Supporting Information

De Novo Prediction of Binders and Non-Binders for T4 Lysozyme by gREST Simulations

Ai Niitsu,[†,‖] Suyong Re,[‡,‖] Hiraku Oshima,[‡] Motoshi Kamiya,[§] and Yuji Sugita*[,†,‡,§]

[†]Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, Hirosawa 2-1, Wako, Saitama 351-0198, Japan

[‡]Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, 6-7-1 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan

[§]Computational Biophysics Research Team, RIKEN Center for Computational Science, 6-7-1 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan

[‖]These authors equally contributed to the work as a first author.

*Corresponding Author: sugita@riken.jp

Table S1. Structures of protein binding cavity. The heavy atoms RMSD values of the protein binding cavity residues calculated for the best predicted binding poses at 300 K for benzene, ethylbenzene, and *n*-hexylbenzene with respect to the corresponding X-ray crystal structures (181L, 4W54, and 4W59 for benzene, ethylbenzene, and *n*-hexylbenzene, respectively). The binding cavity residues include all the SITE residues in the X-ray structures (Leu84, Val87, Ala99, Val103, Met106, Gly110, Val111, Ala112, Leu118, Leu121, and Phe153). The Cα atoms of helices C–J in the simulation trajectories were fitted to the respective X-ray structure. Three RMSD values are calculated using different reference structures, the closed (PDB ID: 181L), intermediate (PDB ID: 4W54 (conformation B)), and open (PDB ID: 4W59 (conformation A)) conformations. For comparison, the ligand RMSDs were also shown, where all carbon atoms for benzene, two carbon atoms of the benzene ring for ethylbenzene and *n*-hexylbenzene were considered for the RMSD calculations.

	Benzene	Ethylbenzene	*n*-Hexylbenzene
Ligand	0.248 Å	0.105 Å	0.070 Å
Binding cavity (closed)	0.934 Å	0.672 Å	1.241 Å
Binding cavity (intermediate)	1.283 Å	1.223 Å	1.054 Å
Binding cavity (open)	1.868 Å	1.865 Å	1.497 Å

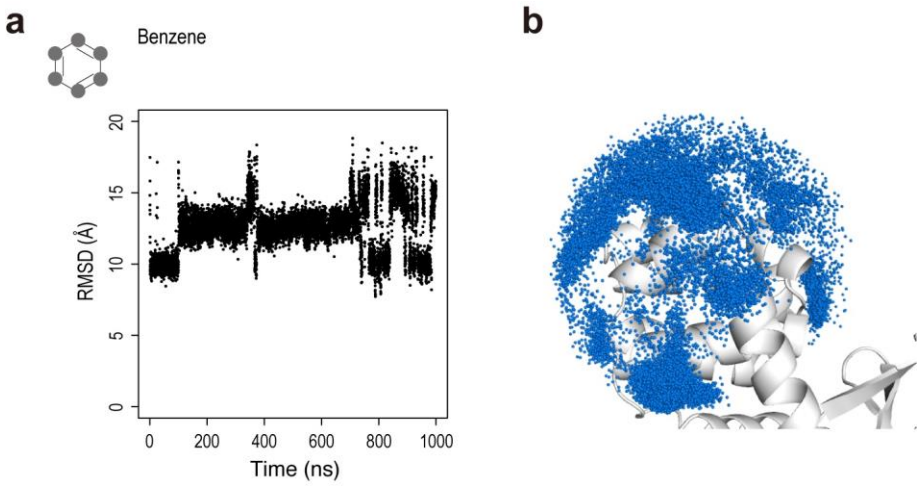
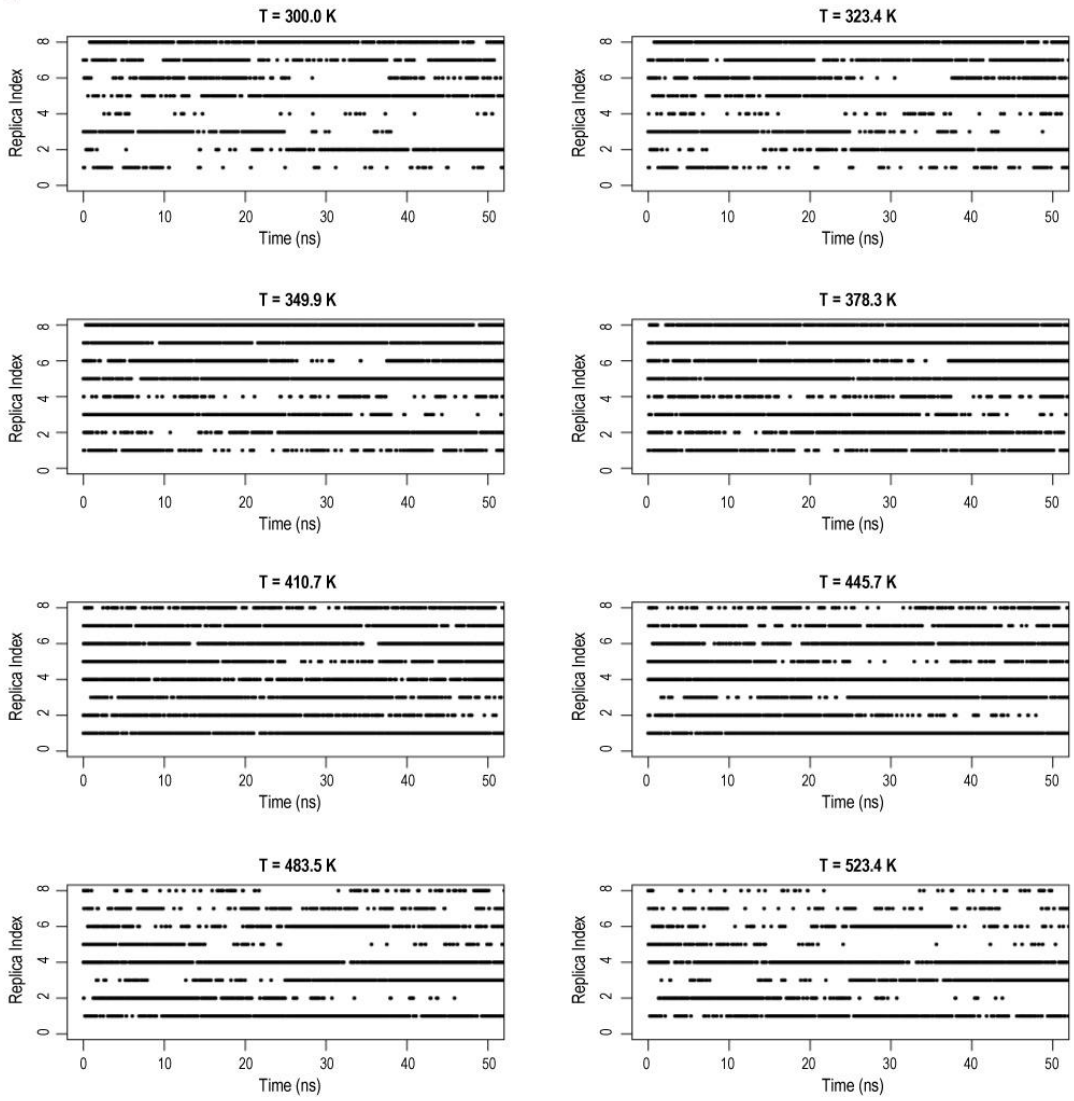


Figure S1. The result of the conventional MD simulation. (a) A time course of the RMSD values from the X-ray structure at 300 K, where carbon atoms of benzene ring are considered for the evaluation. (b) Position of the center of mass (COM) of the ligand (benzene) from the simulation trajectory.

a


T = 300.0 K
T = 323.4 K
T = 349.9 K
T = 378.3 K
T = 410.7 K
T = 445.7 K
T = 483.5 K
T = 523.4 K
Replica Index
Time (ns)
0
2
4
6
8
0
10
20
30
40
50

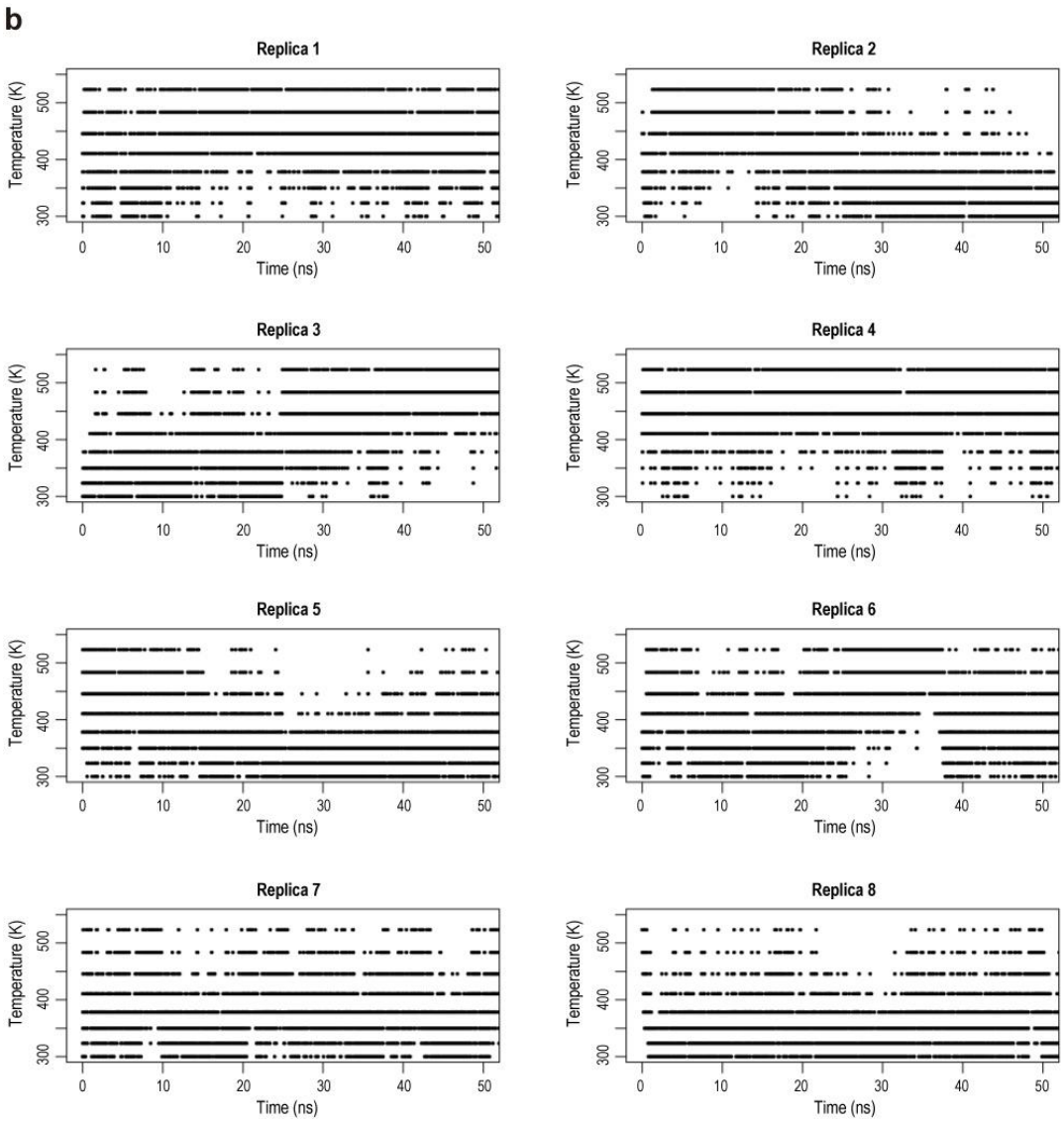


Figure S2. Replica and temperature exchanges. (a) Random walks of replica indices at different solute temperatures and (b) random walks of solute temperatures for eight replicas in the first 50 ns of the benzene-T4L L99A simulation. In both cases, we observe random walks that assure the reliability of current simulations.

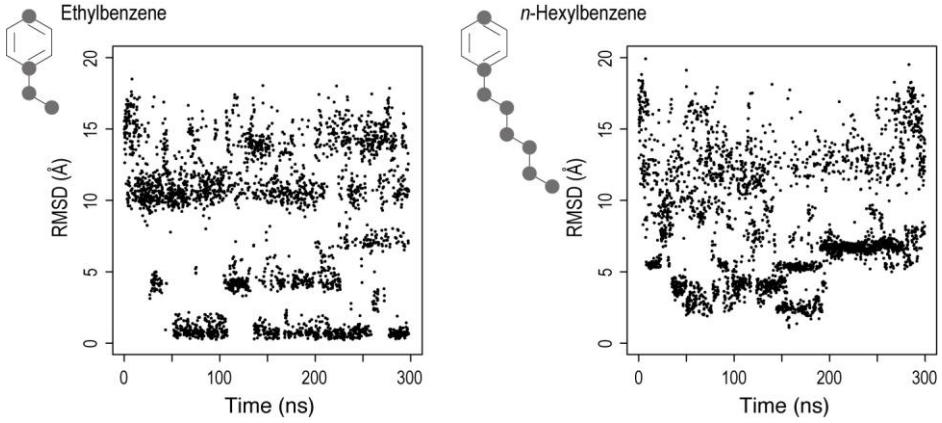


Figure S3. RMSD from the X-ray structure. Time course of the ligand RMSD values from the X-ray structure at 300 K, where carbon atoms of benzene ring and alkyl chain for ethylbenzene and *n*-hexylbenzene are considered for the evaluation.

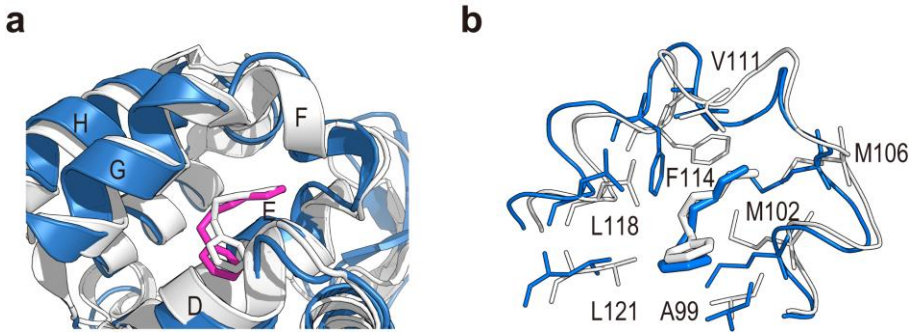


Figure S4. Binding pose in the X-ray structure. (a) The binding pose found in the X-ray structure (PDB ID: 4W59, while) and the closest predicted pose (the ligand RMSD = 0.92 Å, blue and magenta). (b) The comparison of the binding site between the X-ray structure and the predicted pose.

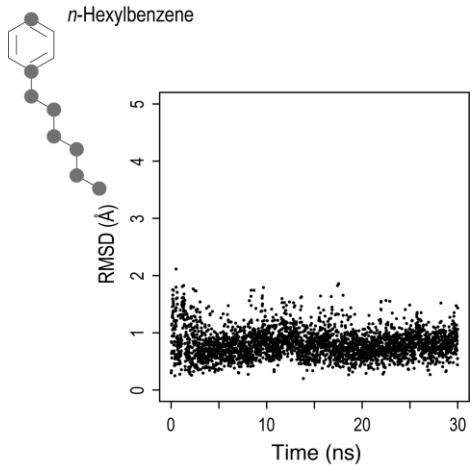


Figure S5. Result of conventional MD simulation of *n*-hexylbenzen. Time course of the ligand RMSD values from the X-ray structure in the conventional MD simulation at 300 K.

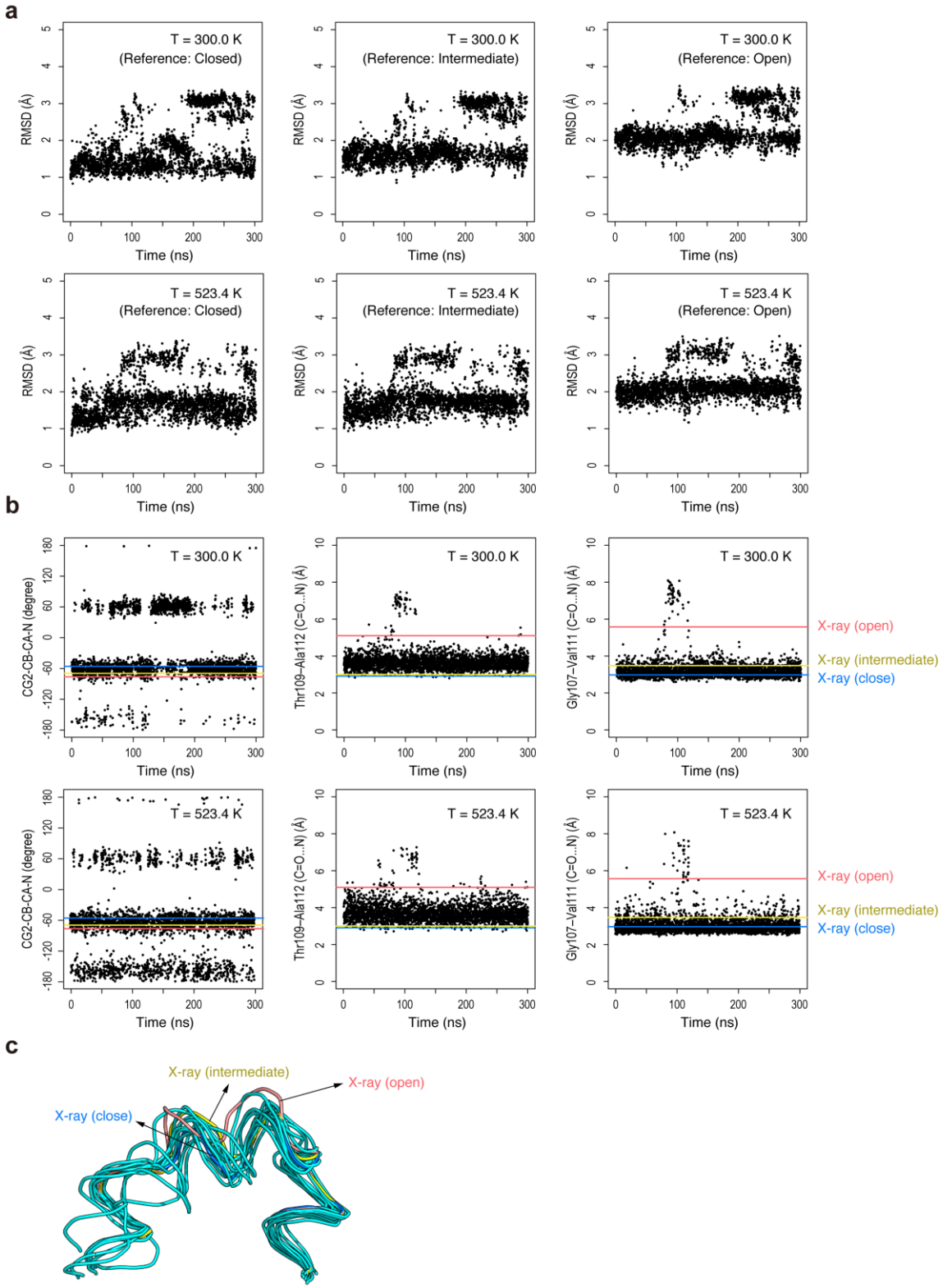

a
T = 300.0 K
(Reference: Closed)
T = 300.0 K
(Reference: Intermediate)
T = 300.0 K
(Reference: Open)
T = 523.4 K
(Reference: Closed)
T = 523.4 K
(Reference: Intermediate)
T = 523.4 K
(Reference: Open)
RMSD (Å)
Time (ns)
b
CG2-CB-CA-N (degree)
Thr109–Ala112 (C=O...N) (Å)
Gly107–Val111 (C=O...N) (Å)
T = 300.0 K
T = 523.4 K
X-ray (open)
X-ray (intermediate)
X-ray (close)
c
X-ray (intermediate)
X-ray (open)
X-ray (close)

Figure S6. Sampling of protein conformations. (a) Time course of the heavy atoms RMSD values of helix F region (residues 105–116) with respect to the three protein conformations (the closed (PDB ID: 181L), intermediate (PDB ID: 4W54 (conformation B)), and open (PDB ID: 4W59 (conformation A)) calculated for the case of *n*-hexylbenzene. RMSDs at the lowest (T = 300.0 K) and the highest temperature (T = 523.4 K) replicas are shown. (b) Time course of the CG2-CB-CA-N dihedral angle of Val111 (left) and the main chain hydrogen bond distances (C=O…NH) between Thr109 and Ala112 (middle) and Gly107 and Val111 (right) at the lowest (T = 300.0 K) and the highest temperature (T = 523.4 K) replicas. The values from the three X-ray crystal structures (the closed (blue), intermediate (yellow), and open (red)) are shown for comparison. (c) Ten snapshots of the helix F region taken from the simulation trajectory at 300 K (cyan). The three X-ray crystal structures (the closed (blue), intermediate (yellow), and open (red)) are shown for comparison.

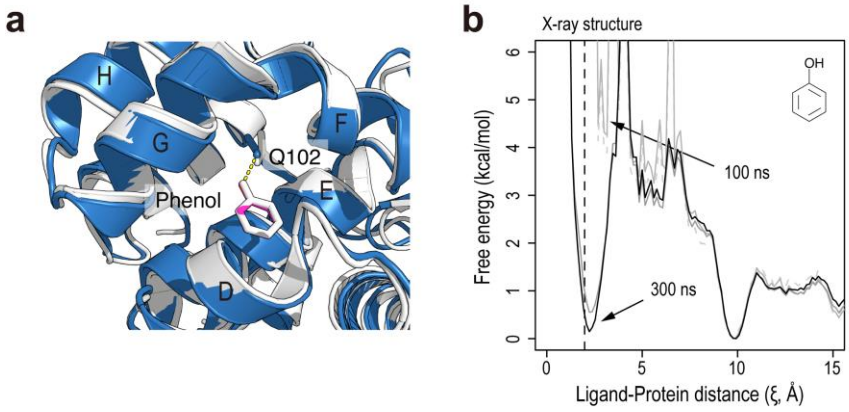


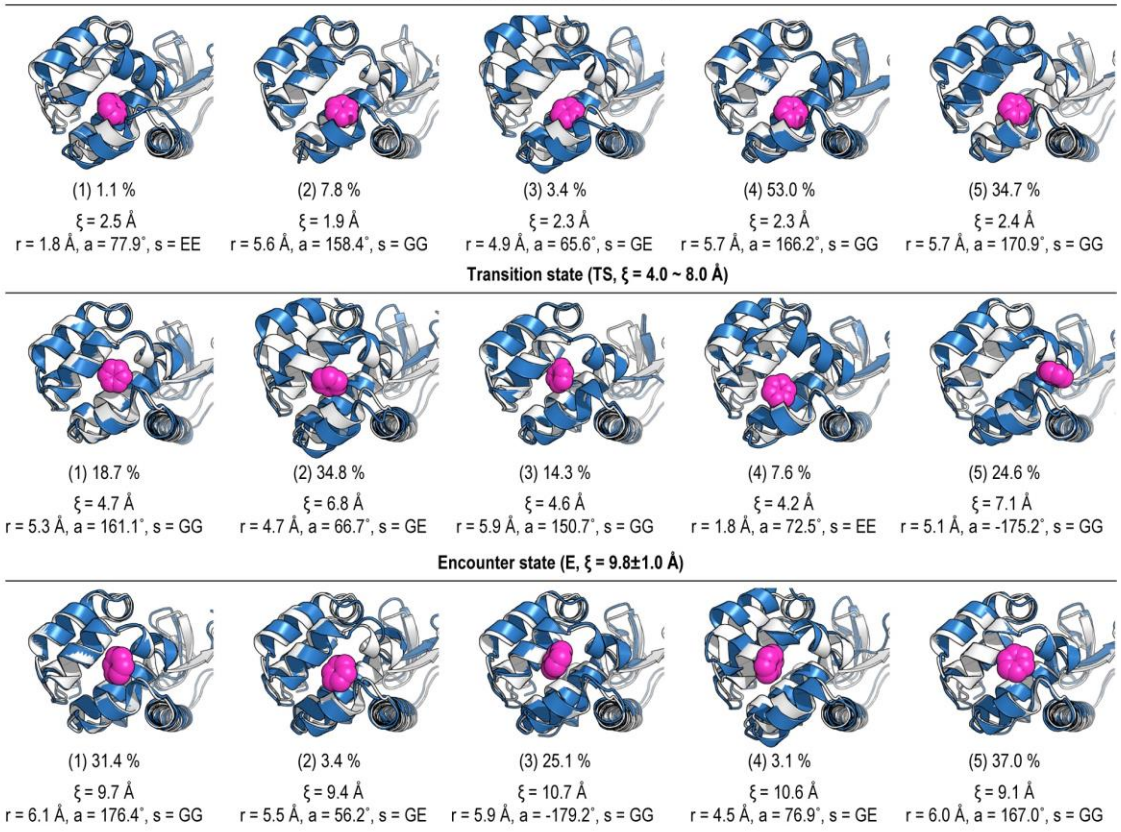
Figure S7. Results of phenol-T4L L99A/M102Q simulation. (a) The predicted binding pose (blue and magenta) of the phenol aligned to the X-ray structure of the phenol-T4L L99A/M102Q double mutant (PDB ID: 1LI2, white). The hydrogen bond between the phenol and Gln102 is shown with yellow dotted line. (b) Free-energy profiles at 300 K along the ligand-protein distance (ξ). The profiles are drawn while appending the data every 50 ns (from dashed light gray (50 ns) to solid black (300 ns)) to show convergence. The ligand-protein distance (ξ) in the X-ray structure (ξ = 1.98 Å) is shown as a dotted line.

a

F114, χ1	G113-S117 distance: 2-3 Å	G113-S117 distance: ~ 6 Å
~ 60°	EE	GE
~ 180°/-180°	EG	GG

b

Benzene

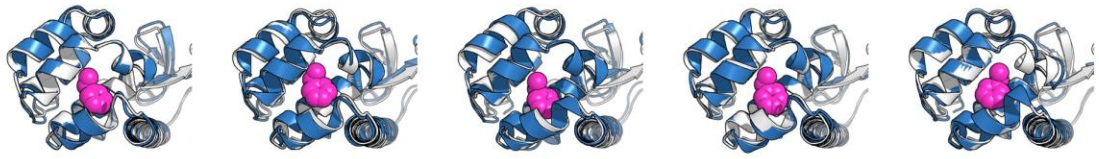


* r: G113-S117 distance, a: χ1 angle, s: state
* Reference X-ray strcuture: PDB ID: 181L

c

Ethylbenzene

Bound state (B, ξ = 2.1±1.0 Å)



(1) 33.0 %
ξ = 1.0 Å, η = 5.5 Å
r = 6.4 Å, a = 174.9˚, s = GG

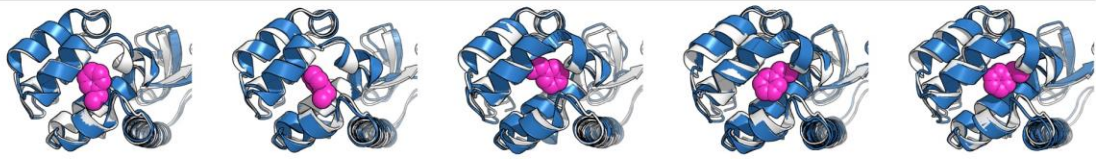
(2) 16.2 %
ξ = 1.4 Å, η = 5.8 Å
r = 5.6 Å, a = -174.3˚, s = GG

(3) 26.3 %
ξ = 1.0 Å, η = 5.2 Å
r = 6.2 Å, a = 167.0˚, s = GG

(4) 16.0 %
ξ = 1.9 Å, η = 5.2 Å
r = 6.3 Å, a = 170.9˚, s = GG

(5) 8.5 %
ξ = 2.1 Å, η = 5.1 Å
r = 5.8 Å, a = 154.6˚, s = GG

Transition state (TS, ξ = 4.0 ~ 6.0 Å)



(1) 15.9 %
ξ = 5.4 Å, η = 1.9 Å
r = 5.2 Å, a = 150.7˚, s = GG

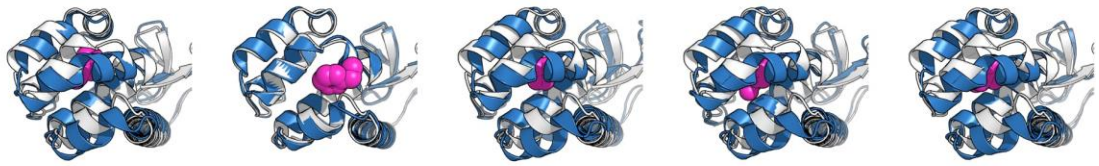
(2) 15.1 %
ξ = 6.1 Å, η = 1.8 Å
r = 5.9 Å, a = 179.9˚, s = GG

(3) 15.0 %
ξ = 5.5 Å, η = 7.0 Å
r = 5.5 Å, a = 163.8˚, s = GG

(4) 29.3 %
ξ = 4.9 Å, η = 6.5 Å
r = 5.6 Å, a = 152.8˚, s = GG

(5) 24.7 %
ξ = 5.1 Å, η = 6.2 Å
r = 5.0 Å, a = 156.6˚, s = GG

Intermediate (I, ξ = 7.3±1.0 Å)



(1) 19.9 %
ξ = 8.5 Å, η = 7.7 Å
r = 2.5 Å, a = 69.8˚, s = EE

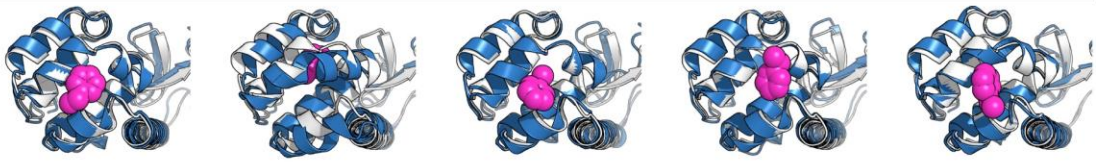
(2) 5.0 %
ξ = 7.1 Å, η = 9.6 Å
r = 5.7 Å, a = -159.2˚, s = GG

(3) 28.1 %
ξ = 8.0 Å, η = 8.7 Å
r = 2.3 Å, a = 49.5˚, s = EE

(4) 27.9 %
ξ = 8.5 Å, η = 7.8 Å
r = 2.7 Å, a = 75.0˚, s = EE

(5) 19.0 %
ξ = 8.5 Å, η = 7.8 Å
r = 2.4 Å, a = 68.5˚, s = EE

Encounter state (E, ξ = 10.1±1.0 Å)



(1) 41.9 %
ξ = 9.7 Å, η = 10.6 Å
r = 6.2 Å, a = 164.2˚, s = GG

(2) 4.2 %
ξ = 9.6 Å, η = 8.5 Å
r = 2.5 Å, a = 56.1˚, s = EE

(3) 16.7 %
ξ = 10.8 Å, η = 9.1 Å
r = 5.7 Å, a = 173.7˚, s = GG

(4) 30.9 %
ξ = 8.8 Å, η = 12.2 Å
r = 6.3 Å, a = -179.5˚, s = GG

(5) 6.4 %
ξ = 10.6 Å, η = 10.8 Å
r = 5.3 Å, a = 158.7˚, s = GG

* ξ: Protein-ligand distance, η: Protein-lignad (terminal alkyl carbon) distance
* r: G113-S117 distance, a: χ1 angle, s: state
* Reference X-ray strcuture: PDB ID: 4W54

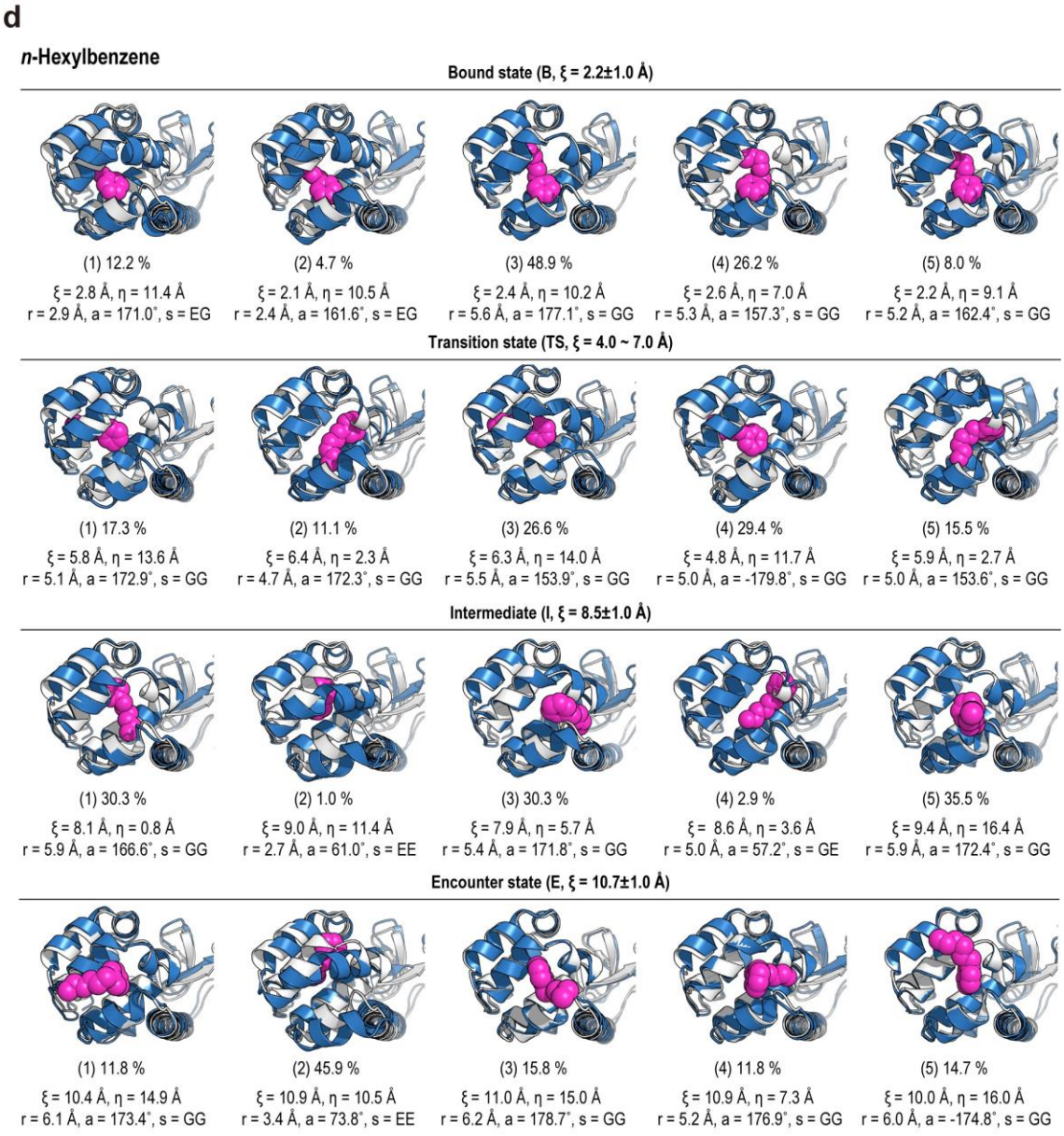


Figure S8. Results of clustering analysis. (a) The conformational states of FG helices are classified into EE (Excited-Excited), EG (Excited-Ground), GE, and GG according to the backbone hydrogen bond distance between G113 and S117 (O–HN distance) and χ1 dihedral angle (C-CA-CB-CG) of Phe114. (b) The snapshots of the representative structures of clusters obtained from *k*-means analysis for each of the bound, transition state (TS), and encounter states of the benzene-T4L

L99A binding. All the snapshots are superimposed to the X-ray structure (PDB ID: 181L) as a reference. The relative population of each cluster, the O–HN distance, and χ1 dihedral angle are also written. The assigned conformational states of FG helices are given in parentheses. (c and d) The results for ethylbenzene and *n*-hexylbenzene.

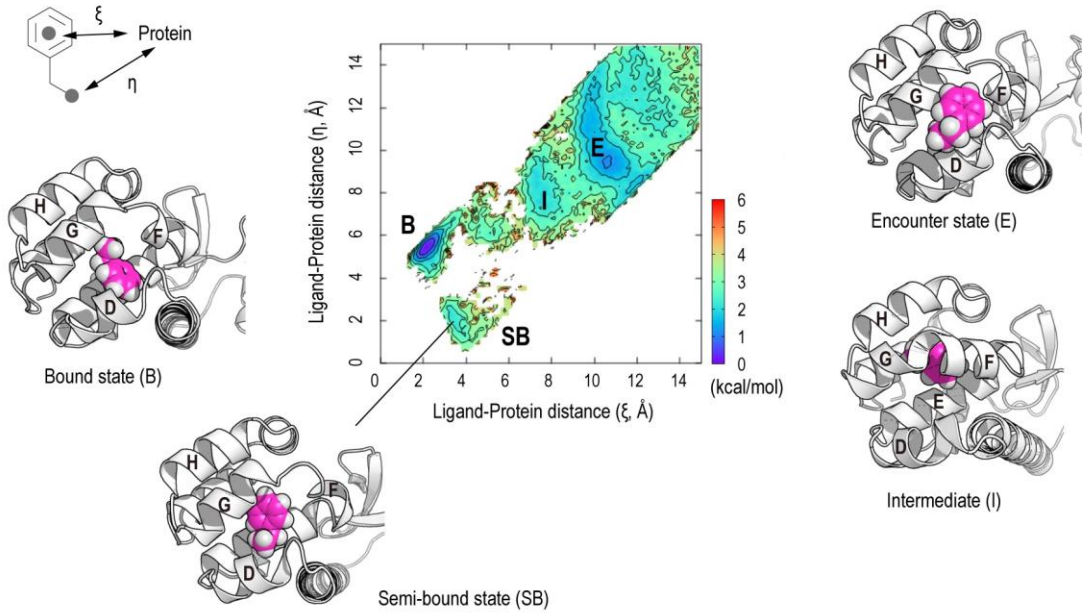


Figure S9. Binding pathway of ethylbenzene–T4L L99A. Free-energy landscape along two ligand-protein distances, ξ and η, for ethylbenzene binding at 300 K. A snapshot of the representative structure for each state is shown: Encounter state (E), intermediate (I), bound (B), and semi-bound states (SB).