Supplementary Information: How robust is the ligand binding transition state?

Samik Bose,[†] Samuel D. Lotz,[†,‡] Indrajit Deb,[†,¶] Megan Shuck,[§] Kin Sing Stephen Lee,[§,‖,⊥] and Alex Dickson[*,†,#]

†*Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States*

‡*Current address: Examol Corporation, Cranberry Township, Pennsylvania 16066, United States*

¶*Current address: Drug Discovery Unit, School of Life Sciences, University of Dundee, Dundee, DD1 4HN, United Kingdom*

§*Department of Pharmacology and Toxicology, Michigan State University, East Lansing, Michigan 48824, United States*

‖*Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States*

⊥*Institute of Integrative Toxicology, Michigan State University, East Lansing, MI 48824, United States*

#*Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, Michigan 48824, United States*

E-mail: alexrd@msu.edu

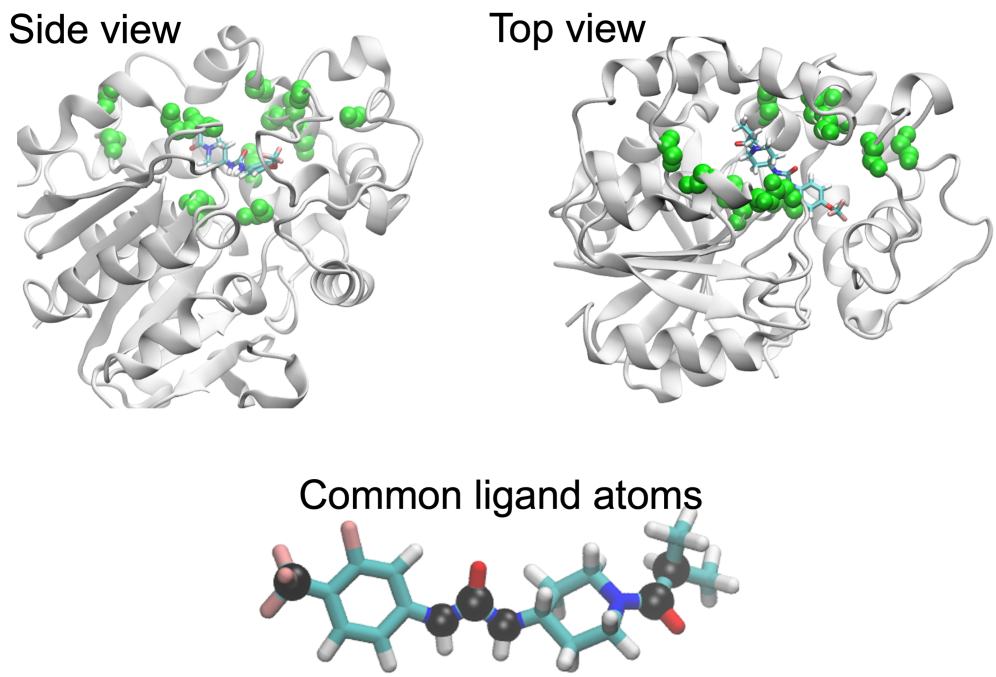

Figure S1: Atoms used to build the ligand-binding site distance vectors for the feature set. Four backbone atoms are used from each of the following 14 amino acids: 267PHE, 335ASP, 336TRP, 363ILE, 381PHE, 383TYR, 384GLN, 387PHE, 419MET, 428LEU, 466TYR, 498VAL, 499LEU, 503MET. These atoms are shown in green. The six common atoms of ligand scaffold are shown in black.

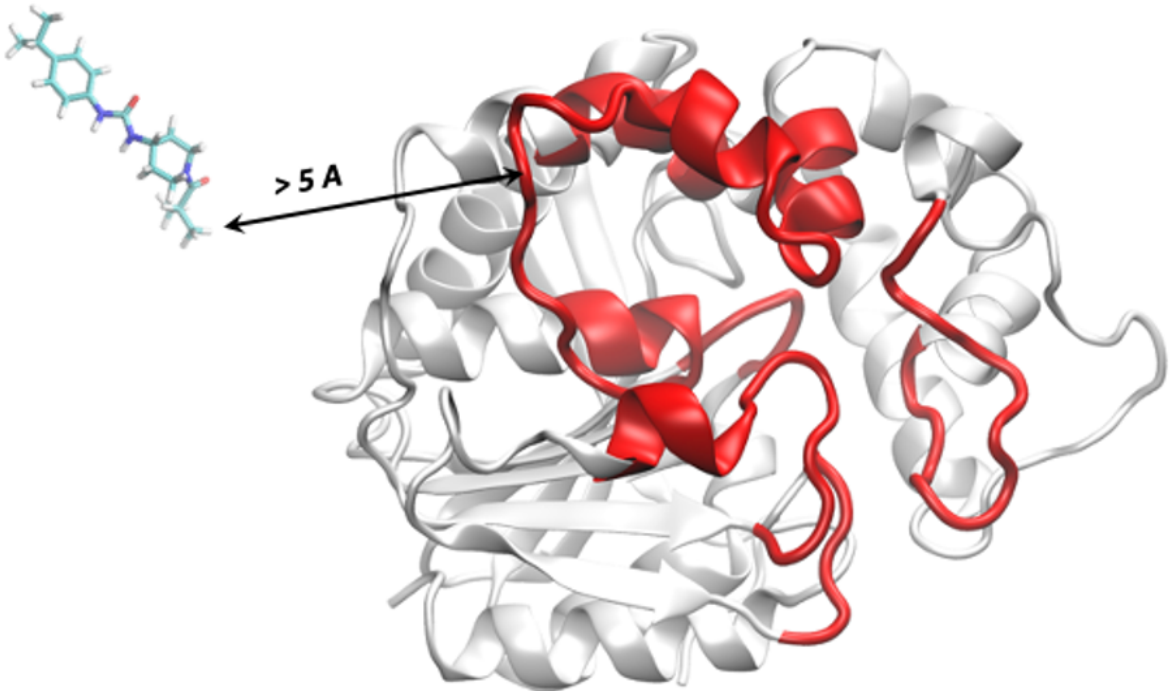

Figure S2: Illustration of a ligand unbound structure: A ligand (ligand 1) is considered to be unbound from the sEH when the minimum distance between the ligand and the binding region is more than 5 Å. The binding region is shown in red.

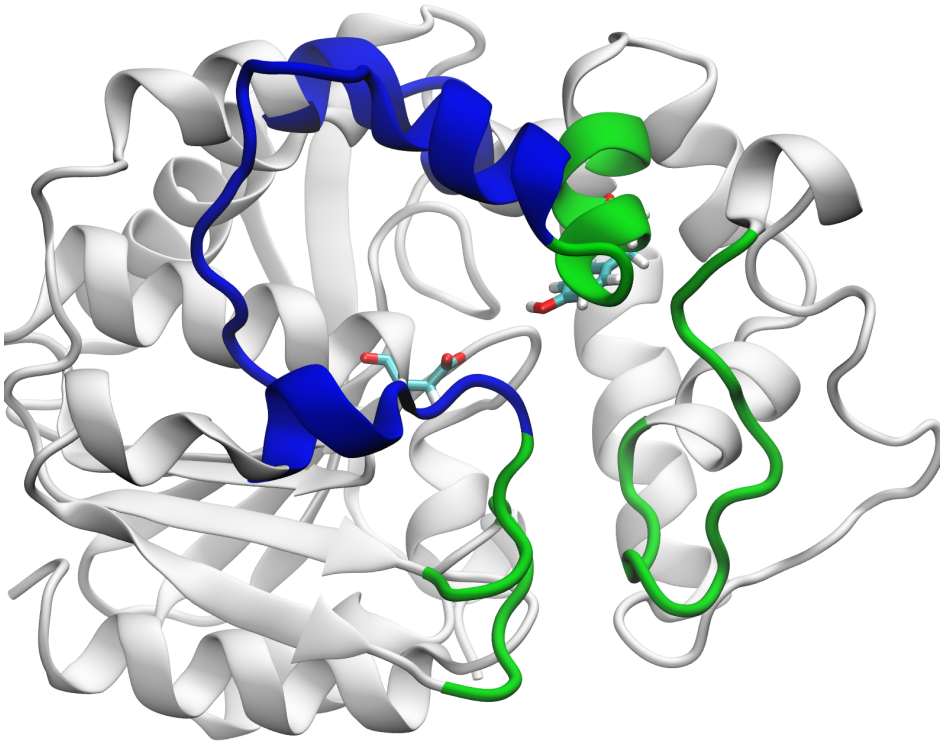

Figure S3: Overview of the sEH cavity: the right side is shown in green and the left side is shown in blue. The binding site residues Asp335 and Tyr383 are highlighted for reference.

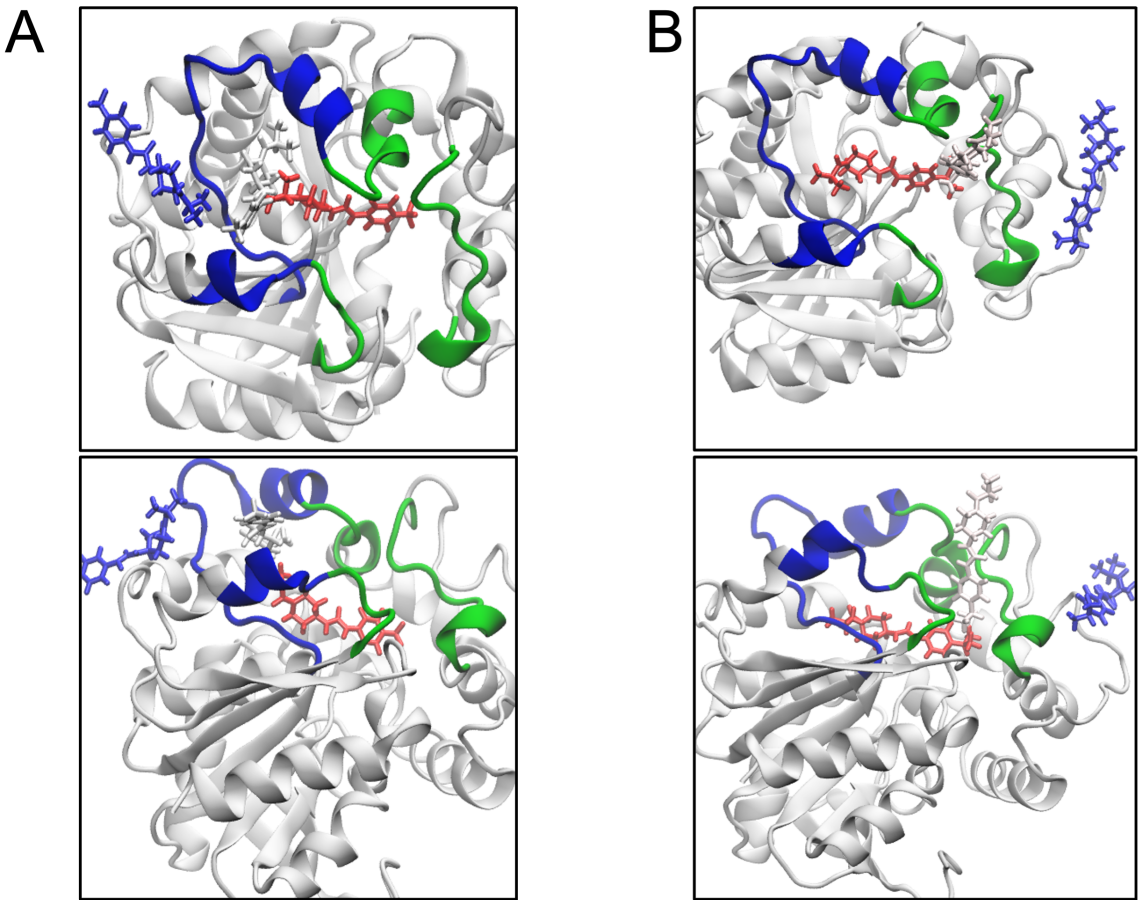

Figure S4: Unbinding transition pathways through the left and right side of the cavity. (A) Ligand 3 unbinding through left side, (B) Ligand 1 unbinding through right side. The top and bottom panels in both figures represent same poses in different points-of-view (top figs: Top POV and bottom figs: Side POV). The blue colored region in the protein indicates the left side of the cavity, while the green colored region indicates the right side. Ligand orientations in three different timesteps during the unbinding process are shown inside a single reference sEH structure. The bound pose is shown in red, the intermediate vertical pose is shown in white and the near-unbound pose is shown in blue.

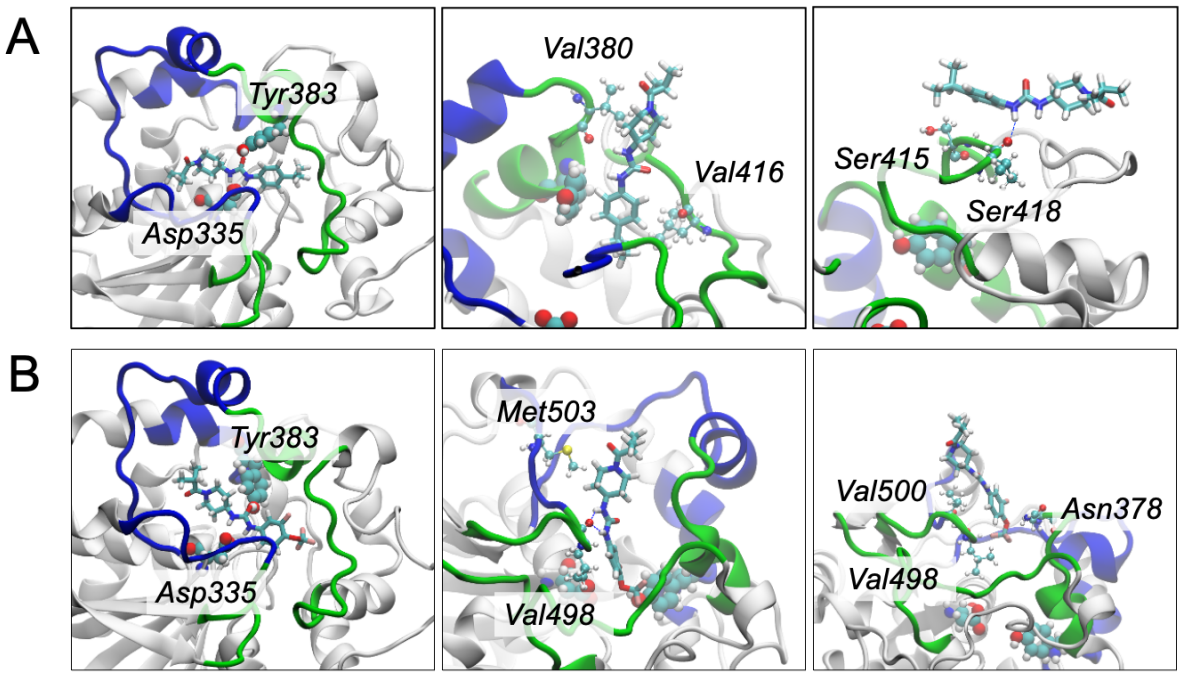

Figure S5: Most probable unbinding pathways for ligand 1 and 2. Three snapshots from the most probable ligand unbinding pathways of Ligand 1 (A) and Ligand 2 (B). In each panel, the ligands are shown in licorice while the amino acid residues within 2.5 Å of the ligands are depicted in CPK representation, with the binding site Asp335 and Tyr383 highlighted in VdW representation for reference. The leftmost snapshot in both (A) and (B) depicts the bound pose of the respective ligands. The middle snapshot is the characteristic vertical pose which has been discussed in detail in the main article. The rightmost snapshots are further along the transition pathway. Both ligand 1 and 2 have their most probabale unbinding pathways through the right side of the cavity (green region).

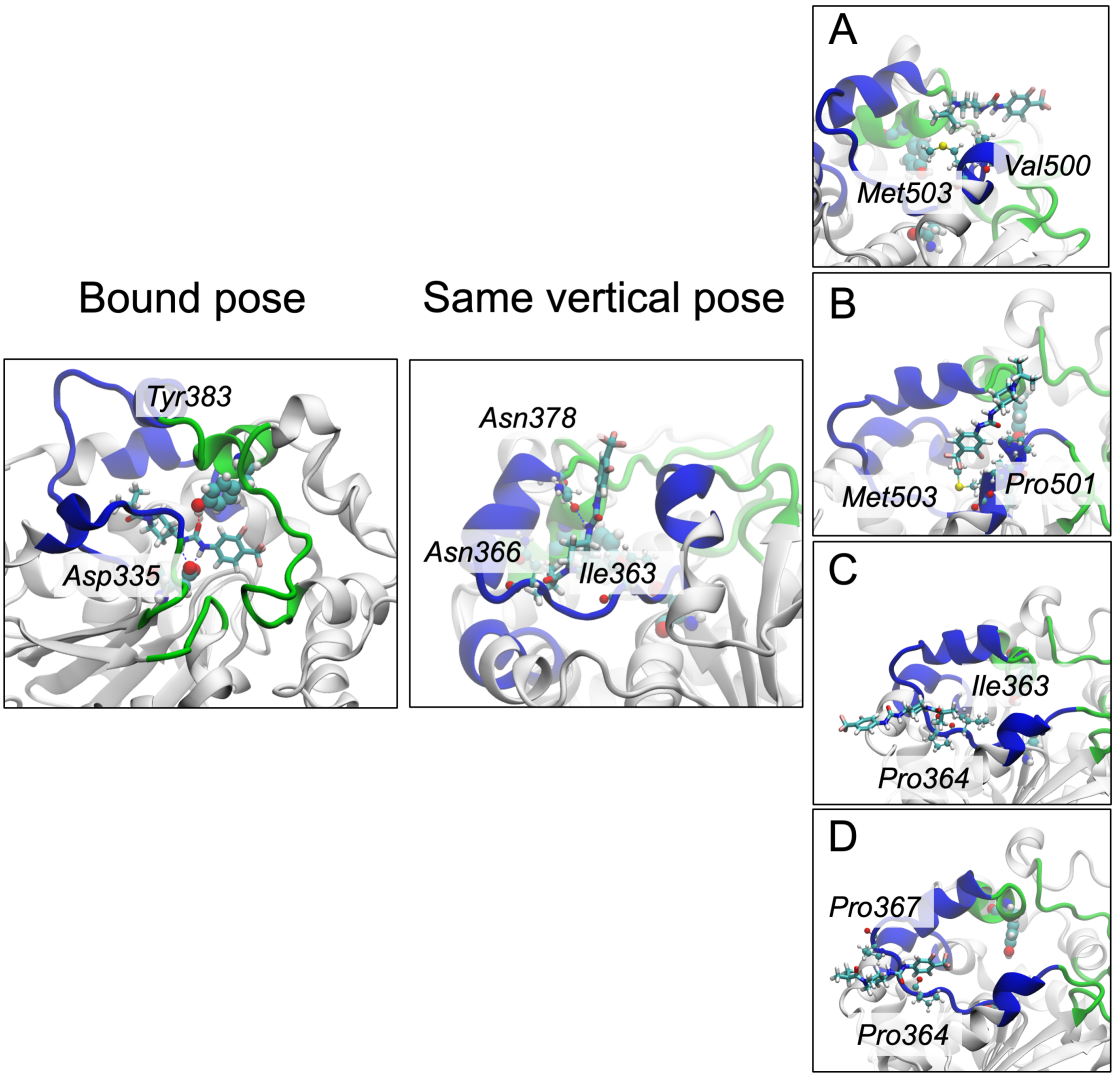

Figure S6: Unbinding pathways for ligand 3: all four transition paths involve same bound and vertical pose. The transition paths branch out from the same vertical pose and finally unbind with equal weights. Transition pathways highlighted in panel (A) and (B) access similar regions in the protein surface while (C) and (D) are more alike in protein-ligand interactions.

Table S1: Performance of Markov state models with different numbers of clusters in terms of agreement with experimental unbinding rates. The Spearman's rho and Kendall's tau are two measurements of rank ordering parameters. Positive values of these quantities indicate a higher predictive power than random ordering, which are only observed for the 1200-cluster MSM.

Number of clusters	Spearman's Rho	Kendall's Tau	RMSE
500	0.00	0.0	0.93
800	0.00	0.0	0.93
1200	0.3	0.2	0.94

Table S2: Average RMSD of TSEs obtained from MSMs with different cluster numbers. The RMSD is computed with respect to the reference bound pose.

Ligand	RMSD (nm) 500	RMSD (nm) 800	RMSD (nm) 1200
1	1.86	1.88	1.92
2	2.01	1.97	2.04
3	1.74	1.97	1.98
4	1.76	1.68	1.86
5	1.47	1.47	1.53

Table S3: Comparing the standard deviation within the TSE, obtained from MSMs with different cluster numbers. This standard deviation is equivalent to the average root mean fluctuation of the ligand atoms, calculated after alignment of the binding site residues.

Ligand	Standard deviation in nm, 500 clusters	Standard deviation in nm, 800 clusters	Standard deviation in nm, 1200 clusters
1	0.33	0.32	0.33
2	0.52	0.45	0.47
3	0.19	0.36	0.28
4	0.45	0.58	0.64
5	0.28	0.32	0.34

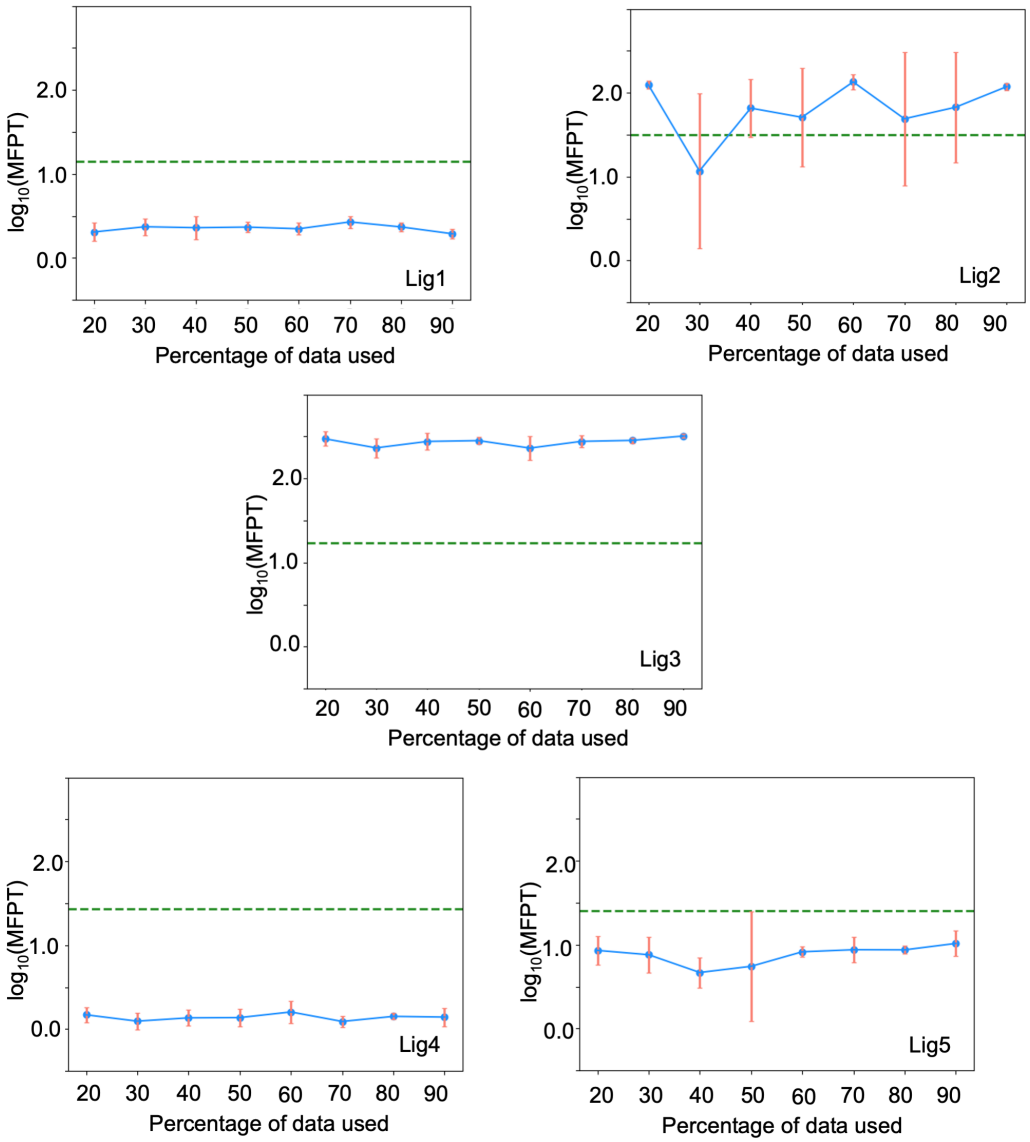

Figure S7: Robustness of the K-means models with respect to data reduction. Smaller subsets (from 90% to 20%) of simulation data are used to fit the K-means model for clustering. The X-axis in all the figures shows the percentage of total simulation data used to fit the K-means algorithm and Y-axis shows the $\log_{10}$(MFPT) in minutes. For each datapoint, 5 iterations are carried out with different randomly chosen subsets to estimate the standard deviations (error-bars in red) and average (blue circles). The full dataset is then used to determine the transition matrix in each case. The experimental residence time is shown by the green dashed line. Similar MFPTs are observed for all cases with as little as 20% of the simulation data used. However, some smaller subsets are observed to have significantly higher standard deviation.

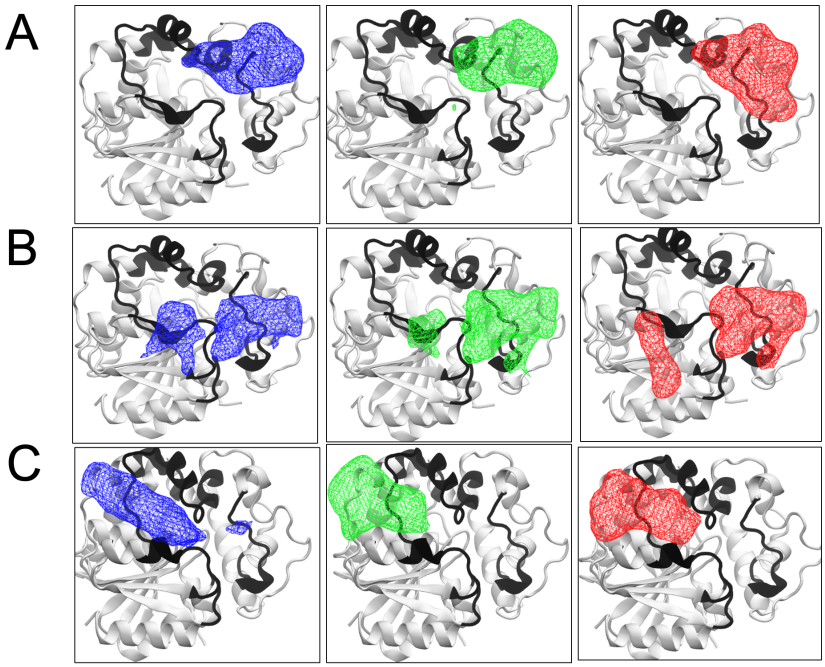

Figure S8: Spatial densities of transition state ensembles obtained from MSMs built with different numbers of clusters. TSE densities of ligand 1 (A), ligand 2 (B) and ligand 3 (C) with cluster numbers of 500 (in blue), 800 (in green) and 1200 (in red). The densities are qualitatively similar in all cases for each ligand.

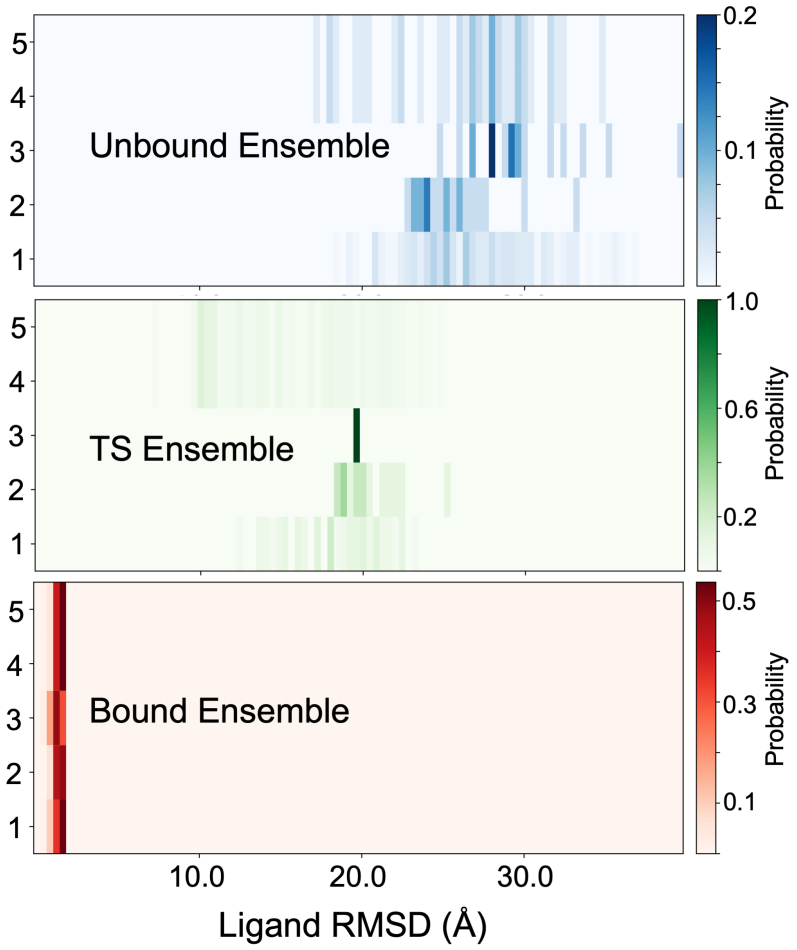

Figure S9: Probability distributions of average ligand RMSDs for clusters corresponding to bound (bottom), TS (middle) and unbound ensembles (top). This gives a clear indication that TSEs are closer to the unbound compared to bound ensembles.

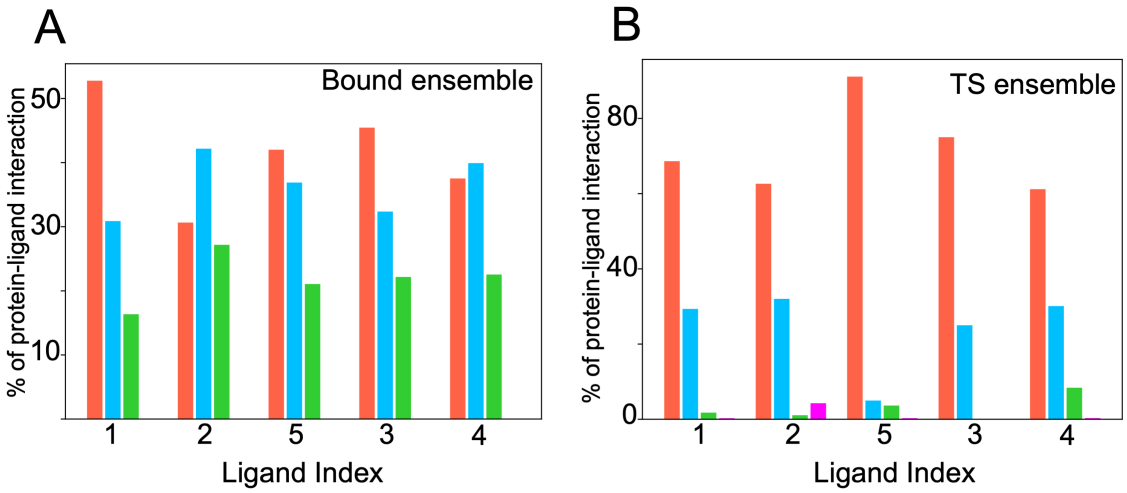

Figure S10: Breakdown of protein-ligand interactions by amino acid type for each ligand. The red bars in each plot show the percentage of interactions with non-polar residues, blue bars show polar residues, green bars show the acidic residues and magenta bars show basic residues. For each ligand we observe a pronounced shift towards a higher percentage of non-polar residues.

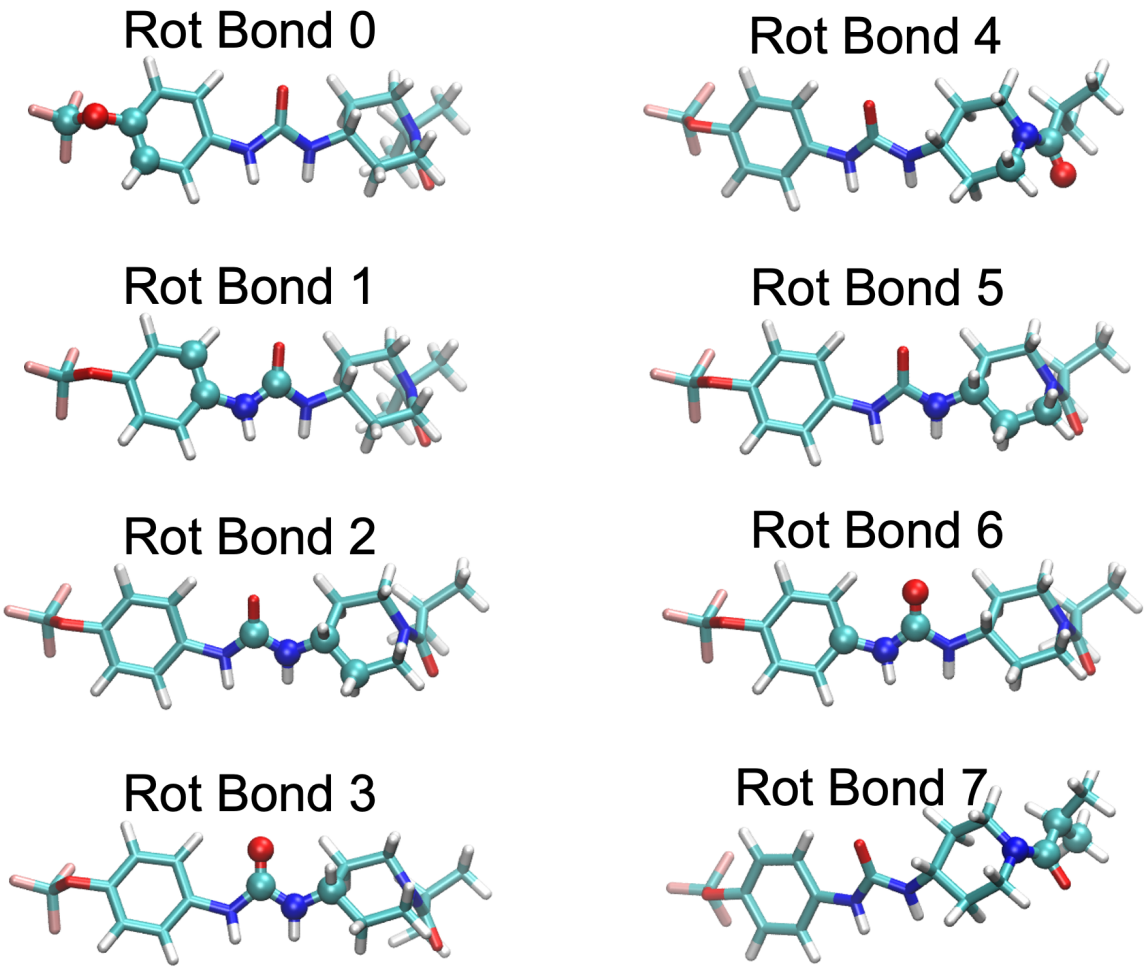

Figure S11: Visual representation of eight common and structurally important rotatable bonds. The atoms making the rotatable bonds are highlighted in VdW scheme.

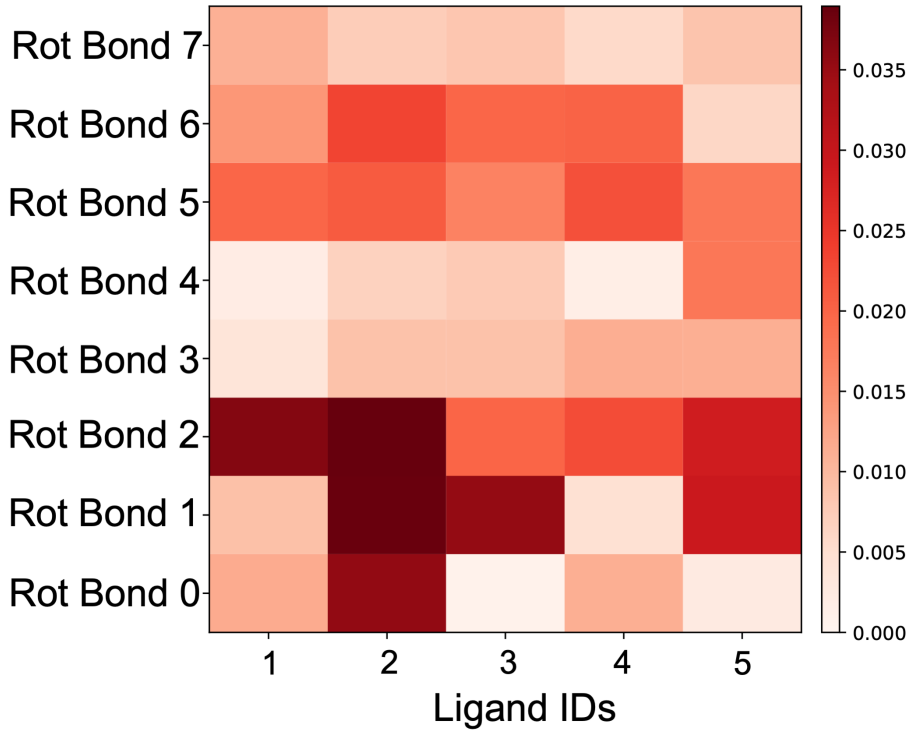

Figure S12: Wasserstein distances between the bound and TSE across all the rotatable bonds for all ligands illustrated as a heatmap. A higher Wasserstein distance indicates a larger difference between the probability distribution of that angle between the bound ensemble and the TSE.

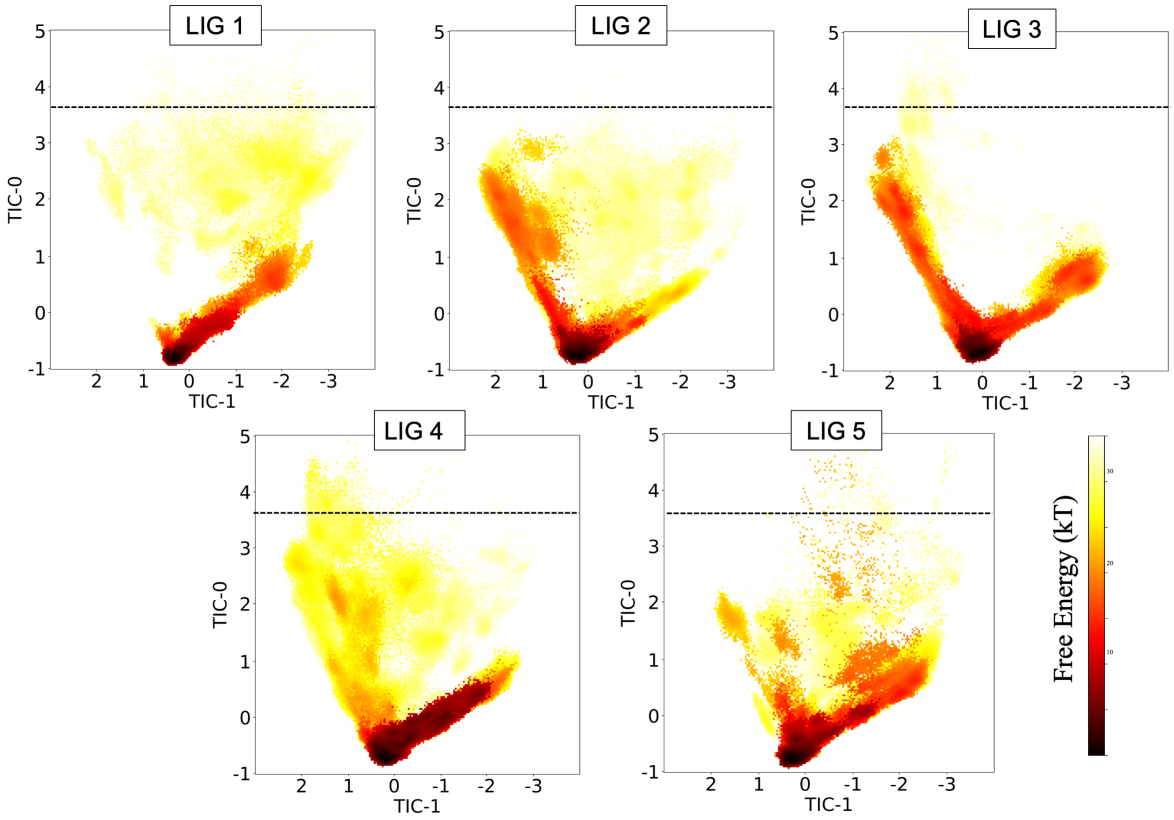

Figure S13: Projections of the unbinding ensemble "free energy" (computed as the negative logorithm of the probability) projected onto tICA components TIC-0 and TIC-1. A single set of tICA components was obtained by training the tICA model on a merged dataset involving data from all of the ligands. The first two tICA components are found to seprate out the bound and unbound ensemble, as well as the cavity preferences. TIC-0 differentiates between the bound (~ 0) and unbound ensemble (> 3.2). TIC-1 separates out the right side of the cavity (< -1.5) and left side of the cavity (> 1.5). The dashed horizontal line indicates the average TIC-1 value of structures in the unbound state.