PNAS
www.pnas.org

Supplementary Information for

Energy Penalties Enhance Flexible Receptor Docking in a Model Cavity

Anna S. Kamenik[1,2,a], Isha Singh[2,a], Parnian Lak[2], Trent E. Balius[2,3,*], Klaus R. Liedl[1,*], Brian K. Shoichet[2,*]

[1] Institute of General, Inorganic and Theoretical Chemistry, Center for Molecular Biosciences Innsbruck, University of Innsbruck, 6020 Innsbruck, Austria

[2] Department of Pharmaceutical Chemistry, University of California, San Francisco, 94158 San Francisco, CA, USA

[3] Current address: National Cancer Institute RAS Initiative, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, MD 21702

[a] These authors contributed equally.

* Corresponding authors.

Trent E. Balius, trent.balius@nih.gov

Klaus R. Liedl, klaus.liedl@uibk.ac.at, 0043-0512-50757100

Brian K. Shoichet, bshoichet@gmail.com, 01-415-541-4126

Supplementary Materials and Methods

Molecular Dynamics Simulation Setup and Analysis

For exhaustive exploration of L99A's conformational space accelerated molecular dynamics (aMD) simulations were performed. The simulations were started from an open state structure (PDB: 4W59) to ensure sufficient sampling of this high-energy area in the conformational space of L99A. We removed the ligand n-hexylbenzene from the cavity, as well as all crystallization agent and water molecules. With the LEaP module implemented in AMBER14 (70), we created topologies and initial coordinate files using the AMBER 99SB-ILDN force field (74). The protein was solvated with a truncated octahedral box of TIP3P water molecules(75) and a minimum wall distance of 12 Å. We further apply an exhaustive equilibration protocol to relax the system in an NPT ensemble prior to productive simulation runs (76). The aMD specific parameters, i.e., threshold energy and boosting parameter, were determined from the final short cMD simulation as described previously (**SI Table S1**) (28, 69).

To maximize computational efficiency all simulations were performed with the GPU implementation of AMBER14s pmemd module (77). Hence, the particle-mesh Ewald (PME) method was used to treat long ranging electrostatic interactions and a non-bonded cutoff of 8 Å (78). We apply a Langevin thermostat (79) with a collision frequency of 2 ps-1 to maintain a simulation temperature of 300K and a Berendsen barostat (80) with a relaxation time of 2 ps to simulate constant atmospheric pressure. To allow for a the timestep of 2 fs all bonds involving hydrogen atoms were constraint using the SHAKE algorithm (81).

We performed five replica aMD simulations of 100 ns length starting from the same coordinates with different velocities. The accumulated simulation time of 500 ns was then clustered using the hierarchical agglomerative clustering implemented in cpptraj (71) using average linkage and a cutoff distance of 0.8 Å. We assigned the resulting clusters to the open, intermediate or closed cavity state based on structural similarity of the representative structure. In **SI Figure S1** we visualize the opening and closing of the cavity color-coded according to the state-populations derived from the clustering of the combined trajectory. The cluster populations where then reweighted with an approximation of the exponential term using a Maclaurin series of the 10th order (82). To estimate the uncertainty of these reweighted state populations we applied the same procedure for each 100 ns aMD trajectory individually (**SI Table S2**).

Furthermore, we performed extensive sampling with cMD simulations to derive unbiased thermodynamic and kinetic information with the aid of an MSM. Here, most cMD simulations were seeded from representative structures of the aMD ensemble, regardless of the conformational state of the binding cavity. Several cMD simulations were also started from an open state crystal structure, as described above, to increase the statistical robustness. In total we accumulated 7.75 µs of simulation time.

From the unbiased cMD simulation data we constructed an MSM using PyEMMA 2.5.7.(72) Based on structural characteristics in the ligand-bound crystal structures, we selected the distance between the buried residue ALA99 and the Cα atoms of the F-helix as well as the rotation of residue V112 (chi1) as input features a time-lagged independent component analysis (TICA)(83) with a lag-time of 0.5 ns. After projecting the structural information, we divide the TICA space into 100 microstates using a k-means clustering. Based on the discretized trajectory we then built a Bayesian MSM (62) with a lag time of 0.5

ns. We perform a Perron-cluster cluster analysis+ (PCCA+) to coarse-grain our model into four states, as deduced from gap between successive eigenvalues (**SI Figure S2**). A Chapman-Kolmogorov test displaying the robustness of the model is shown in **SI Figure S3**. A visualization of the thermodynamics and kinetics calculated from the MSM is depicted in the **SI Figure S4**.

Flexible Receptor Docking

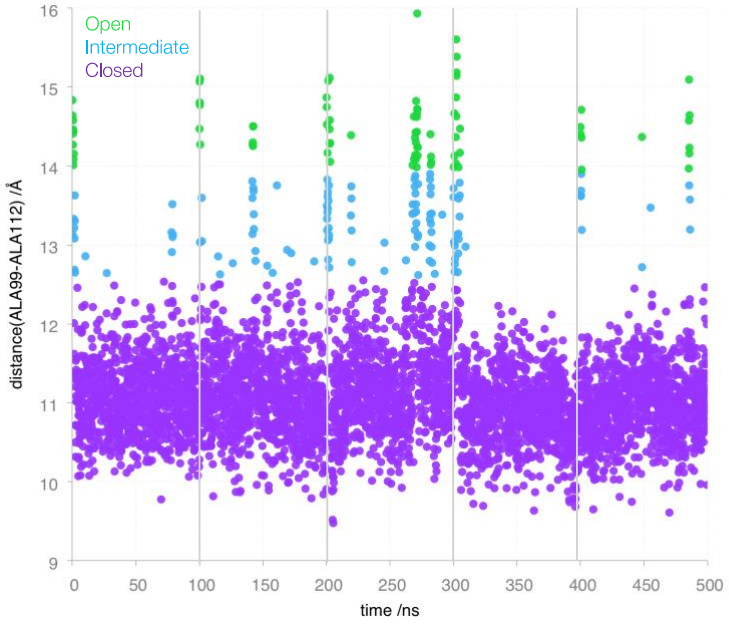
The flexible receptor docking protocol, scripts, and programs implemented in DOCK3.7 were used to calculate and score ligand poses with each receptor conformation (4). The crystal structure of n-butyl-benzene bound to L99A (PDB 4W57) was prepared using REDUCE (84) to add hydrogens. For the orientation of docked ligands in the binding site, the crystallographic ligand atoms were converted into "spheres", which are pseudoatoms that are used to orient new docked ligands (85). The generation of these spheres was accomplished as implemented in the Blastermaster script distributed with DOCK3.7 using the SPHGEN program. Using QNIFFT (86), electrostatic potentials were calculated by solving the Poisson-Boltzmann equation, and stored on a lattice for scoring look-up. Van der Waals interactions were calculated using CHEMGRID (87), and ligand desolvation grids were calculated using SOLVMAP (65).

Retrospective testing was based on 68 known ligands (32, 34, 35, 37, 50, 88, 89), for which we generated property-matched decoys using the DUD-E (90) workflow. For prospective screening, we docked a library of 985,201 molecules from a subset of ZINC15 (91) with a cLogP up to 4 and a molecular weight up to 300 Da.

Supplementary Figures and Tables

Supplementary Table S 1. Parameters applied in aMD simulations as calculated from average energies, particles and residue number.

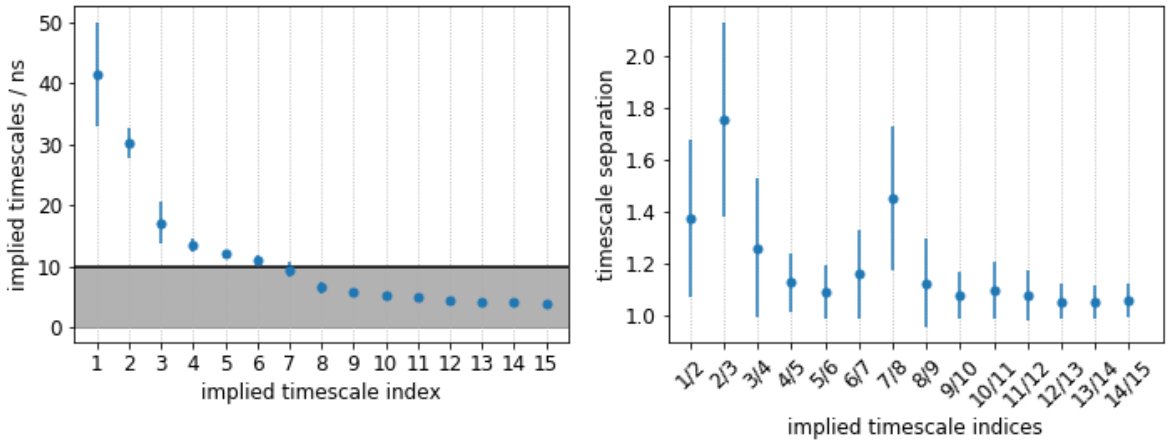
E_{dihed}	2382
α_{dihed}	66
E_{tot}	-80587
α_{tot}	4189



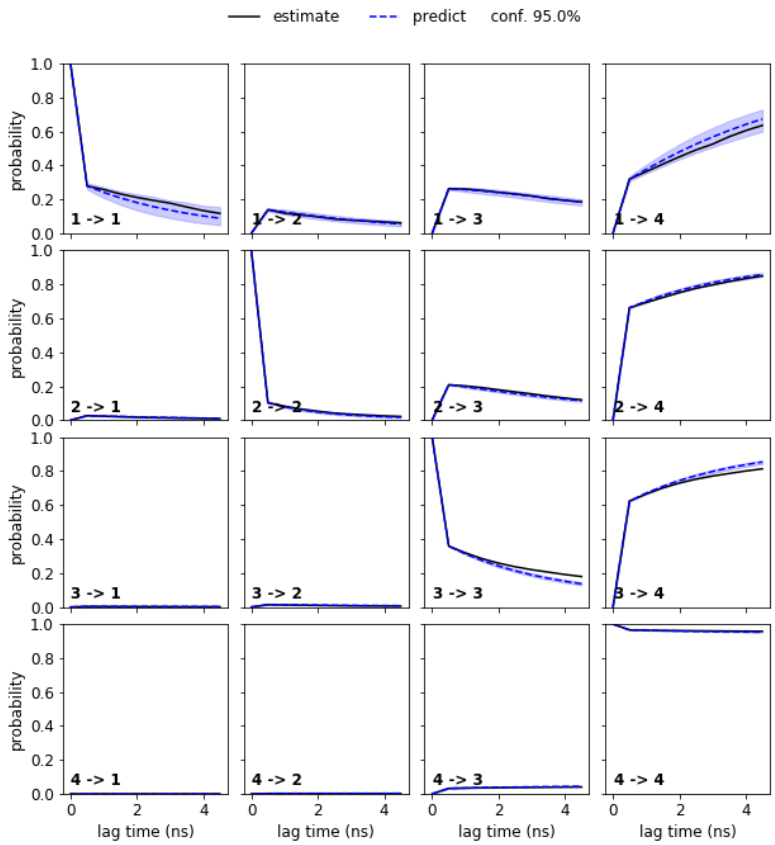
Supplementary Figure S 1. Cavity opening and closing during aMD simulations. The distance between the bottom of the cavity (ALA99) and ALA112 in the F-helix is colored according to the conformational state as defined by the clustering.

Supplementary Table S 2. Error estimation of state populations. The reweighted population of closed, intermediate and open state was calculated from five 100 ns aMD trajectories individually to estimate the associated uncertainty.

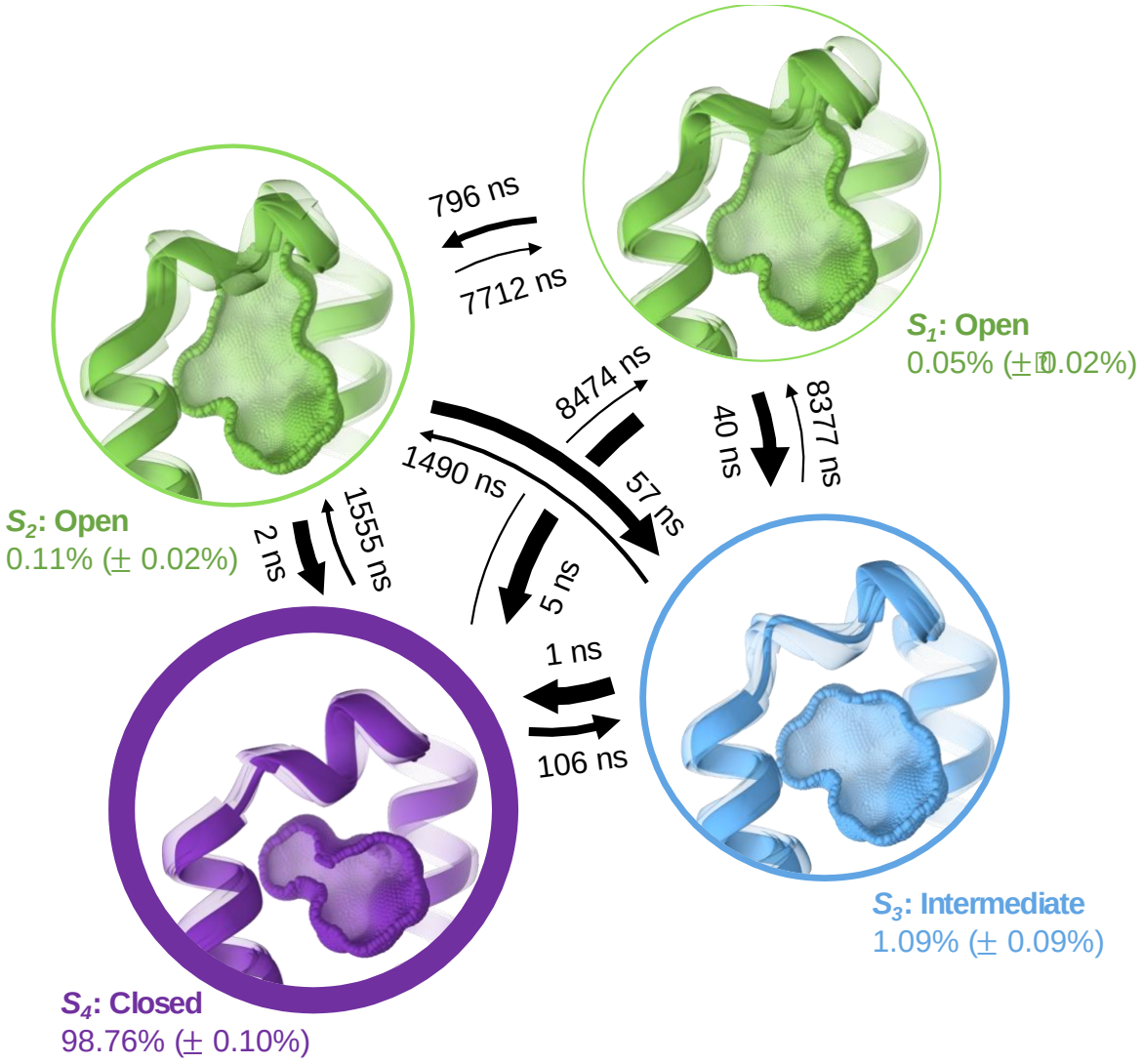
State	aMD1	aMD2	aMD3	aMD4	aMD5	Average	Standard Deviation
Closed	96.7%	97.6%	95.4%	97.3%	98.7%	97.1%	± 1.2%
Intermediate	2.8%	1.8%	3.8%	1.4%	1.2%	2.2%	± 1.1%
Open	0.5%	0.6%	0.9%	1.2%	0.1%	0.7%	± 0.4%



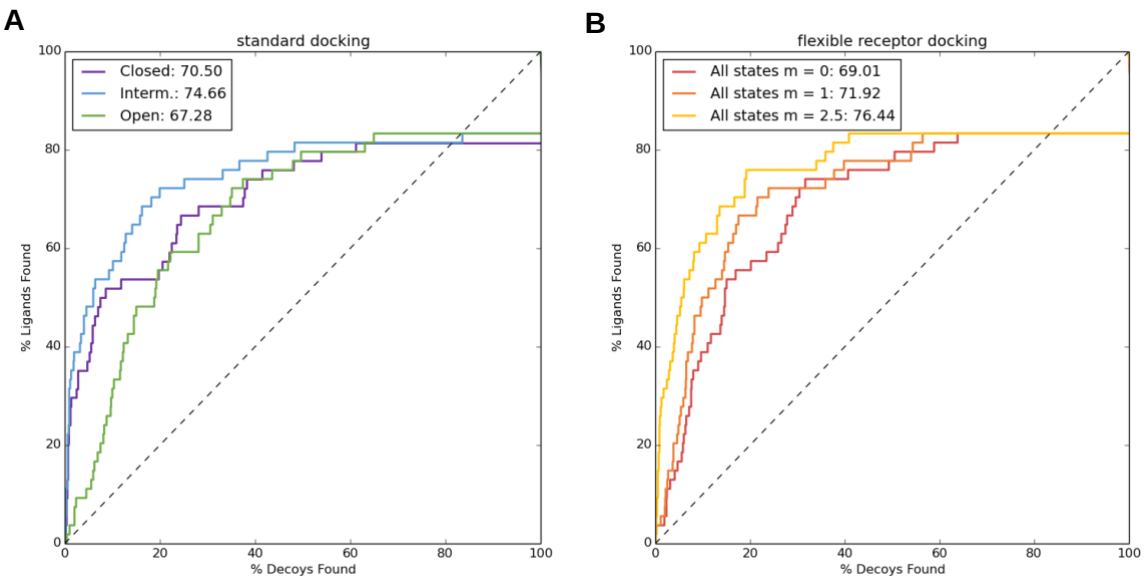
Supplementary Figure S 2. Successive Eigenvalues of the TICA suggesting four macrostates for the PCCA+ coarse-graining of the MSM.



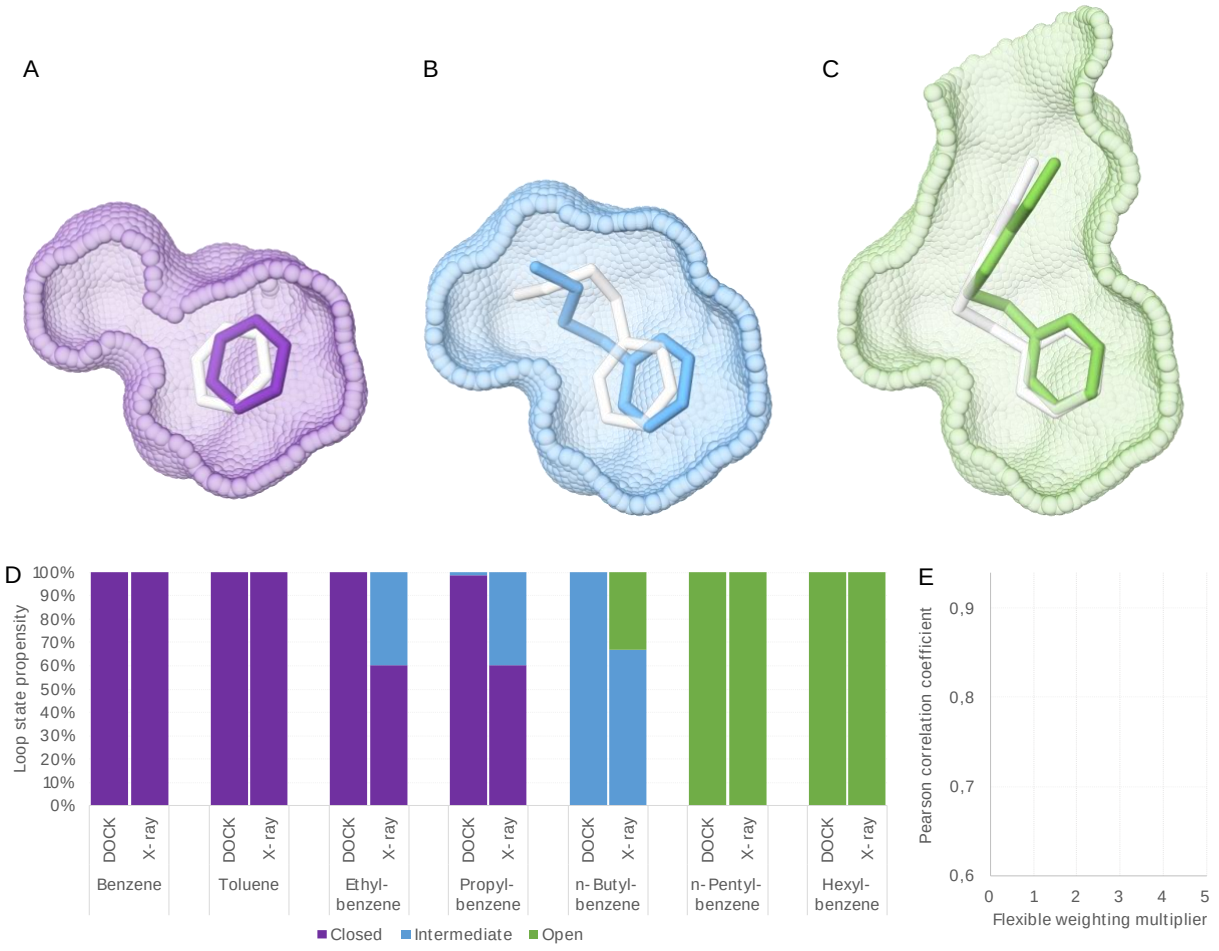
Supplementary Figure S 3. Chapman-Kolmogorov test of the MSM. The test is depicting good agreement between predicted results using the applied lag time of 0.5 ns (dashed line) and estimated at lag times up to 5ns (continuous line).



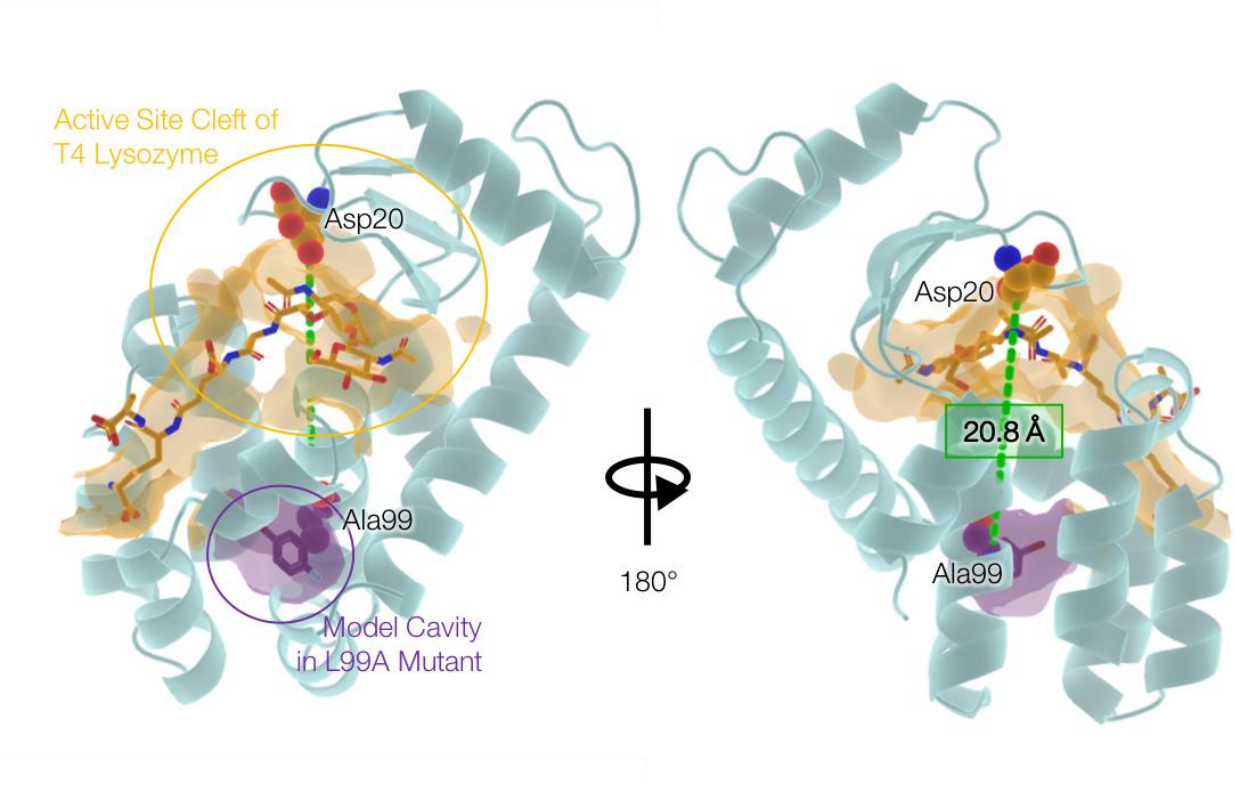
Supplementary Figure S 4. Distinct conformational states of the ligand binding site in L99A. A Bayesian Markov state model of the apo protein ensemble estimates the population of each state in the absence of a ligand. The MSM states S1 (0.05 %) and S2 (0.11 %) both resemble the open state, S3 (1.09 %) represents an intermediate conformational state and the closed conformation is characterized by S4 (98.76 %).



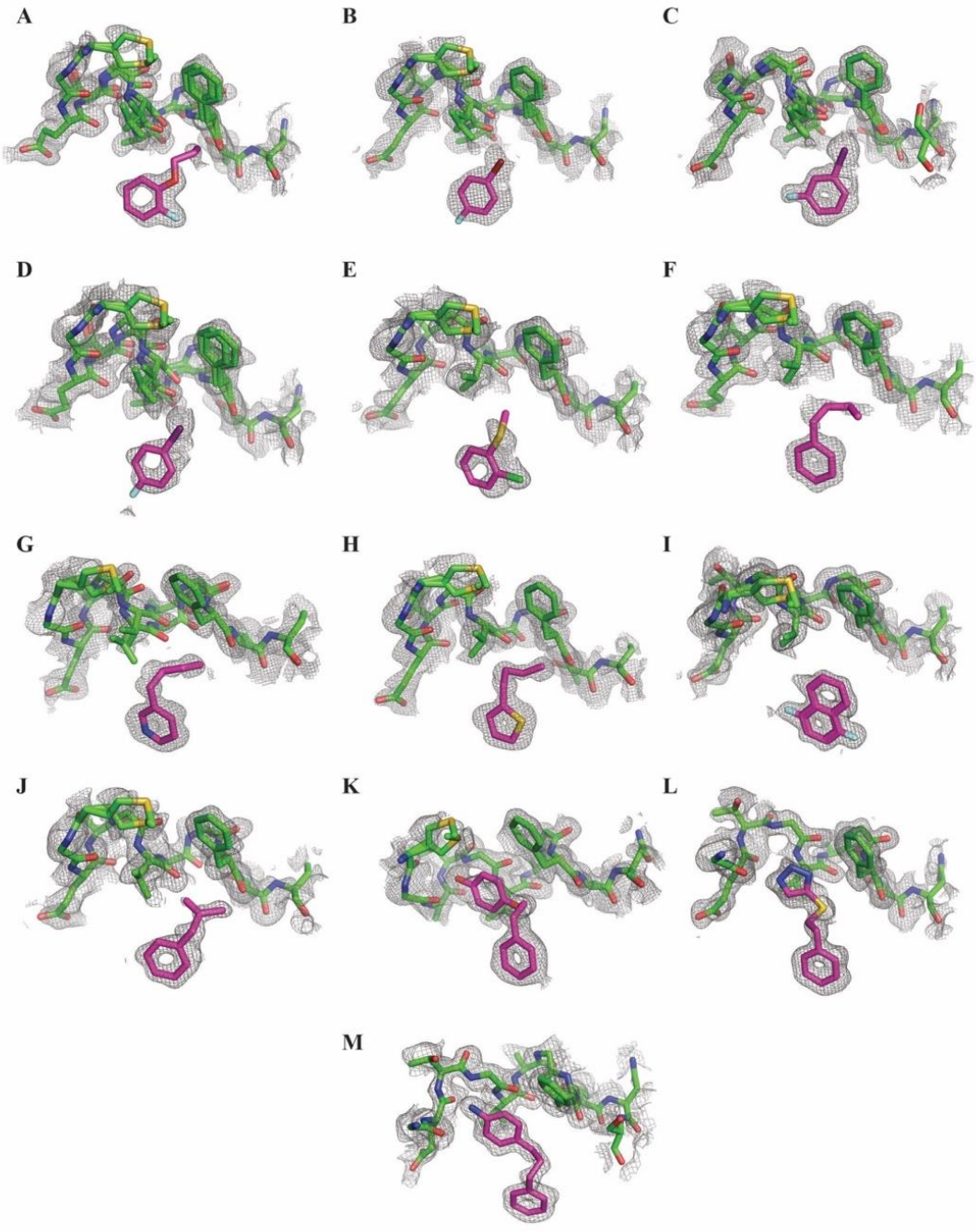
Supplementary Figure S 5. Ligand enrichment in top-scoring poses. The enrichment of ligands over decoys in the top-scoring poses are depicted as ROC curves and quantified via the AUC for (A) standard docking and (B) flexible receptor docking.



Supplementary Figure S 6. Crystallographic and predicted geometries of known ligands and conformational preferences. Docking poses for (A) Benzene, a closed state binder (purple), (B) n-Butylbenzene, an intermediate state binder (blue), and (C) n-Hexylbenzene, an open state binder (green) in overlay with the crystallographic ligand geometries (white). (D) Conformational preferences predicted from DOCK compared to crystallographic occupancies for a homologous ligand series. (E) Pearson correlation coefficient of crystallographic occupancies and state probabilities calculated from the DOCK score as function of the flexible weighting multiplier.



Supplementary Figure S 7. Enzymatic active site cleft of T4 Lysozyme (orange) and model cavity in L99A mutant (purple) are structurally far apart. The distance between the catalytic Asp20 and the mutated Ala99 measures 20.8 Å. For visualization of the distinct sites, the structure of substrate (carbons in gold) bound T4 Lysozyme (PDB 148L) is overlaid with a ligand (carbons in purple) bound structure of its L99A mutant (PDB 1LOC).



Supplementary Figure S 8. Electron Density Maps for L99A binders and F-helix: The initial Fo-Fc electron density map contoured at 1.6σ around the ligand and F-helix (density in grey) for L99A lysozyme-ligand complexes A. 7LOB, B. 7LOC, C. 7LOA, D. 7LOD, E. 7LX8, F. 7LX9, G. 7LOG, H. 7LOF, I. 7LOE, J. 7LXA, K. 7LX7, L. 7LX6 and M. 7LOJ. Ligand carbons in magenta and protein carbons in green, oxygens red, nitrogens blue, sulfurs yellow and chlorides green.

Supplementary Table S 3 Crystallographic Statistics

PDB ID	7LOB	7LOC	7LOA
Data collection			
Space group	P 32 2 1	P 32 2 1	P 32 2 1
Cell dimensions			
a, *b*, *c* (Å)	60.412, 60.412, 96.278	60.266, 60.266, 96.411	60.057, 60.057, 96.276
a, b, g (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	45.97 - 1.1 (1.139 - 1.1)*	52.19 - 1.16 (1.202 - 1.16)	52.01 - 1.07 (1.108 - 1.07)
Total reflections	1607952 (151824)	140805 (13413)	177215 (17042)
Unique reflections	83069 (8240)	70437 (6741)	88627 (8536)
Multiplicity	19.4 (18.4)	2.0 (2.0)	2.0 (2.0)
R_{sym} or R_{merge}	0.08578 (4.381)	0.01331 (0.4135)	0.01322 (0.6274)
I / sI	20.07 (0.69)	25.08 (1.88)	24.17 (1.24)
Completeness (%)	99.89 (99.41)	99.69 (96.98)	99.43 (96.73)
CC1/2	1 (0.391)	1 (0.733)	1 (0.527)
Refinement			
Resolution (Å)	52.318 - 1.100	52.192 - 1.160	52.011 - 1.070
Reflections used in refinement	82977 (8192)	70424 (6737)	88609 (8529)
Reflections used for R-free	4176 (449)	3528 (338)	4481 (435)
R-work/R-free (%)	20.90/22.18	19.76/20.80	20.76/21.25
No. atoms			
Protein	1332	1345	1353
Ligand/ion	23	24	28
Water	116	124	122
***B*-factors**			
Protein	16,05	15,39	16,05
Ligand/ion	21,1	24,03	24,88

Water	23,47	22,91	23,96
R.m.s. deviations			
Bond lengths (Å)	0,004	0,004	0,004
Bond angles (°)	0,76	0,75	0,78

Supplementary Table S 4 continued

PDB ID	7LOD	7LX8	7LX9
Data collection			
Space group	P 32 2 1	P 32 2 1	P 32 2 1
Cell dimensions			
a, *b*, *c* (Å)	60.209, 60.209, 96.185	60.289, 60.289, 96.328	60.2464, 60.2464, 96.317
a, b, g (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	35.35 - 1.02 (1.056 - 1.02)	45.9 - 1.03 (1.067 - 1.03)	52.17 - 1.19 (1.233 - 1.19)
Total reflections	201396 (16383)	198405 (17450)	130681 (12679)
Unique reflections	100815 (8282)	99284 (8780)	65390 (6389)
Multiplicity	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)
R_{sym} or R_{merge}	0.01013 (0.8608)	0.008513 (0.787)	0.02252 (1.172)
I / sI	29.37 (0.95)	22.11 (0.90)	16.70 (0.72)
Completeness (%)	97.72 (80.74)	98.71 (88.61)	99.32 (93.40)
CC1/2	1 (0.478)	1 (0.479)	1 (0.312)
Refinement			
Resolution (Å)	45.840 - 1.020	45.903 - 1.030	52.175 - 1.190
Reflections used in refinement	100734 (8232)	99239 (8757)	65039 (6053)
Reflections used for R-free	5070 (425)	5034 (445)	3136 (286)
R-work/R-free (%)	21.22/21.35	20.63/21.22	21.48/22.26
No. atoms			
Protein	1348	1295	1292
Ligand/ion	28	17	18
Water	114	116	103
***B*-factors**			
Protein	15,79	16,51	16,91
Ligand/ion	25,26	19,08	18,76

Water	23,25	24,11	24,54
R.m.s. deviations			
Bond lengths (Å)	0,004	0,004	0,004
Bond angles (°)	0,73	0,75	0,76

Supplementary Table S 5 continued

PDB ID	7LOG	7LOF	7LOE
Data collection			
Space group	P 32 2 1	P 32 2 1	P 32 2 1
Cell dimensions			
a, *b*, *c* (Å)	60.2482, 60.2482, 96.119	60.181, 60.181, 96.346	60.0782, 60.0782, 96.266
a, b, g (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	45.86 - 0.99 (1.025 - 0.99)	45.84 - 1.05 (1.088 - 1.05)	45.77 - 1.01 (1.046 - 1.01)
Total reflections	216941 (16204)	182471 (14771)	205681 (16484)
Unique reflections	108752 (8263)	91335 (7448)	102989 (8331)
Multiplicity	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)
R_{sym} or R_{merge}	0.007583 (0.3653)	0.009087 (0.6067)	0.00924 (0.7696)
I / sI	31.03 (2.28)	20.73 (1.27)	24.59 (1.08)
Completeness (%)	96.47 (74.36)	96.43 (79.33)	97.27 (79.24)
CC1/2	1 (0.786)	1 (0.497)	1 (0.507)
Refinement			
Resolution (Å)	52.176 - 0.990	52.118 - 1.050	52.029 - 1.010
Reflections used in refinement	108714 (8253)	91305 (7421)	102914 (8286)
Reflections used for R-free	5610 (436)	4713 (401)	5258 (439)
R-work/R-free (%)	19.98/20.31	19.75/20.06	20.39/21.88
No. atoms			
Protein	1314	1304	1308
Ligand/ion	30	21	34
Water	151	117	126
***B*-factors**			
Protein	13,17	15,58	13,64

Ligand/ion	17,32	17,48	14,84
Water	20,03	22,77	21,01
R.m.s. deviations			
Bond lengths (Å)	0,004	0,004	0,004
Bond angles (°)	0,78	0,76	0,75

Supplementary Table S 6 continued

PDB ID	7LXA	7LX7	7LX6	7LOJ
Data collection				
Space group	P 32 2 1	P 32 2 1	P 32 2 1	P 32 2 1
Cell dimensions				
a, *b*, *c* (Å)	60.1543, 60.1543, 96.293	60.2546, 60.2546, 96.212	60.2724, 60.2724, 96.5	60.4019, 60.4019, 95.855
a, b, g (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	45.82 - 1.07 (1.108 - 1.07)	45.87 - 1.05 (1.088 - 1.05)	45.91 - 1.05 (1.088 - 1.05)	52.31 - 1.5 (1.554 - 1.5)
Total reflections	178577 (17485)	184577 (15540)	190136 (18730)	65467 (6183)
Unique reflections	89299 (8752)	92363 (7823)	95080 (9374)	32955 (3193)
Multiplicity	2.0 (2.0)	2.0 (2.0)	2.0 (2.0)	2.0 (1.9)
R_{sym} or R_{merge}	0.009848 (0.8013)	0.009464 (0.6015)	0.009314 (0.5639)	0.009915 (0.08336)
I / sI	15.55 (1.07)	23.60 (1.37)	21.14 (1.36)	31.95 (6.90)
Completeness (%)	99.75 (98.28)	97.44 (83.70)	99.93 (99.56)	99.63 (97.83)
CC1/2	1 (0.504)	1 (0.509)	1 (0.65)	1 (0.971)
Refinement				
Resolution (Å)	52.095 - 1.070	52.182 - 1.050	52.197 - 1.050	52.310 - 1.500
Reflections used in refinement	89185 (8692)	92346 (7821)	95045 (9365)	32953 (3194)
Reflections used for R-free	4517 (451)	4781 (428)	4906 (498)	1608 (150)
R-work/R-free (%)	20.43/21.06	19.37/20.21	19.91/20.18	18.53/20.18
No. atoms				
Protein	1289	1290	1297	1299
Ligand/ion	18	24	22	31
Water	120	139	137	140
***B*-factors**				
Protein	16,5	14,21	15,25	15,24

Ligand/ion	18,83	15,04	17,56	23,09
Water	23,81	21,52	22,62	22,2
R.m.s. deviations				
Bond lengths (Å)	0,004	0,004	0,004	0,005
Bond angles (°)	0,85	0,76	0,75	0,79

(One crystal for each structure)

*Values in parentheses are for the highest-resolution shell.

Supplementary References

1. D. A. Case *et al.* (AMBER (2014), University of California, San Francisco.
2. K. Lindorff-Larsen *et al.*, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958 (2010).
3. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926-935 (1983).
4. H. G. Wallnoefer, K. R. Liedl, T. Fox, A challenging system: free energy prediction for factor Xa. *J Comput Chem* **32**, 1743-1752 (2011).
5. L. C. Pierce, R. Salomon-Ferrer, F. d. O. C. Augusto, J. A. McCammon, R. C. Walker, Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J Chem Theory Comput* **8**, 2997-3002 (2012).
6. A. S. Kamenik, U. Kahler, J. E. Fuchs, K. R. Liedl, Localization of Millisecond Dynamics: Dihedral Entropy from Accelerated MD. *J Chem Theory Comput* **12**, 3449-3455 (2016).
7. R. Salomon-Ferrer, A. W. Gotz, D. Poole, S. Le Grand, R. C. Walker, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **9**, 3878-3888 (2013).
8. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. *J Chem Phys* **98**, 10089-10092 (1993).
9. S. Adelman, J. Doll, Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *J Chem Phys* **64**, 2375-2388 (1976).
10. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J Chem Phys* **81**, 3684-3690 (1984).
11. J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics* **23**, 327-341 (1977).
12. D. R. Roe, T. E. Cheatham, 3rd, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **9**, 3084-3095 (2013).
13. Y. Miao *et al.*, Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J Chem Theory Comput* **10**, 2677-2689 (2014).
14. M. K. Scherer *et al.*, PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J Chem Theory Comput* **11**, 5525-5542 (2015).
15. G. Perez-Hernandez, F. Paul, T. Giorgino, G. De Fabritiis, F. Noe, Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* **139**, 015102 (2013).
16. J. H. Prinz *et al.*, Markov models of molecular kinetics: generation and validation. *J Chem Phys* **134**, 174105 (2011).
17. M. Fischer, R. G. Coleman, J. S. Fraser, B. K. Shoichet, Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat Chem* **6**, 575-583 (2014).
18. J. M. Word, S. C. Lovell, J. S. Richardson, D. C. Richardson, Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *Journal of molecular biology* **285**, 1735-1747 (1999).

19. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin, A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **161**, 269-288 (1982).
20. K. A. Sharp, Polyelectrolyte electrostatics: Salt dependence, entropic, and enthalpic contributions to free energy in the nonlinear Poisson–Boltzmann model. *Biopolymers: Original Research on Biomolecules* **36**, 227-243 (1995).
21. E. C. Meng, B. K. Shoichet, I. D. Kuntz, Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry* **13**, 505-524 (1992).
22. M. M. Mysinger, B. K. Shoichet, Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* **50**, 1561-1573 (2010).
23. M. Merski, M. Fischer, T. E. Balius, O. Eidam, B. K. Shoichet, Homologous ligands accommodated by discrete conformations of a buried cavity. *Proc Natl Acad Sci U S A* **112**, 5039-5044 (2015).
24. A. Morton, W. A. Baase, B. W. Matthews, Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry* **34**, 8564-8575 (1995).
25. A. Morton, B. W. Matthews, Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry* **34**, 8576-8588 (1995).
26. H. Lee, M. Fischer, B. K. Shoichet, S. Y. Liu, Hydrogen Bonding of 1,2-Azaborines in the Binding Cavity of T4 Lysozyme Mutants: Structures and Thermodynamics. *J Am Chem Soc* **138**, 12021-12024 (2016).
27. A. I. Su *et al.*, Docking molecules by families to increase the diversity of hits in database screens: computational strategy and experimental evaluation. *Proteins* **42**, 279-293 (2001).
28. D. L. Mobley *et al.*, Predicting absolute ligand binding free energies to a simple model site. *J Mol Biol* **371**, 1118-1134 (2007).
29. A. P. Graves *et al.*, Rescoring docking hit lists for model cavity sites: predictions and experimental testing. *J Mol Biol* **377**, 914-934 (2008).
30. M. M. Mysinger, M. Carchia, J. J. Irwin, B. K. Shoichet, Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* **55**, 6582-6594 (2012).
31. J. J. Irwin, B. K. Shoichet, ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **45**, 177-182 (2005).