PNAS
www.pnas.org

Supplementary Information for

Near-Experimental Accuracy in Protein Structure Refinement of Homology Models via Molecular Dynamics Simulations

Lim Heo and Michael Feig*

*Corresponding author
E-mail: mfeiglab@gmail.com

This PDF file includes:

Detailed methodology
Figs. S1 to S38
Tables S1 to S7
Description of Movies S1 to S4
Additional references

Detailed methodology

Molecular dynamics simulations

Protein structures were solvated in a cubic box with at least 9 Å distance to the edge of the box. The CHARMM c36m force field (1) was used to describe the proteins. Explicit water was modeled via the CHARMM version of the TIP3 model (2). Sodium or chloride ions were added to neutralize the system. Lennard-Jones interactions and direct electrostatic interactions were truncated at 10 Å with switching applied from 8 to 10 Å. Particle-mesh Ewald summation was used to calculate the full electrostatic energy term. Each system was equilibrated via energy minimization and subsequent heating to 298 K. Simulations were then carried out as in previous refinement simulations in the NVT ensemble at 298 K using a Langevin thermostat with a friction coefficient of 0.01/ps. A 2 fs integration time step was used together with holonomic constraints on bonds involving hydrogens. Structures were saved from the generated trajectories at 50-ps intervals after excluding the first 5 ns following equilibration. The CHARMM program (3), version c42a1, was used for the initial equilibration and production runs used CHARMM in combination with openMM (4), version 7.1.1, to take advantage of GPU hardware.

Markov-state modeling and identification of transition pathways

MSM analysis was applied using MSMBuilder (5), version 3.8.0, to identify refinement pathways based on the final set of trajectories for each system. Distance matrices for Cα-Cα atom pairs were evaluated and utilized via tICA analysis to reduce the dimensionality of the system. Conformations were grouped by using k-medoids clustering analysis into 30–200 microstates depending on total simulation time, protein size and the initial model quality. Microstates were lumped into macrostates according to lag times that were determined based on implied time scales for both micro- and macro-state MSMs. Once a MSM was built, refinement pathways from the initial template-based model state to the native state were identified by transition path theory (TPT) analysis. One or more representative paths with the highest fluxes were selected for further analysis. Detailed conformational changes were tracked along subsampled trajectories for the refinement pathways. A refinement trajectory was generated by subsampling each state and transition between states alternatively until it reached the native state from the initial state. Structure changes were characterized in terms of backbone dihedral angles, side-chain χ-angles, inter-atomic distances, and Cα-RMSDs and differences in rigid body orientations using mdtraj (6) and custom-written python scripts.

Conformational reweighting and scoring

Conformational snapshots were reweighted and scored by calculating potential energies with the CHARMM program (3) or the RWplus (7) and dDFIRE (8) programs in combination with the MMTSB Tool Set. Explicit solvent snapshots were reweighted simply by recalculating energies with different force field input files. MMGB/SA scores were calculated by adding solute potential energies with different force fields to implicit solvent free energies with the Generalized Born Molecular Volume (GBMV) method (9) including a solvent-accessible surface area (SASA) term with γ=5.4 cal/mol/Å^2.

Selection of alternative models

Alternative models were obtained from the server predictions generated during CASP11 and CASP12 and available from the CASP web site (http://predictioncenter.org/). Most of the submitted server models were considered except for models that deviated by more than twice the RMSD from the best models. In practice, these inferior models would be easily identified via scoring and/or quality assessment methods. The resulting models were then clustered based on RMSD with a radius of less than 2 Å to group substantially similar models together. This resulted in 17-65 clusters for each target except for TR837 where we identified only 3 clusters. Representative models were then extracted for each cluster for further analysis. Three models for each target were further chosen to conduct additional MD simulations (see Table S7). For each initial model, five independent trajectories of 100 ns-long MD simulations were performed in the same manner described above.

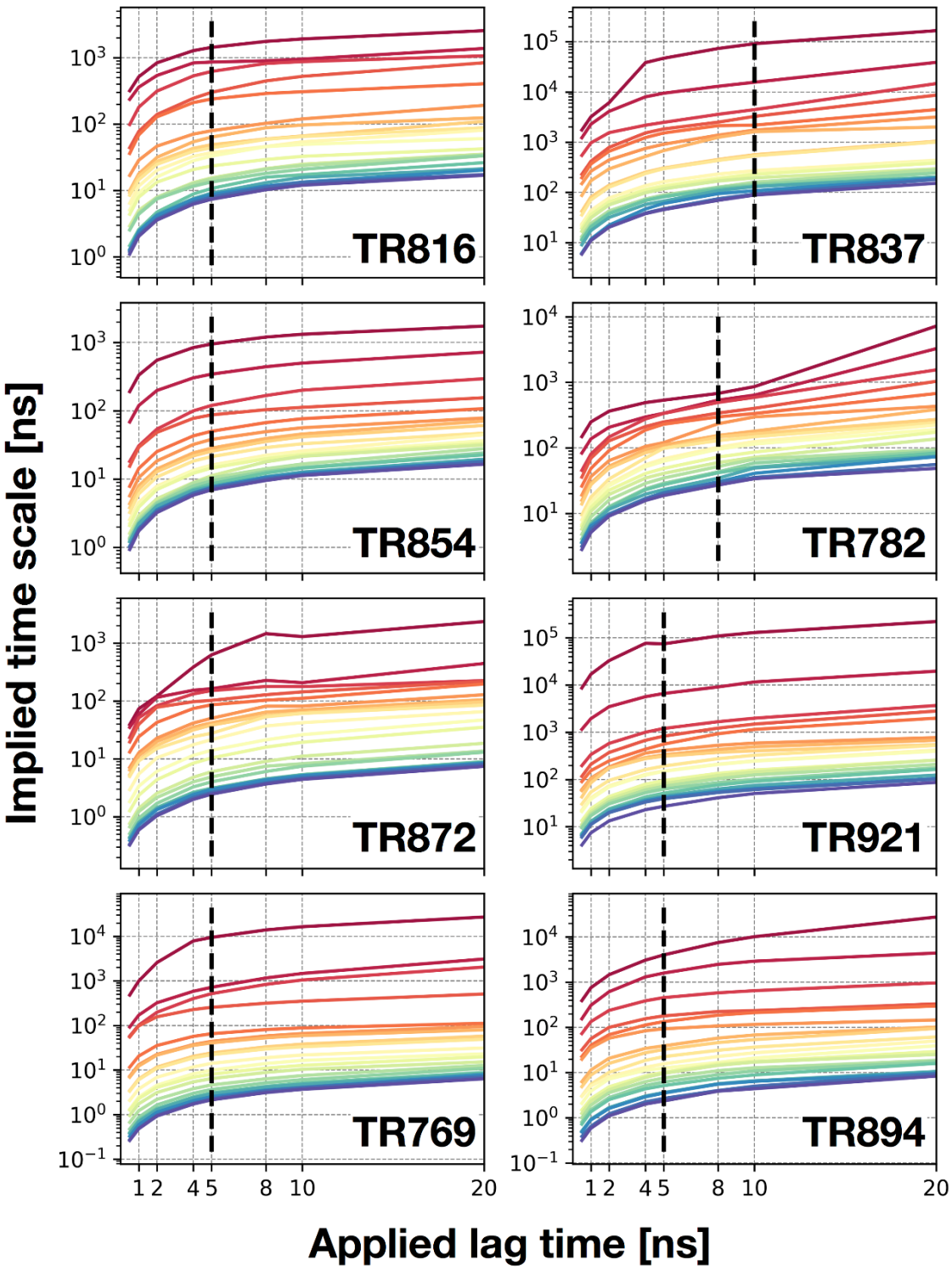

Figure S1. Implied time scales vs. lag times in Markov state model generation. Different colors indicate curves for different states. The dashed line indicates the chosen lag times in the final Markov state models.

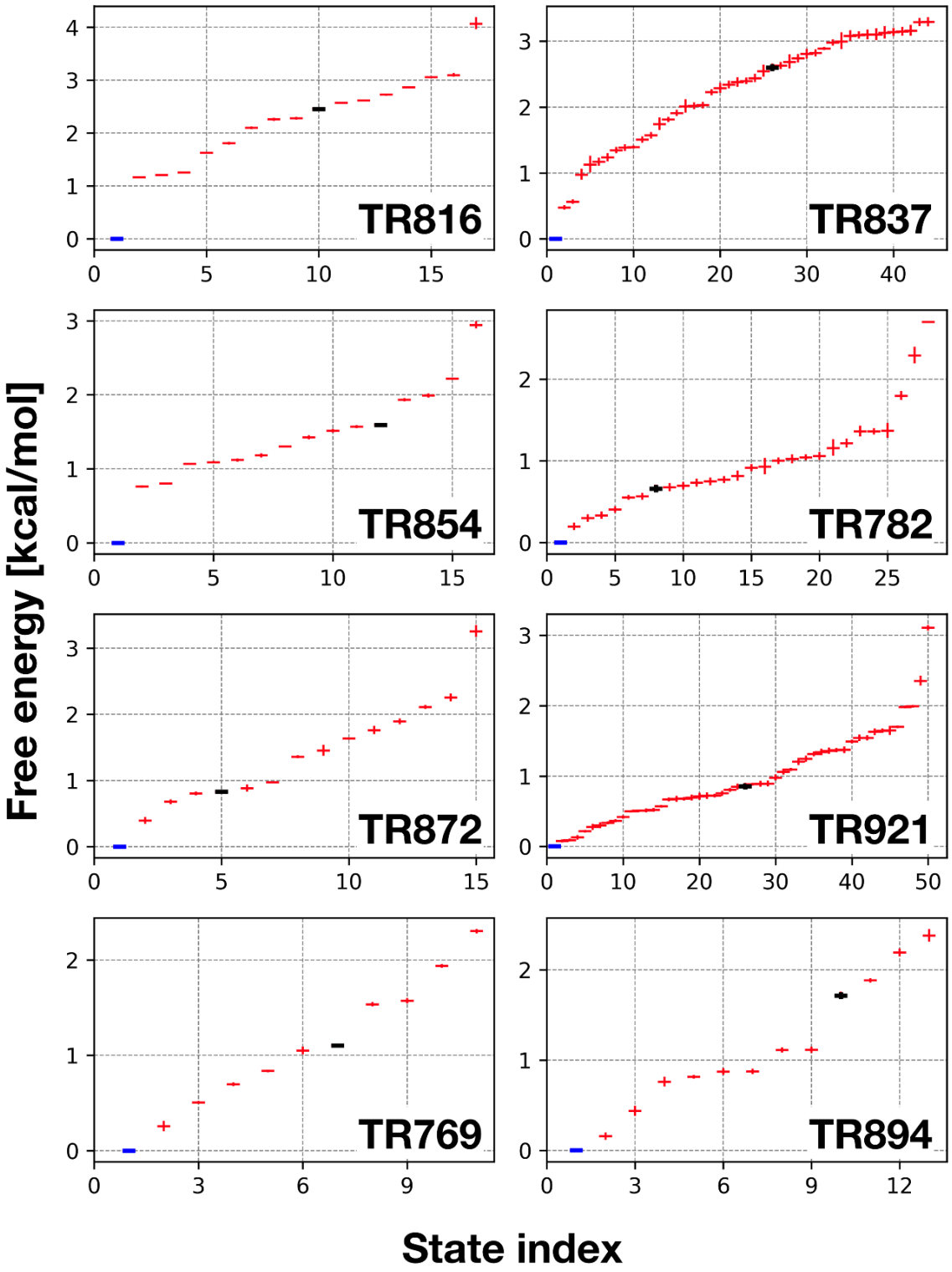

Figure S2. Free energies for different Markov model states. The states are ordered by free energy and vertical bars indicate uncertainties. The state closest to the experimental structure (the native state) is indicated in blue and the state closest to the initial homology model is shown in black.

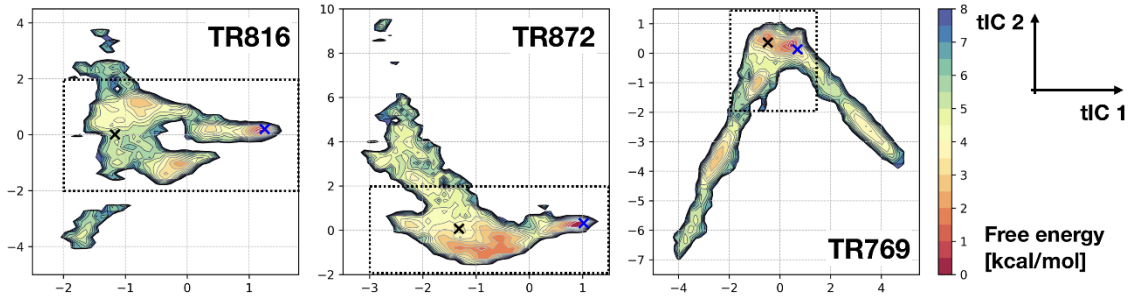

Figure S3. Complete free energy landscapes for TR816, TR872, and TR769.
Potentials of mean force projected onto the first two tICA principal coordinates according to the color bar. Contour lines are drawn for every 0.5 kcal/mol up to 8.0 kcal/mol. Projections of the experimental structures and initial homology models are indicated with blue and black Xs, respectively.

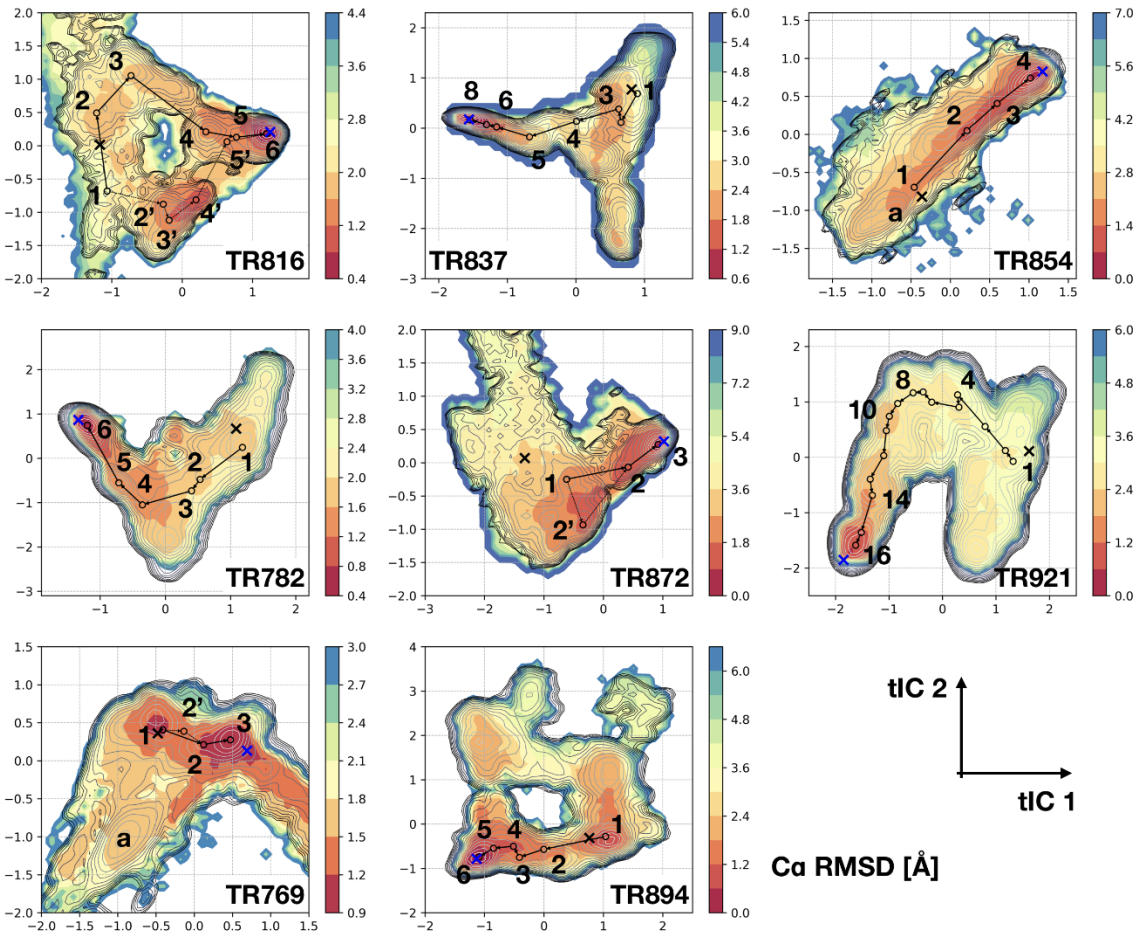

Figure S4. Deviations from experimental structures. Cα RMSD deviations with respect to the experimental reference structures are mapped onto the energy landscapes as a function of the first two tICA coordinates. Minimum RMSD values were chosen where multiple structures map onto the same tICA coordinates. Contour lines correspond to free energies as in Fig. 2 but colors indicate RMSD deviations. MSM states and refinement transition paths are shown as in Fig. 2.

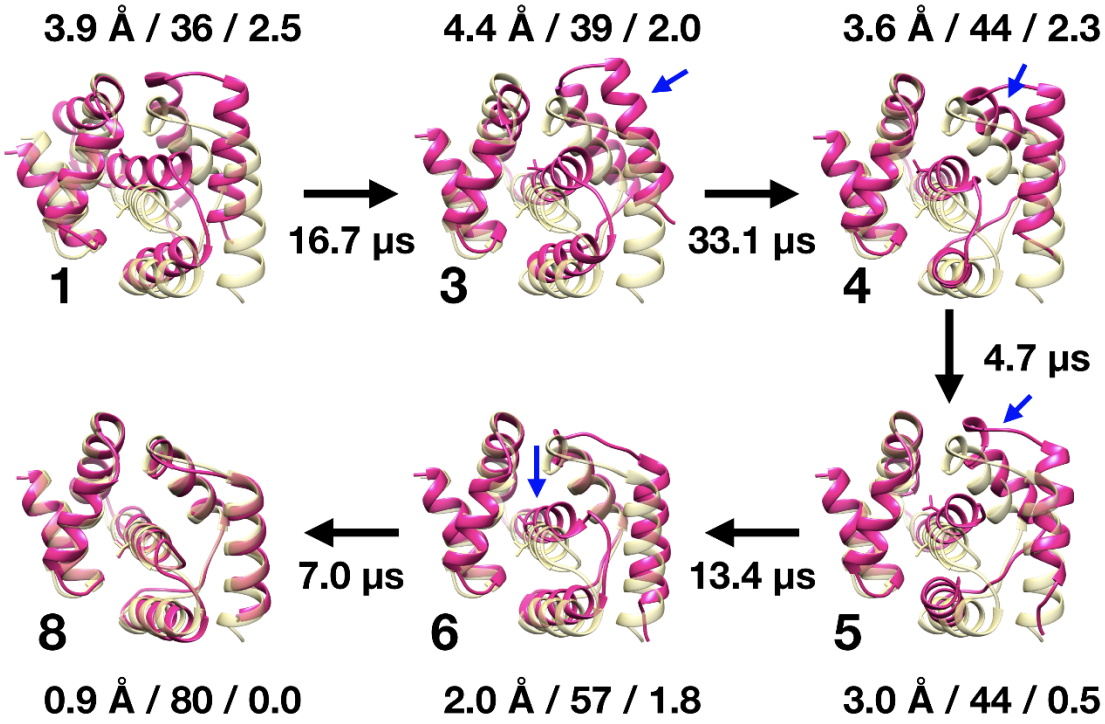

Figure S5. Conformational transitions in TR837. Ensemble-averaged structures for MSM states involved in the refinement transition are depicted in magenta cartoon representations and compared with the experimental structures in yellow. The numbering of states corresponds to the states identified in the free energy surfaces in Fig. 2. Cα-RMSD values, GDT-HA scores, and free energies are given for each state either above or below each structure. Mean first passage times (MFPT) for transitions between states from the MSM are given for the transitions towards the native state. Blue arrows indicate key structural changes after each transition.

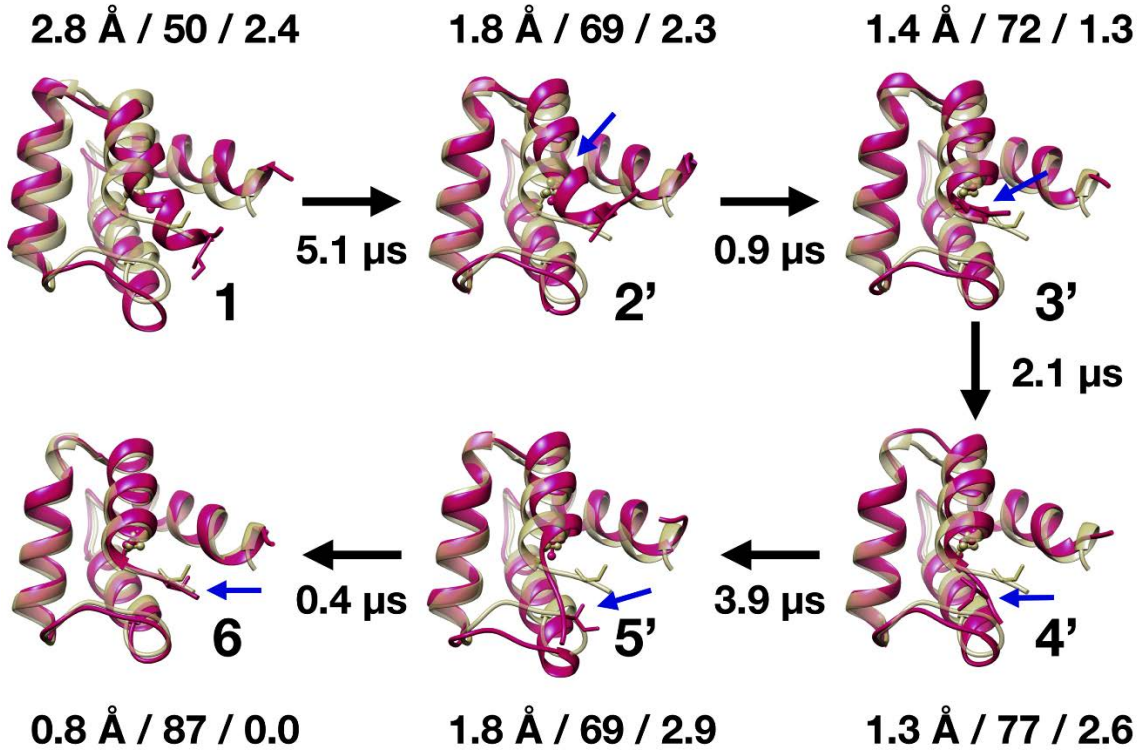

Figure S6. Conformational transitions in TR816. Alternative transition path between ensemble-averaged structures for MSM states involved in the refinement transition path as in Fig. S5.

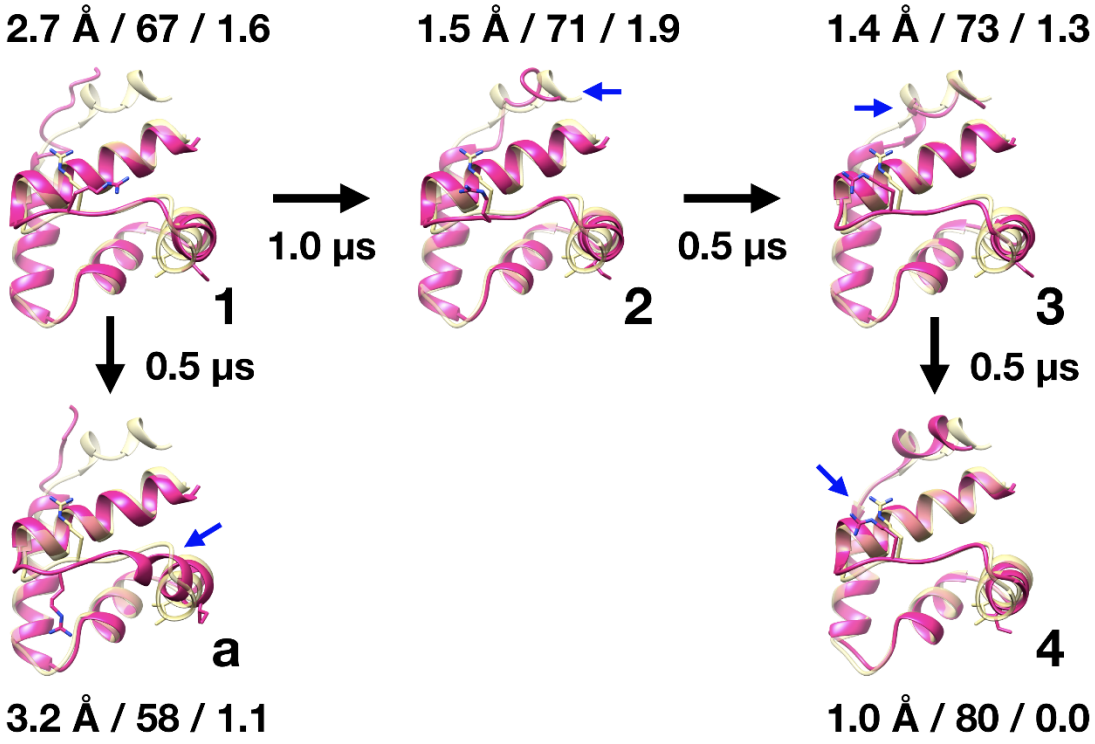

Figure S7. Conformational transitions in TR854. Transitions between ensemble-averaged structures for MSM states involved in the refinement transition path as well as to the off-pathway state ‘a’ as in Fig. S5.

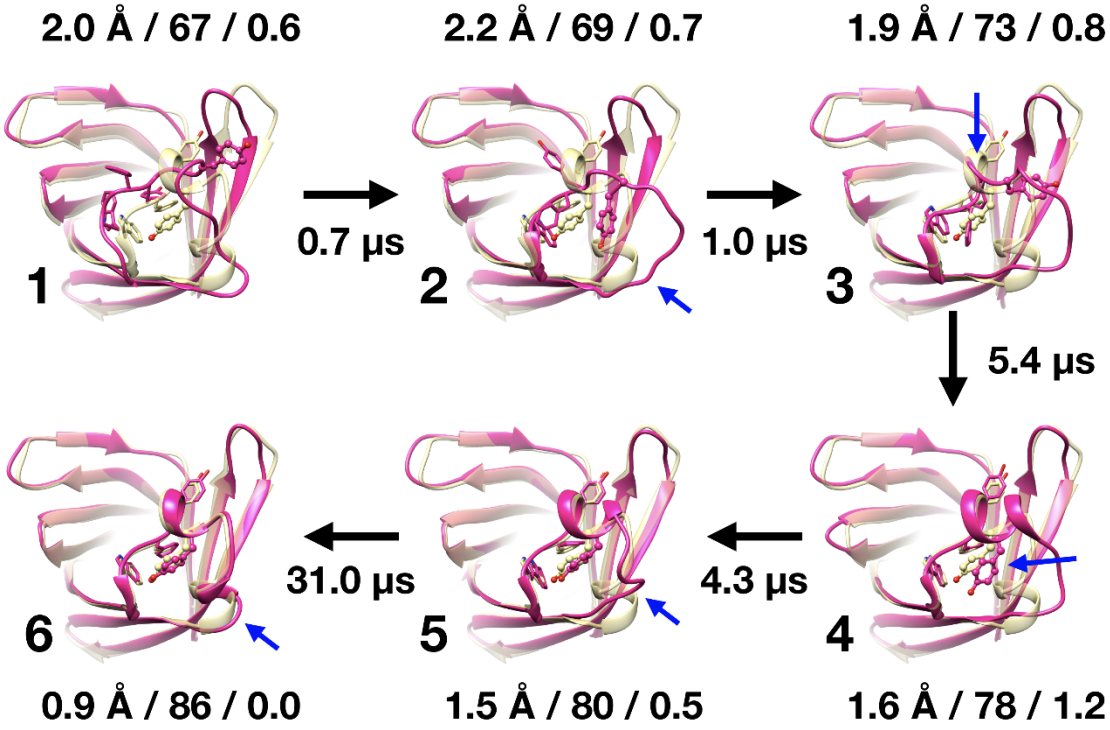

Figure S8. Conformational transitions in TR782. Transitions between ensemble-averaged structures for MSM states involved in the refinement transition path as in Fig. S5.

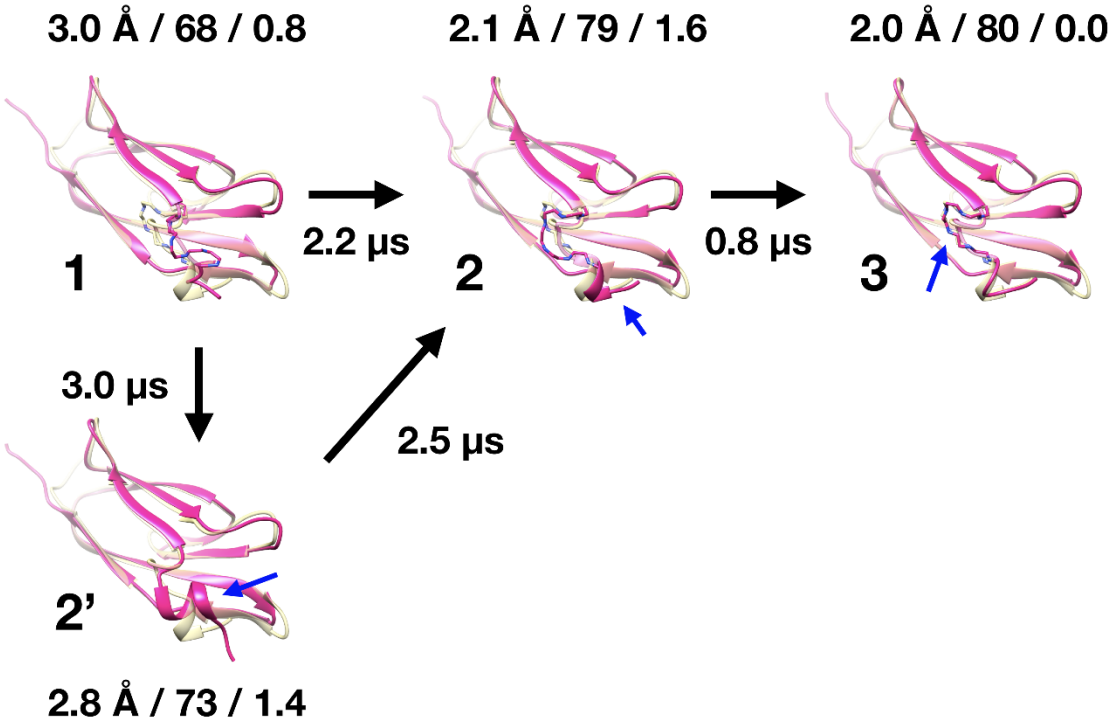

Figure S9. Conformational transitions in TR872. Transitions between ensemble-averaged structures for MSM states involved in refinement transition paths as in Fig. S5.

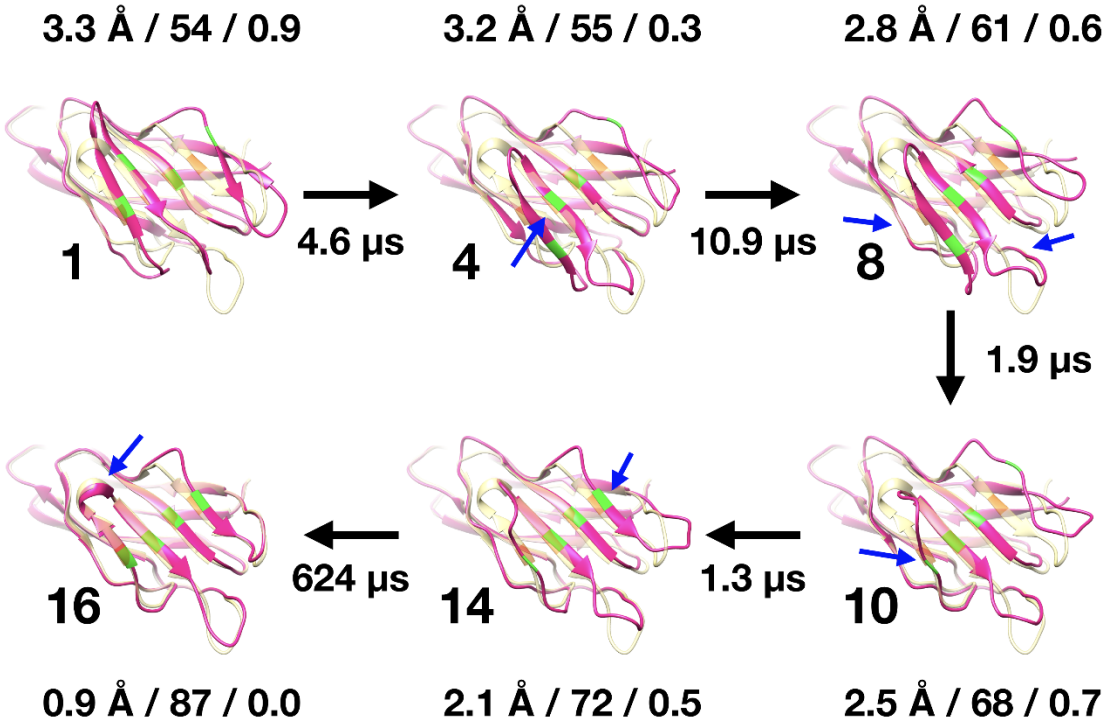

Figure S10. Conformational transitions in TR921. Transitions between ensemble-averaged structures for MSM states involved in refinement transition paths as in Fig. S5. Additional minor transitions are not shown. Four aligned residues (M66, N75, D40, and E119) are highlighted in orange and green for the native and ensemble averaged structures, respectively, to indicate changes in β-sheet residue pairing,

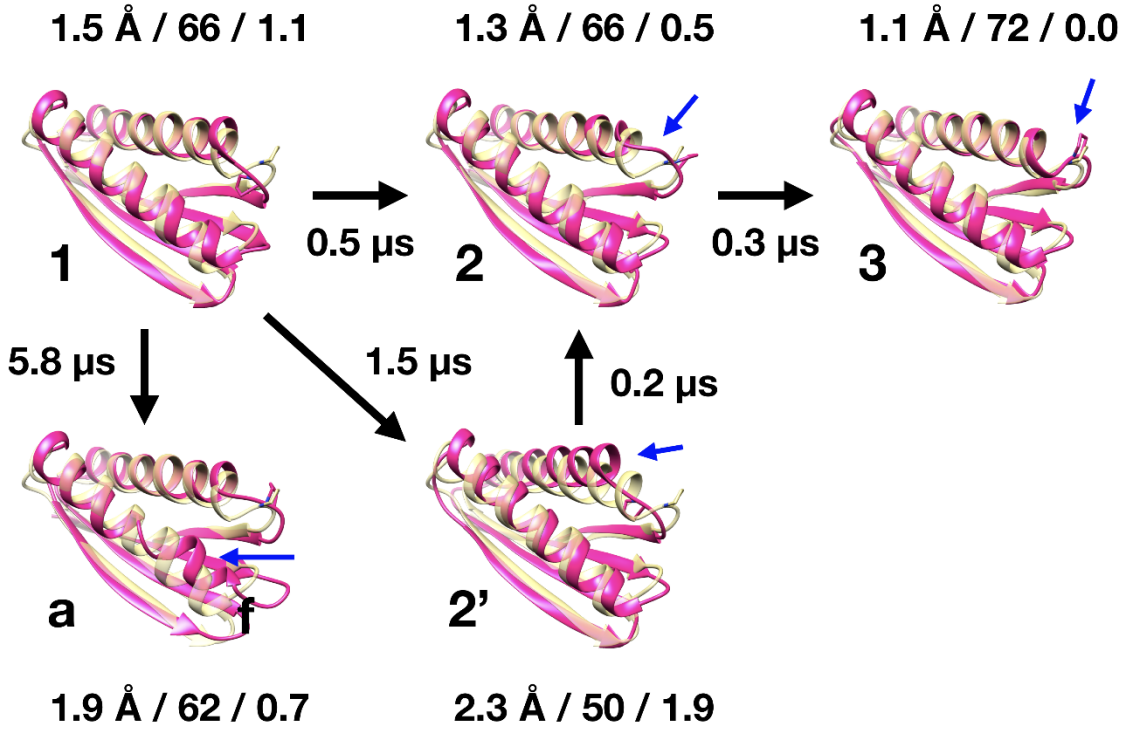

Figure S11. Conformational transitions in TR769. Transitions between ensemble-averaged structures for MSM states involved in the refinement transition paths as well as to the off-pathway state 'a' as in Fig. S5.

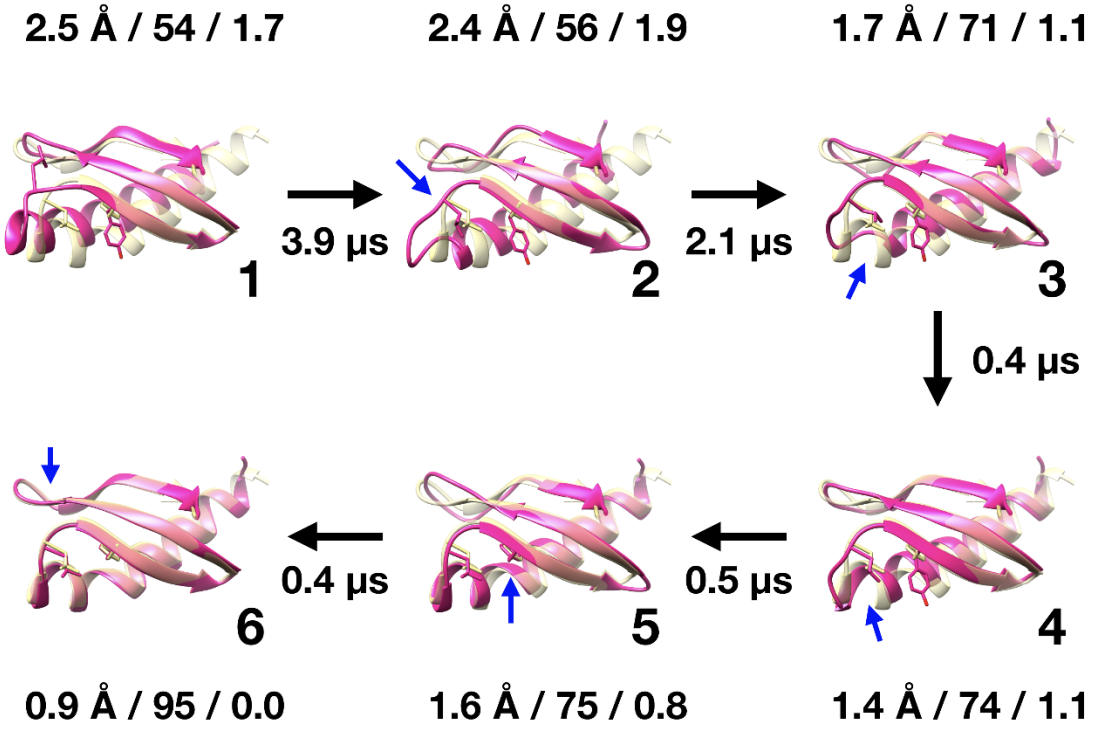

Figure S12. Conformational transitions in TR894. Transitions between ensemble-averaged structures for MSM states involved in the refinement transition path as in Fig. S5.

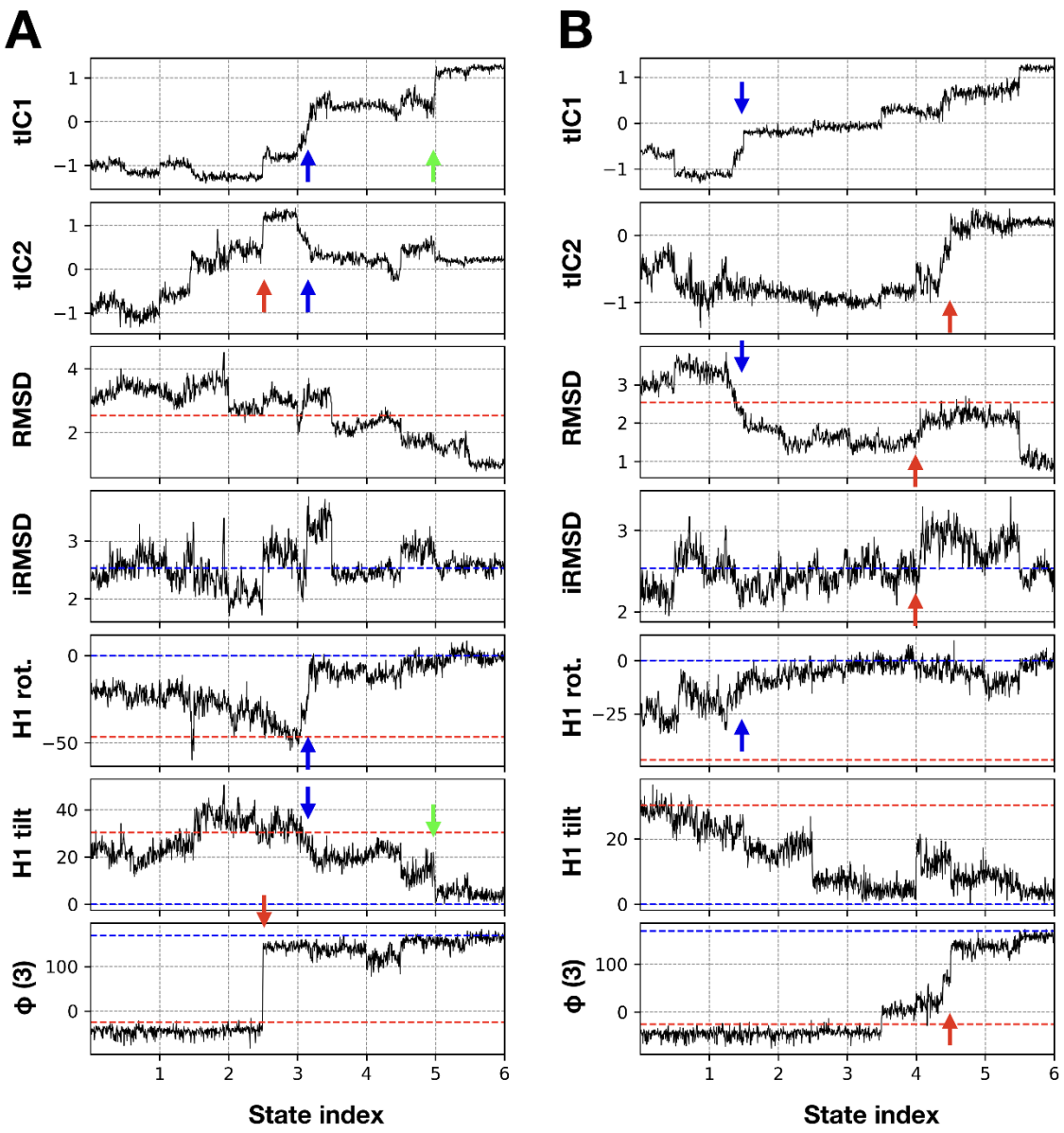

Figure S13. Structural transitions during refinement of TR816. Progress along the principal major tICA coordinates, in terms of Cα-RMSDs (in Å) with respect to the experimental structure (RMSD) and the initial model (iRMSD), the rotational and tilt angles between the helical axis in the model and the experimental structures for the N-terminal helix (H1, residues 4-15), and the backbone φ torsion angle for residue 3 as a function of refinement progress from subsampled trajectories. Key transitions are indicated by arrows. Arrows with the same color refer to a correlated event. Blue and red dashed lines indicate the values in the experimental structure and initial models, respectively. Progress along two different refinement paths (A and B) is shown and state indices correspond to Figs. 2 and 3.

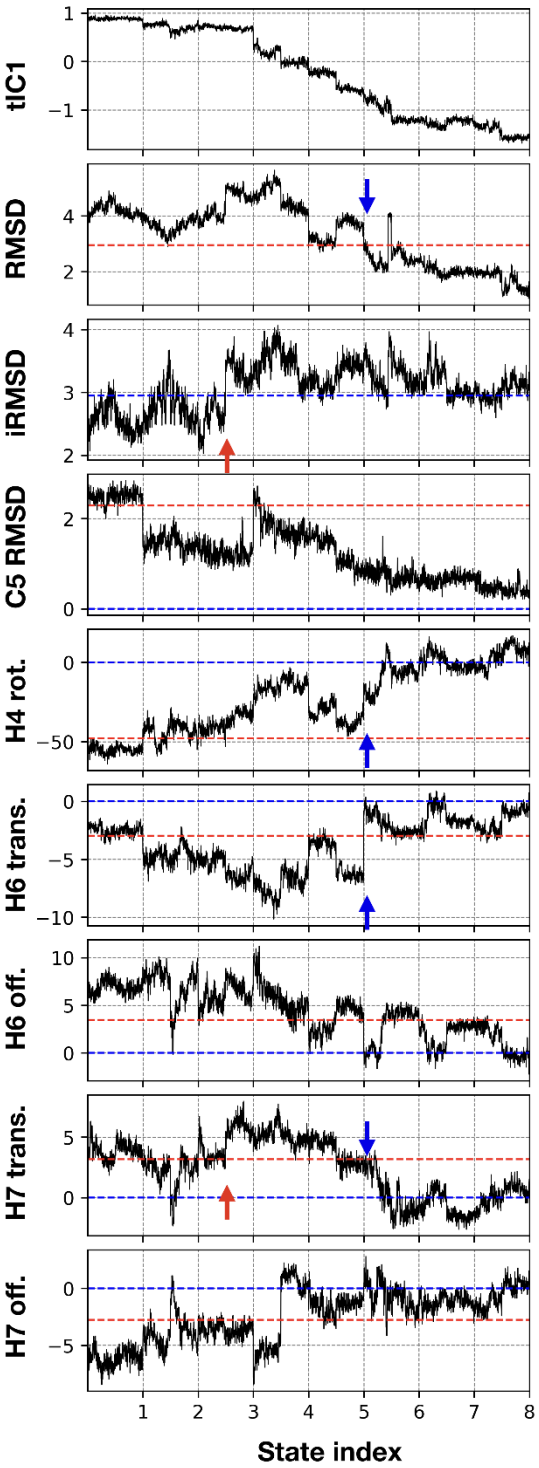

Figure S14. Structural transitions during refinement of TR837. As Fig. S13 with Cα RMSD (in Å) of loop C5 (residues 80-87), helical rotation for helix H4 (residues 45-59), and rigid body translation along the helical axis (***trans.***) and perpendicular away from the helix axis (***off.***) for helices H6 (88-96) and H7 (105-121)

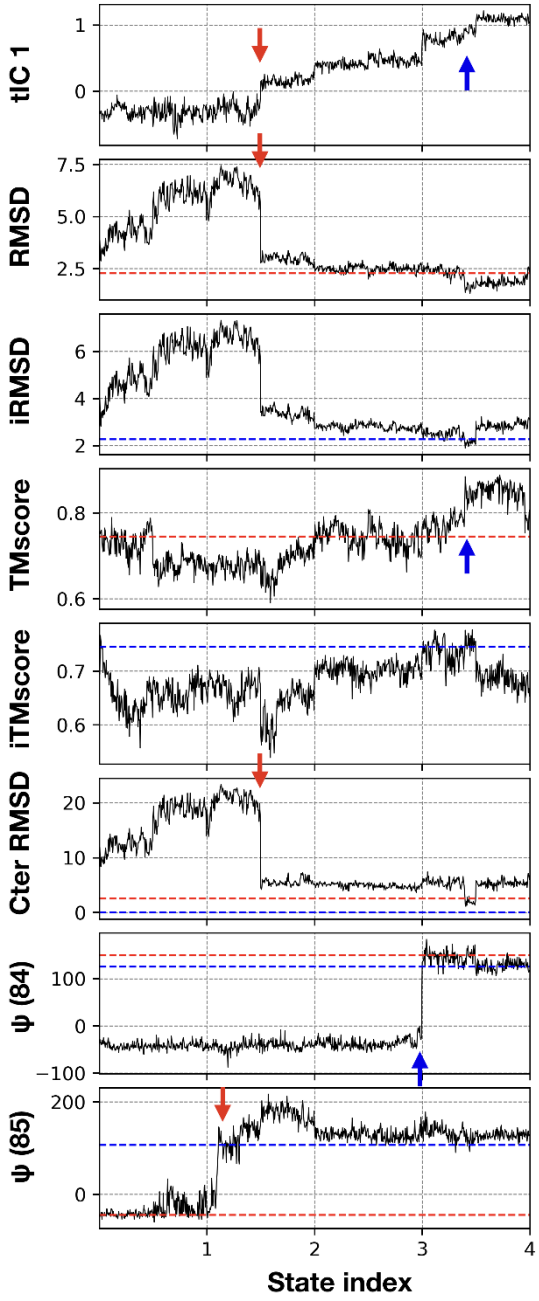

Figure S15. Structural transitions during refinement of TR854. As Fig. S13 with TM-score (10) from the experimental structure (TMscore) and initial model (iTMscore), Cα RMSD (in Å) of the C-terminus (residues 84-93) and backbone ψ torsion angles for residues 84 and 85.

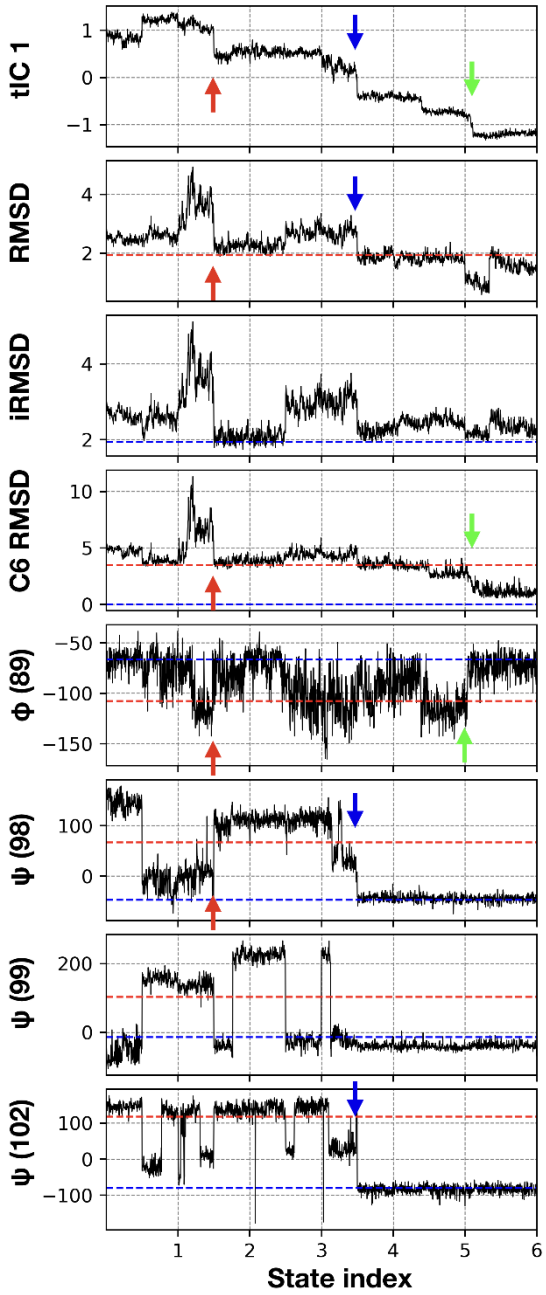

Figure S16. Structural transitions during refinement of TR782. As Fig. S13 with Cα RMSD (in Å) of loop C6 (residues 88-109) and backbone ψ torsion angles for residues 89, 98, 99, and 102.

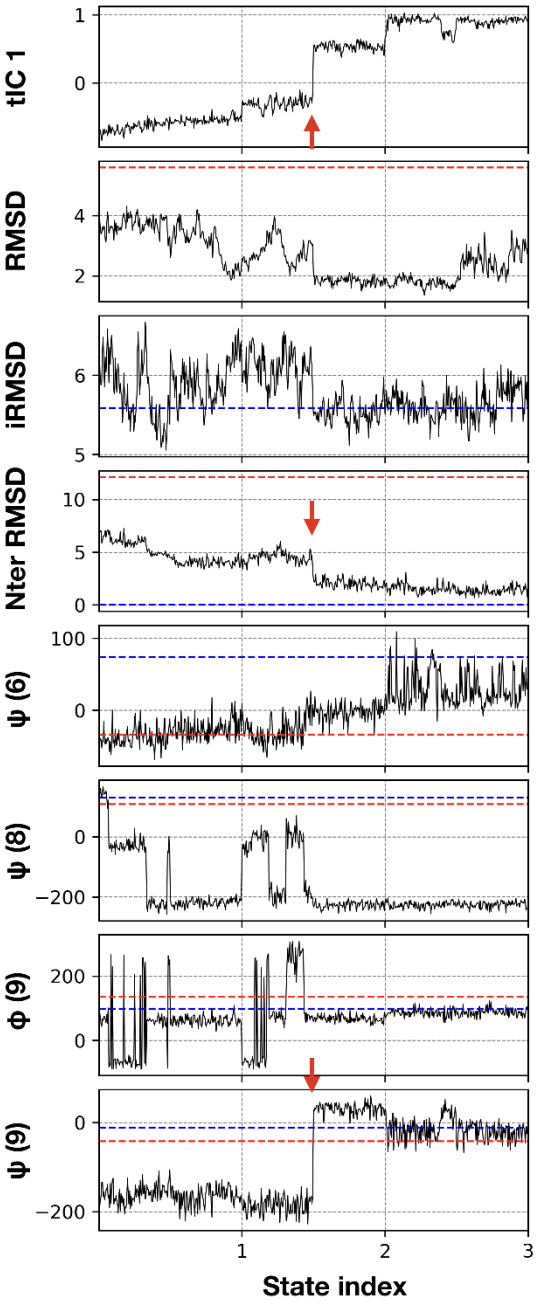

Figure S17. Structural transitions during refinement of TR872. As Fig. S13 with Cα RMSD (in Å) of the N-terminus (residues 1-11), backbone ψ torsion angles for residues 6, 8, and 9, and backbone φ torsion angle for residue 9.

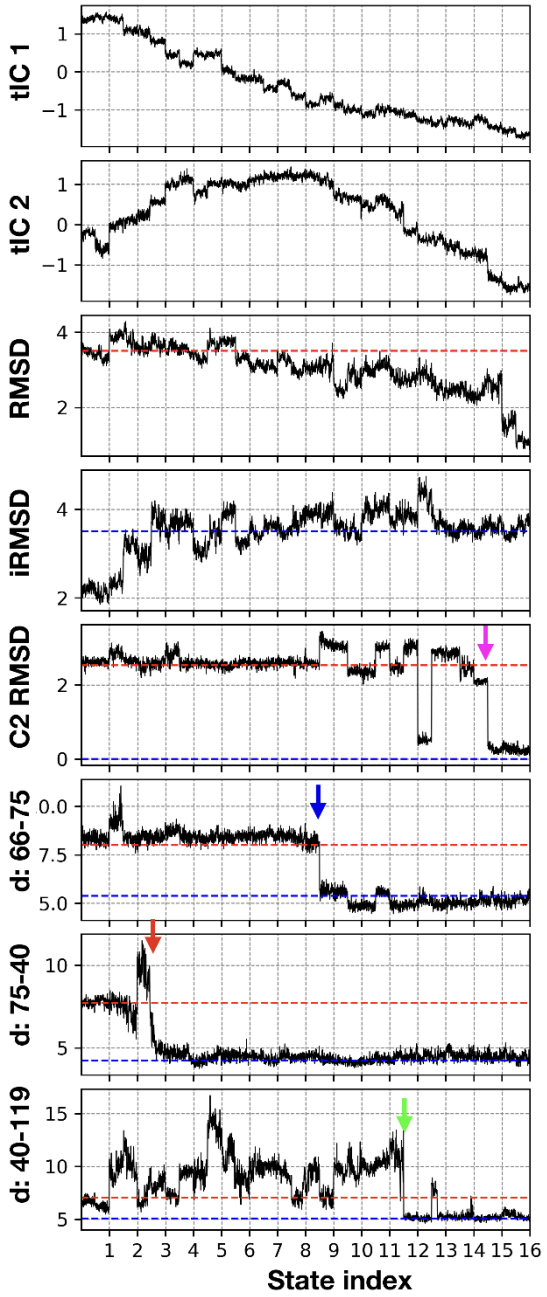

Figure S18. Structural transitions during refinement of TR921. As Fig. S13 with Cα RMSD (in Å) of loop C2 (68-73) and distances (in Å) between selected residue pairs.

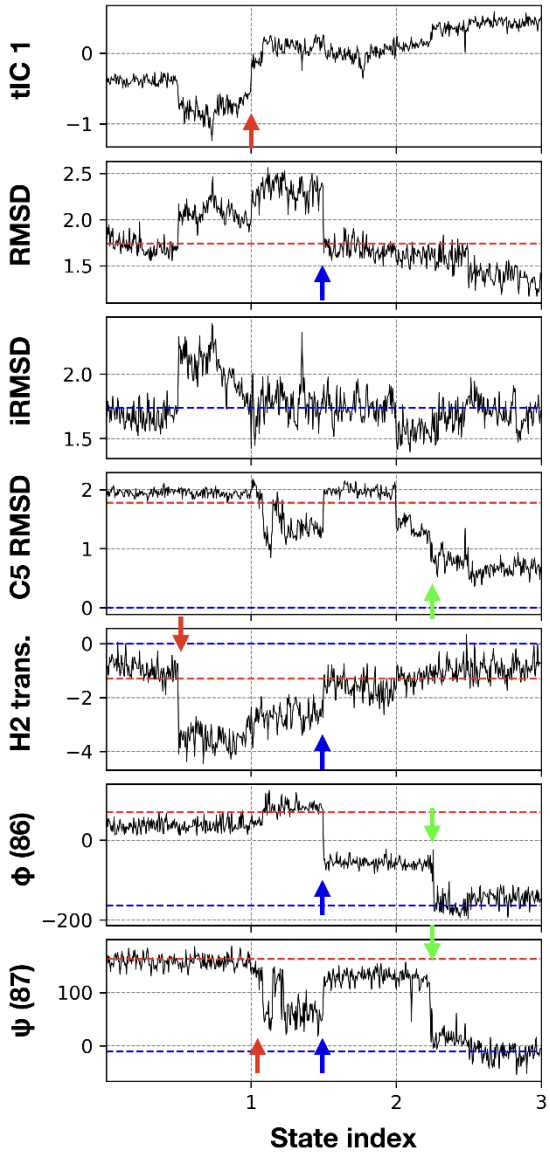

Figure S19. Structural transitions during refinement of TR769. As Fig. S13 with Cα RMSD (in Å) for loop C5 (residues 85-90), translation of helix H2 (residues 64-85), backbone ψ torsion angle for residue 87, and backbone φ torsion angle for residue 86.

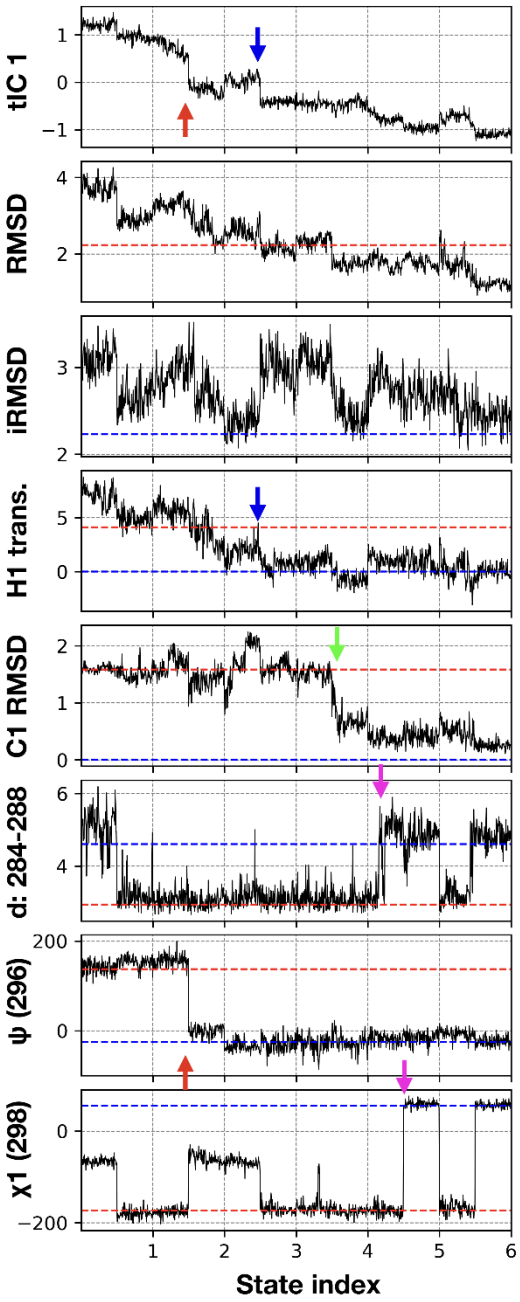

Figure S20. Structural transitions during refinement of TR894. As Fig. S13 with Cα RMSD (in Å) for loop C1 (residues 292-297), translation of helix H1 (residues 271-286), the pairwise distance between residues 284 and 288 (in Å), the backbone ψ torsion angle for residue 296, and the side chain χ_1 torsion angle for residue 298.

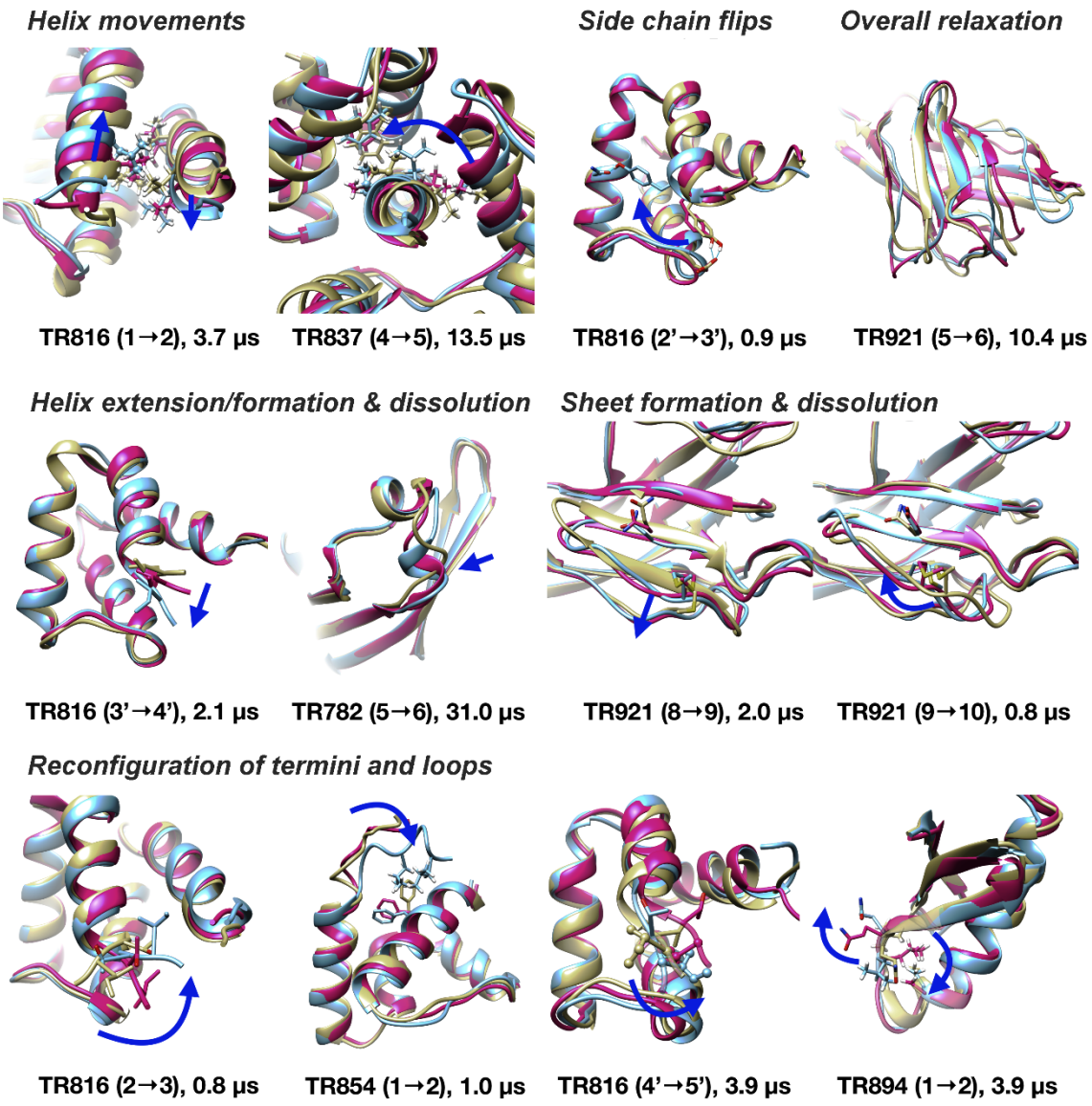

Figure S21. Structural basis of kinetic barriers during refinement transitions. Examples of transitions during refinement categorized by transition type. Transition state structure are shown in magenta, structures before and after transitions are shown in yellow and light blue, respectively. Key structure changes are indicated by blue arrows. The system, state indices between which the transition occurs, and the mean first passage times (MFPT) from the MSM analysis are indicated below each structure.

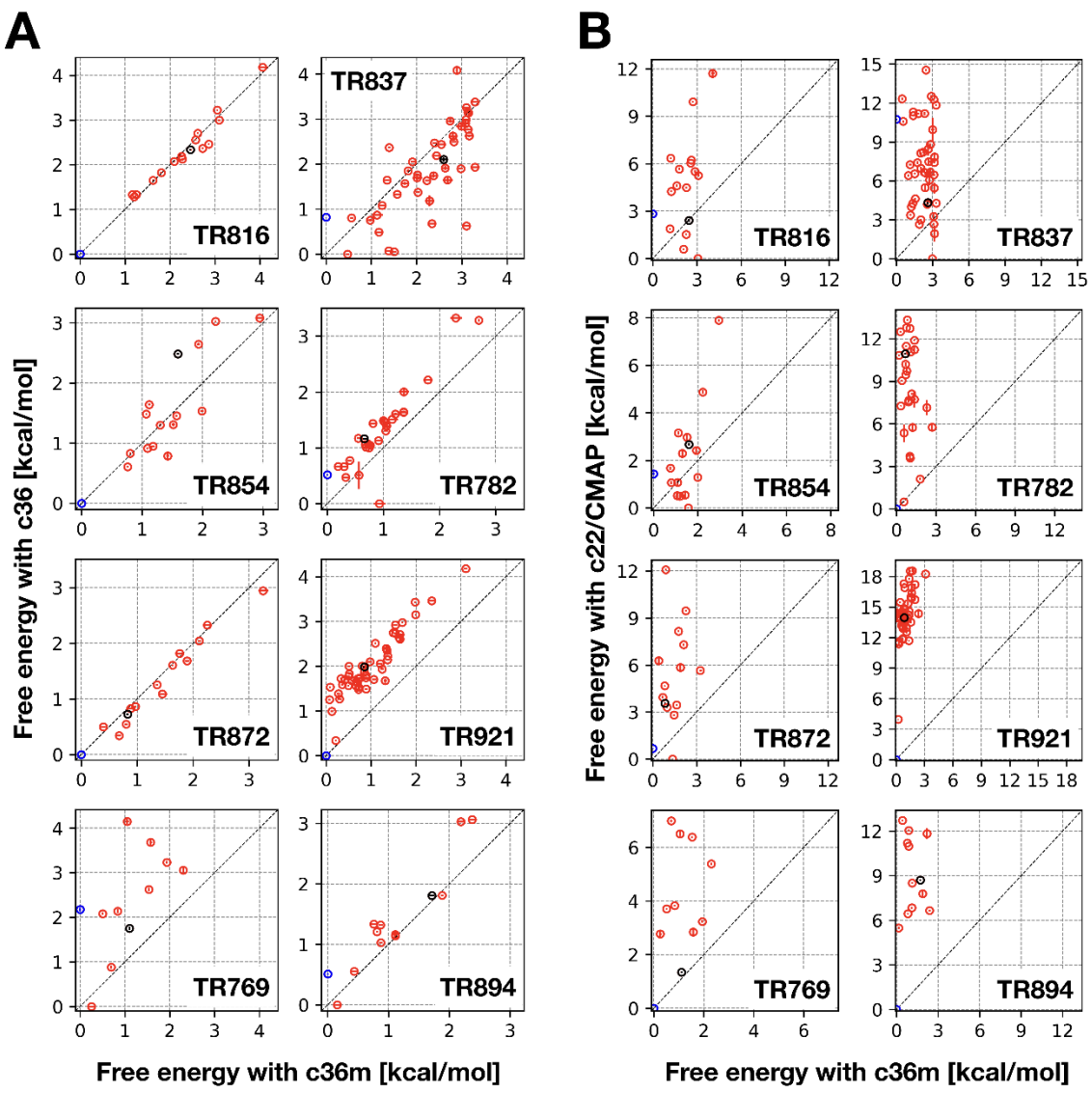

Figure S22. Free energies of MSM states with different force fields. Free energies were reweighted based on the energy difference between force fields between CHARMM c36m and c36 (A) and CHARMM c36m and c22/CMAP (B). The native states, the initial states, and other states are shown as blue, black, and red circles, respectively. Error bars were obtained by using 20-fold cross-validations with 95% subsets of the trajectories.

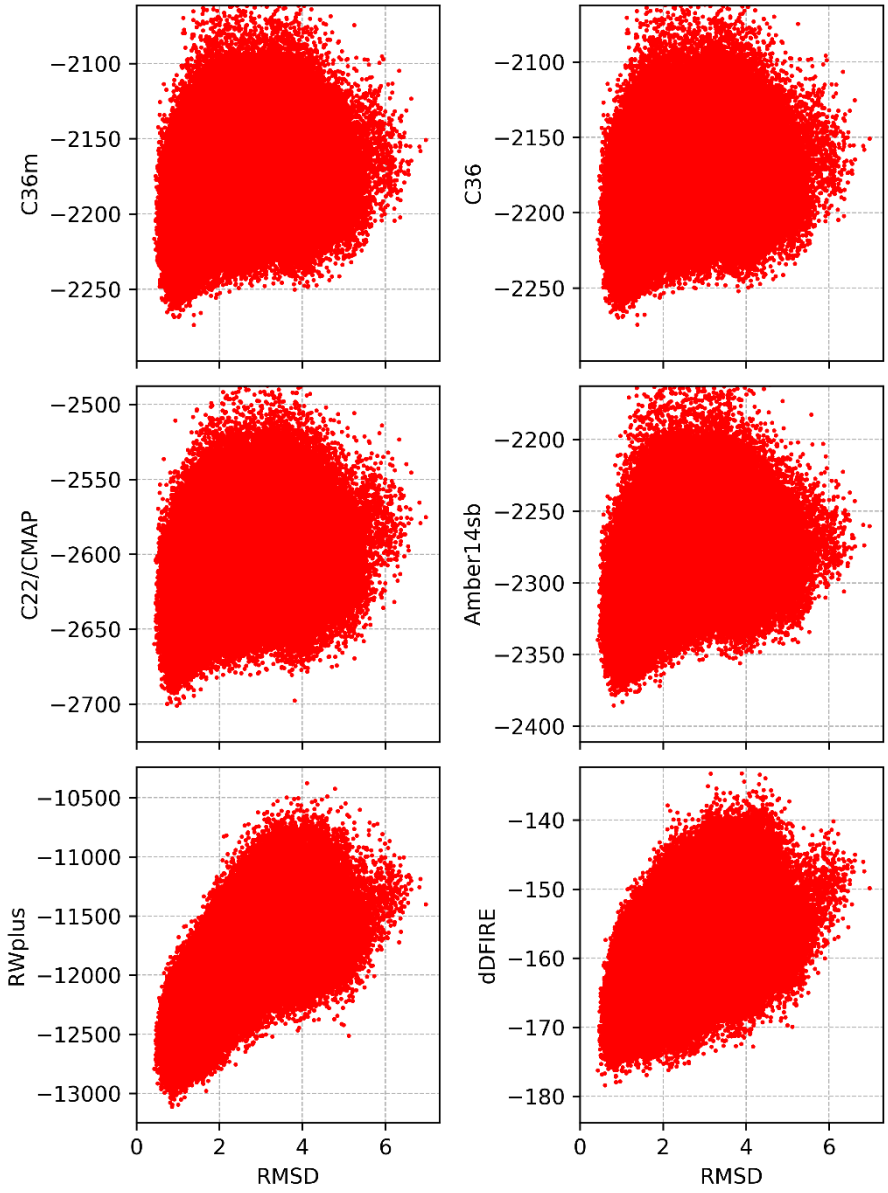

Figure S23. Scoring of TR816 snapshots. Comparison of different scoring functions with the Ca RMSD from the experimental structure for the entire set of snapshots generated by the iterative MD simulations. MMGB/SA scores with four force fields (CHARMM c36m, c36, c22/CMAP, and Amber ff14sb) are compared with the statistical potentials (RWplus and dDFIRE). Data points where scores are more than five standard deviations away from the mean are not shown to focus on the low-score regions.

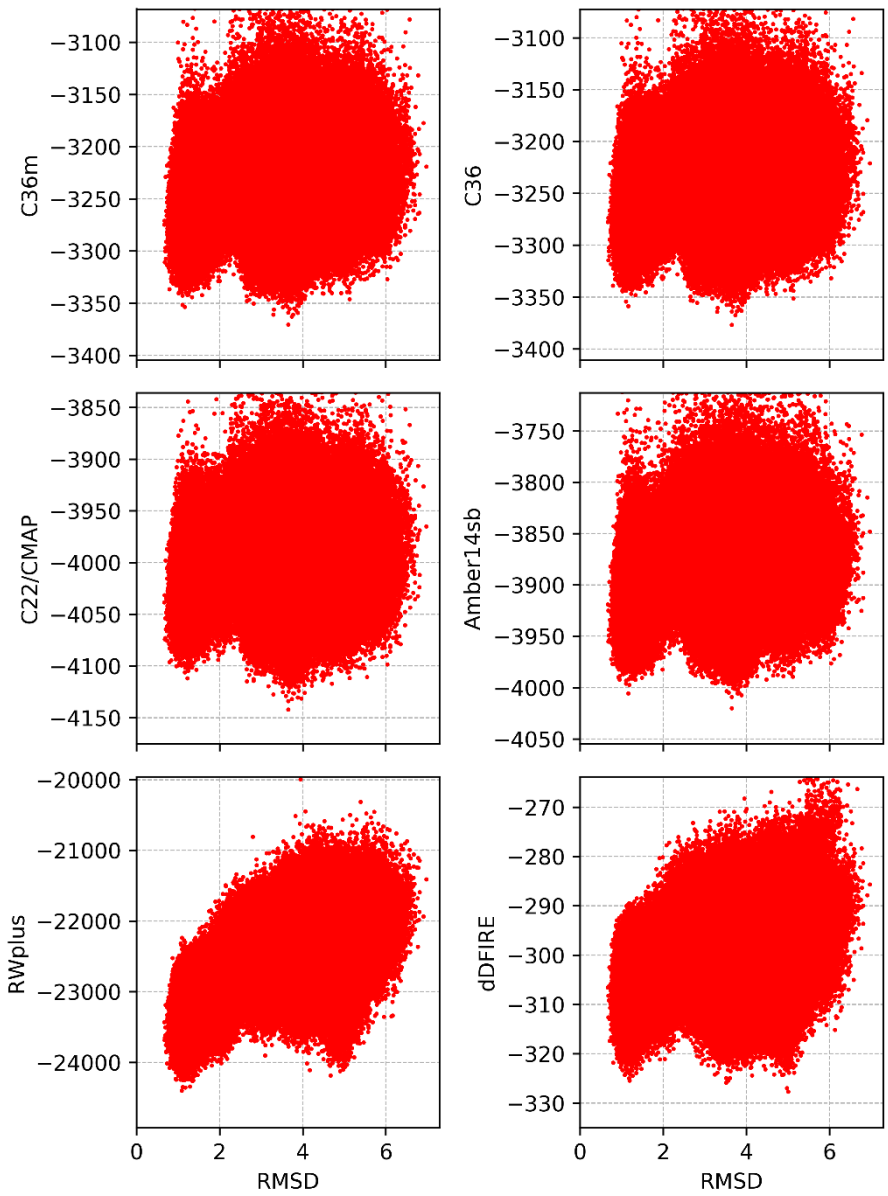

Figure S24. Scoring of TR837 snapshots. See Fig. S23.

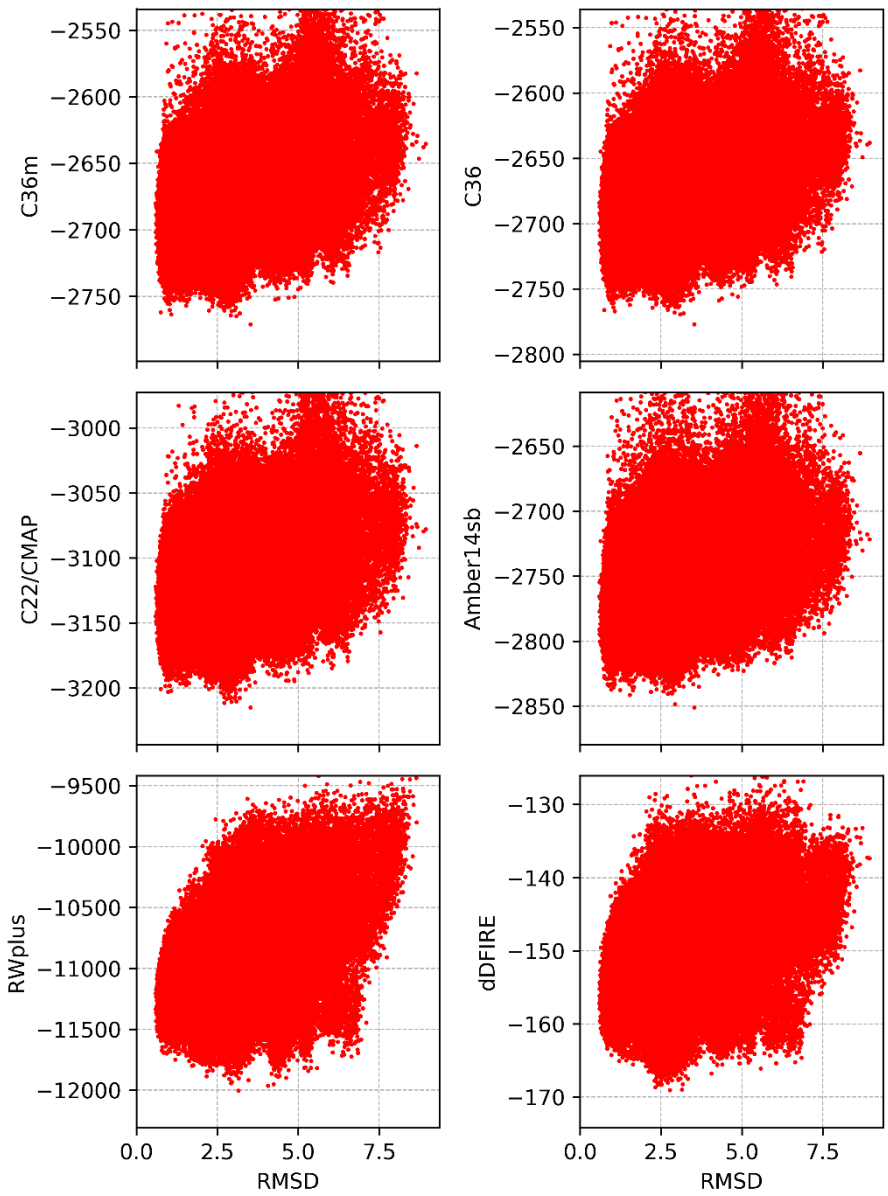

Figure S25. Scoring of TR854 snapshots. See Fig. S23.

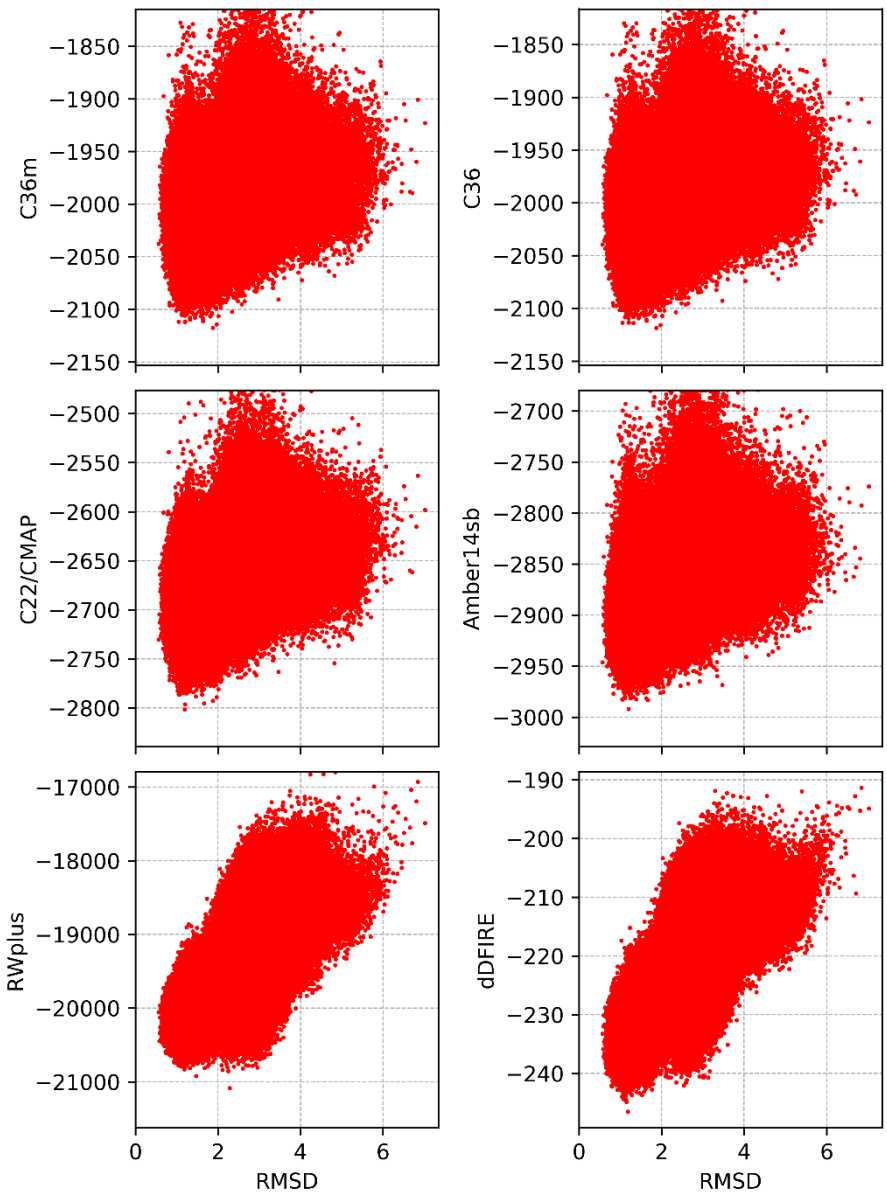

Figure S26. Scoring of TR782 snapshots. See Fig. S23.

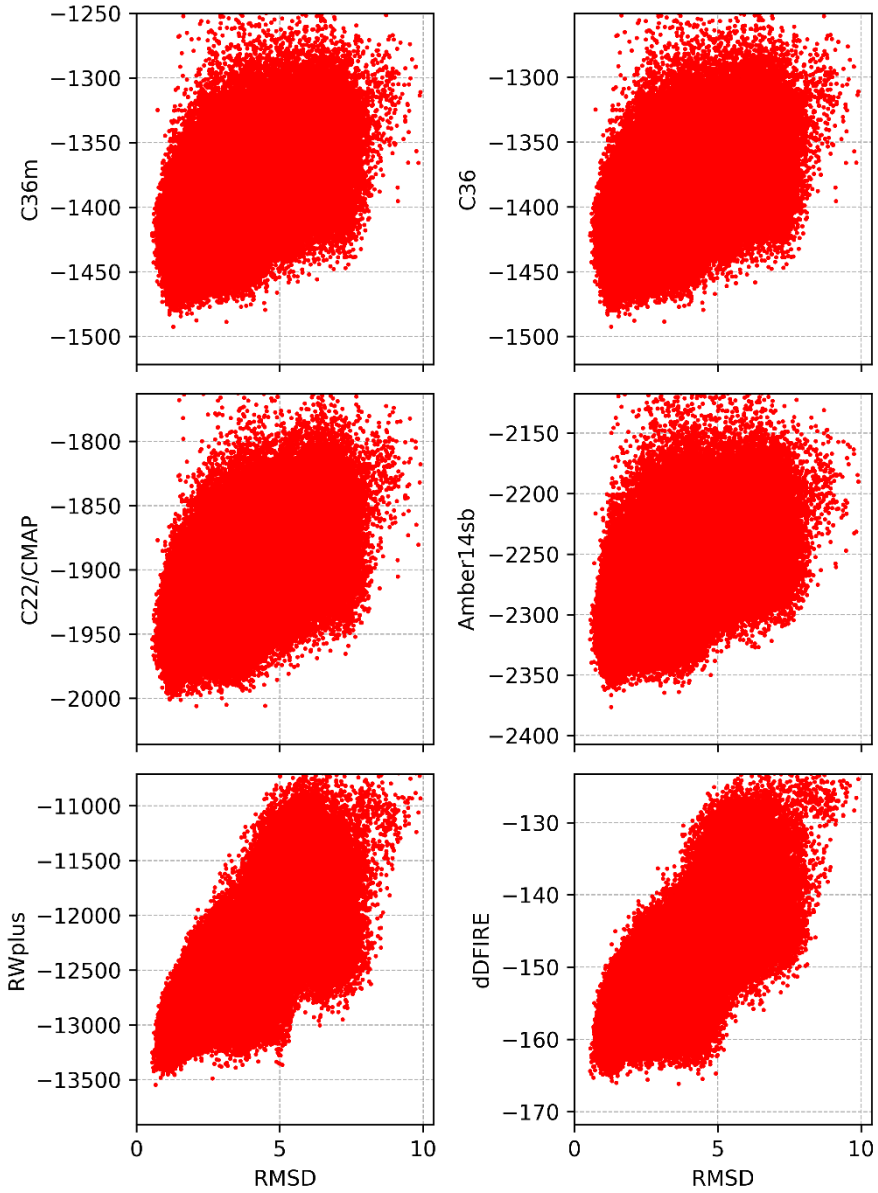

Figure S27. Scoring of TR872 snapshots. See Fig. S23.

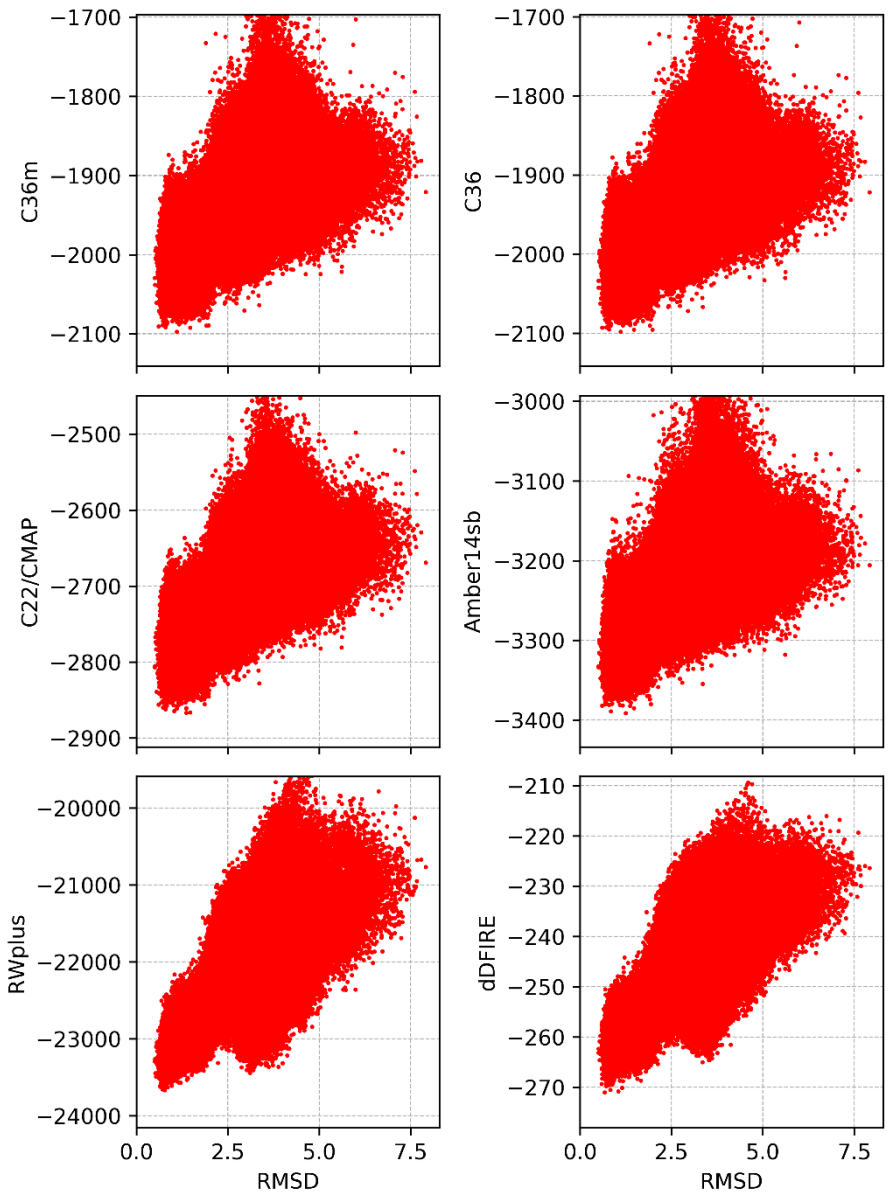

Figure S28. Scoring of TR921 snapshots. See Fig. S23.

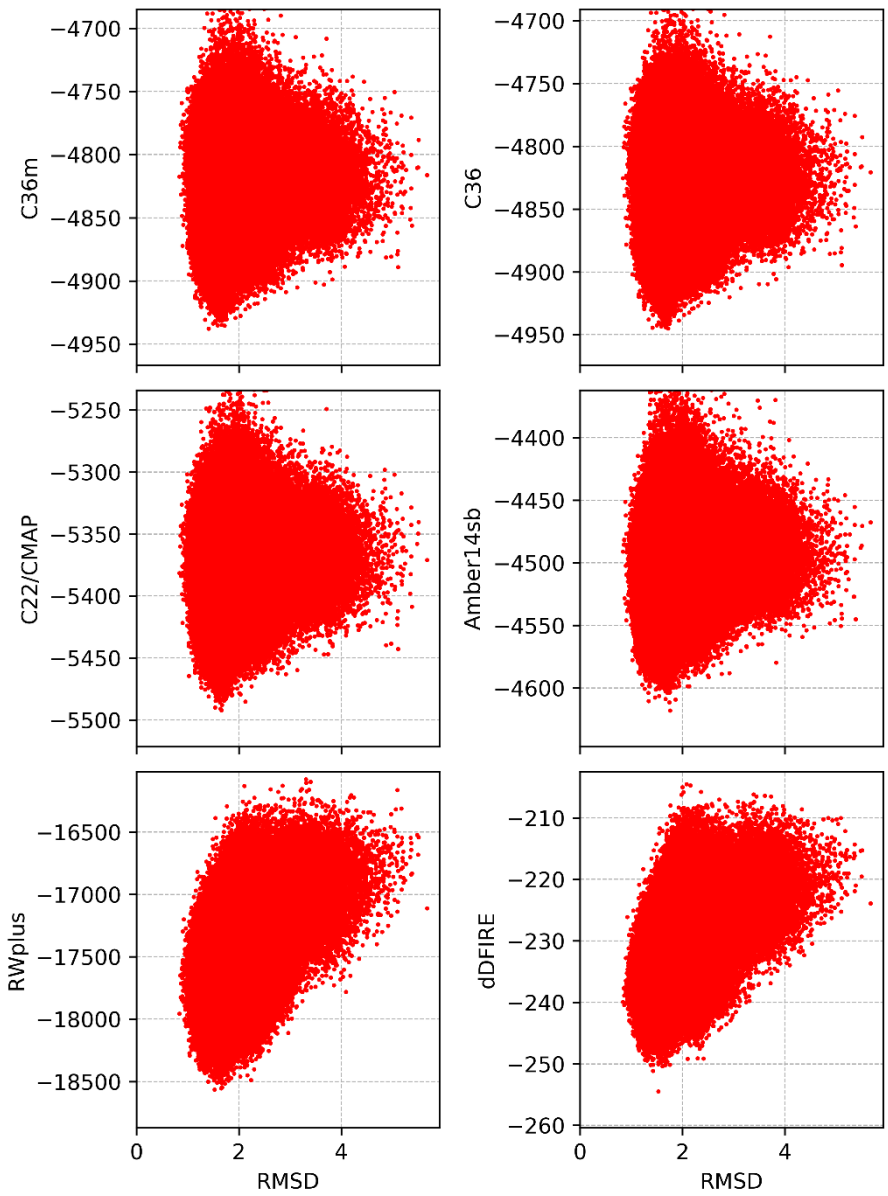

Figure S29. Scoring of TR769 snapshots. See Fig. S23.

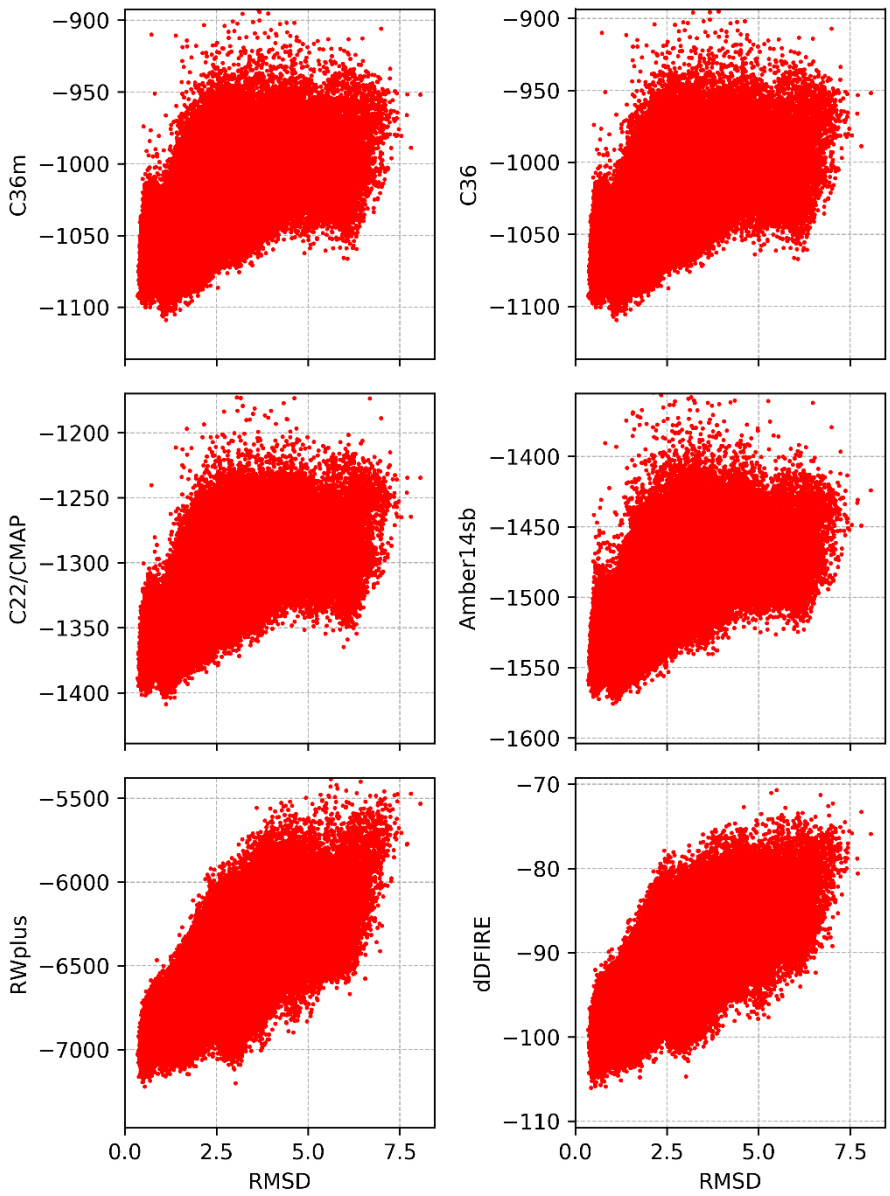

Figure S30. Scoring of TR894 snapshots. See Fig. S23.

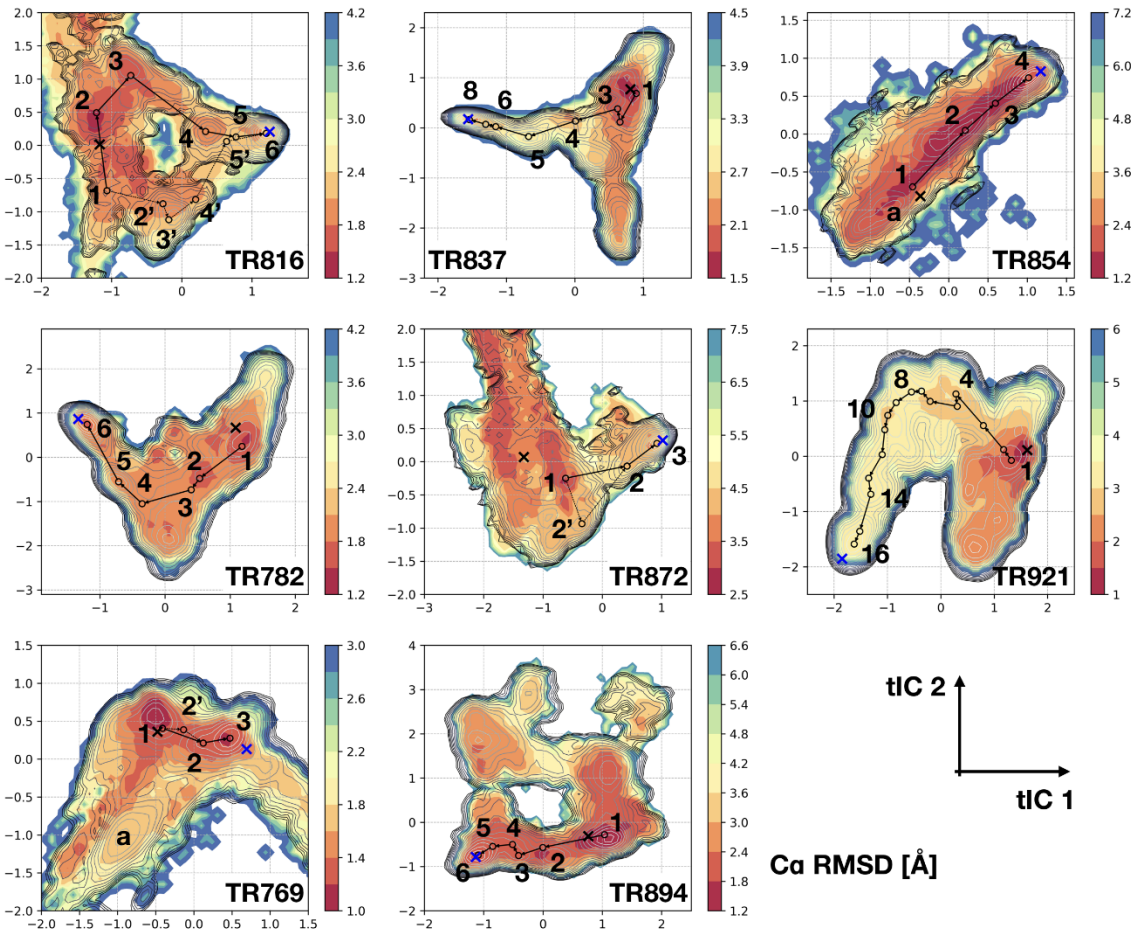

Figure S31. Deviations from homology models. Cα RMSD deviations with respect to the initial homology model are mapped onto the energy landscapes as a function of the first two tICA coordinates. Minimum RMSD values were chosen where multiple structures map onto the same tICA coordinates. Contour lines correspond to free energies as in Fig. 2 but colors indicate RMSD deviations. MSM states and refinement transition paths are shown as in Fig. 2.

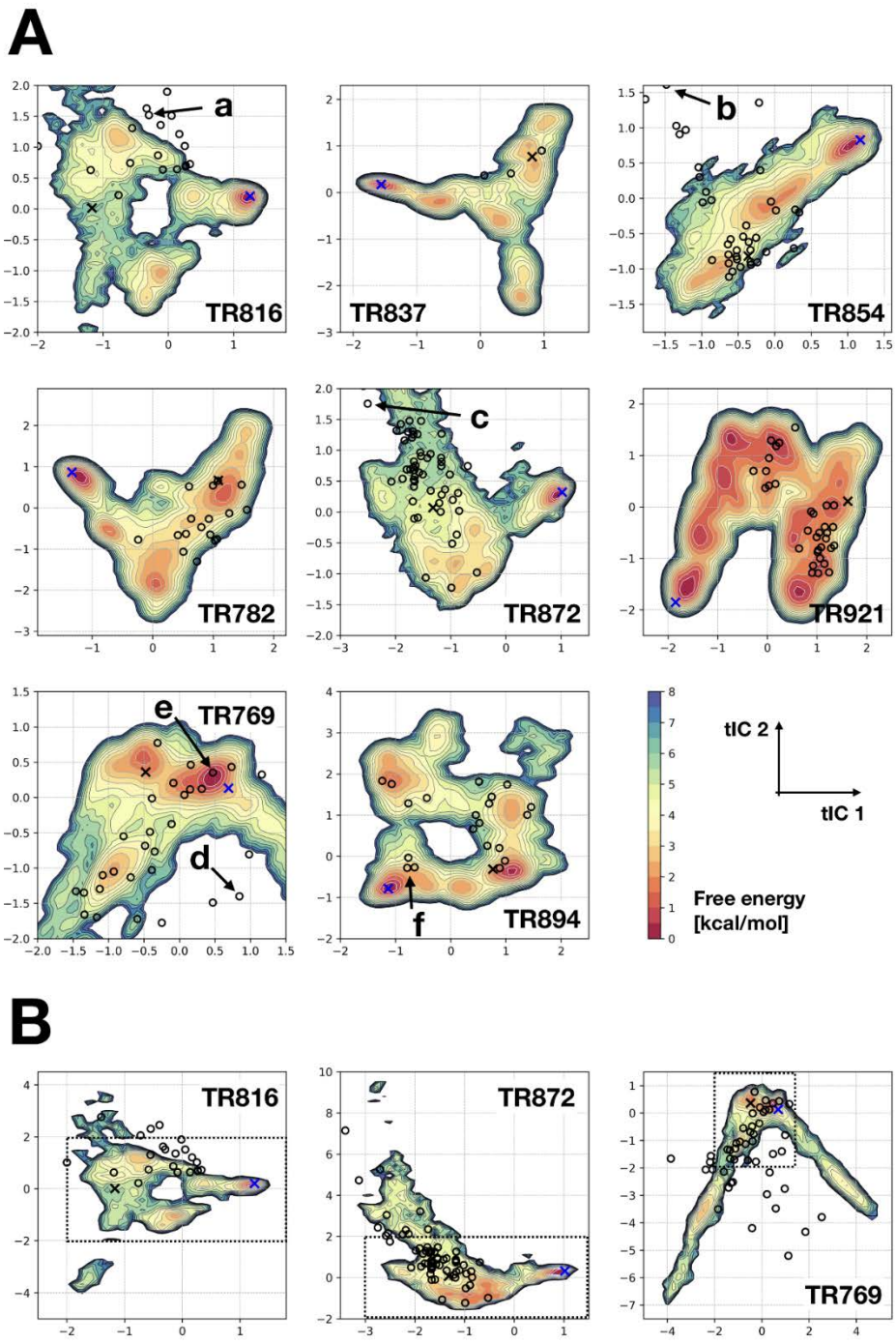

Figure S32. Projections of alternative initial models onto free energy landscapes. Alternative initial models are projected onto free energy landscapes as in Fig. 2 (A). Additional complete maps are shown for TR817, TR872, and TR769 as in Fig. S3 (B). Projections of the experimental structures and initial homology models are shown as blue and black Xs, while those of alternative initial models are indicated with black circles. For additional information about the free energy landscape maps, see Fig. 2. Selected outliers that are outside the 8 kcal/mol contour lines (a-d) and structures that appear very close to the native state (e and f) are shown in Fig. S2 and discussed in the text.

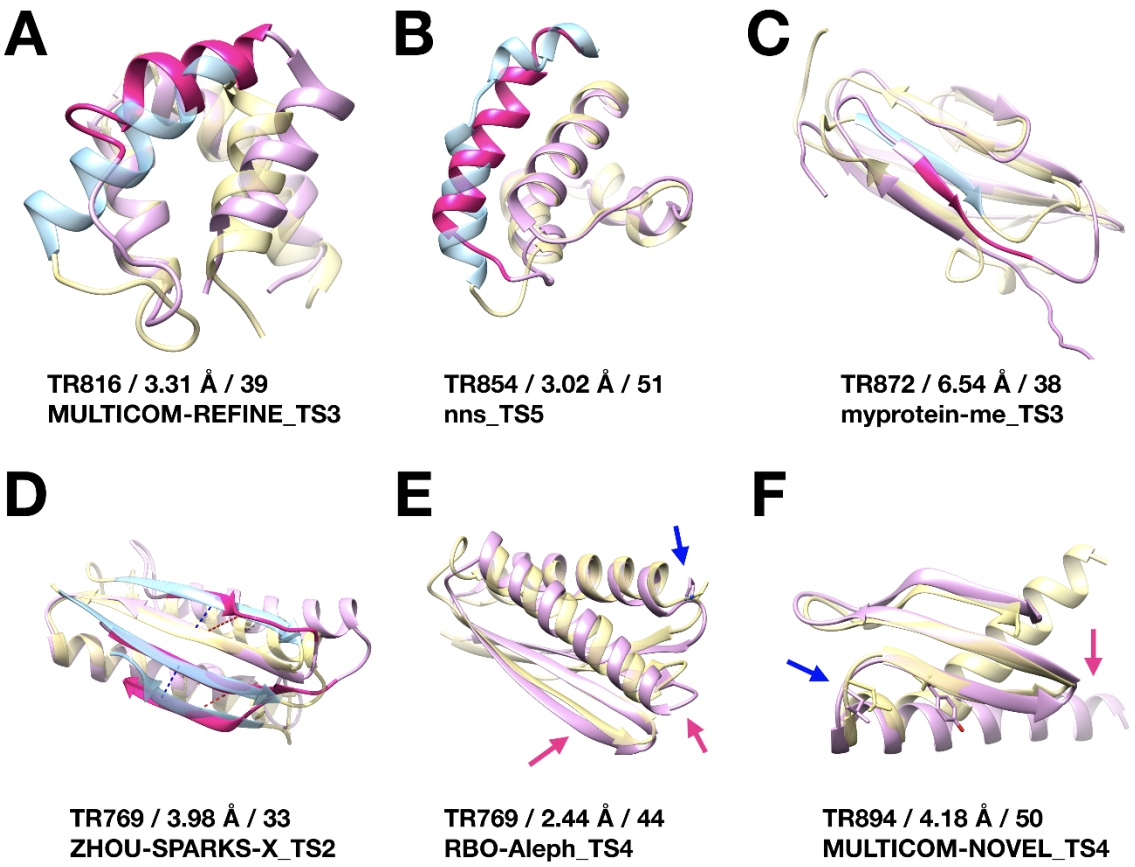

Figure S33. Selected initial models. Experimental structures (yellow) and alternative initial models (pink) with CASP target ID, deviation in Cα-RMSD and GDT-HA, and corresponding model names are shown in each panel. Structures A–D correspond to outliers in high free energy regions (see Fig. S32). The most different regions are depicted in light blue and magenta for the experimental and model, respectively. In panel D, the correct beta-pairing residues in the experimental structure are shown in blue dashed lines, while corresponding pairs in the model structure are indicated as red dashed lines. Structures E and F show initial models that appear very close to the native states in Fig. S32. Blue arrows indicating features that are significant modeling error in the original initial models but not in these models while magenta arrows indicating other modeling errors leading to overall larger structural deviations.

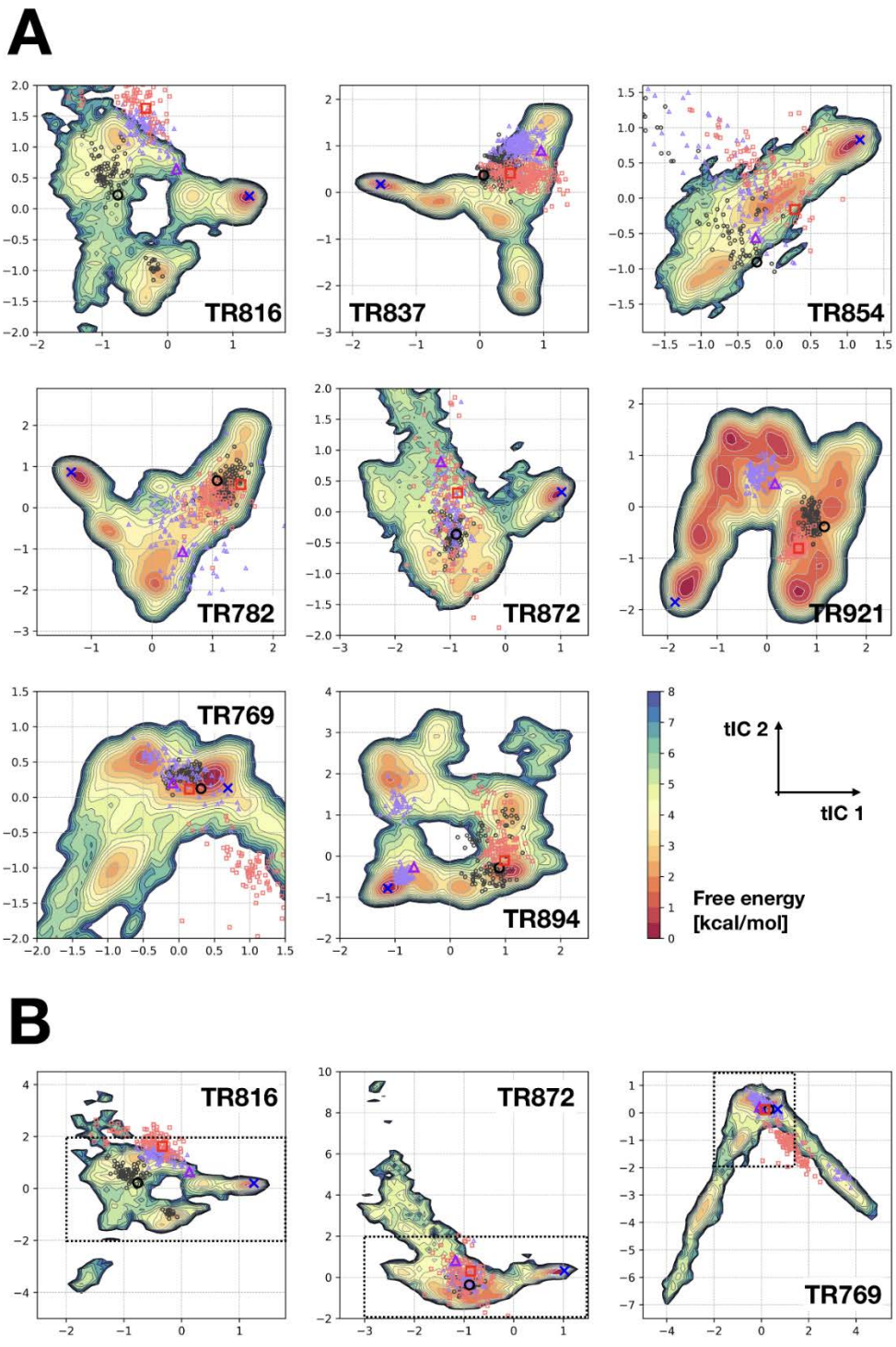

Figure S34. Projections of sampled conformations from three selected alternative initial models. Selected alternative initial models are shown as larger black circles, purple triangles, and red squares. Sampled conformations for each alternative initial model are shown with smaller shapes and lighter colors as the initial model. Other descriptions are the same as in Fig. S32.

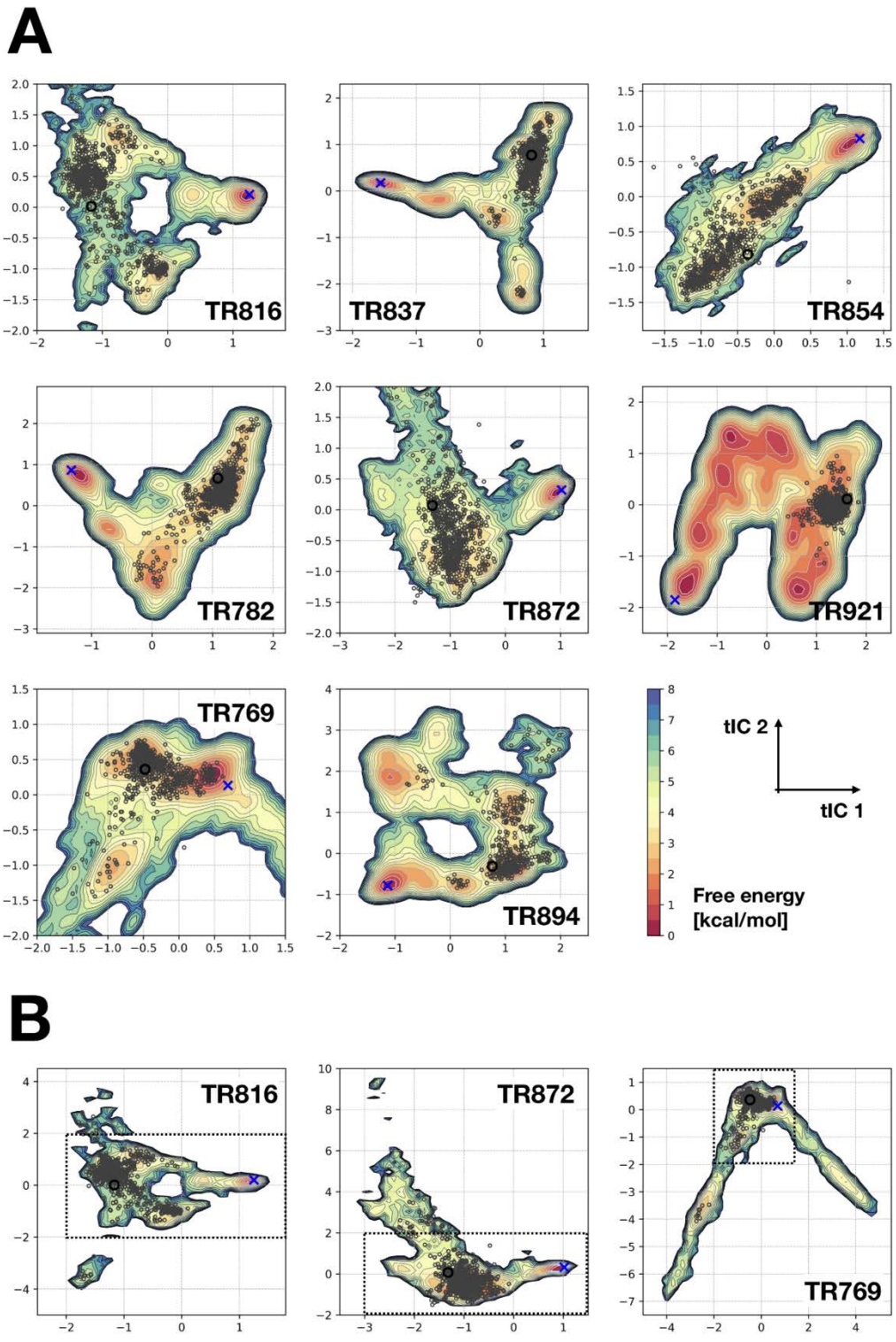

Figure S35. Projections of sampled conformations from the initial models. Conformations are projected onto the free energy landscapes at every 5 ns. They are shown in dark grey circles, while the initial models are shown as large black circles. The other descriptions are the same as in Fig. S32.

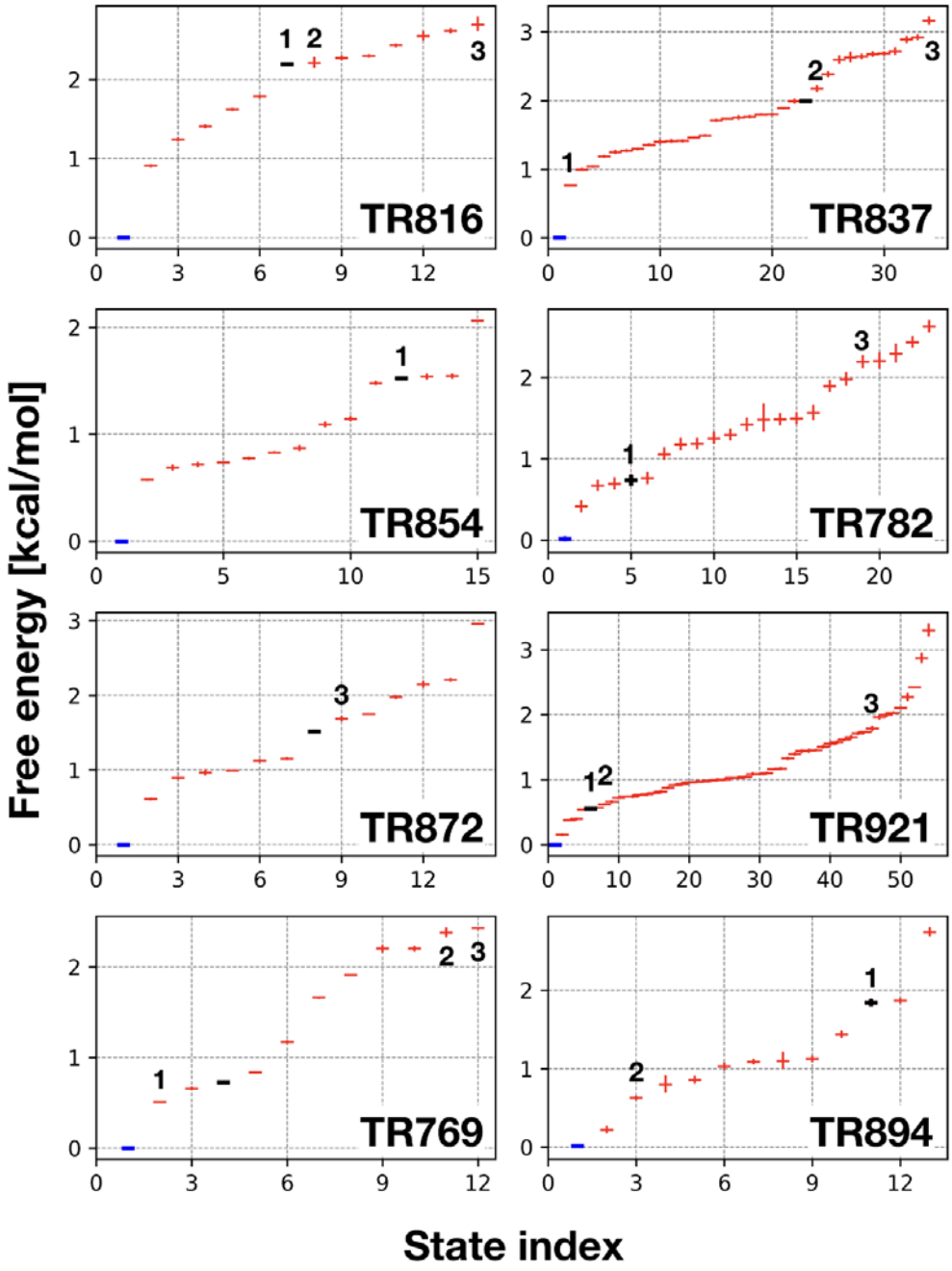

Figure S36. Free energies for different Markov model states with sampling from additional initial models. The states are ordered by free energy and vertical bars indicate uncertainties as in Fig. S2. Conformations that unfolded significantly were excluded from the analysis for TR837, TR854, TR782, and TR894. Sampling from two alternative initial models for TR872 did not overlap with the rest of the sampling so that a complete Markov model could only be built with one of the alternative models. The state closest to the experimental structure (the native state) is indicated in blue and the state closest to the initial homology model is shown in black. The states closest to the additional initial models are marked with 1–3.

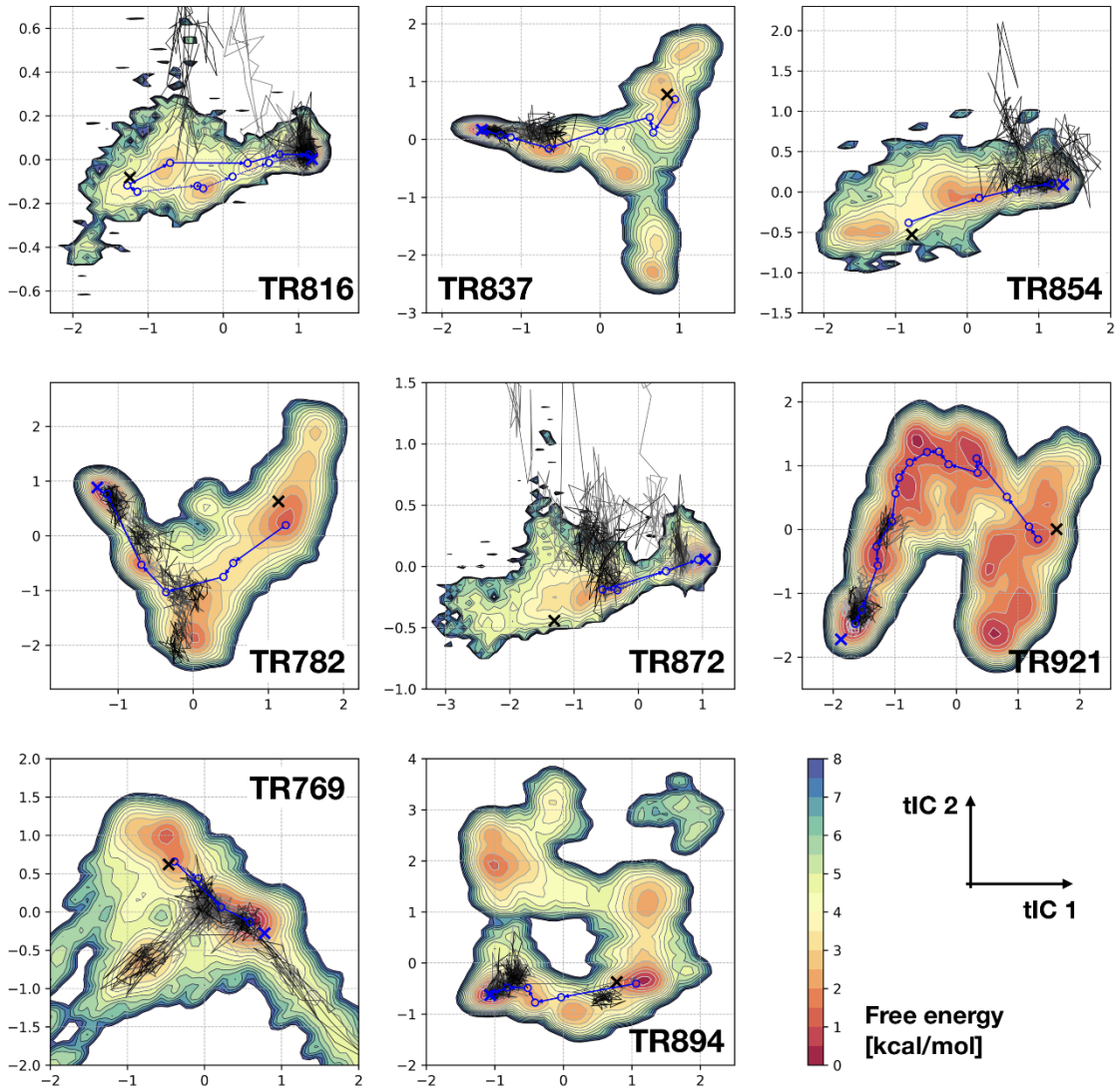

Figure S37. Projections of unfolding simulation trajectories. Trajectories from unfolding simulations with the experimental structures in 380 K are projected onto the free energy landscapes. tICA projection parameters were reoptimized by using the original simulation trajectories and unfolding trajectories together to incorporate the dynamics of unfolding. The free energy landscape is drawn only based on the simulation data from the original sampling at 298 K but for the new tICA coordinates. Since the projection coordinates are different, the free energy landscape maps look slightly different from Fig. 2. Results from five independent 200-ns long trajectories are projected at every 5 ns in black lines. The lines are depicted from light grey to black according to their simulation time. The refinement pathways are shown in blue circles and arrows.

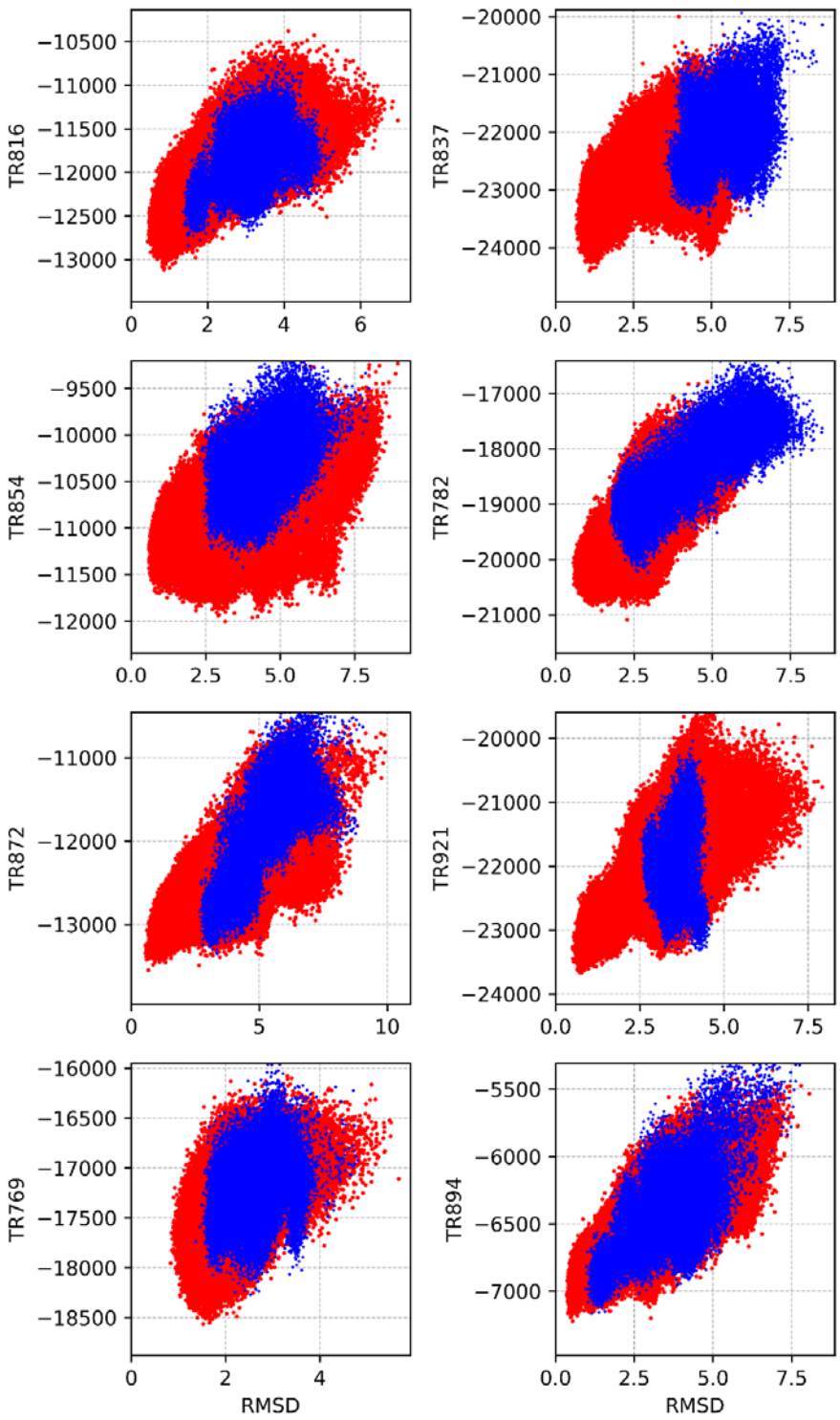

Figure S38. RWplus scores for conformations from alternative initial models. Snapshots from additional initial models (blue dots) are overlaid on the snapshots generated by the iterative MD simulations (red dots). Data points where scores are more than five standard deviations away from the mean are not shown to focus on the low-score regions.

Table S1. Summary of systems

Target	PDB ID	N_{res}	Fold	Cα RMSD [Å]	GDT-HA	Modeling Server	Modeling errors
TR816	5a1qA (1–68)	68	α	2.53	51.8	Zhang-server	Helix orientations, N-terminal structure
TR837	5tf3A (22–142)	121	α	2.95	43.8	QUARK	Helix orientations
TR854[1]	4rn3A (24–93)	70	α	2.27	60.4	Rosetta	Helix orientations, loop structure
TR782	4qrlA (26–135)	110	β	1.93	65.2	Rosetta	Long loop structure
TR872	5jmbA (43–130)	88	β	5.59	56.8	Pcons-net	N-terminal structure and its orientation
TR921	5aozA (402–539)	138	β	3.51	48.4	GOAL	Sequence alignment, loop structures
TR769	2mq8A (1–97)	97	α/ β	1.74	59.8	Zhang-server	loop structure, packing
TR894[2]	5hkqA (270–323)	54	α/ β	2.23	54.2	Bates-BMM	Helix/sheet orientations, loop structure

[1] N-terminal domain.
[2] C-terminal domain; this protein forms a hetero-complex with another protein (5hkqB).

Table S2. MD simulation details

Target	Number of iterations	Total simulation time (μs)	Number of atoms
TR816	6	35.2	11,810–27,464
TR837	6	37.6	18,174–29,874
TR854	3	24.7	12,921–25,620
TR782	5	30.6	17,966–31,925
TR872	4	22.2	27,720–64,089
TR921	13	55.9	26,421–53,964
TR769	3	26.5	18,102–26,928
TR894	4	15.0	12,339–19,509

Table S3. Summary of Markov state models

Target	Lag time [ns]	Number of microstates	Number of macrostates	Pair similarity[1]
TR816	5	40	17	2.38 / 63.9
TR837	10	200	44	2.94 / 49.2
TR854	5	100	16	2.99 / 72.9
TR782	8	100	28	1.84 / 78.2
TR872	5	100	12	3.25 / 60.7
TR921	5	200	50	2.60 / 63.9
TR769	5	30	11	1.67 / 73.4
TR894	5	40	13	2.42 / 60.1

[1] Average of pairwise Cα RMSD and GDT-HA between macrostates after structural averaging.

Table S4. Deviations of initial and native states from initial model

Target	Native state[1] (MD[2])	Initial state[1] (MD[2])
TR816	2.42 / 51.5	1.88 / 58.1
TR837	2.88 / 45.7	2.12 / 52.5
TR854	2.26 / 58.6	3.00 / 55.0
TR782	1.96 / 53.0	1.93 / 62.3
TR872	5.24 / 61.1	5.18 / 58.2
TR921	3.41 / 49.8	1.73 / 65.0
TR769	1.39 / 67.8	1.41 / 67.5
TR894	2.29 / 51.4	1.95 / 68.1

[1] structure similarity between structures in Cα RMSD and GDT-HA.
[2] ensemble averaged structure corresponding to the MSM macrostates closest to the experimental structure (native state) and the initial homology model (initial state).

Table S5. Correlation between average scores and free energies for MSM states.

Target	*MMGB/SA*				*Statistical Potentials*	
	c36m	**c36**	**c22/CMAP**	**ff14SB**	**RWplus**	**dDFIRE**
TR816	0.785[1] / 0.060[1]	0.785 / 0.061	0.734 / 0.052	0.784 / 0.055	0.509 / 0.002	0.583 / 0.157
TR837	0.317 / 0.014	0.309 / 0.014	0.221 / 0.010	0.380 / 0.016	0.228 / 0.0007	0.112 / 0.020
TR854	0.592 / 0.017	0.591 / 0.017	0.594 / 0.018	0.662 / 0.018	0.495 / 0.002	0.462 / 0.092
TR782	-0.108 / -0.004	-0.104 / -0.003	-0.151 / -0.004	-0.079 / -0.003	-0.396 / -0.0007	-0.316 / -0.037
TR872	0.399 / 0.015	0.396 / 0.015	0.373 / 0.013	0.369 / 0.014	0.370 / 0.0008	0.262 / 0.030
TR921	0.404 / 0.011	0.404 / 0.011	0.410 / 0.010	0.399 / 0.012	0.524 / 0.001	0.548 / 0.072
TR769	-0.031 / -0.002	-0.044 / -0.003	-0.076 / -0.004	0.035 / 0.002	0.051 / 0.0002	0.059 / 0.011
TR894	0.473 / 0.023	0.483 / 0.023	0.422 / 0.018	0.393 / 0.018	0.185 / 0.0009	0.232 / 0.059
Avg.	*0.354 / 0.017*	*0.352 / 0.017*	*0.316 / 0.014*	*0.368 / 0.017*	*0.246 / 0.0008*	*0.243 / 0.051*

[1]Pearson's correlation coefficients and slopes from linear regression fits.

Table S6. Summary of refinement transitions

Transition type	**Time-scales [μs]**	**Rate-determining factors**
Helix movement	0.8-33.1	steric hindrance of side chains
α-helix formation	4.7, 5.4, 31.0	backbone torsion conversion
α-helix dissolution	2.1, 33.1	breaking of hydrogen bonds
β -sheet formation	0.8-1.3	steric hindrance
β-sheet dissolution	2.0-3.0	breaking of hydrogen bonds
Loop conversion	0.8-5.4, 31.0, 623	backbone torsion conversion, steric hindrance, topological constraints
Terminal conversion	0.8-2.2, 3.9	backbone torsion conversion
Sidechain flip	0.5, 0.9	side chain torsion conversion
Relaxation	0.4-1.0, 3.0, 4.0	protein size
Overall adjustment	10.4	steric hindrance

Table S7. Summary of the selected alternative initial models

Target	Model name	Similarity to the experimental structure[1]	Similarity to the initial model[1]
TR816	Zhang-Server_TS2	2.53 / 53.8 / 22.0	-[2]
	RBO_Aleph_TS2	2.44 / 48.9 / 22.9	2.28 / 55.2 / 24.5
	RaptorX-FM_TS2	2.85 / 41.6 / 19.9	3.45 / 38.2 / 14.1
	MULTICOM-REFINE_TS5	2.91 / 46.0 / 8.3	2.77 / 45.6 / 11.9
TR837	QUARK_TS1	2.95 / 43.8 / 21.1	-[2]
	Zhang-Server_TS2	3.66 / 39.5 / 19.2	2.11 / 49.0 / 29.2
	TASSER-VMT_TS4	5.24 / 23.8 / 10.2	4.34 / 32.2 / 12.6
	RaptorX-FM_TS1	5.76 / 24.8 / 8.4	5.25 / 25.6 / 9.3
TR854	BAKER-ROSETTASERVER_TS3	2.27 / 60.4 / 28.2	-[2]
	QUARK_TS1	2.22 / 60.0 / 16.4	1.95 / 63.6 / 21.9
	RBO_Aleph_TS2	2.39 / 53.6 / 22.9	2.13 / 58.6 / 25.6
	Raghavagps-tsppred_TS1	2.54 / 56.1 / 25.7	2.78 / 52.9 / 25.4
TR782	BAKER-ROSETTASERVER_TS3	1.93 / 65.2 / 37.1	-[2]
	myprotein-me_TS5	1.93 / 65.0 / 37.6	0.21 / 99.6 / 77.1
	FALCON_EnvFold_TS5	4.08 / 44.7 / 18.8	4.04 / 44.3 / 20.3
	SAM-T08-server_TS1	4.37 / 49.6 / 24.8	4.12 / 51.4 / 27.5
TR872	Pcons-net_TS3	5.59 / 56.8 / 38.1	-[2]
	BAKER-ROSETTASERVER_TS1	3.55 / 54.3 / 34.6	6.32 / 54.0 / 31.2
	IntFOLD4_TS3	4.29 / 44.0 / 19.5	6.88 / 44.0 / 20.0
	myprotein-me_TS2	4.83 / 45.7 / 22.7	7.91 / 44.6 / 20.0
TR921	GOAL_TS1	3.51 / 48.4 / 27.1	-[2]
	HHGG_TS4	3.10 / 48.2 / 29.3	1.84 / 68.5 / 41.8
	slbio_TS4	3.23 / 45.1 / 20.5	2.52 / 57.1 / 27.7
	BAKER-ROSETTASERVER_TS3	3.29 / 48.6 / 24.6	2.67 / 58.0 / 31.5
TR769	Zhang-Server_TS1	1.74 / 59.8 / 33.0	-[2]
	MULTICOM-NOVEL_TS1	2.01 / 55.4 / 27.9	2.35 / 58.5 / 26.9
	STRINGS_TS1	2.30 / 48.7 / 23.3	2.02 / 61.6 / 29.4
	nns_TS5	2.33 / 53.6 / 25.6	2.40 / 52.3 / 21.6
TR894	Bates-BMM_TS1	2.23 / 54.2 / 23.6	-[2]
	Zhang-Server_TS1	2.59 / 54.2 / 17.8	0.97 / 75.5 / 64.3
	RBO_Aleph_TS1	2.97 / 52.3 / 22.8	3.35 / 43.1 / 22.1
	BAKER-ROSETTASERVER_TS1	3.64 / 46.8 / 22.8	3.04 / 55.1 / 20.8

[1] Structure similarities are measured in Cα-RMSD, GDT-HA, and GDC-SC measures.
[2] They are the initial models themselves.

Movie S1. Structure transition for TR816 via the first path, state 1 → 2.

Movie S2. Structure transition for TR816 via the first path, state 2 → 3.

Movie S3. Structure transition for TR816 via the second path, state 2 → 3.

Movie S4. Structure transition for TR816 via the second path, state 4 → 5.

References

1. Huang J, *et al.* (2017) CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **14**:71-73.
2. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **79**:926-935.
3. Brooks BR, *et al.* (2009) CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **30**:1545-1614.
4. Eastman P, *et al.* (2017) OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *Plos Comp Biol* **13**:e1005659.
5. Harrigan MP, *et al.* (2017) MSMBuilder: Statistical Models for Biomolecular Dynamics. *Biophys J* **112**:10-15.
6. McGibbon Robert T, *et al.* (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J* **109**:1528-1532.
7. Zhang J & Zhang Y (2010) A Novel Side-Chain Orientation Dependent Potential Derived from Random-Walk Reference State for Protein Fold Selection and Structure Prediction. *Plos One* **5**:e15386.
8. Yang YD & Zhou YQ (2008) Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins* **72**:793-803.
9. Lee MS, Feig M, Salsbury FR, & Brooks CL (2003) New Analytic Approximation to the Standard Molecular Volume Definition and its Application to Generalized Born Calculations. *J Comput Chem* **24**:1348-1356.
10. Zhang Y & Skolnick J (2004) Scoring Function for Automated Assessment of Protein Structure Template Quality. *Proteins* **57**:702-710.