Biophysical Journal, Volume *112*

Supplemental Information

Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore

Alex Dickson and Samuel D. Lotz

Multiple Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore
Supplemental Information

Alex Dickson
Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI 48824 and Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824

Samuel D. Lotz
Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI 48824
(Dated: January 26, 2017)

TABLE S1: Mean first passage times of unbinding as measured here (*) and in previous simulations. Also tabulated is the method used and the total simulation time.

	Experiment [1]	*	Plattner and Noé [2]	Tiwary et al [3]	Doerr et al [4]	Buch et al [5]	Teo et al [6]
MFPT (μs)	1700	180	76	110000	100	11	3800
Methodology	–	WExplore	MSM	Metadynamics	Adaptive MSM	MSM	AMS [a]
Observables	–	k_{off} only	k_{off} and k_{on}	k_{off} and k_{on}	k_{off} and k_{on}	k_{off} and k_{on}	k_{off} only
Total sim. time (μs)	–	4.1	150	5.0	10	50	2.3

[a] Adaptive Multilevel Splitting method

[1] F. Guillain and D. Thusius, Journal of American Chemical Society **92**, 5534 (1970).
[2] N. Plattner and F. Noé, Nature Communications **6**, 7653 (2015).
[3] P. Tiwary, V. Limongelli, M. Salvalaglio, and M. Parrinello, Proceedings of the National Academy of Sciences **112**, E386 (2015).
[4] S. Doerr and G. De Fabritiis, Journal of Chemical Theory and Computation **10**, 2064 (2014).
[5] I. Buch, T. Giorgino, and G. De Fabritiis, Proceedings of the National Academy of Sciences of the United States of America **108**, 10184 (2011).
[6] I. Teo, C. G. Mayne, K. Schulten, and T. Lelièvre, Journal of Chemical Theory and Computation p. acs.jctc.6b00277 (2016).

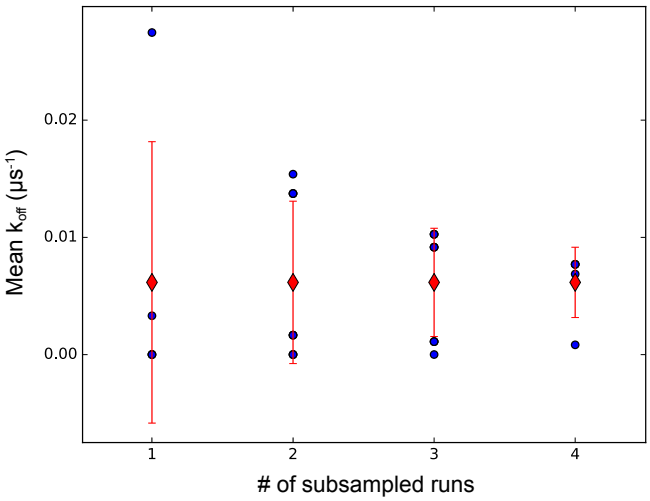

FIG. S1: **Subsampling error analysis.** Groups of 1-4 runs are used to calculate a series of average k_{off} values. The average of these averages is constant and shown using red diamonds. The standard deviations of these averages are shown as the error bars, and the individual averages are shown as blue circles. There are 5, 20, 20 and 5 points for the groups of 1, 2, 3 and 4, respectively, although in many cases the points are so close together that they are indistinguishable.

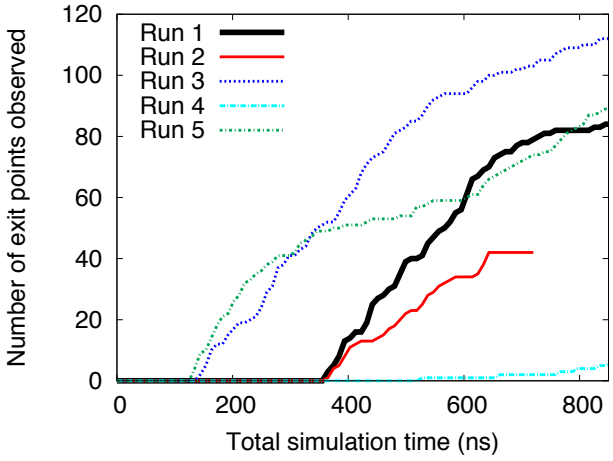

FIG. S2: **Total number of exit points.** For each of the five WExplore simulations, the total number of exit points is shown as a function of simulation time.

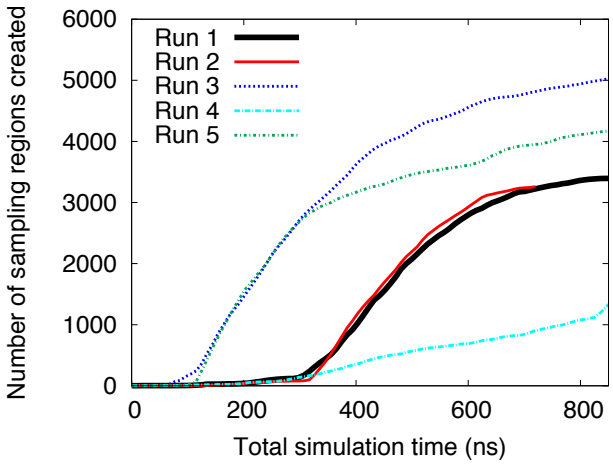

FIG. S3: **Total number of regions.** The number of regions discovered by each WExplore simulation is shown as a function of simulation time. Only the lowest-level regions are counted, which have a radius of 1.7 Å.

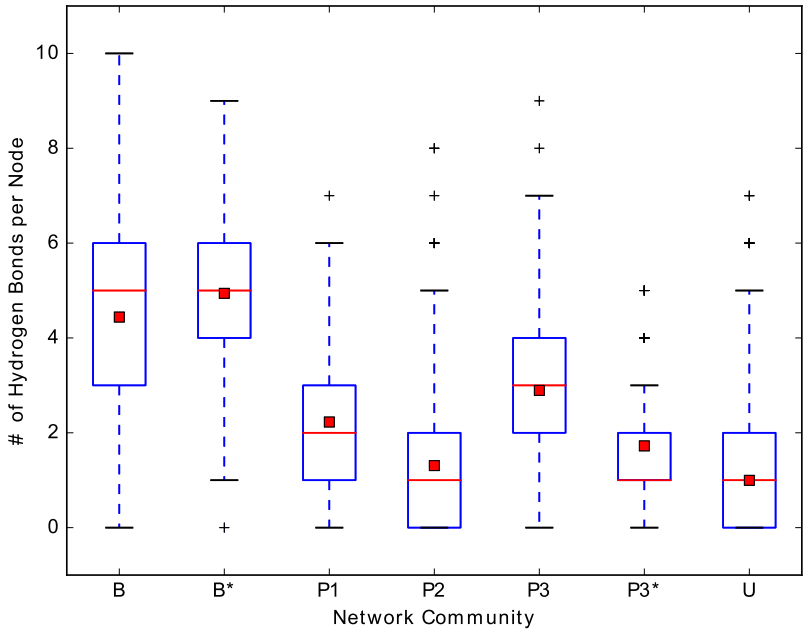

FIG. S4: **Box-plots of the number of hydrogen bonds per node by community.** The vertical axis gives the number of hydrogen bonds per node. The red square indicates the average for the community and the red line is the 50th percentile. The box bounds the 25th and 75th percentiles and the dashed line indicates the range with outliers are drawn as crosses.

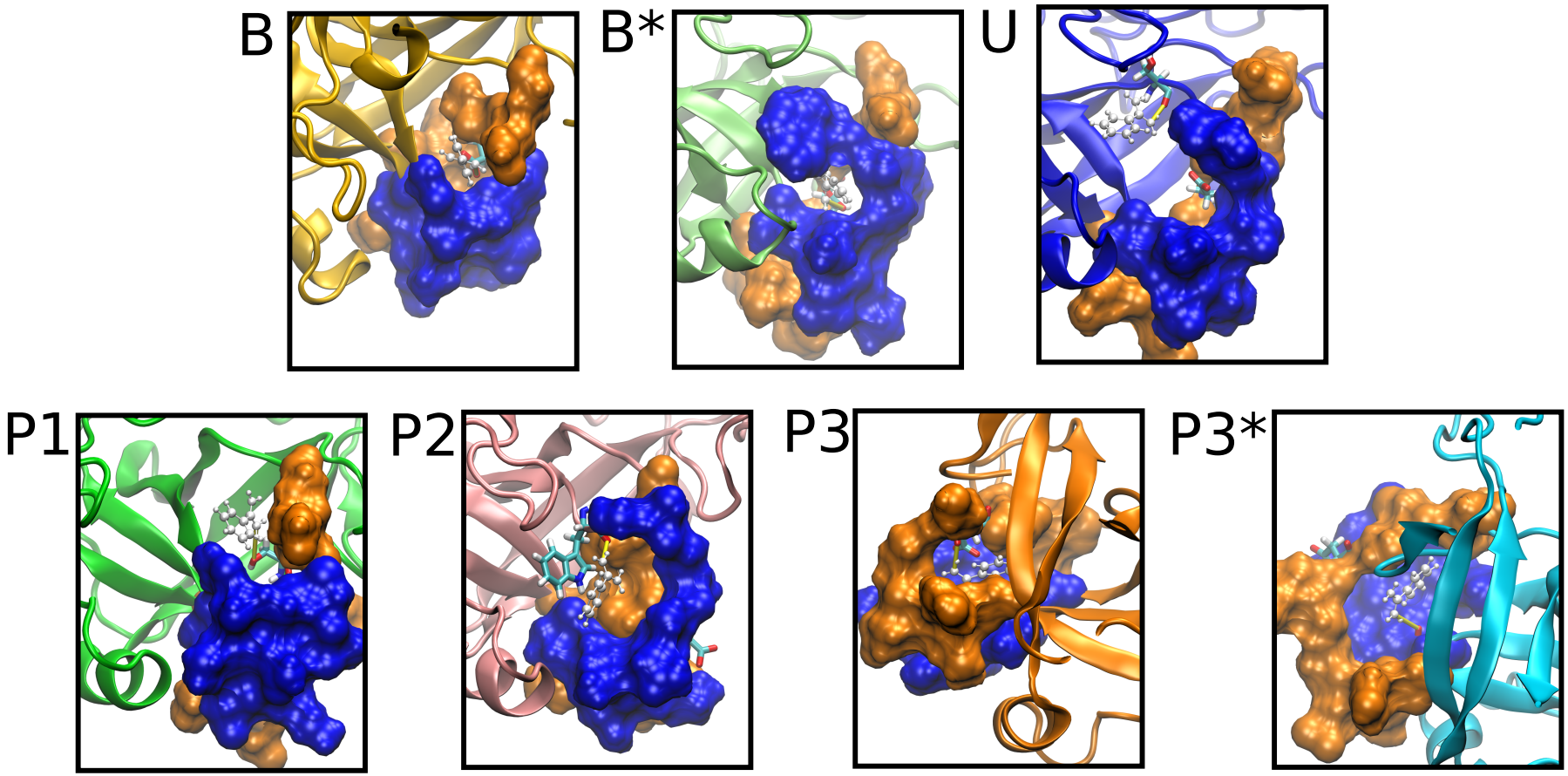

FIG. S5: **Binding poses of community representative structures.** The panels correspond to the representative structures that were chosen for each community based on the highest weighted conformation that contains the highest frequency hydrogen bond (shown as a yellow cylinder). The ligand is shown in white CPK representation and the hydrogen bonding partner is shown in licorice representation. The blue and orange loops in surface representation correspond to residue indices 209-218 and 179-190, and are the same as in Figs. 2 and S5. The native binding site residue ASP189 is shown in licorice representation as a point of reference. Structures for B, B*, U, P1, and P2 are shown from the binding site side of the protein (blue loop), while P3 and P3* show the alternative exit pathway through the orange loop.

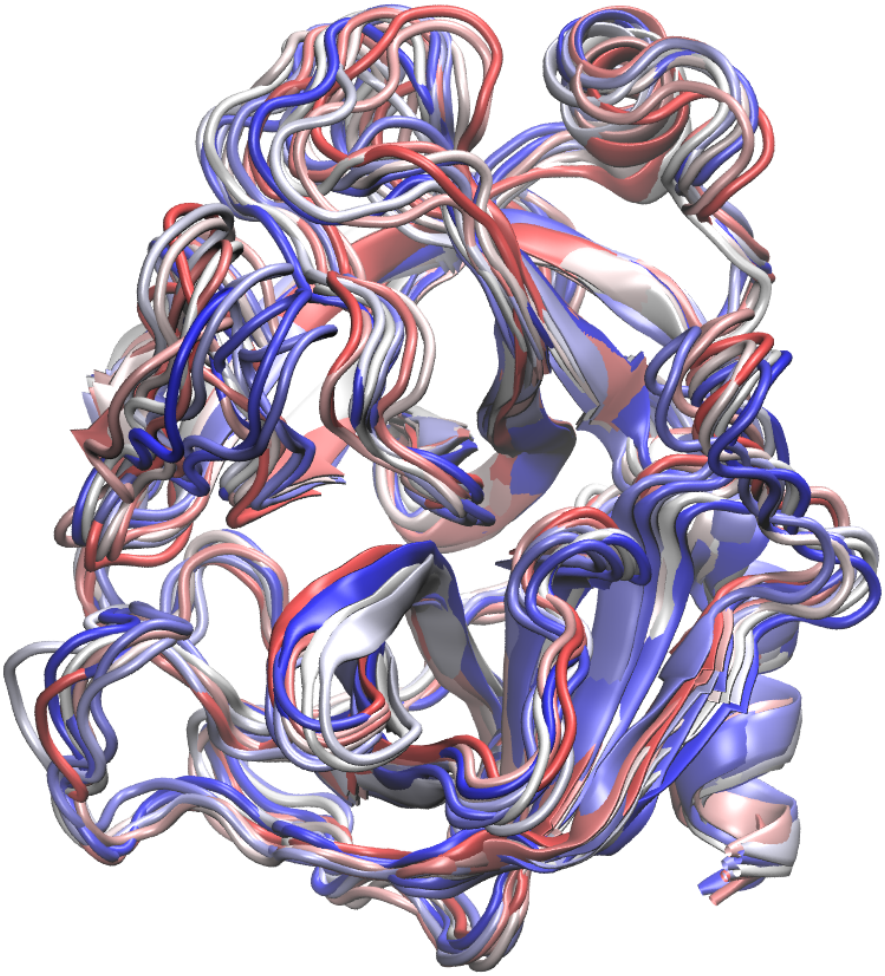

FIG. S6: **Set of 10 apo states** The set of apo states used for comparison of interaction energies in Figure 3 and Figure S7. These states are identified from the network as being the highest probability states with the minimum protein-ligand distance being greater than 5 Å.

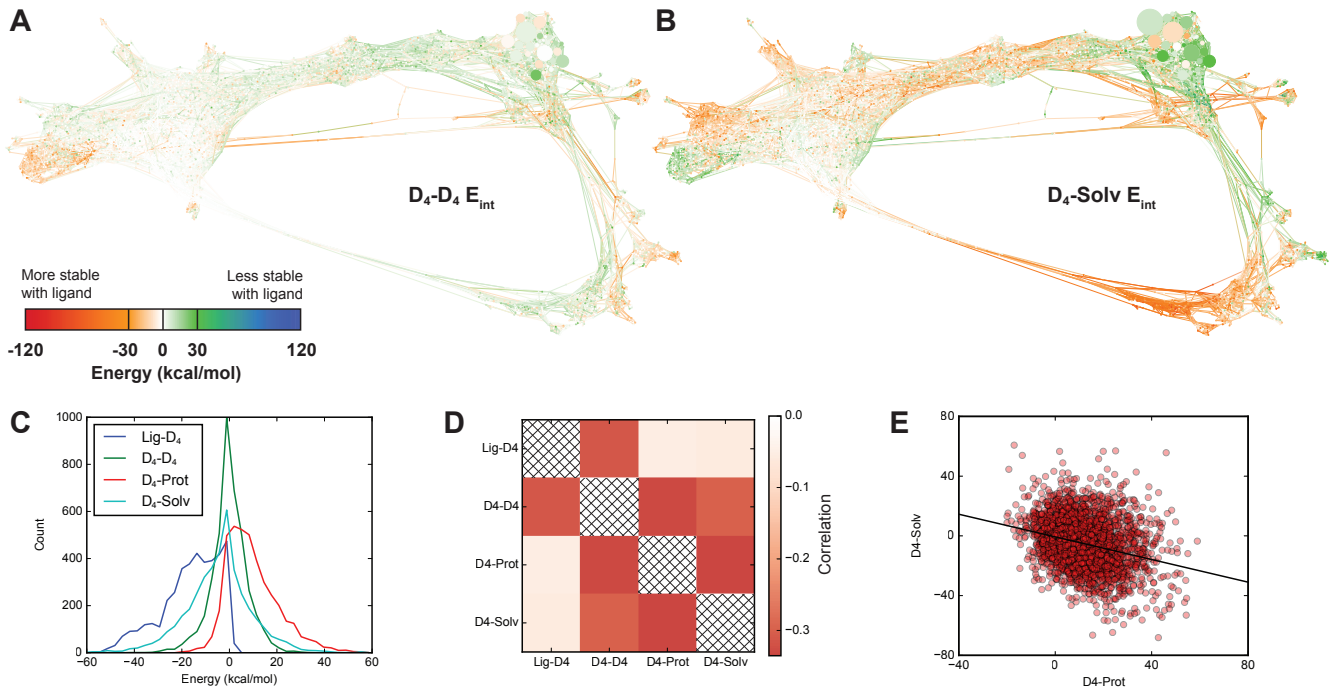

FIG. S7: **Extended analysis of D_4 interactions.** (A) A conformation space network colored by the difference in D_4-D_4 interaction energies between a given node and the set of apo nodes, as described in the main text. Green indicates that the D_4-D_4 interaction energy is more stabilizing in the apo structures, on average, while orange indicates the D_4-D_4 interaction energy is more stabilizing with the ligand bound. The results are mixed overall, and there is no clear trend that we observe. The color scale is the same as that used in Figure 3 in the main text. (B) A conformation space network colored by the difference in D_4-solvent interaction energies between a given node and the set of apo nodes. The near-crystallographic state (top right) shows that the ligand destabilizes D_4-solvent interactions, although the opposite is true in other places in the network, presenting a mixed picture. (C) Histograms for the four energy differences, comparing the distributions of interaction energies. D_4-D_4 and D_4-Solvent are roughly centered at zero, while Ligand-D_4 interactions are mostly stabilizing, and the ligand predominantly destabilizes D_4-Protein interactions, as described in the main text. (D) The correlation between the four energy differences is shown. Weak but significant (p-value $< 10^{-6}$) anti-correlations are found between four pairs of interactions, involving all four energy differences that we examined here. (E) A scatter plot showing the energy differences for each node in the network for the pair of quantities that showed the largest anti-correlation in (D): D_4-Protein and D_4-Solvent. This relationship implies that D_4-Protein interactions are likely to be formed at the expense of D_4-Solvent interactions, and vice-versa.

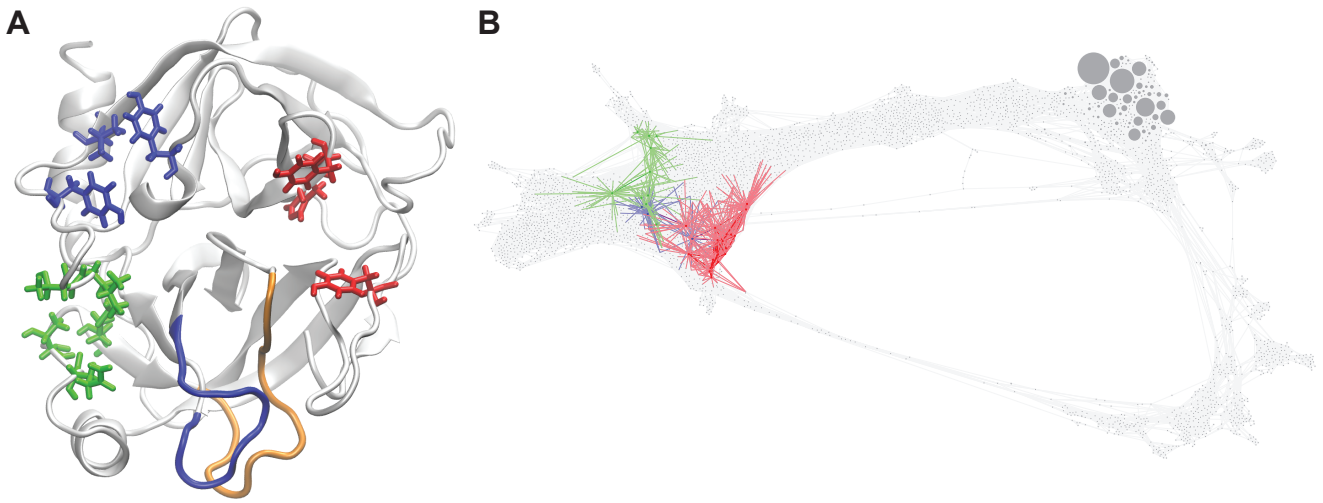

FIG. S8: **Comparison with states found along binding pathways by Buch et al.** (A) Residues in the binding sites of metastable states identified in a prior study by Buch et al [5] are highlighted. In blue are residues 57, 87 and 91, which comprise the S1 pocket. Red residues 37, 38 and 146 comprise the S2 pocket, and the S3 pocket involves interactions with the residues 95, 96, 170, 172 and 175, shown in green. (B) A conformation space network showing the locations of states S1, S2 and S3, as defined in Ref [5].

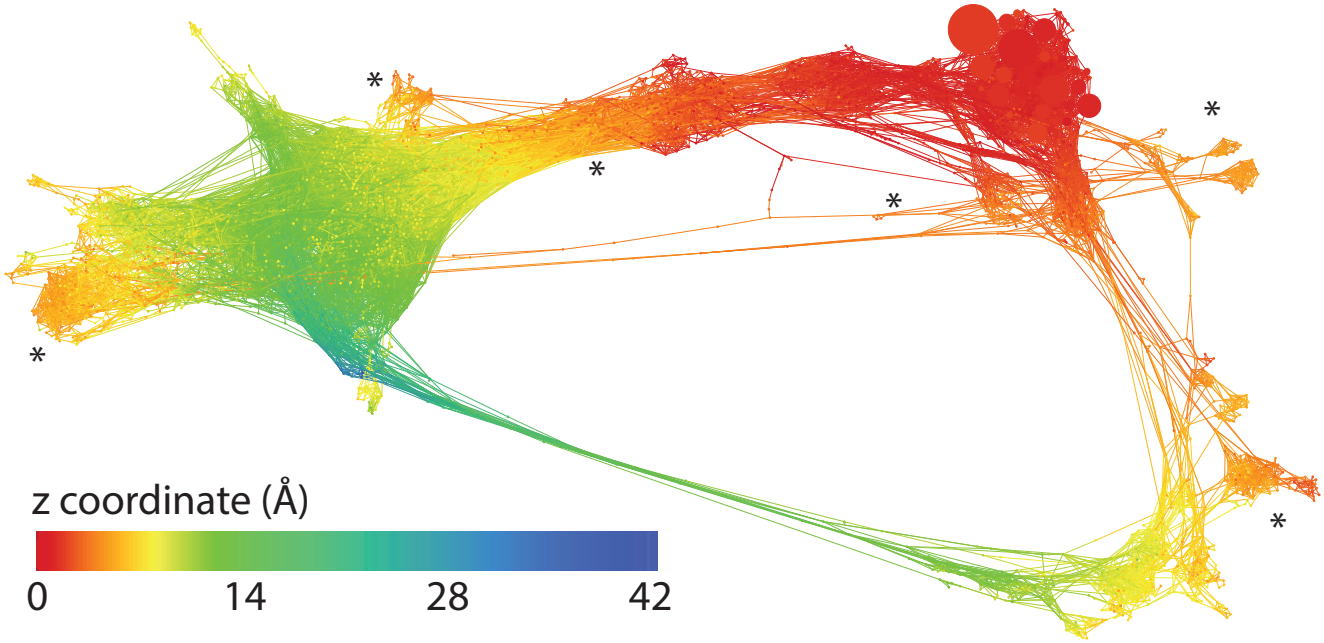

FIG. S9: **Projection onto the z coordinate used by Teo et al.** The progress coordinate z used by Teo et al [6] in the Adaptive Multilevel Splitting algorithm is computed for a representative structure from each state and is shown on the network. This is computed as the distance between the center of mass of a set of Cα atoms close to the binding site, and the center of mass of the benzamidine ligand. We use the same binding site residues as in the original work, which in our numbering (corresponding to PDBID 3PTB) are residues $186-192$, $206-209$, $211-213$, 219 and 220. The asterisks are used to draw attention to regions which have z values approximately equal to 5 Å.

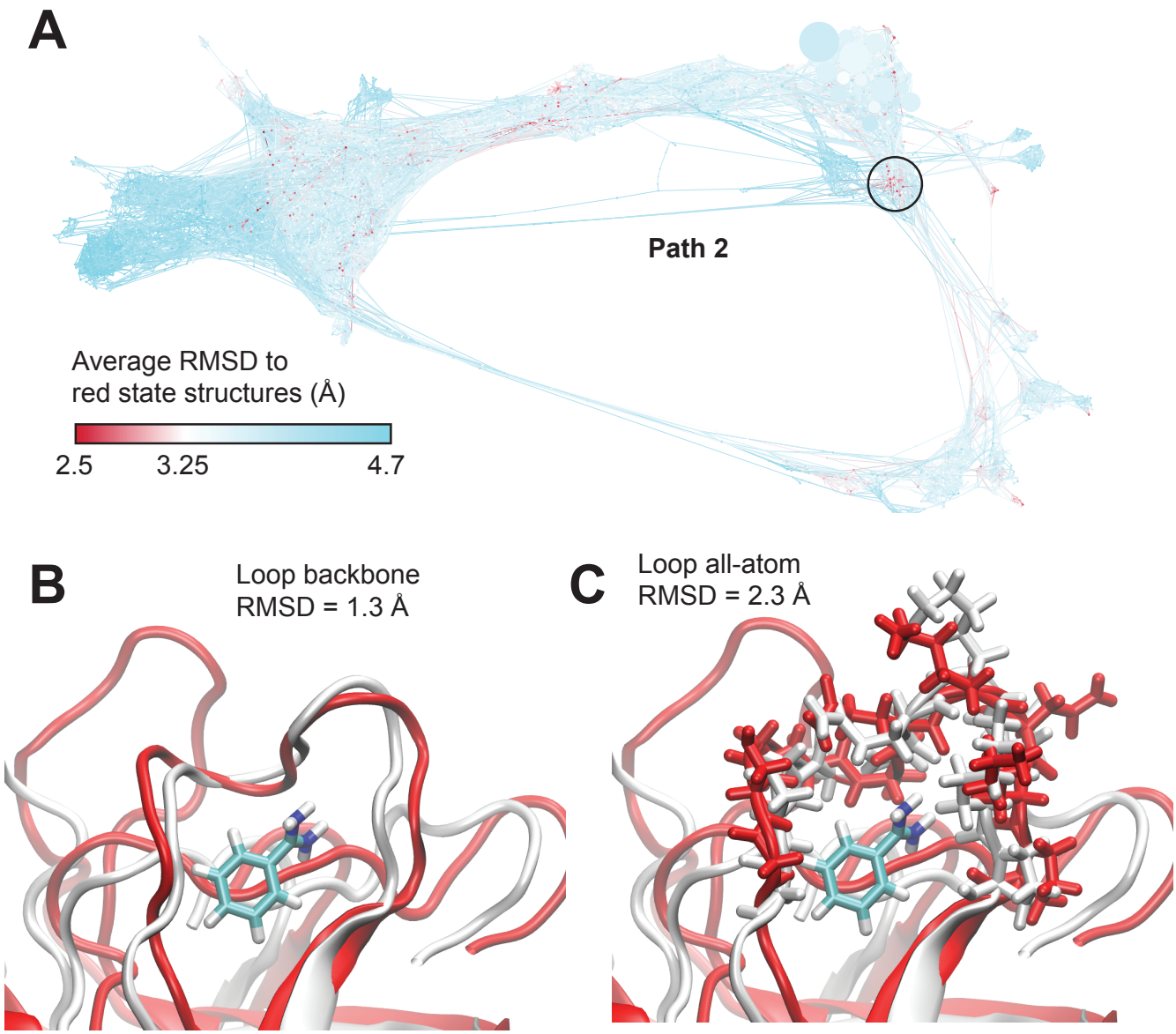

FIG. S10: **Comparison with representative "red" states from Plattner and Noé.** (A) A conformation space network showing the average RMSD to the loop regions (residues 209-218) of three representative red state structures from the work of Plattner and Noé [2]. The largest cluster of low RMSD regions are circled. They lie at the foot of Path 2, along which benzamidine exits through the 209-218 loop. (B) An alignment of one of the circled states (State 934) with the best matching representative red state structure. The backbone RMSD in the loop region is 1.3 Å. The benzamidine conformation from State 934 is shown for reference. (C) The same alignment with the sidechains in the loop region shown. In both structures the loop is open wide enough to allow for the exit of benzamidine through the newly formed exit channel. The RMSD including sidechains is 2.3 Å.