Supplementary Information

Molecular dynamics simulations elucidate the role of the F–F′ loop in substrate entry into CYP3A4

Junfang Yan[1] and Hajime Hirao[1*]

[1] *Warshel Institute for Computational Biology, School of Medicine, The Chinese University of Hong Kong, Shenzhen, Guangdong, 518172, P. R. China*

*hirao@cuhk.edu.cn; Tel.: +86-0755-23519030

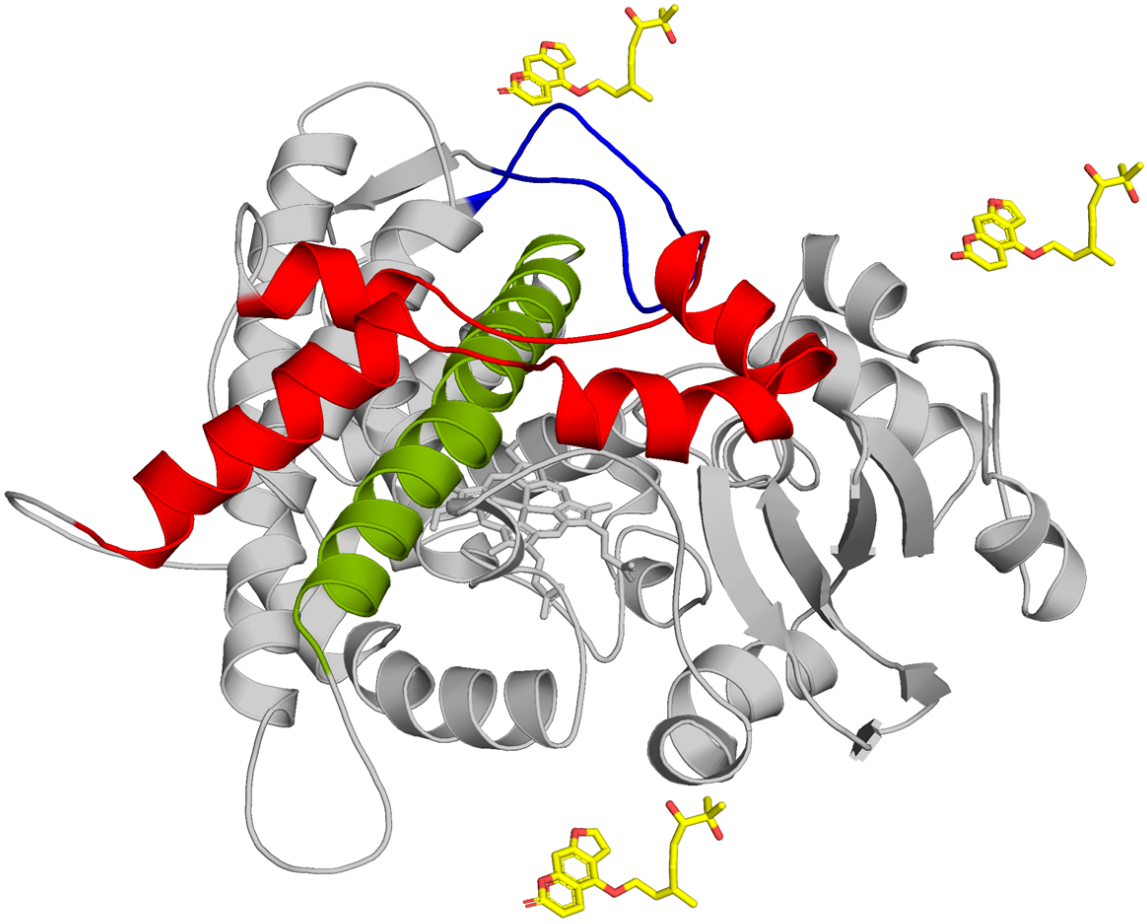
Contents

1. Protonation states of histidine residues

Supplementary Table 1. Histidine protonation states in CYP3A4 used in this study. Protonation states were determined based on pK_a values estimated using the PROPKA3 web tool[1] (assuming a pH of 7.5), and further refined through visual inspection of the local environments of residues.

Residue	Protonation state	Residue	Protonation state
H28	HIE	H267	HIE
H30	HIE	H287	HIE
H54	HID	H324	HIE
H65	HIE	H402	HID

2. Model construction



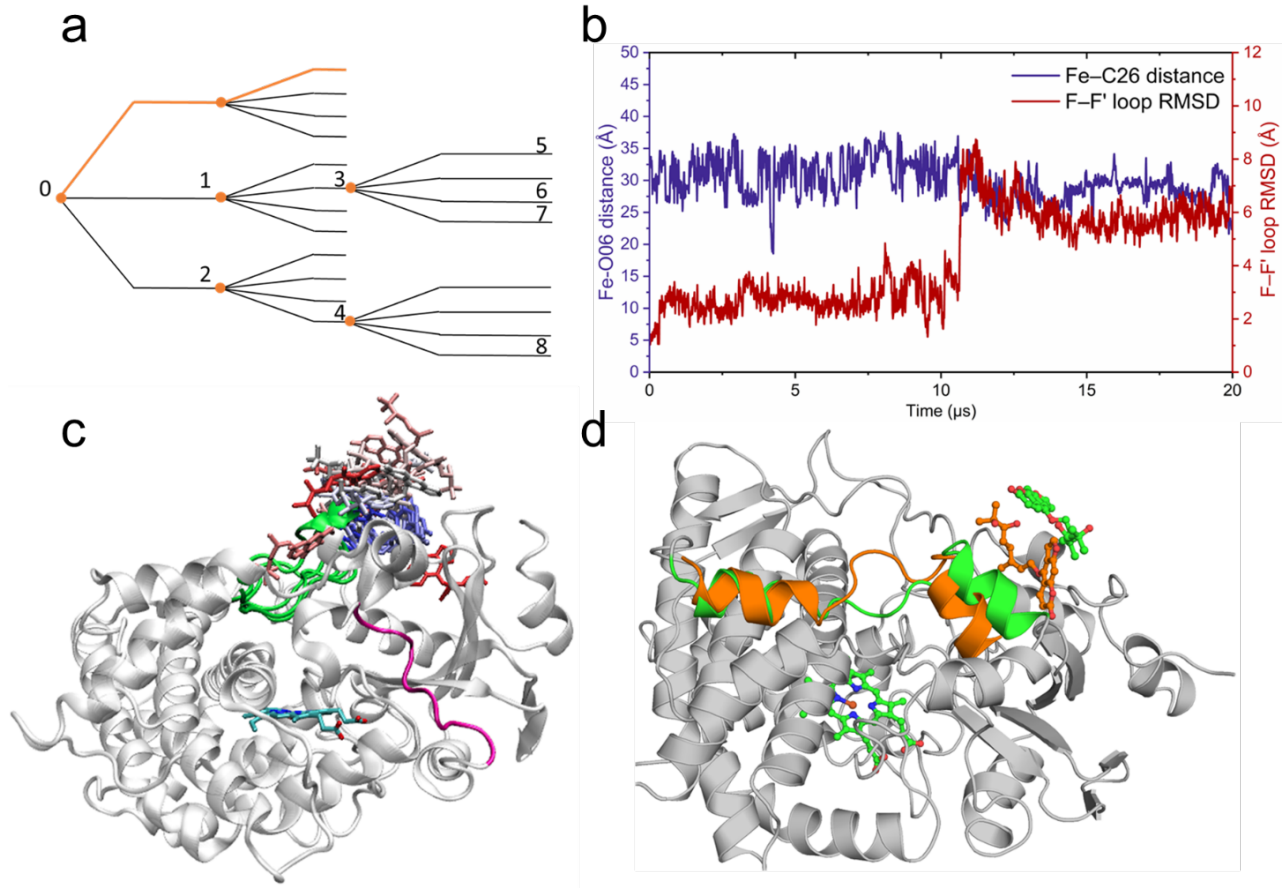
Supplementary Figure 1. Model construction. The initial structure was prepared for classical MD simulations prior to the addition of ions and water molecules. Three DHB molecules were randomly placed in the solvent, ensuring no direct contact with the protein. The protein structure is shown in dark gray with key regions highlighted: the F–G region (residues 202–258) in red, helix I (residues 291–323) in green, and the C-terminal loop (residues 464–490) in blue.

3. MD simulation data

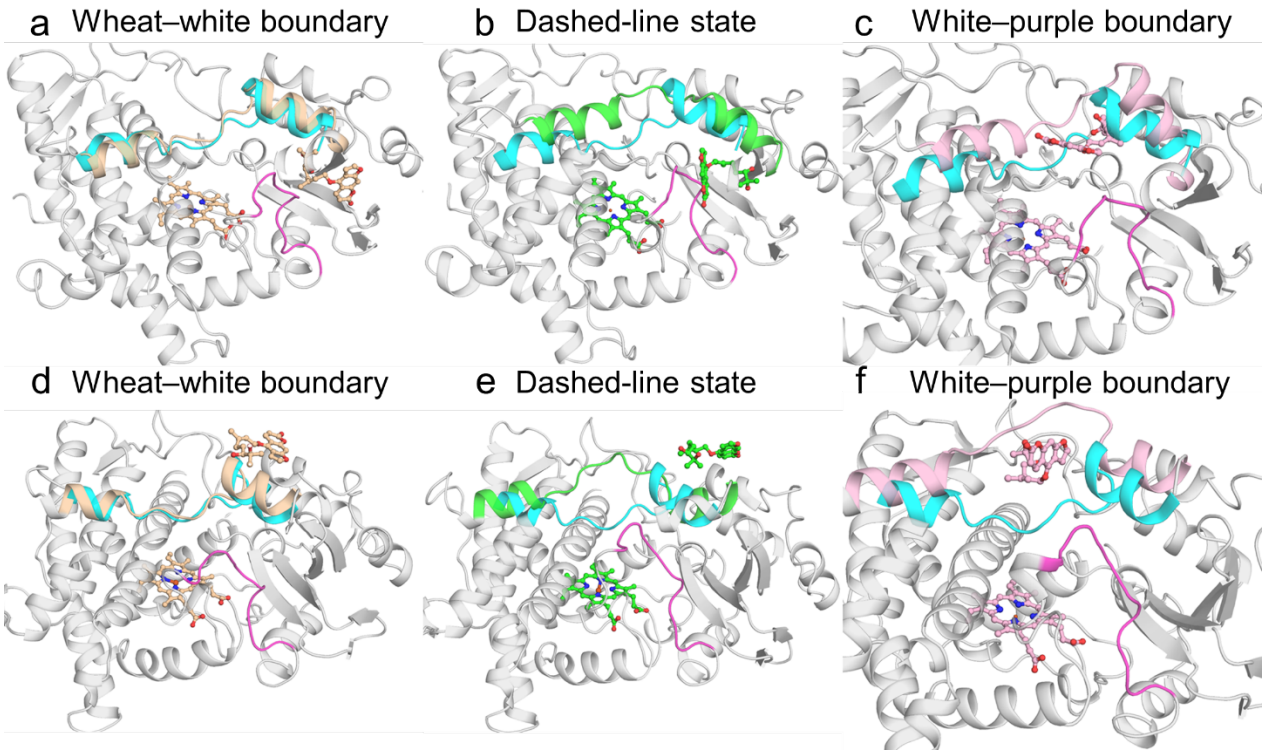
Supplementary Table 2. Cartesian coordinates of all atoms in DHB (in Å) and their RESP charges (in atomic units).

number	atom	x	y	z	charge
1	C01	-19.573594	-14.379129	-14.871737	-0.276612
2	C02	-20.324353	-15.397046	-14.044867	0.075543
3	C03	-21.762856	-15.038125	-13.726760	-0.130634
4	C04	-22.725028	-15.363223	-14.889921	-0.078819
5	C05	-24.215834	-15.116004	-14.619564	0.251050
6	O06	-24.897531	-15.744141	-15.716509	-0.710747
7	C07	-24.679032	-13.634231	-14.478269	0.525667
8	C08	-26.208820	-13.572360	-14.331588	-0.394565
9	C09	-24.217533	-12.756649	-15.647678	-0.394565
10	O10	-24.079444	-13.186239	-13.245067	-0.798027
11	C11	-19.713736	-16.529443	-13.651691	-0.355050
12	C12	-20.311133	-17.647557	-12.861103	0.263746
13	O13	-19.512217	-17.826463	-11.654746	-0.305790
14	C14	-19.839387	-18.900281	-10.870284	0.098727
15	C15	-20.365525	-18.659034	-9.582546	-0.247378
16	C16	-20.606784	-17.333762	-9.073081	0.181996
17	C17	-21.115279	-17.146057	-7.831703	-0.558697
18	C18	-21.436863	-18.269013	-6.961235	0.893739
19	O19	-21.897738	-18.202134	-5.842783	-0.585616
20	O20	-21.182755	-19.539902	-7.479644	-0.432580
21	C21	-20.669609	-19.754525	-8.729773	0.462378
22	C22	-20.475384	-21.080282	-9.106173	-0.515031
23	C23	-19.947659	-21.263940	-10.376615	0.334885
24	C24	-19.614532	-20.223407	-11.272424	0.084224
25	C25	-19.065672	-20.878837	-12.439275	-0.372715
26	C26	-19.113959	-22.207793	-12.178825	0.010221
27	O27	-19.647484	-22.473401	-10.935551	-0.225811
28	H1	-18.551503	-14.701812	-15.088172	0.090009
29	H2	-19.527059	-13.416658	-14.345993	0.090009
30	H3	-20.078745	-14.190781	-15.827468	0.090009
31	H4	-21.821448	-13.971862	-13.490828	0.102513
32	H5	-22.106379	-15.561941	-12.828633	0.102513
33	H6	-22.641318	-16.428978	-15.130946	0.076887
34	H7	-22.429462	-14.826059	-15.797922	0.076887
35	H8	-24.487639	-15.613791	-13.676044	0.018197
36	H9	-25.816934	-15.906159	-15.472311	0.415243
37	H10	-26.719301	-13.835700	-15.263307	0.097294

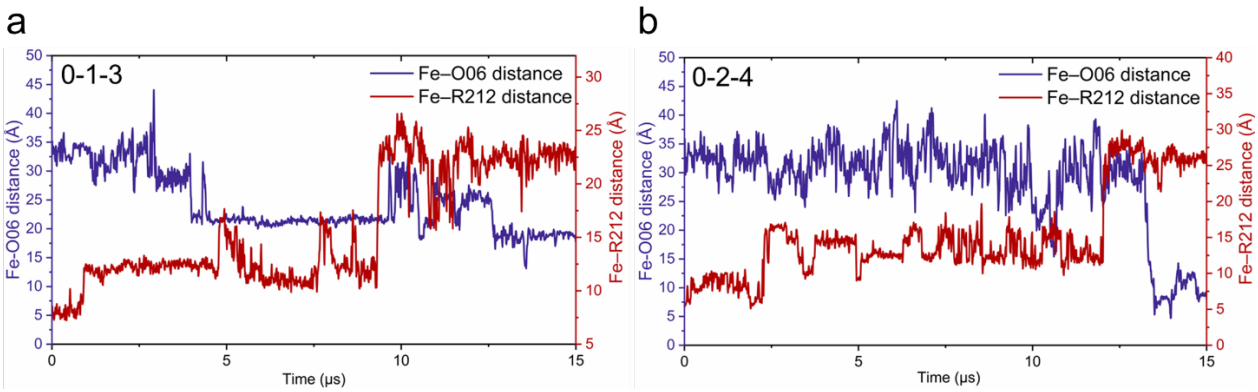
38	H11	-26.550643	-14.235823	-13.529158	0.097294
39	H12	-26.518748	-12.552343	-14.074910	0.097294
40	H13	-24.648284	-11.751636	-15.564163	0.097294
41	H14	-23.129266	-12.655819	-15.655621	0.097294
42	H15	-24.537903	-13.185548	-16.602089	0.097294
43	H16	-24.390694	-12.290770	-13.056075	0.449015
44	H17	-18.672058	-16.684654	-13.930766	0.159978
45	H18	-20.293445	-18.582554	-13.435296	0.009455
46	H19	-21.348013	-17.453241	-12.567391	0.009455
47	H20	-20.362217	-16.490095	-9.710129	0.100795
48	H21	-21.305640	-16.161758	-7.421415	0.202599
49	H22	-20.714164	-21.898159	-8.438092	0.243346
50	H23	-18.674041	-20.420922	-13.335176	0.219016
51	H24	-18.816353	-23.076960	-12.744684	0.160769



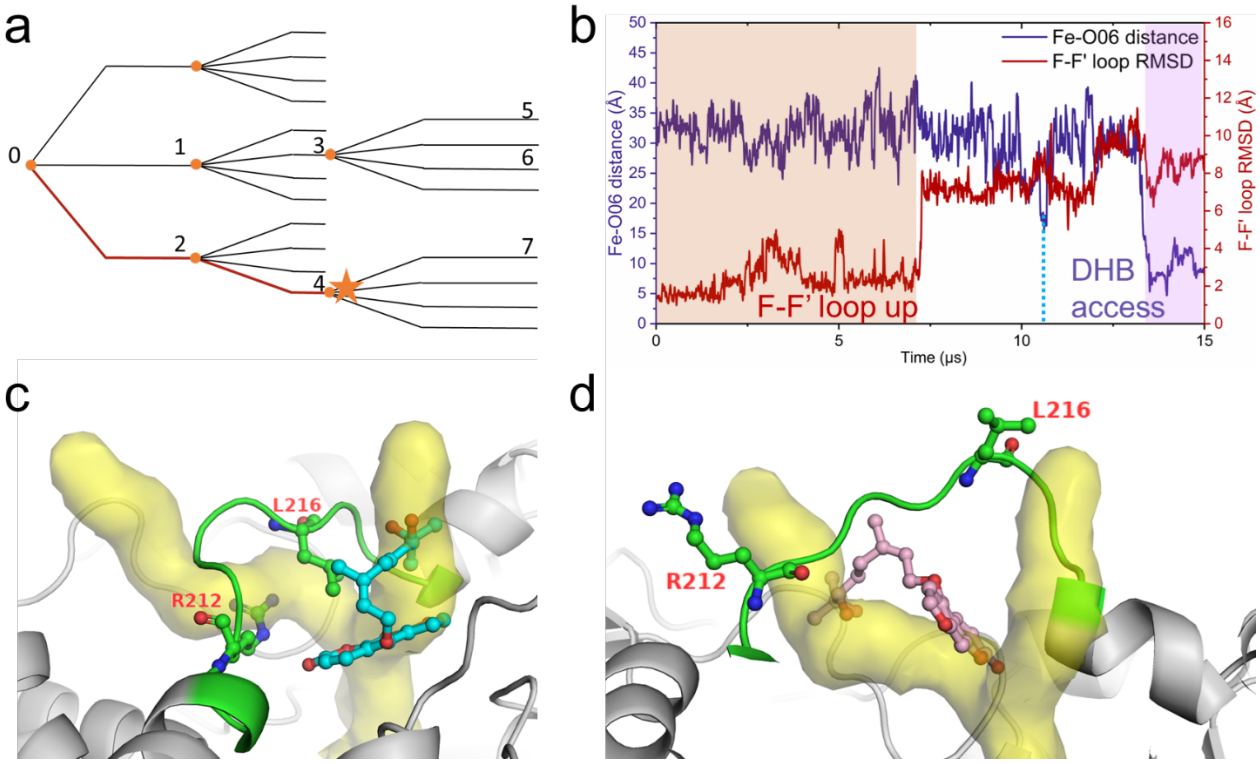
Supplementary Figure 2. (a) DHB binding processes observed in long, unbiased MD simulations. The orange line indicates the specific simulation trajectory examined in detail. (b) Time evolution of Fe–C26 distance and F–F′ loop RMSD along the orange trajectory shown in panel a. Trajectory frames were aligned to the crystal structure before RMSD calculations. The F–F′ loop remained in a downward-oriented conformation for the first 11 μs, after which it shifted upward and adopted a raised orientation. (c) The color gradient (red to blue) represents the DHB molecule's positional changes over simulation time. Multiple F–F′ loop conformations (shown in green) illustrate its dynamic behavior. One DHB molecule remained near the entrance of channel 2a (between helices A′ and F′), consistent with the binding site observed in trajectory 0-2-4. (d) Structural comparison of representative conformations, highlighting differences in the F–F′ region between the first 11 μs (green) and the subsequent 9 μs (orange) of the simulation.



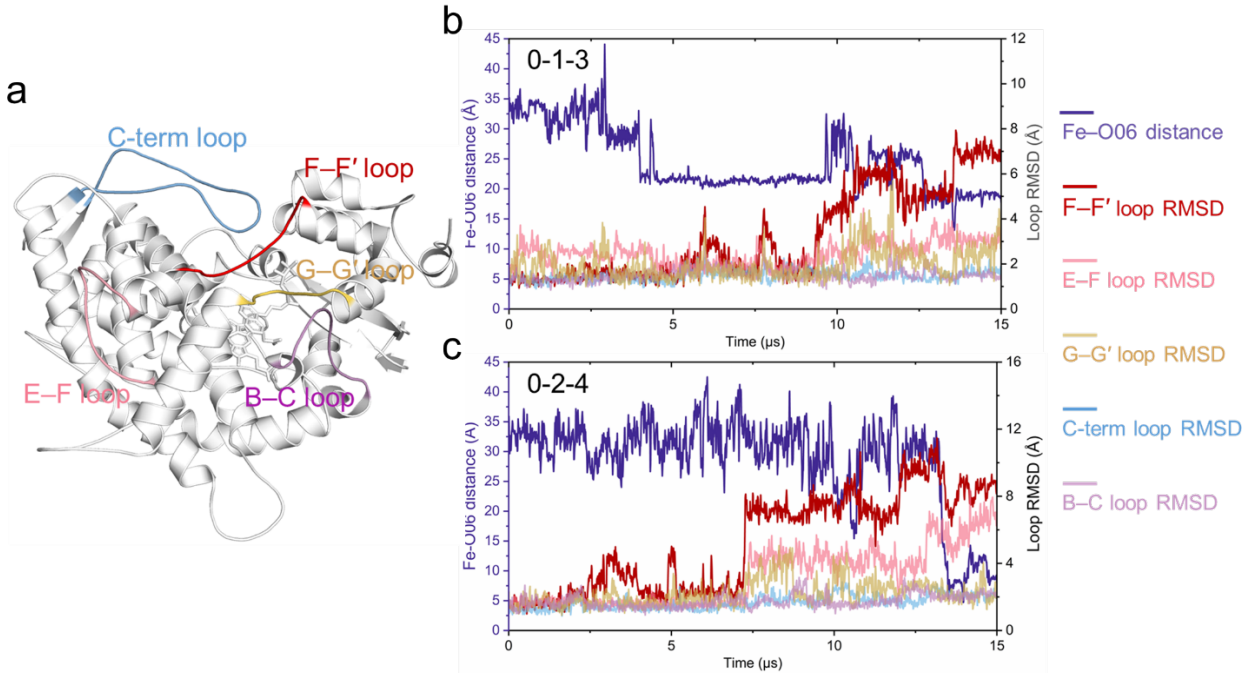
Supplementary Figure 3. Comparison of the F–F′ regions between the crystal structures (cyan) and the simulated structures along trajectories 0-1-3 (a–c) and 0-2-4 (d–f). A portion of the B–C loop is highlighted in magenta. (a/d) The wheat-colored F–F′ loops and DHB molecule represent the structures at the wheat–white boundary in Fig. 4a,b of the main text. (b/e) The green-colored F–F′ loops and DHB molecule represent the structures at the dashed-line state in Fig. 4a,b of the main text. (c/f) The pink-colored F–F′ loops and DHB molecule represent the structures at the white–purple boundary in Fig. 4a,b of the main text.



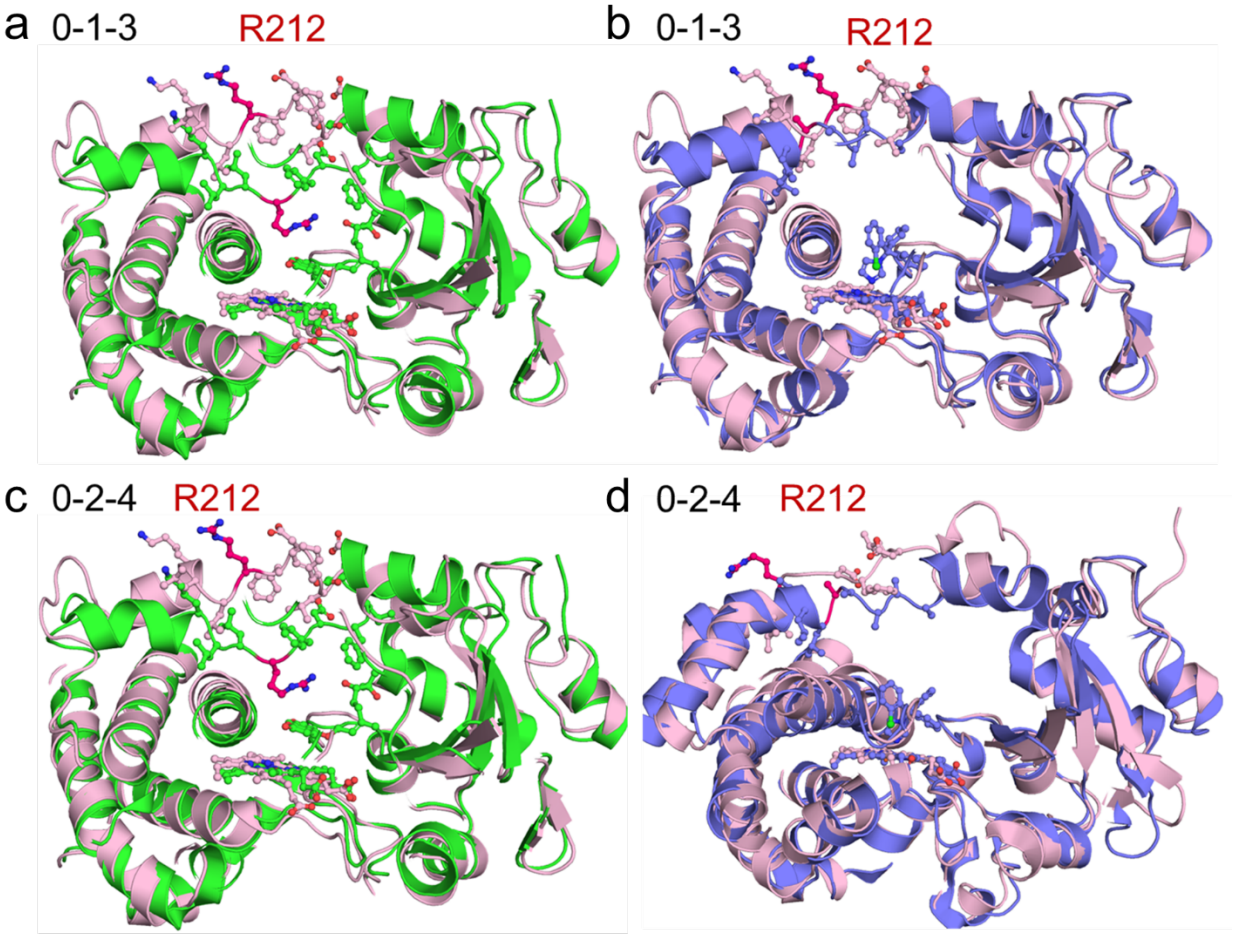
Supplementary Figure 4. Key geometric parameters during the DHB binding process, highlighting the upward movement of the side chain of R212 in trajectories 0-1-3 and 0-2-4. (a/b) Time evolution of the Fe–O06 and Fe–R212 distances in trajectories 0-1-3 (a) and 0-2-4 (b). The Fe–R212 distance was calculated as the distance between Fe and the center of mass of the NE, CZ, NH1, and NH2 atoms in the side chain.



Supplementary Figure 5. (a) DHB binding processes observed in long-timescale, unbiased MD simulations. The red line indicates trajectory 0-2-4, which was analyzed in detail. (b) Time evolution of the Fe–O06 distance and F–F′ loop RMSD in trajectory 0-2-4. Trajectory frames were aligned to the crystal structure prior to RMSD calculation. (c) Structure corresponding to the cyan dashed-line state in panel b. The downward orientation of the side chains of residues R212 and L216 impeded further progression of DHB toward the active site. (d) Structure corresponding to the white–purple boundary in panel b. The upward orientation of the side chains of R212 and L216 enabled DHB entry into the active site.



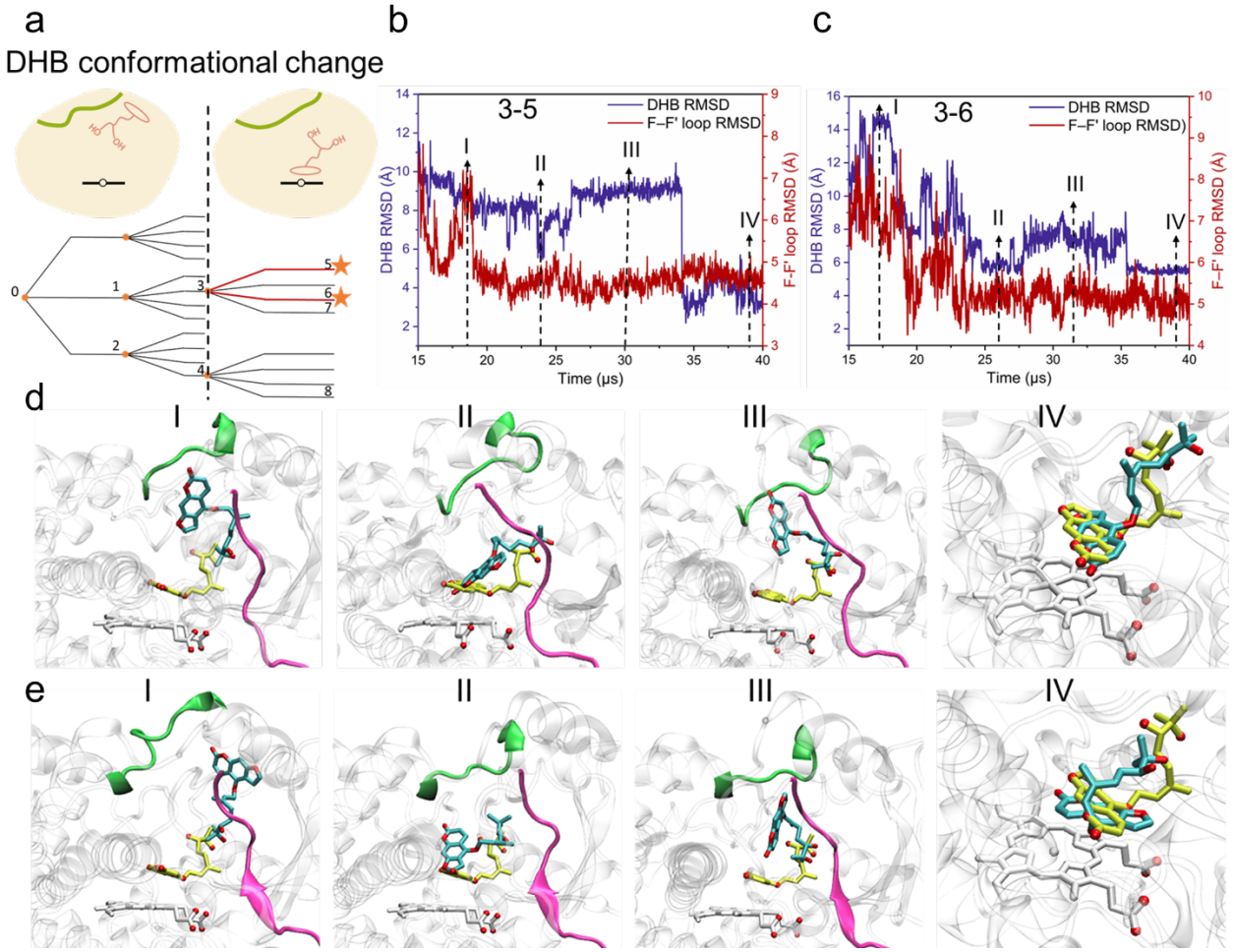
Supplementary Figure 6. (a) Locations of key loops in the CYP3A4 structure. (b-c) Time evolution of the Fe–O06 distance and RMSD changes of key loops during simulations in trajectories 0-1-3 (b) and 0-2-4 (c). All trajectory frames were aligned to the crystal structure prior to RMSD calculations. Among the loops analyzed, the F–F′ loop exhibited the largest RMSD fluctuations.



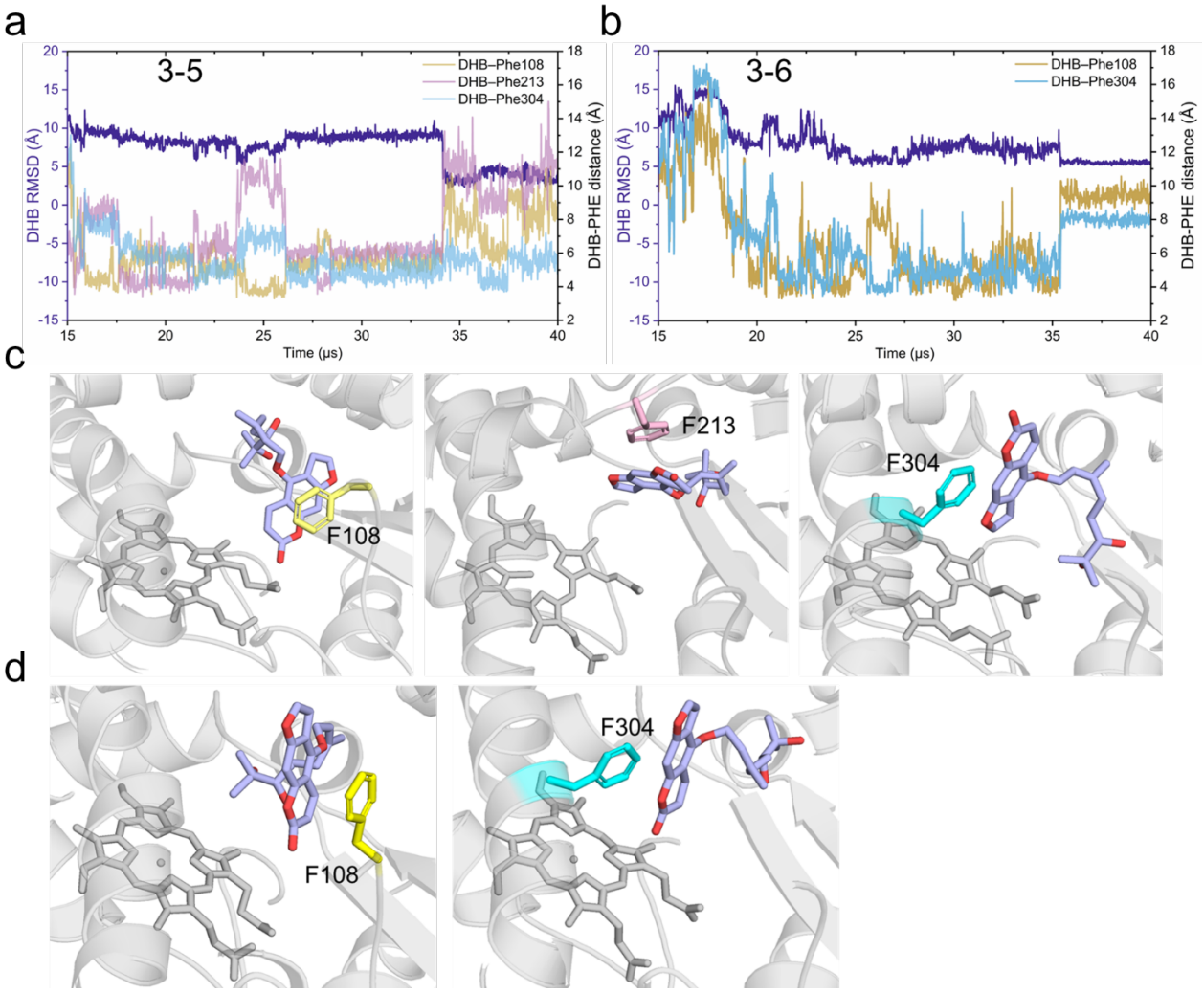
e

RMSD (Å)	Panel a	Panel b	Panel c	Panel d
All structure	1.05	1.17	1.10	1.20
F–F′ loop	6.76	4.80	11.68	6.52

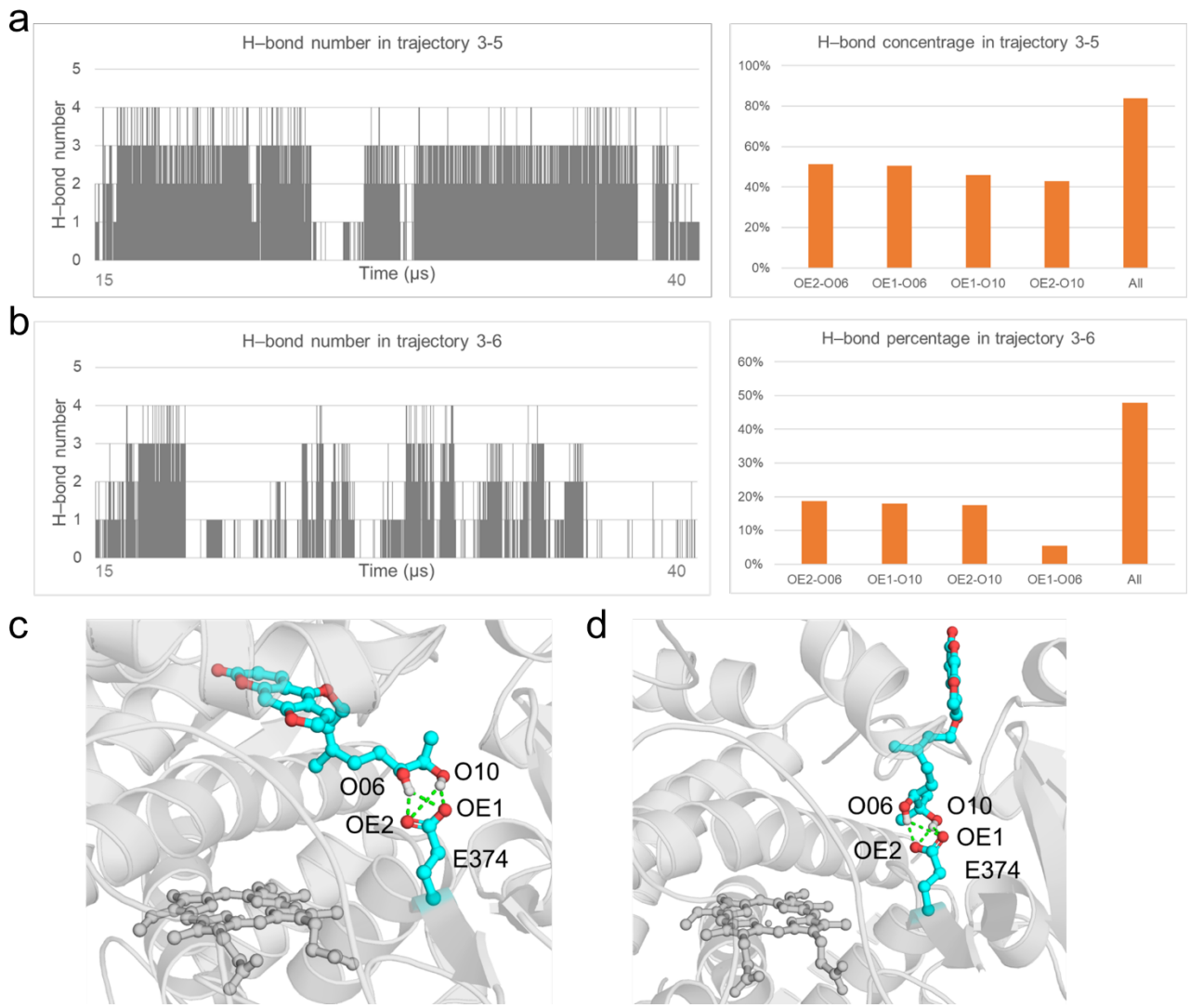
Supplementary Figure 7. (a/c) Structural comparisons between the crystal structure (PDB ID: 6OOB) and representative structures from the purple regions in Fig. 4a,b of the main text. The crystal structure is shown in green, while the representative structures from the purple regions are colored pink. (b/d) Structural comparisons between the crystal structure (PDB ID: 8SPD) and representative structures from the purple regions in Fig. 4a,b of the main text. The crystal structure is shown in purple, and the representative structures from the purple regions are colored pink. To aid visualization, side chains of residues in the F–F′ loop are displayed in licorice representation. Note that the side chain of residue R212 (in red) is partially missing in the 8SPD crystal structure, with only the β-carbon of the side chain of R212 resolved; to avoid introducing artificial bias, the raw structure was used without manual modification. (e) RMSD values of backbone atoms are shown for both the entire structure and the F–F′ loop regions in panels a-d.



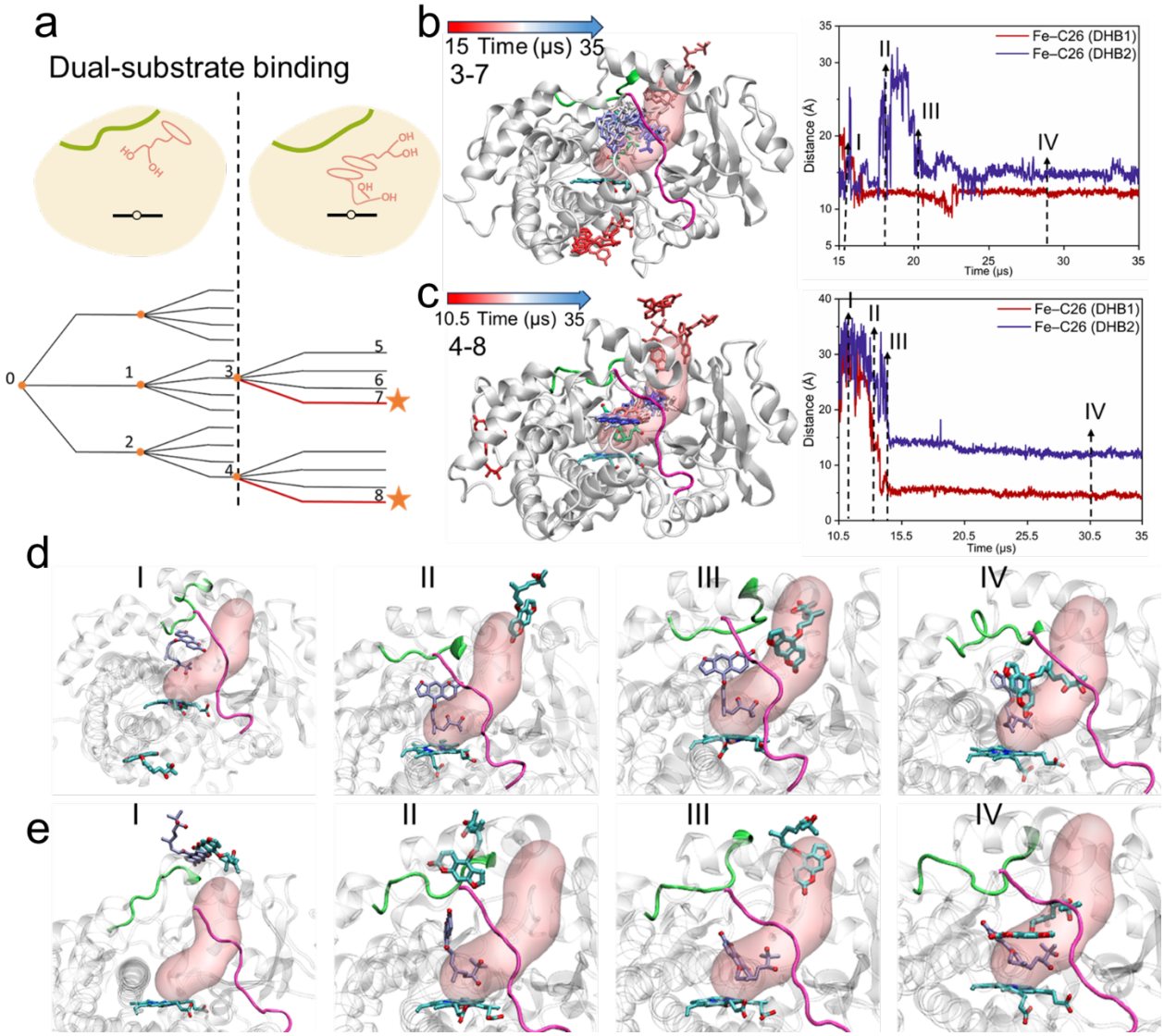
Supplementary Figure 8. DHB conformational changes in CYP3A4. (a) DHB conformational changes in the active site of CYP3A4 observed in two independent unbiased MD simulations. The trajectories are named 3-5 and 3-6, respectively. (b/c) Time evolution of DHB and F–F′ loop RMSD values during MD simulations 3-5 (b) or 3-6 (c). (d/e) Snapshots of DHB binding poses in trajectories 3-5 (d) and 3-6 (e) corresponding to states I, II, III, and IV in panels b and c. The DHB molecule from the crystal structure is shown in yellow. The F–F′ loop is colored green, and a portion of the B–C loop is colored magenta.



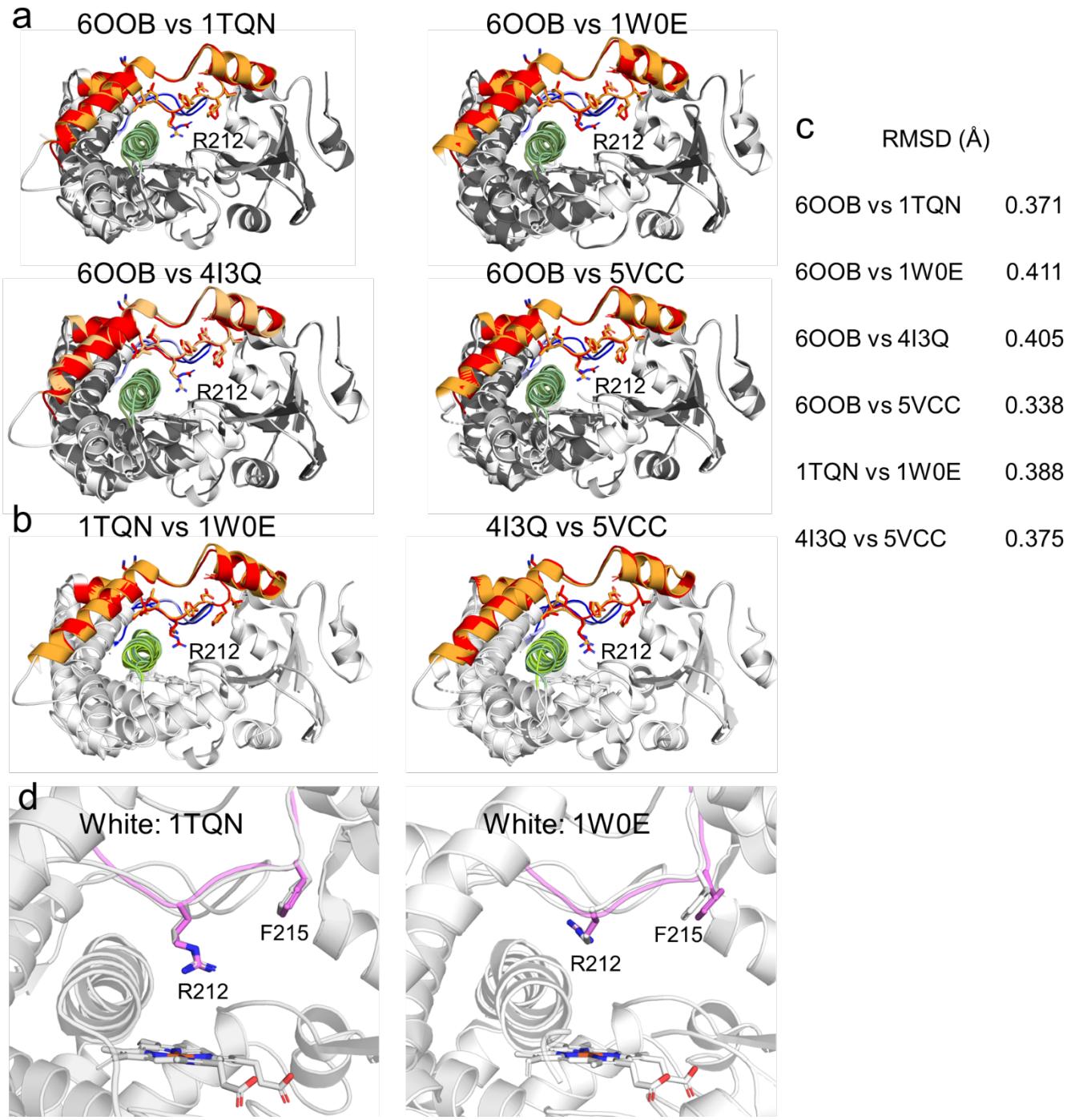
Supplementary Figure 9. Aromatic (π–π stacking-like) contacts between the psoralen moiety of DHB molecule and phenylalanine residues. (a/b) Time evolution of DHB RMSD and the distance between the central benzene ring of DHB and the benzene ring of selected phenylalanine residues in trajectory 3-5 (a) or 3-6 (b). (c/d) Representative structures illustrating aromatic contacts with phenylalanine residues in trajectories 3-5 (c) and 3-6 (d).



Supplementary Figure 10. (a/b) Hydrogen bond interactions between the tail of DHB and E374. Left: Time-dependent changes in the number of hydrogen bonds between DHB and E374 in trajectories 3-5 (a) and 3-6 (b). Right: Hydrogen bond percentages for OE2(E374)–O06(DHB), OE1(E374)–O06(DHB), OE2(E374)–O10(DHB), and OE1(E374)–O10(DHB). "All" refers to frames in which any hydrogen bond was formed between the tail of DHB and E374. Percentages were calculated as the number of frames in which the corresponding hydrogen bonds were present, divided by the total number of frames in each simulation. (c/d) Representative structures showing hydrogen bond interactions between DHB and E374 in trajectories 0-1-3 (c) and 0-2-4 (d).

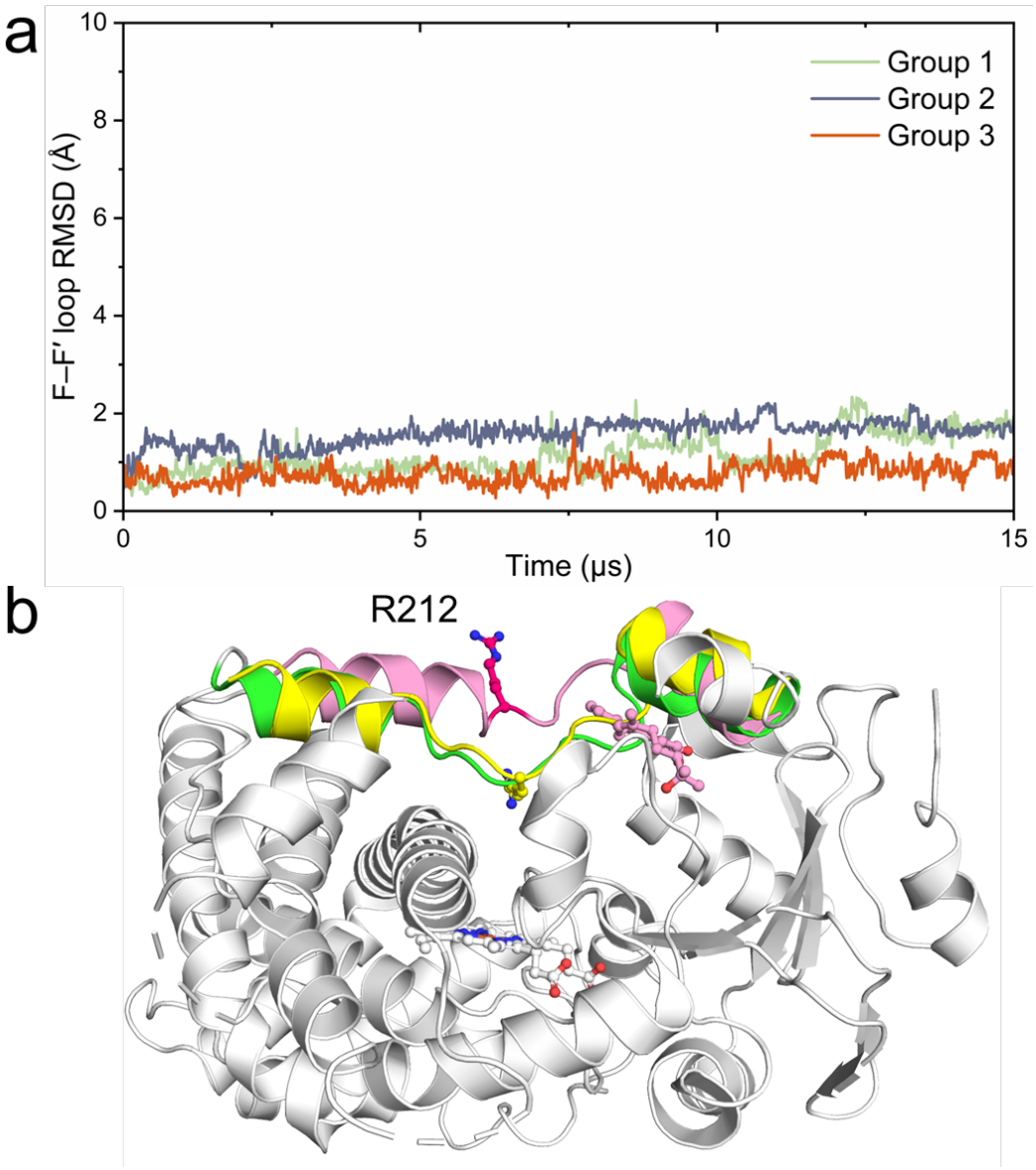


Supplementary Figure 11. (a) Dual-substrate binding scheme of DHB in CYP3A4. (b/c) Left: Positional changes of the second bound DHB molecule in the active site of CYP3A4 in trajectory 3-7 (b) and 4-8 (c). Right: Time evolution of the Fe–C26 distance for the first (DHB1) and second (DHB2) bound DHB molecules. (d/e) Snapshots of dual DHB molecule binding in trajectories 3-7 (b) and 4-8 (c), corresponding to states I, II, III, and IV in panels b and c. The first bound DHB molecule is colored purple, and the second bound DHB molecule is colored cyan. For consistency, the F–F′ loop is colored green, and a portion of the B-C loop is colored magenta.

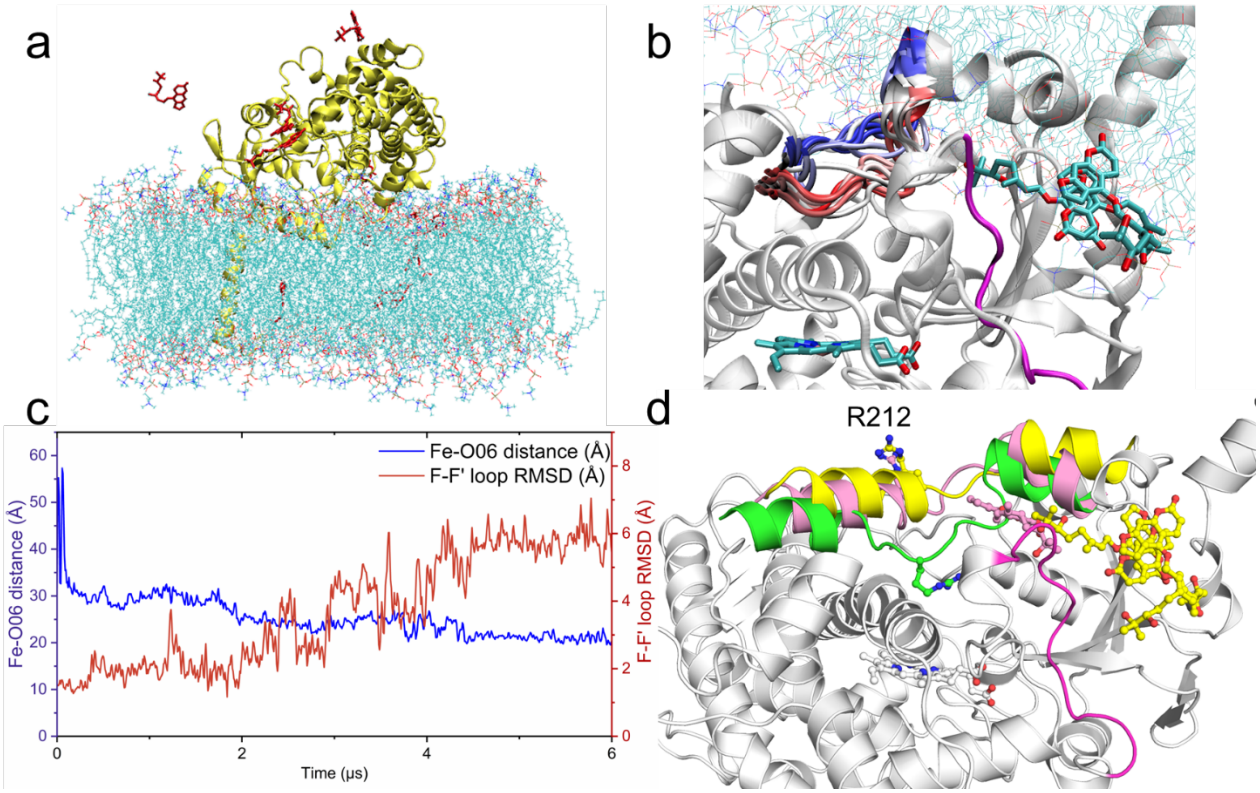


Supplementary Figure 12. (a) Structural comparison of DHB-bound CYP3A4 (PDB ID: 6OOB) with apo structures (1TQN, 1W0E, 4I3Q, 5VCC). The DHB-bound structure is shown in gray, with the F–G region (residues 202–258) highlighted in red, helix I (residues 291–323) in green, and the C-terminal loop (residues 464–490) in blue. The corresponding regions in the apo structures (shown in white) are rendered with lighter shading for clarity. The side chains of the F–F′ loop are shown, revealing close alignment between the DHB-bound and apo structures. (b) Comparison among apo CYP3A4 structures. Structures from 1TQN and 4I3Q are depicted with darker shading, while those

from 1W0E and 5VCC are shown with lighter shading. (c) RMSD values corresponding to the comparisons shown in panels a and b. The DHB-bound CYP3A4 structure closely resembles the apo structures, with only minor differences in the orientations of F–F′ loop residues (including R212), a variation also observed among different apo structures. (d) Comparison of the R212 and F215 side chains between apo structures and MD simulation structures. The overall white-colored structures represent apo structures (1TQN and 1W0E), with the R212 and F215 side chains shown in licorice representation. The pink-colored structures correspond to different snapshots from the MD simulations.



Supplementary Figure 13. (a) Time evolution of the F–F′ loop RMSD in three MD simulation sets conducted in the absence of DHB. (b) Structural comparison of the F–G region between a representative structure from MD simulations with DHB (pink) and one from MD simulations without DHB (yellow). Both representative structures were extracted from the final frames of 15-μs simulations. The overall CYP3A4 structure is shown based on the crystal structure (PDB ID 6OOB), with the F–G region highlighted in green.



Supplementary Figure 14. (a) Construction of a membrane-bound CYP3A4 model. The N-terminal transmembrane helix was added based on the AlphaFold-predicted structure, and the full-length protein was embedded in a lipid bilayer with ten DHB molecules distributed around its surface. (b) F–F′ loop dynamics. The positional changes of the F–F′ loop during the simulation are shown with a red-to-blue color gradient, while a portion of the B–C loop is highlighted in magenta. Several DHB molecules remained stabilized at the entrance of channel 2b, consistent with observations from trajectory 0-1-3 (Fig. 3 in the main text). (c) Time evolution of the Fe–O06 distance for a selected DHB molecule and the RMSD of the F–F′ loop. (d) Structural comparison of the F–G region. The F–G regions from MD simulations without a membrane (pink), with a membrane (yellow), and from the crystal structure (PDB ID 6OOB, green) are shown, with a portion of the B–C loop highlighted in magenta. The overall CYP3A4 structure is based on the model containing the N-terminal helix. Representative structures were extracted from the final frames of the 15-μs simulation without a membrane and the 6-μs simulation with a membrane. The orientations in panels b and d are identical to those in Fig. 2 of the main text, with the heme group positioned horizontally and the F–F′ loop located above it.

Supplementary Table 3. MM-PBSA-calculated binding affinities (in kcal/mol) for complexes between the pentacoordinate ferric-heme and DHB in different conformations in trajectory 3-5.

	Binding energy (kcal/mol)
Pose I	-21 ± 7.8
Pose II	-25 ± 5.0
Pose III	-23 ± 5.6
Pose IV	-30 ± 5.4

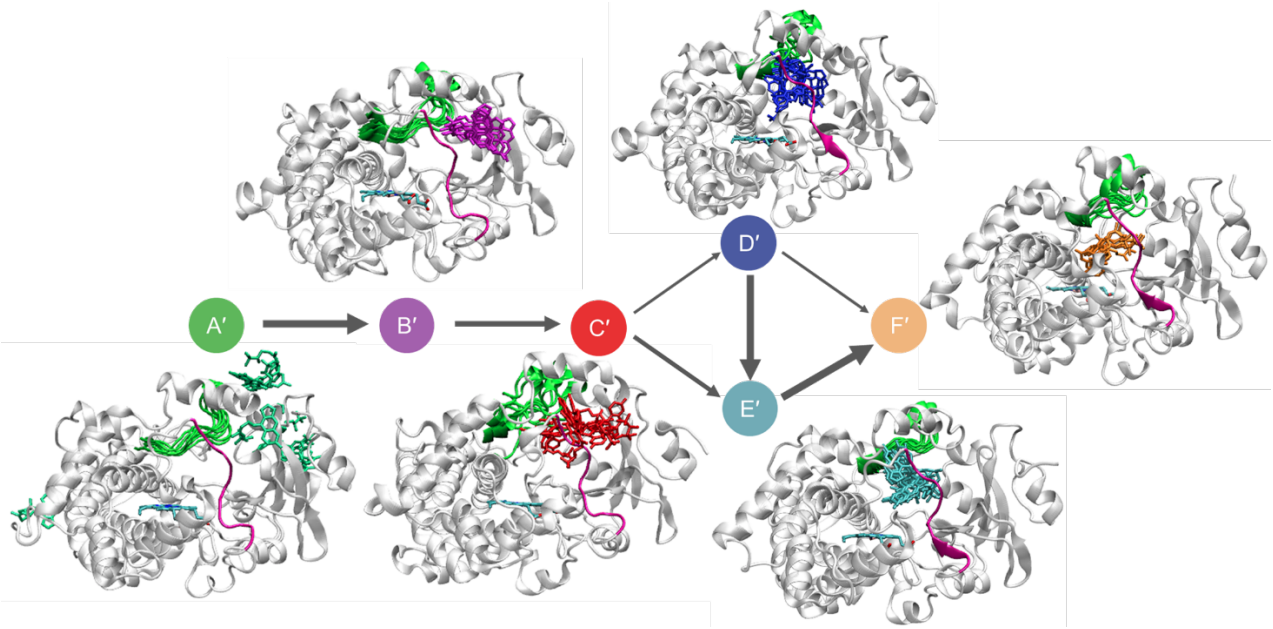
Supplementary Table 4. The simulation duration time for each stage.

	Simulation duration time (μs)
Trajectory 0-1, 0-2	11
Trajectory 1-3, 2-4	4
Trajectory 3-5, 3-6	25
Trajectory 3-7, 4-8	20
Unlabeled trajectory in Supplementary Fig. 2	20
Trajectory without DHB	15
Trajectory in the membrane environment	6

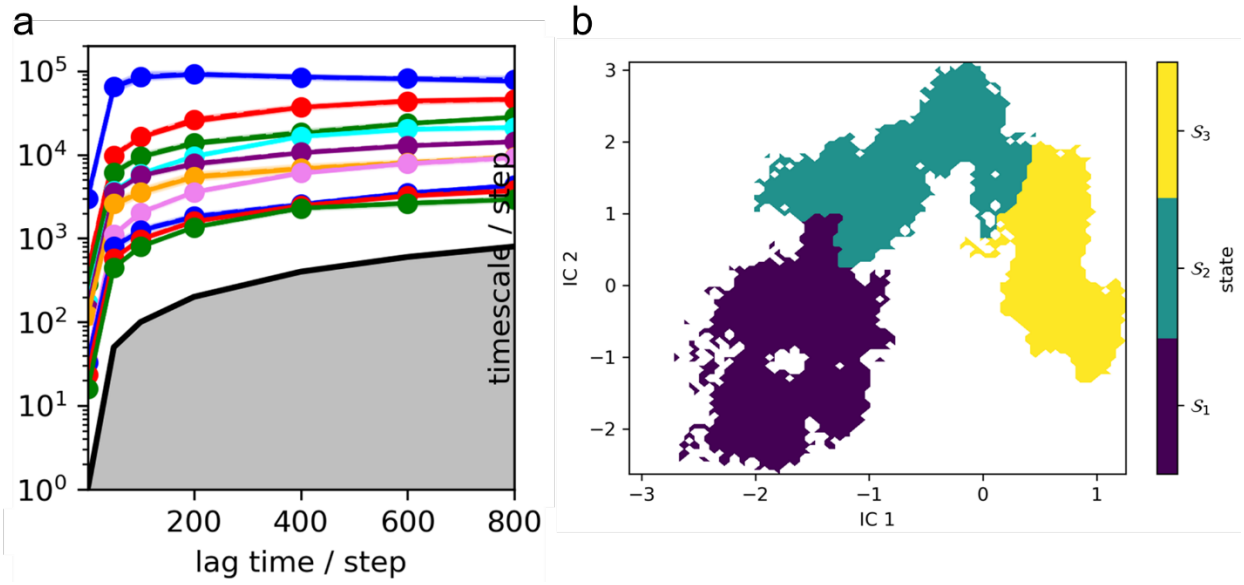
Supplementary Table 5. Summary of the Supplementary Movies.

Movie name	Description
Supplementary Movie 1	CYP3A4–DHB binding in trajectory 0-1-3
Supplementary Movie 2	CYP3A4–DHB binding in trajectory 0-2-4
Supplementary Movie 3	DHB conformational change in CYP3A4 (trajectory 3-5)
Supplementary Movie 4	DHB conformational change in CYP3A4 (trajectory 3-6)
Supplementary Movie 5	Dual-substrate binding in CYP3A4 (trajectory 3-7)
Supplementary Movie 6	Dual-substrate binding in CYP3A4 (trajectory 4-8)

4. MSM data



Supplementary Figure 15. A six-macrostate MSM characterizing the DHB binding process. All conformations were extracted from the MSM constructed along the trajectory leading to the native binding pose (trajectory 0-1-3-5). Ten representative snapshots of the F–F′ loops (colored green) and DHB molecules from each macrostate are shown. Each node in the network represents a macrostate, with the thickness of the arrows indicating the net flux between them. Model A′: DHB molecules are randomly distributed near the surface of CYP3A4. Model B′: DHB is positioned at the entrance of channel 2b. Model C′: DHB enters the active site, accompanied by an upward movement of the F–F′ loop. Models D′ and E′: DHB undergoes conformational changes within CYP3A4. Model F′: DHB adopts the final bound state in CYP3A4 with the F–F′ loop beginning to move downward.



Supplementary Figure 16. (a) The top 10 slowest implied timescales plotted on a logarithmic scale as a function of lag time. The trajectory step size is 0.1 ns. The final MSM was constructed using lag time of 60 ns (600 steps). (B) Mapping of the coarse-grained MSM with three macrostates onto the IC1–IC2 coordinate space.

5. Supplementary references

(1) Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7**, 525-537(2011**)**.