	0.65
1wdn_A → 1ggg_A	220	1369.03	109.18	5.34	0.63
2lao_A → 1lst_A	238	1707.31	115.12	4.7	0.39
1lst_A → 2lao_A	238	1448.57	114.31	4.7	0.35
3pjr_A → 1qhh_B	261	1813.44	135.3	8.31	0.64

1qhh_B → 3pjr_A	261	1929.56	128.72	8.31	0.55
1urp_D → 2dri_A	271	2292.07	147.4	4.2	0.36
2dri_A → 1urp_D	271	2096.51	129.76	4.2	0.38
1ram_B → 1lei_A	273	1916.94	133.83	3.07	0.47
1lei_A → 1ram_B	273	2168.85	132.73	3.07	0.46
5at1_C → 8atc_C	310	2166.19	146.9	2.36	0.68
8atc_C → 5at1_C	310	2485.8	145.75	2.36	0.68
1ckm_A → 1ckm_B	317	2227.09	159.69	3.49	0.52
1ckm_B → 1ckm_A	317	2115.98	151.3	3.49	0.53
3dap_B → 1dap_A	320	2228.18	169.47	4.28	0.41
1dap_A → 3dap_B	320	2019.66	146.78	4.28	0.38
1eyk_A → 1nuz_A	327	2283.2	188.2	4.54	0.76
1nuz_A → 1eyk_A	327	2203.42	150.46	4.54	0.82
1bp5_B → 1a8e_A	329	2299.09	187.58	6.78	0.45
1a8e_A → 1bp5_B	329	2289.36	153.91	6.78	0.47
1jqj_A → 2pol_A	366	2495.64	204.34	2.05	0.74
2pol_A → 1jqj_A	366	2404.76	169.51	2.05	0.78
1omp_A → 1anf_A	370	2606.05	203.76	3.77	0.48
1anf_A → 1omp_A	370	2503.97	170.32	3.77	0.4
8adh_A → 6adh_B	374	2514.05	215.74	1.35	0.7
6adh_B → 8adh_A	374	2545.7	219.51	1.35	0.63
9aat_A → 1ama_A	401	2735.77	237.72	1.66	0.48
1ama_A → 9aat_A	401	2814.81	242.44	1.66	0.51
1ux5_A → 1y64_B	411	2714.49	252.58	10.33	0.73
1y64_B → 1ux5_A	411	2779.33	257.7	10.33	1.51
1qf5_A → 1hoo_B	431	3964.11	287.62	2.17	0.68
1hoo_B → 1qf5_A	431	2952.76	256.8	2.17	0.74
1yyo → 1yyw	438	1733.47	114.92	17.46	0.59
1yyw → 1yyo	438	1706.23	119.44	17.46	0.56
1bnc_A → 1dv2_B	452	3130.23	207.14	3.92	0.48
1dv2_B → 1bnc_A	452	2976.04	274.85	3.92	0.56
1rkm_A → 2rkm_A	517	3777.25	298.17	3.08	0.42
2rkm_A → 1rkm_A	517	3598.92	332.81	3.08	0.38

1sx4_G → 1oel_F	524	3478.49	247.78	12.39	0.78
1oel_F → 1sx4_G	524	3296.45	323	12.39	0.83
1hp1_A → 1hpu_C	525	3710.32	279.07	10.01	0.41
1hpu_C → 1hp1_A	525	3356.42	246.47	10.01	0.49
2hmi_A → 3hvt_A	556	3664.31	282.81	3.45	1.43
3hvt_A → 2hmi_A	556	3672.59	266.81	3.45	1.32
1i7d_A → 1d6m_A	620	4234.92	299.65	3.4	0.61
1d6m_A → 1i7d_A	620	4207.54	421.49	3.4	0.58
8ohm_A → 1cu1_B	645	2964.66	195.97	4.49	0.6
1cu1_B → 8ohm_A	645	2951.66	227.06	4.49	0.6
1lfg_A → 1lfh_A	691	4814.59	399.16	6.43	0.56
1lfh_A → 1lfg_A	691	4711.94	488.74	6.43	0.58
1qvi_A → 1kk8_A	837	5185.55	647.53	27.4	1.14
1kk8_A → 1qvi_A	837	5317.37	594.08	27.4	1.89
1q9x_B → 1q9y_A	899	6244.06	764.38	5.43	0.48
1q9y_A → 1q9x_B	899	6323.5	664.86	5.43	0.47
1ih7_A → 1ig9_A	903	6169.42	808.31	6.49	0.57
1ig9_A → 1ih7_A	903	6303.24	731.22	6.49	0.56
1su4_A → 1iwo_A	994	6445.33	768.81	13.97	1.15
1iwo_A → 1su4_A	994	6445.56	928.44	13.97	1.21

Table S1: running times, memory usage and final RMSD for all 94 morphing simulations.

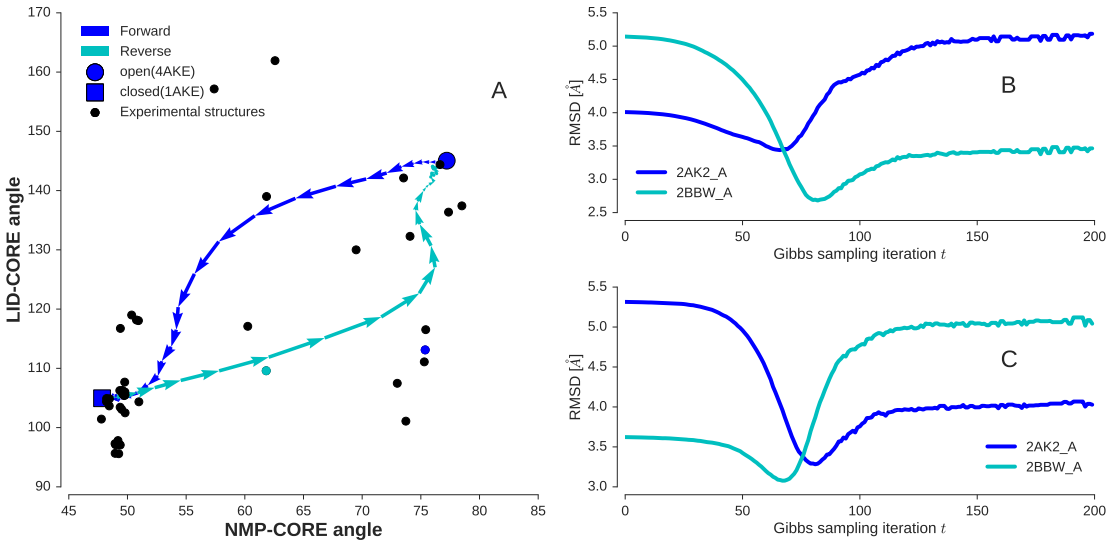

Figure S1: Forward and reverse transition in adenylate kinase. **(A)** The forward pathway connecting the open state (blue circle, PDB code 4ake) with the closed state (blue square, PDB code 1ake) is shown in blue, the reverse transition is indicated by cyan arrows. The pathways were projected onto the LID-CORE and NMP-CORE angle. Black circles indicate experimental structures. The blue and cyan circles mark intermediate structures (PDB entries 2ak2 and 2bbw). Panels **(B)** and **(C)** show the evolution of the global RMSD during the forward and reverse transition between our generated structures and the intermediate structures from PDB entries 2ak2 and 2bbw.

Detailed analysis of conformational transitions

Because structural transitions follow paths in a very high dimensional space, it is a non-trivial task to compare the transition paths generated by our Gibbs sampling algorithm with other pathways reported in the literature. We projected the transitions onto various reaction coordinates such intra-domain angles as well as principal components. We studied four examples in more detail.

Adenylate kinase

Adenylate kinase (AdK) is a phosphotransferase that catalyzes the reaction converting ATP and AMP into 2 ADP molecules. AdK is composed of three domains: NMP binding domain (residues 30-60), LID (residues 115-160) and the CORE domain (residues 1-

30, 61-114, and 161-214). Figure S1 presents the conformational change between the open (PDB code 4ake, chain A) and the closed state (PDB code 1ake, chain B). This large-scale structural transition can be captured by two intra-domain angles θ_{NMP} and θ_{LID} (Beckstein *et al.*, 2009). The NMP-CORE angle θ_{NMP} is the angle between the centers of mass of two segments L115-V125 and L35-A55 relative to I90-G100 based on Cα positions. The LID-CORE angle θ_{LID} is the angle between the centers of mass based on Cα positions of segments I179-E185 and V125-L153 relative to L115-V125. The forward and reverse transitions generated by our Gibbs sampler follow different pathways. Visual comparison with the analysis by Seyler *et al.* (2015) reveals that our transition path is close to the path generated by GOdMD (Sfriso *et al.*, 2013) in that the LID-CORE angle changes first and is followed by a transition in the NMP-CORE angle. Our reverse transition is close to the pathway generated with ANMPathway (Das *et al.*, 2014). Among 45 experimental structures of AdK deposited in the PDB, we identified several structures that are close to our transition paths in angular space. The closest intermediate structures based on global RMSD are PDB entries 2bbw (chain A) and 2ak2 (chain A).

GroEL

To illustrate the transition path between the T state (PDB code 1oel, chain F) and R" state (PDB code 1sx4, chain G) in GroEL, we used a reaction coordinate (RC) similar to the one defined by Zheng and Wen (2017). We defined the reacion coordinate RC_{S} to measure the movement of some domain S as follows:

$$RC_{\mathrm{S}} = (\delta X_{\mathrm{S}} \delta X_{\mathrm{S,obs}}) / |\delta X_{\mathrm{S,obs}}|^2 \tag{1}$$

where δX_{S} is the displacement vector from the center of mass of domain S in the initial structure to the center of mass of domain S in the intermediate structure. Figure S2 shows the movement of the Apical (A) and Intermediate (I) domain relative to the Equatorial (E) domain measured by RC_{AE} and RC_{IE}.

5'-nucleotidase

Escherichia coli 5'-nucleotidase (5'-NTase) is an enzyme composed of an N-terminal domain (residues 26-351) and a C-terminal domain (residues 365-550) that move relative to each other. To elucidate the transition pathway of 5'-NTase, we used two angles

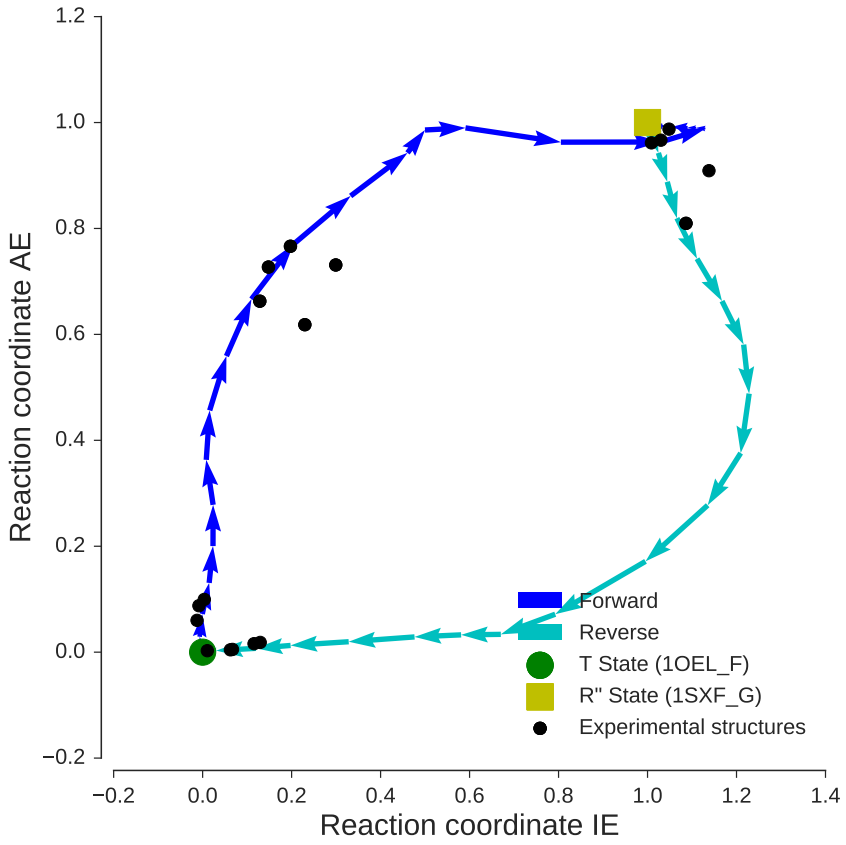

Figure S2: Analysis of the transition path of the T state (green circle) and R" state (yellow square) in GroEL. To visualize the transition, we used two reaction coordinates, RC_{IE} and RC_{AE}, where the A, I, and E domain were defined as in Xu *et al.* (1997) and computed by using equation 1. The black dots indicate experimental structures from the following PDB entries 4aaq (chains B, C), 4ab2 (chains A, B), 4aar (chains B, C), 4aau (chains B, C, J, K), 4pko (chains A, D, L, M), 3wvl (chains G, J), 1pf9 (chains A, D), 1sx4 (chain J), 2eu1 (chain F), 2c7e (chains L, M), 1xck (chain B), 1mnf (chain I).

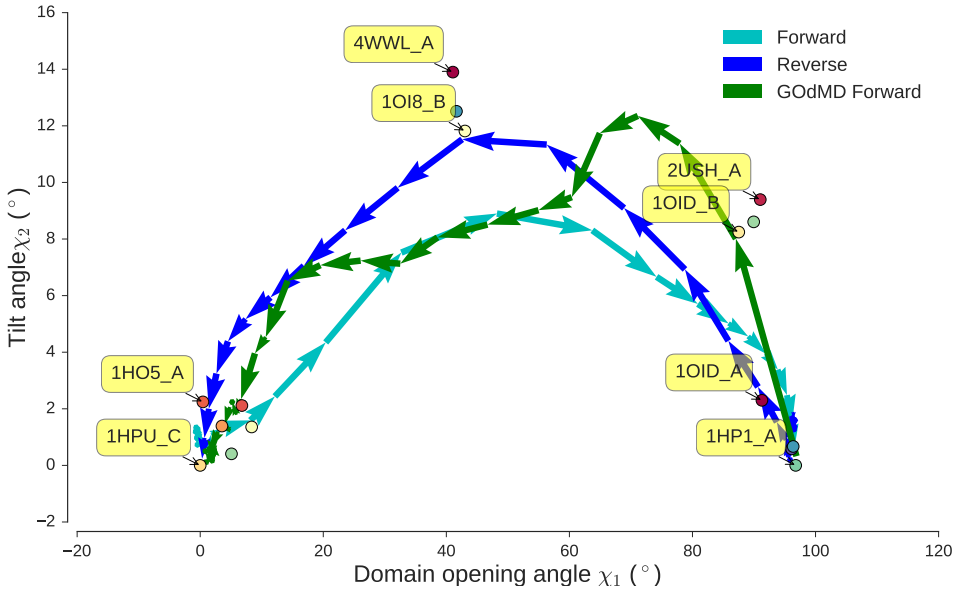

Figure S3: Transition paths in 5'-nucleotidase mapped onto two angles χ_1 and χ_2. Experimental structures (colored dots) are taken from PDB entries 1hpu (chains A, B, C, D), 1ho5 (chains A, B), 4wwl, 1oi8 (chains A, B), 2ush (chains A, B), 1oid (chains A, B), 1oie (chain A), 1ush (chain A), 1hp1 (chain A).

χ_1 (domain opening angle) and χ_2 (tilt angle) defined by Knöfel and Sträter (2001) and Krug *et al.* (2016). Figure S3 shows the evolution of χ_1 and χ_2 during the transition paths generated with our Gibbs sampler and highlights that the generated paths find an experimentally characterized intermediate state (PDB code 1oi8). For comparison, we also show the transition path for the forward direction generated with GOdMD by Sfriso *et al.* (2013), which steps through a similar sequence of collective variables (the GOdMD simulation for the reverse direction failed, producing a final structure with an RMSD of $\sim$7 Å to the target).

Ribonuclease III

Ribonuclease III (RNase III) is a ribonuclease that plays an important role in RNA processing. RNase III recognizes and cleaves *ds*RNA at several target locations to create mature RNAs (Gan *et al.*, 2005). We investigated the transition path of RNase III and compared our results with pathways generated by Orellana *et al.* (2016). We used principal component analysis (PCA) to project the transition path onto the first two principal components (PCs).
Figure S4 shows the forward and reverse transition starting from non-catalytic complex (PDB code 1yyo) and targeting the pre-catalytic complex (PDB code 1yyw). The forward transition comes close to the inactive dsRNA-bound state (PDB entries 1yyk, 2nue). The reverse transition comes close to the Mg2+ bound catalytic state (PDB entries 4m30, 4mz2).

Supplementary movies

Our supplementary movies show forward and reverse transitions in cartoon representation generated by Pymol (DeLano, 2002).

- Movies Adk_forward.avi and Adk_reverse.avi show the forward and reverse transition in Adenylate kinase wher NMP, LID and CORE domain are colored in red, green and blue. Structures were superimposed onto the CORE domain.
- Movies Groel_forward.avi and Groel_reverse.avi show the forward and reverse transition in GroEL where the Apical, Intermediate and Equatorial domain are col-

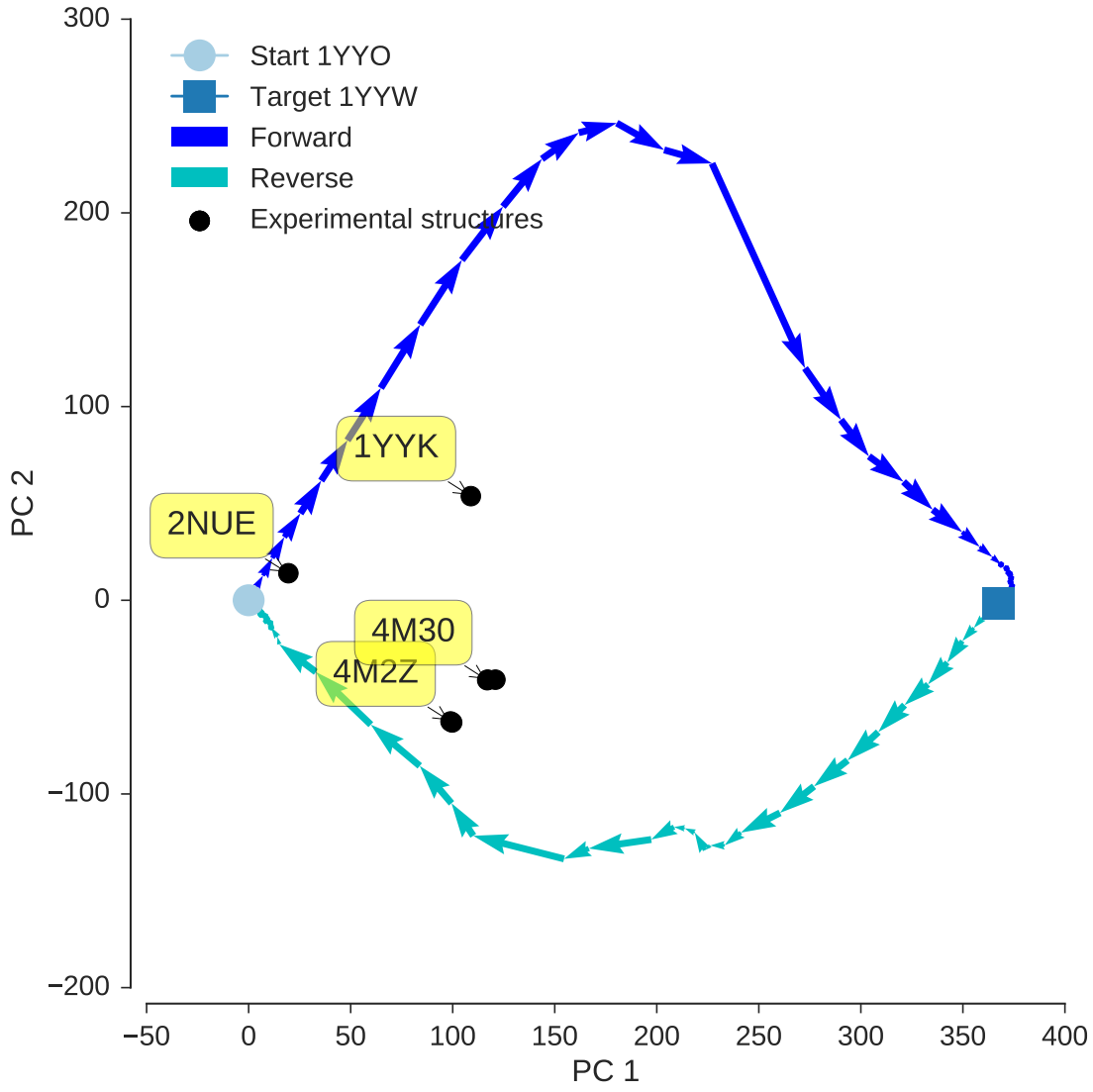

Figure S4: Conformational transition in RNase III starting from the non-catalytic complex (PDB code 1yyo) and targeting the pre-catalytic complex (PDB code 1yyw).

ored in red, green and blue. Structures were superimposed onto the Equatorial domain.

References

Beckstein, O., Denning, E. J., Perilla, J. R., and Woolf, T. B. (2009). *J. Mol. Biol.*, **394**(1), 160–176.

Das, A., Gur, M., Cheng, M. H., Jo, S., Bahar, I., and Roux, B. (2014). *PLoS Comput Biol*, **10**(4), e1003521.

DeLano, W. L. (2002).

Gan, J., Tropea, J. E., Austin, B. P., Court, D. L., Waugh, D. S., and Ji, X. (2005). *Structure*, **13**(10), 1435–1442.

Knöfel, T. and Sträter, N. (2001). *Journal of molecular biology*, **309**(1), 255–266.

Krug, U., Alexander, N. S., Stein, R. A., Keim, A., Mchaourab, H. S., Strater, N., and Meiler, J. (2016). *Structure*, **24**(1), 43–56.

Orellana, L., Yoluk, O., Carrillo, O., Orozco, M., and Lindahl, E. (2016). *Nature communications*, **7**.

Seyler, S. L., Kumar, A., Thorpe, M. F., and Beckstein, O. (2015). *PLoS Comput Biol*, **11**(10), e1004568.

Sfriso, P., Emperador, A., Orozco, M., *et al.* (2013). *Bioinformatics*, **29**(16), 1980–1986.

Xu, Z., Horwich, A. L., and Sigler, P. B. (1997). *Nature*, **388**(6644), 741–750.

Zheng, W. and Wen, H. (2017). *Current Opinion in Structural Biology*, **42**, 24–30.