Supporting Information

CHARMM-GUI Free Energy Calculator for Practical Ligand Binding Free Energy Simulations with AMBER

Han Zhang[1], Seonghoon Kim[2], Timothy J. Giese[3], Tai-Sung Lee[3], Jumin Lee[1], Darrin M. York[3], and Wonpil Im[1]*

[1]Department of Biological Sciences, Chemistry, Bioengineering, and Computer Science and Engineering, Lehigh University, Pennsylvania 18015, USA

[2]School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Republic of Korea

[3]Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine, and Department of Chemistry and Chemical Biology, Rutgers, the State University of New Jersey, New Jersey 08854, USA

*Corresponding author: wonpil@lehigh.edu

Free energy calculation systems preparation for more challenging cases

Preparation of free energy calculation systems is a time-consuming and error-prone process, and different systems may require different preparation and present different problems. Challenging systems including ions, co-factors, missing side chains and residues, ambiguous water positions, etc. *Free Energy Calculator* is able to set up free energy calculation systems for these cases and provide free energy calculation inputs for AMBER, NAMD, CHARMM, and GENESIS. The details of generated challenging systems are described below.

It was estimated that at least one-third of all enzyme-catalyzed reactions require metal ions for catalytic activity.[1] Serine/threonine protein phosphatases (PPs), including PP1, PP2A, PP2B, PP4, PP5, PP6, and PP7, are responsible for the dephosphorylations of phospho-serine/-threonine residues.[2] A cocrystal structure of okadaic acid and PP1 (PDB: 1JK7) can be used to study the interaction between the inhibitor and protein.[3] Two metal ions, usually two Mn^{2+} ions, are located in the PP1 active site center, and a water molecule (W1) bridges these two metal ions.[4] Okadaic acid inactivates the enzyme via binding to its binuclear metal center.[4] Therefore, it is crucial to keep the metal ions and water molecules. As shown in **Figure S1A**, CHARMM-GUI is able to keep all key elements during the system building.

Cofactors are crucial to help a protein or enzyme to function appropriately. N-terminal acetyltransferase (NAT) catalyzes acetylation through the mechanism in which cofactor acetyl-CoA binds first.[5] The NAT complex catalyzes the transfer of the acetyl group from acetyl-CoA to the N-terminal α-amino group of newly synthesized proteins, and the expression of NAT is abnormal (increased or decreased) in cancer cells. The Nα-terminal acetyltransferase (Naa50) enzyme is one of the members of NAT protein family, and therefore it could be a potential target for drug discovery. Inhibitor 4a has a potent binding affinity to Naa50 and the cocrystal structure (PDB: 6WFN) can be used for further drug design.[6] As shown in **Figure S1B**, CHARMM-GUI was able to keep the cofactor acetyl-CoA in its site and successfully generated the free energy calculation system.

It is very common to find missing residues in crystal structures, especially for the flexible loop or side chains. There are many PDB files of Tropomyosin receptor kinases (TRK) in the PDB, but most of them have missing residues and need to be added before running molecular dynamics simulation. TrkA, TrkB, and TrkC are activated by hormones of the neurotrophin family, including nerve growth factor (NGF), brain derived neurotrophic factor (BDNF), neurotrophin 3 (NT3), and neurotrophin 4 (NT4).[7] PDB: 6PMB is the holo structure of TrkA and ligand 1a, but there are missing residues: 485 to 499 (N-terminal missing residues), 549 to 551, and 609 to 613.[8] CHARMM-GUI uses GalxyFill to model the missing residues if users prefer to add the missing amino acids. The binding site of TrkA and the added missing residues are shown in **Figure S1C**.

Hydrogen bonds are critical interactions determining binding affinity between ligand and protein and thus fundamental to any biological process.[9] Serine proteases such as trypsin specifically recognize substrates containing basic residues prior to the peptide bond cleavage.[10] Benzamidine is an alkaline inhibitor that blocks proteolytic function and associates with Asp189 in an arginine-analogous manner. Schiebel et al showed that water molecule W1 is trapped upon ligand (Benzamidine or N-amidinopiperidine) binding with suboptimal energetic properties.[10] CHARMM-GUI provides options to keep the crystal water molecules in the PDB files. In this specific case, we generated the free energy calculation system by using PDB: 5MNH and kept the crystal water molecules (**Figure S1D**).

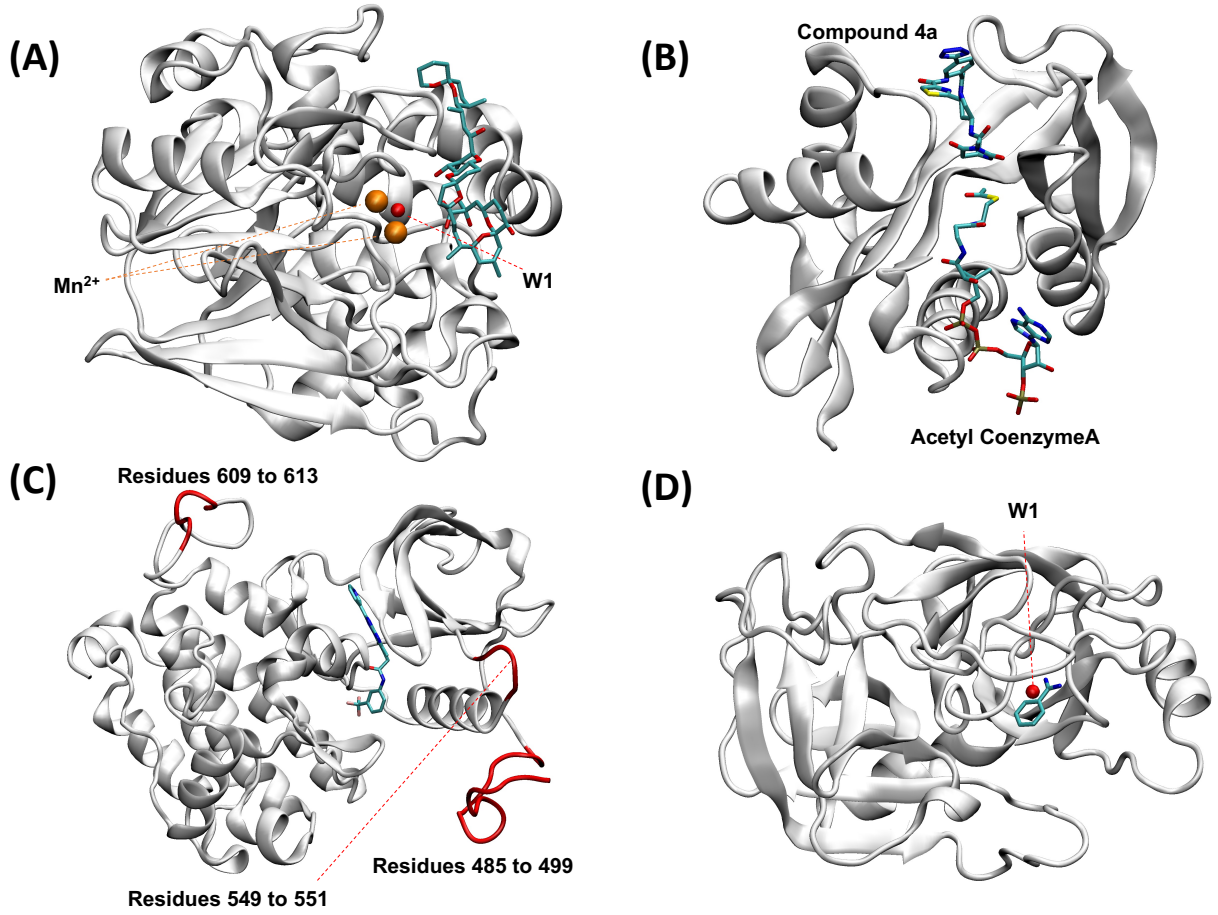
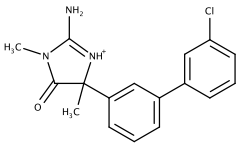
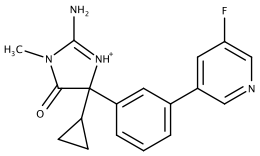
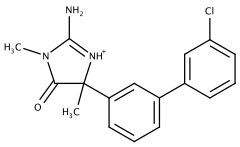
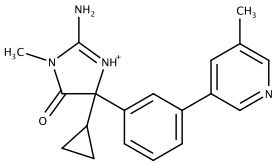
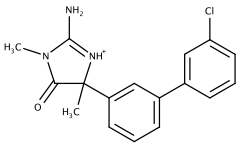
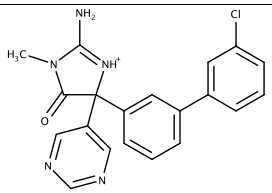
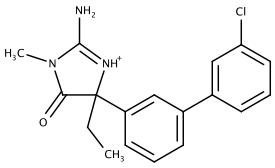
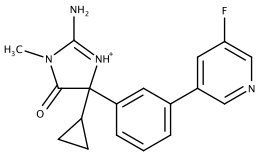
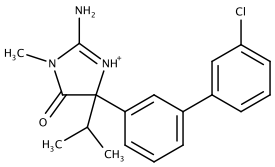
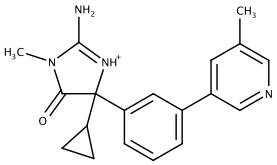
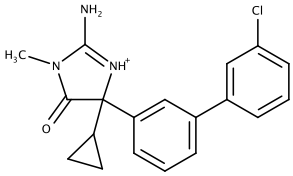
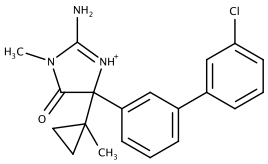
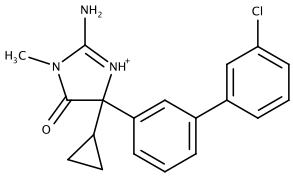
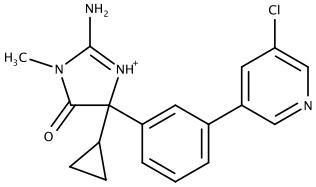
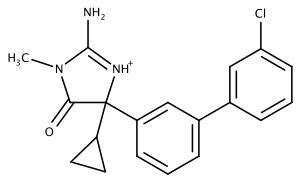
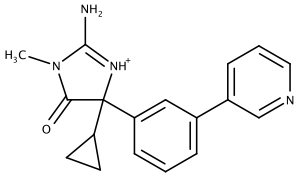
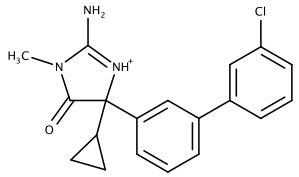
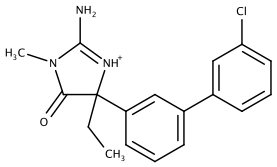
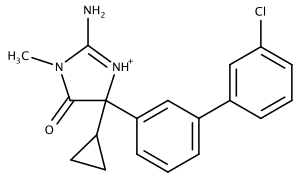
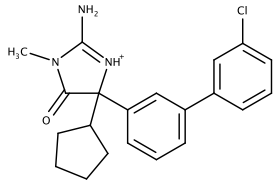
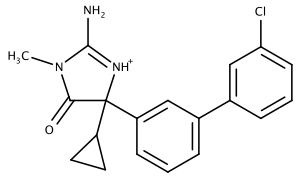
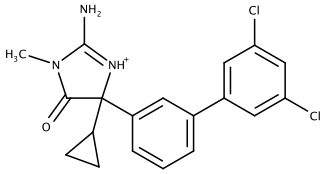
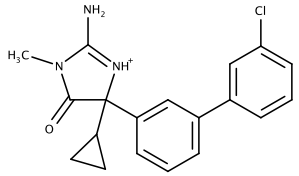
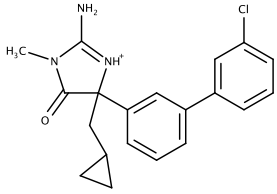
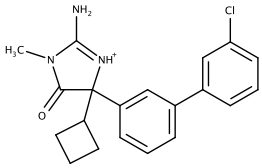
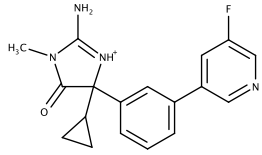
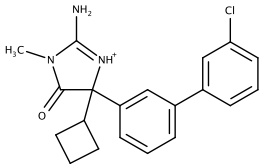
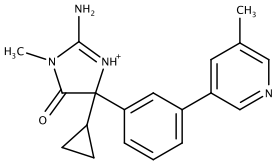
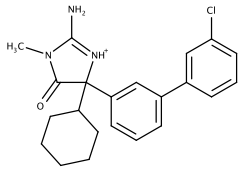
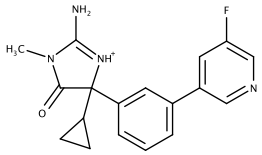
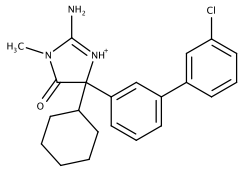
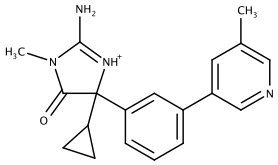
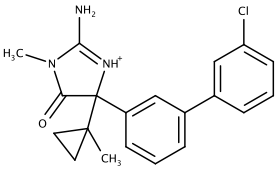
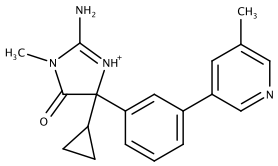
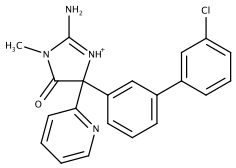
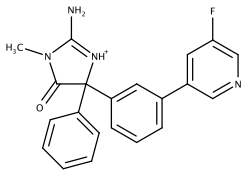
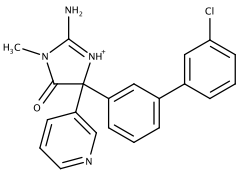
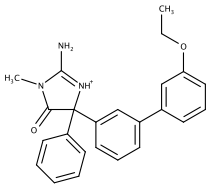
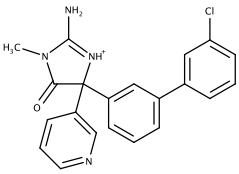
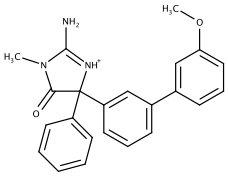
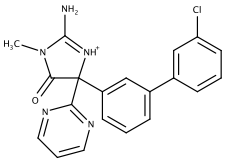
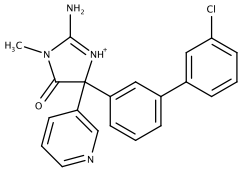


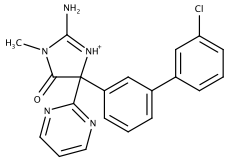
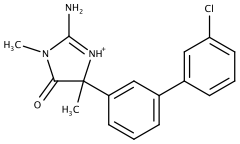
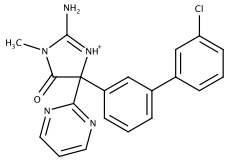
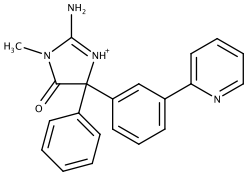
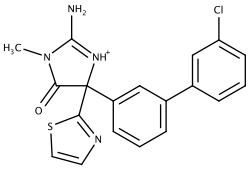
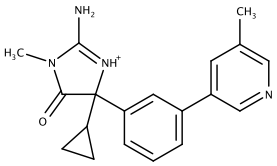
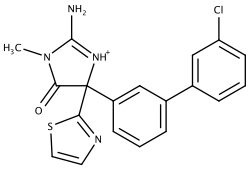
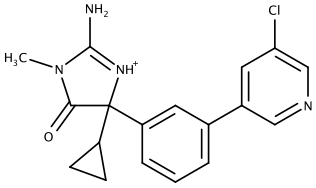
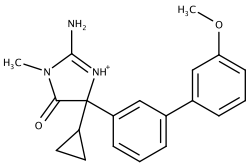
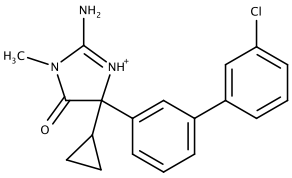
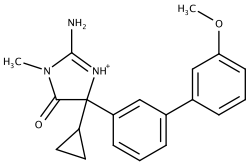
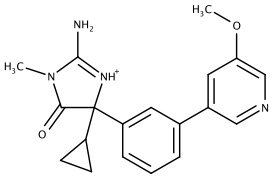
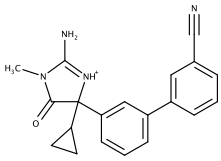
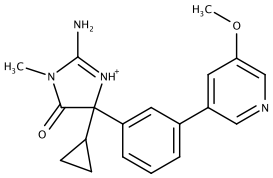
Figure S1. (A) PP1 and okadaic acid complex structure. Two Mn^{2+} ions and W1 water molecules are shown in orange and red, respectively. (B) Naa50, cofactor acetyl coenzyme A, and compound 4a complex structure. (C) TrkA and ligand 1a complex structure. All missing residues of PDB: 6PMB are added and shown in red. (D) Trypsin and benzamidine complex. W1 water molecule is shown in red.

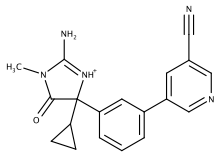
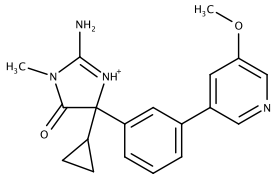
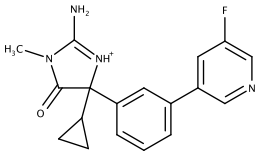
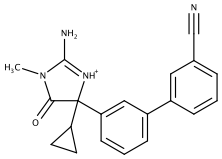
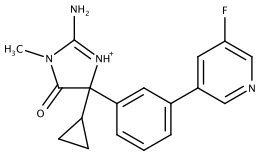
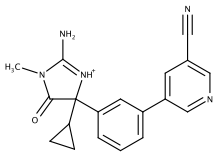
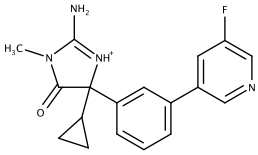
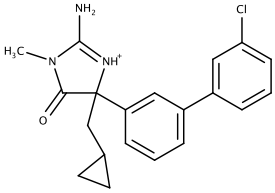
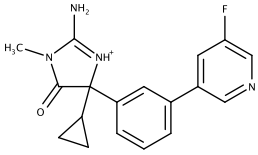
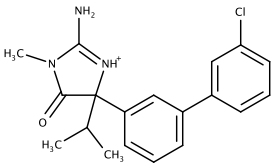
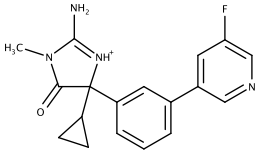
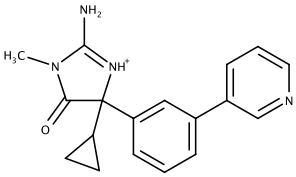
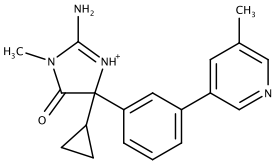
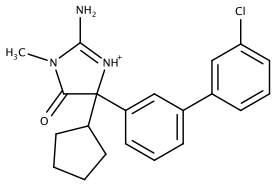
Table S1. 58 pairs are used for $\Delta\Delta G_{bind}$ calculations

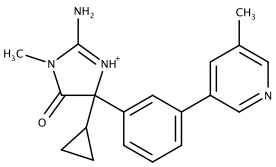
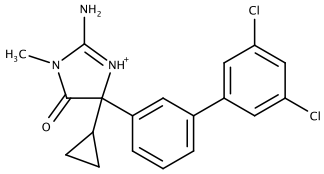
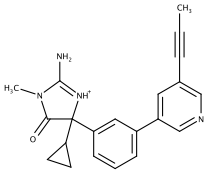
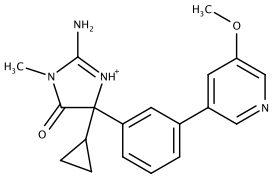
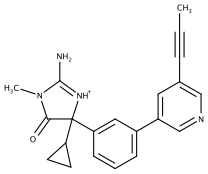
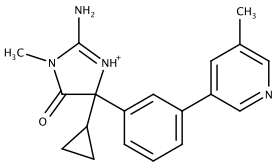
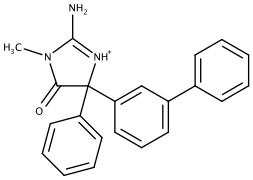
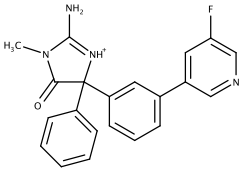
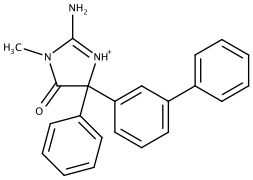
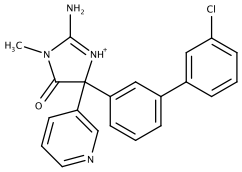
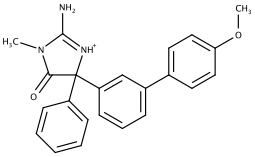
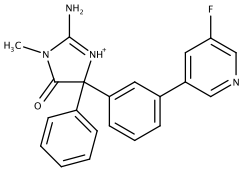
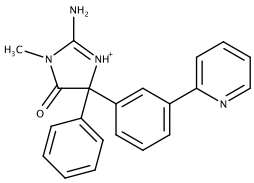
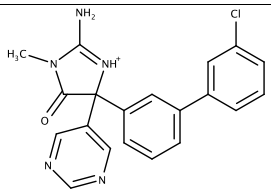
	Ligand 0	Ligand 1
1	 CAT-13a	 CAT-17g
2	 CAT-13a	 CAT-17i
3	 CAT-13a	 CAT-13m
4	 CAT-13b	 CAT-17g
5	 CAT-13c	 CAT-17i
6	 CAT-13d	 CAT-13h
7	 CAT-13d	 CAT-17h

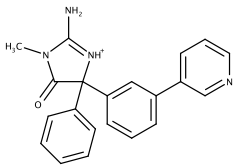
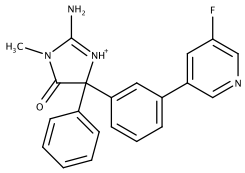
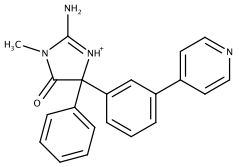
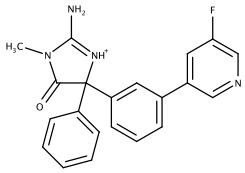
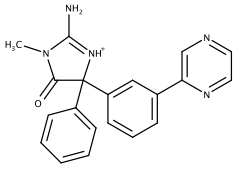
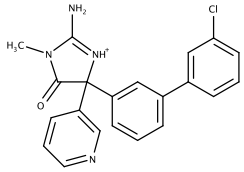
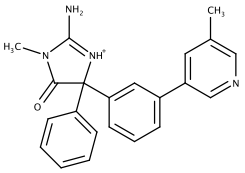
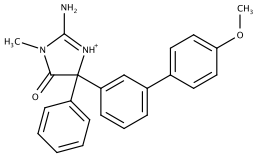
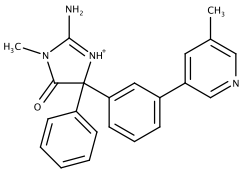
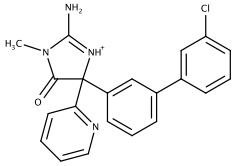
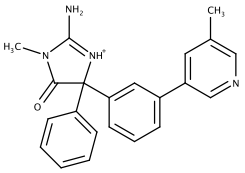
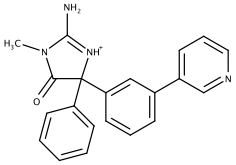
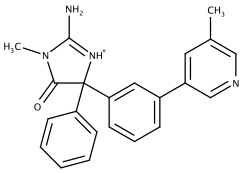
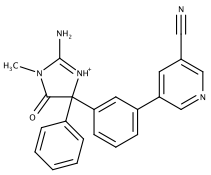
8	 CAT-13d	 CAT-17d
9	 CAT-13d	 CAT-13b
10	 CAT-13d	 CAT-13f
11	 CAT-13d	 CAT-17a
12	 CAT-13d	 CAT-13i
13	 CAT-13e	 CAT-17g
14	 CAT-13e	 CAT-17i

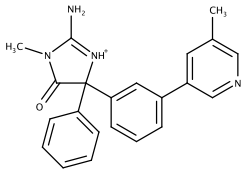
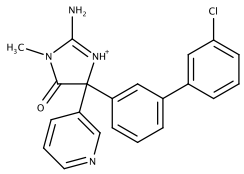
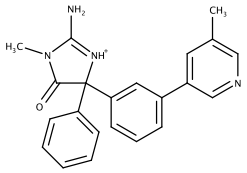
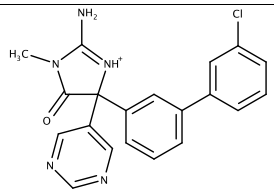
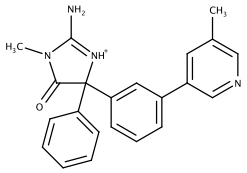
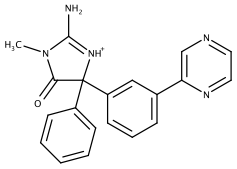
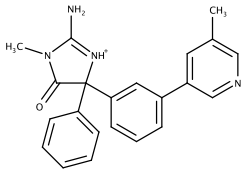
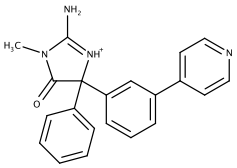
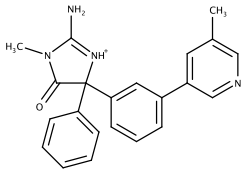
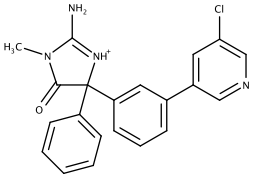
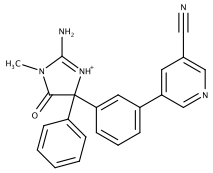
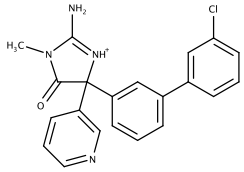
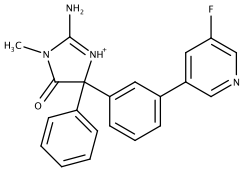
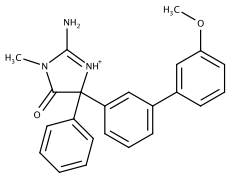
15	 CAT-13g	 CAT-17g
16	 CAT-13g	 CAT-17i
17	 CAT-13h	 CAT-17i
18	 CAT-13j	 CAT-4o
19	 CAT-13k	 CAT-4d
20	 CAT-13k	 CAT-4b
21	 CAT-13n	 CAT-13k

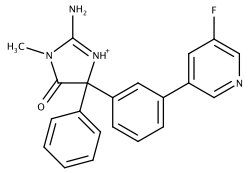
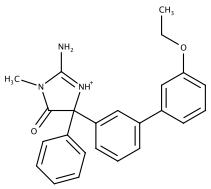
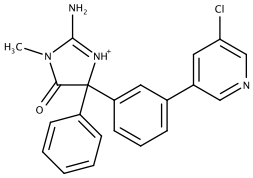
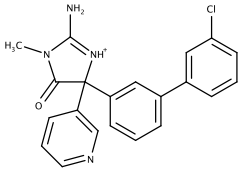
22	 CAT-13n	 CAT-13a
23	 CAT-13n	 CAT-4i
24	 CAT-13o	 CAT-17i
25	 CAT-13o	 CAT-17h
26	 CAT-17b	 CAT-13d
27	 CAT-17b	 CAT-17e
28	 CAT-17c	 CAT-17e

29	 CAT-17f	 CAT-17e
30	 CAT-17g	 CAT-17c
31	 CAT-17g	 CAT-17f
32	 CAT-17g	 CAT-13i
33	 CAT-17g	 CAT-13c
34	 CAT-17g	 CAT-17d
35	 CAT-17i	 CAT-13f

36	 CAT-17i	 CAT-17a
37	 CAT-24	 CAT-17e
38	 CAT-24	 CAT-17i
39	 CAT-4a	 CAT-4o
40	 CAT-4a	 CAT-13k
41	 CAT-4c	 CAT-4o
42	 CAT-4i	 CAT-13m

43	 CAT-4j	 CAT-4o
44	 CAT-4k	 CAT-4o
45	 CAT-4l	 CAT-13k
46	 CAT-4m	 CAT-4c
47	 CAT-4m	 CAT-13j
48	 CAT-4m	 CAT-4j
49	 CAT-4m	 CAT-4n

50	 CAT-4m	 CAT-13k
51	 CAT-4m	 CAT-13m
52	 CAT-4m	 CAT-4l
53	 CAT-4m	 CAT-4k
54	 CAT-4m	 CAT-4p
55	 CAT-4n	 CAT-13k
56	 CAT-4o	 CAT-4b

57	 CAT-4o	 CAT-4d
58	 CAT-4p	 CAT-13k

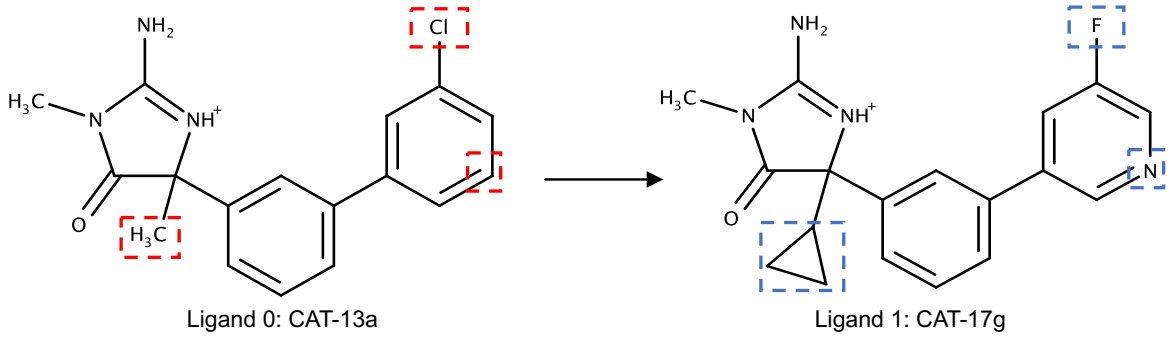


Figure S2. Alchemical transformation of CAT-13a to CAT-17g. The softcore region of “scmask1” and “scmask2” are highlighted by red and blue dashed rectangles, respectively.

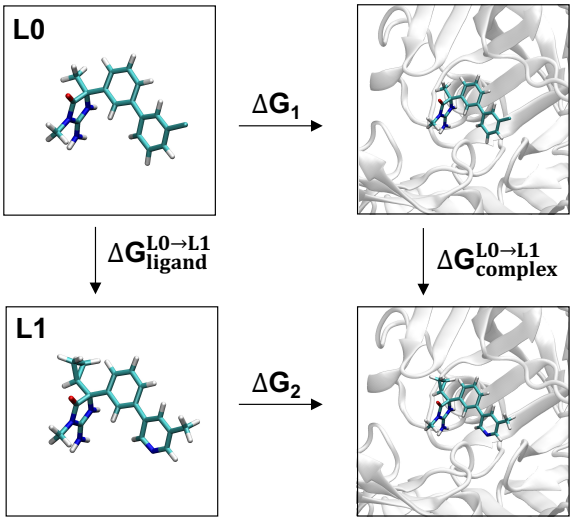


Figure S3. Thermodynamic cycle for relative binding free energy ($\Delta\Delta G_{bind}$) calculations. L0 and L1 are two different ligands.

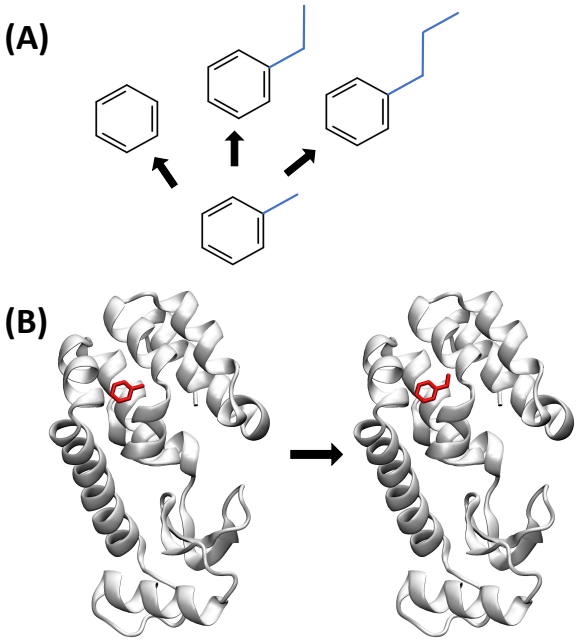


Figure S4. (A) Four ligands and three alchemical transformations used in T4-lysozyme system. (B) Alchemical transformation of toluene to ethylbenzene in the binding site of T4-lysozyme.

References

1. Riordan, J. F., The role of metals in enzyme activity. *Ann Clin Lab Sci* **1977,** *7* (2), 119-29.
2. Liu, F.; Grundke-Iqbal, I.; Iqbal, K.; Gong, C. X., Contributions of protein phosphatases PP1, PP2A, PP2B and PP5 to the regulation of tau phosphorylation. *Eur J Neurosci* **2005,** *22* (8), 1942-50.
3. Maynes, J. T.; Bateman, K. S.; Cherney, M. M.; Das, A. K.; Luu, H. A.; Holmes, C. F.; James, M. N., Crystal structure of the tumor-promoter okadaic acid bound to protein phosphatase-1. *J Biol Chem* **2001,** *276* (47), 44078-82.
4. Zhang, M. M.; Yogesha, S. D.; Mayfield, J. E.; Gill, G. N.; Zhang, Y., Viewing serine/threonine protein phosphatases through the eyes of drug designers. *Febs J* **2013,** *280* (19), 4739-4760.
5. Drazic, A.; Myklebust, L. M.; Ree, R.; Arnesen, T., The world of protein acetylation. *Biochim Biophys Acta* **2016,** *1864* (10), 1372-401.
6. Kung, P. P.; Bingham, P.; Burke, B. J.; Chen, Q.; Cheng, X.; Deng, Y. L.; Dou, D.; Feng, J.; Gallego, G. M.; Gehring, M. R.; Grant, S. K.; Greasley, S.; Harris, A. R.; Maegley, K. A.; Meier, J.; Meng, X.; Montano, J. L.; Morgan, B. A.; Naughton, B. S.; Palde, P. B.; Paul, T. A.; Richardson, P.; Sakata, S.; Shaginian, A.; Sonnenburg, W. K.; Subramanyam, C.; Timofeevski, S.; Wan, J.; Yan, W.; Stewart, A. E., Characterization of Specific N-alpha-Acetyltransferase 50 (Naa50) Inhibitors Identified Using a DNA Encoded Library. *ACS Med Chem Lett* **2020,** *11* (6), 1175-1184.
7. Bagal, S. K.; Omoto, K.; Blakemore, D. C.; Bungay, P. J.; Bilsland, J. G.; Clarke, P. J.; Corbett, M. S.; Cronin, C. N.; Cui, J. J.; Dias, R.; Flanagan, N. J.; Greasley, S. E.; Grimley, R.; Johnson, E.; Fengas, D.; Kitching, L.; Kraus, M. L.; McAlpine, I.; Nagata, A.; Waldron, G. J.; Warmus, J. S., Discovery of Allosteric, Potent, Subtype Selective, and Peripherally Restricted TrkA Kinase Inhibitors. *J Med Chem* **2019,** *62* (1), 247-265.
8. Subramanian, G.; Vairagoundar, R.; Bowen, S. J.; Roush, N.; Zachary, T.; Javens, C.; Williams, T.; Janssen, A.; Gonzales, A., Synthetic inhibitor leads of human tropomyosin receptor kinase A (hTrkA). *RSC Med Chem* **2020,** *11* (3), 370-377.
9. Chen, D.; Oezguen, N.; Urvil, P.; Ferguson, C.; Dann, S. M.; Savidge, T. C., Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Sci Adv* **2016,** *2* (3), e1501240.
10. Schiebel, J.; Gaspari, R.; Wulsdorf, T.; Ngo, K.; Sohn, C.; Schrader, T. E.; Cavalli, A.; Ostermann, A.; Heine, A.; Klebe, G., Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nat Commun* **2018,** *9* (1), 3559.