P450-BM3 Catalyzed Sulfoxidation versus Hydroxylation: A Common or Two Different Catalytically Active Species?

Jian-bo Wang‡ [a,b,*], Qun Huang‡ [a,b], Wei Peng,[c] Peng Wu,[c] Da Yu [a,b], Bo Chen[a,b], Binju Wang[c,*] and Manfred T. Reetz[d,e*]

[a] Key Laboratory of Phytochemistry R&D of Hunan Province, College of Chemistry and Chemical Engineering, Hunan Normal University, 410081 Changsha, People's Republic of China

[b] Key Laboratory of Chemical Biology and Traditional Chinese Medicine Research (Ministry of Education), College of Chemistry and Chemical Engineering, Hunan Normal University, 410081 Changsha, People's Republic of China

[c] State Key Laboratory of Physical Chemistry of Solid Surfaces and Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 360015, P. R. China

[d] Chemistry Department, Philipps-University, Hans-Meerwein-Strasse 4, 35032 Marburg, Germany

[e] Max-Planck Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim, Germany

Table of Contents

Experimental Procedures

Bacterial strains, enzymes, chemicals and instruments

BL21 (DE3) was used to express P450-BM3 and glucose dehydrogenase (GDH).

The KOD hot start DNA polymerase was obtained from Novagen.

All chemicals were purchased from Geel Belgium, Bide Pharmtech Ltd, or Tokyo Chemical.

SHIMADZU 20 A high performance liquid chromatography (HPLC) with UV detector and SHIMADZU Nexis GC-2030 gas chromatograph (GC) with a flame-ionization detector (FID).

General information of directed evolution

Primers used in this work are listed in Table S1.

Table S1. List of primers used in mutagenesis.

Primer	Sequence
L75NNKF	GATAAAAACTTAAGTCAAGCGNNKAAATTTGTACGTGATTTTGCA
L75WTF	GATAAAAACTTAAGTCAAGCGCTCAAATTTGTACGTGATTTTGCA
F87NNKF	GATTTTGCAGGAGACGGGTTANNKACAAGCTGGACGCATGAAAAAA
I263NNKR	GACCACTTGTTGTTTCGTGTCCCGCMNNTAAGAATGTAATAATTTGATAG
A264NNKR	GACCACTTGTTGTTTCGTGTCCMNNAATTAAGAATGTAATAATTTGATAG
A328NNKR	TTTGCATATAGGGAAAACGCAGGMNNAGTTGGCCATAAGCGCAGCGCTTC
L437NNKR	CACAAAGCCTTCAGGTTTTAACGTMNNAGTTTCTTTAATATCCAGCTCG
T438NNKR	CCACAAAGCCTTCAGGTTTTAAMNNTAAAGTTTCTTTAATATCCAGC
A328WTR	TTTGCATATAGGGAAAACGCAGGAGCAGTTGGCCATAAGCGCAGCGCTTC

General procedure of directed evolution

Seven residues were first selected to do NNK-based saturation mutagenesis separately. The 2-step PCR method was applied to construct all of the libraries,[1] first amplifying the megaprimer containing all the diversity at target sites, second using the megaprimer to amplify the whole plasmid to express all the mutants. The plasmid pRSF-BM3 WT was used as template,

primers L75NNKF + A328WTR, F87NNKF + A328WTR, L75WTF + I263NNKR, L75WTF + A264NNKR, L75WTF + A328NNKR, L75WTF + L437NNKR, L75WTF + T438NNKR were used to amplify megaprimers. After all the megaprimers were confirmed by DNA agarose electrophoresis, they were used directly to amplify the whole plasmid. Templates were erased by *Dpn*I digestion at 37 ℃ for 7 hours after the PCR products were confirmed by electrophoresis. To get the best mutants, mutant F87L was used as template. Primers L75NNKF and A328NNKR were used to amplify the megaprimer. This led to saturation mutagenesis at L75 and A328 simultaneously. As demonstrated above, the obtained megaprimer was used to amplify the whole plasmid, then it was incubated with *Dpn*I to digest the template.

Library screening was performed as in previous work.[2]

All P450-BM3 variants in this work are listed in Table S2.

Table S2. NNK-based saturation mutagenesis derived P450-BM3 variants as catalysts in the asymmetric sulfoxidation of thiochroman-4-one (**1**).

P450-BM3 variants	geno-types	er product
WT		75:25 (*S*)
L75N	L75N	80:20 (*S*)
L75S	L75S	80:20 (*S*)
F87L	F87L	53:47 (*R*)
F87V	F87V	57:43 (*S*)
A328L	A328L	62:38 (*S*)
A328I	A328I	67:33 (*S*)
L75F-F87L-A328F	L75F-F87L-A328F	92:8 (*S*)

Expression and purification of P450-BM3

P450-BM3 was expressed and purified as described.[3]

The preparation of GDH lysate

Single colonies were carefully picked into a 10 mL culture tube with 3-6 mL LB media containing 35 μg/mL chloramphenicol, incubated at 37 ℃, 220 rpm overnight. 5 mL seed culture was transferred into the 500 mL TB media containing 35 μg/mL

chloramphenicol, incubated at 37 ℃, 220 rpm for about 2-3 h. When the optical density (OD_{600}) of cells reached 0.6-0.8, 250 μL of 1 M IPTG (isopropyl β-D-1-thiogalactopyranoside) were added into the expressed system, and the protein was expressed at 18 ℃, 220 rpm for 18h. Cells were gathered by centrifuge (8000 rpm, 2 min), washed with PBS buffer (pH 8.0), and concentrated to 10~15 mL. Lysozyme was added, stored at -80 ℃ until lysis.

To prepare the lysate powder, cells were taken out from the -80 ℃ refrigerator and thawed in ice-water at 4 ℃ overnight, and then disrupted by sonication for 15 min (400 W 2 s pulse and 4 s pause; SCIENTZ-IID). The cell debris was removed by ultracentrifugation (12000 rpm, 1 h, 4 ℃). The supernatant was quick-frozen by liquid nitrogen and then freeze-dried (SCIENTZ-10ND) to obtain lysate.

Kinetic experiments

The general reaction was performed in 100 mM phosphate buffer (pH=8.0). The system contained 100 mM glucose, 200 μM $NADP^+$, 10 mg/mL GDH, 2 mM substrate and 10 μM P450-BM3 enzyme. Reactions were performed at 30 ℃, 800 rpm for 9 h (TON calculation) or 2 h (TOF calculation). To quench the reactions, 30 μL concentrated hydrochloride were added into the systems, substrate and product were extracted by ethyl acetate (300 μL). The concentrations of substrates and products were determined by HPLC or GC. For the kinetic tests in the presence of catalase, additional high concentration of catalase (>1200 U/mL) was used. As the activity of wild type and mutant enzymes are greatly different for two different substrates, enzyme concentration and reaction time were slightly adjusted for some compounds.

Detection methods

The bioconversions of substrate **1** was detected by HPLC using substrate standards and 6-point calibration curves. Sample injection volume was 5 μL. Analysis was performed in triplicate on a SHIMADZU 20 A high performance liquid chromatography with UV detector. Direct analysis of the extract was performed on a chiral CHIRALPAK® column (DAICEL CORPORATION). Mobile phase is ethanol (A) and *n*-heptane (B). The HPLC method: 0.01~6 min 70% B, 6~6.01min 70~5% B, 6.01~13min 5~5% B.

The bioconversions of substrate **3** were detected by GC using substrate standards and 6-point calibration curves. Sample injection volume was 2 μL. Analysis was performed in triplicate on a SHIMADZU Nexis GC-2030 gas chromatograph with a flame ionization detector (FID). Direct analysis of the extract was performed on a chiral LIPODEX®E column

(MACHEREY-NAGE). The GC method: 200 ℃ inlet, 220 ℃ detector, 60 ℃ oven for 0 min, then 20 ℃/min gradient to 200 ℃, holding for 6 min.

Graphs of reactions

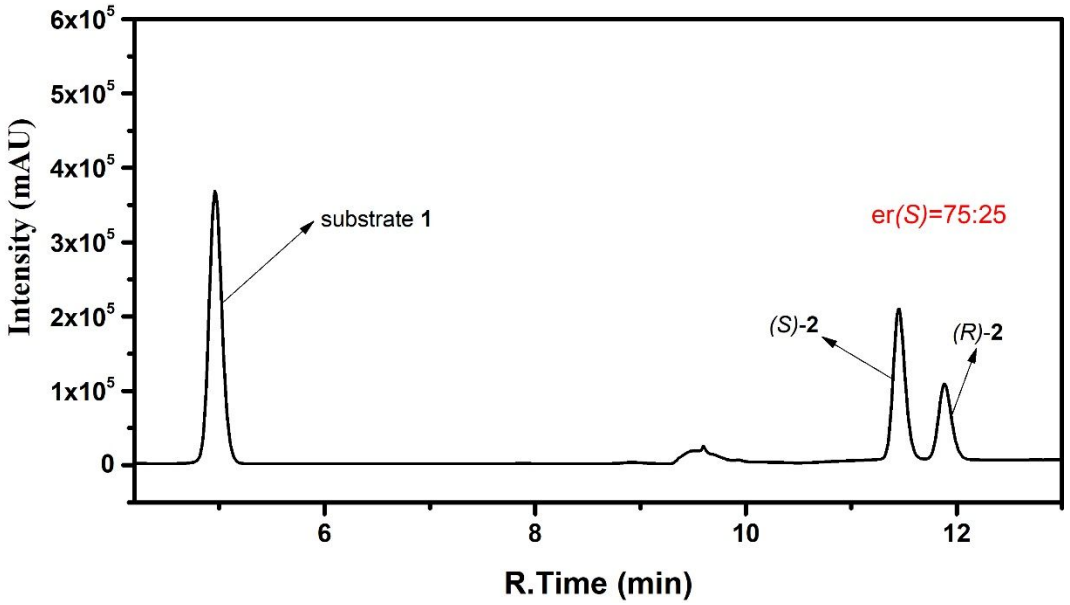

Figure S1. HPLC graph of WT-catalyzed reaction of substrate **1**.

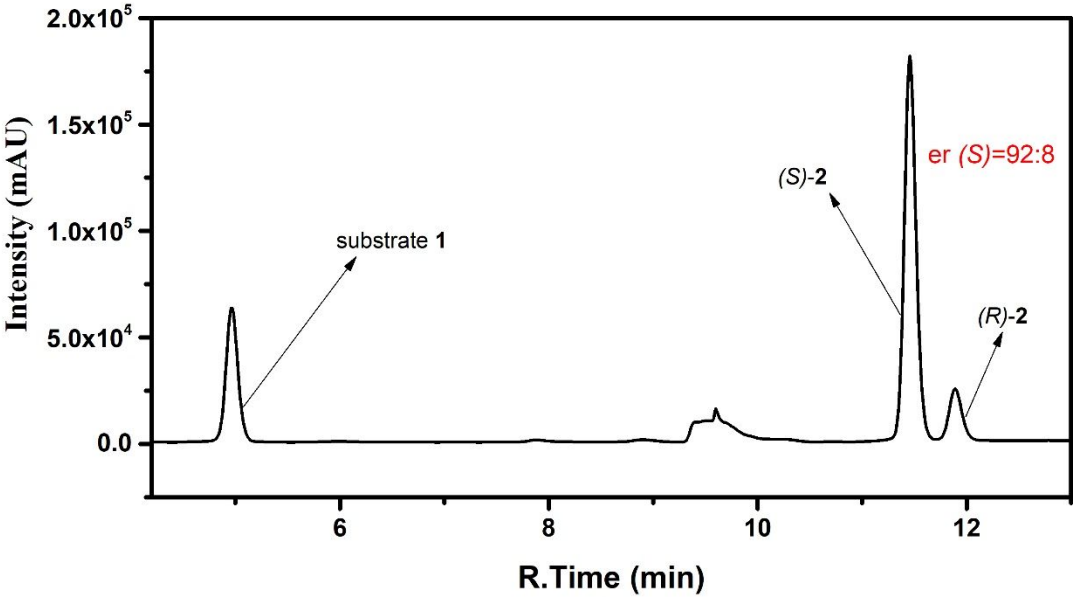

Figure S2. HPLC graph of WAJ-9-catalyzed reaction of substrate **1**.

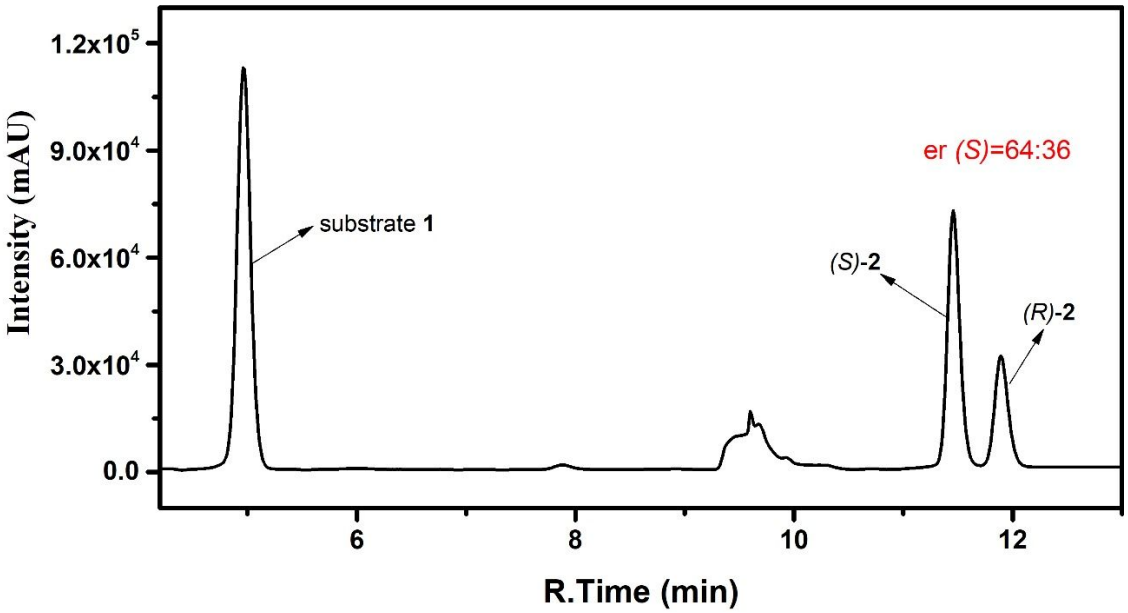

Figure S3. HPLC graph of A328F-catalyzed reaction of substrate **1**.

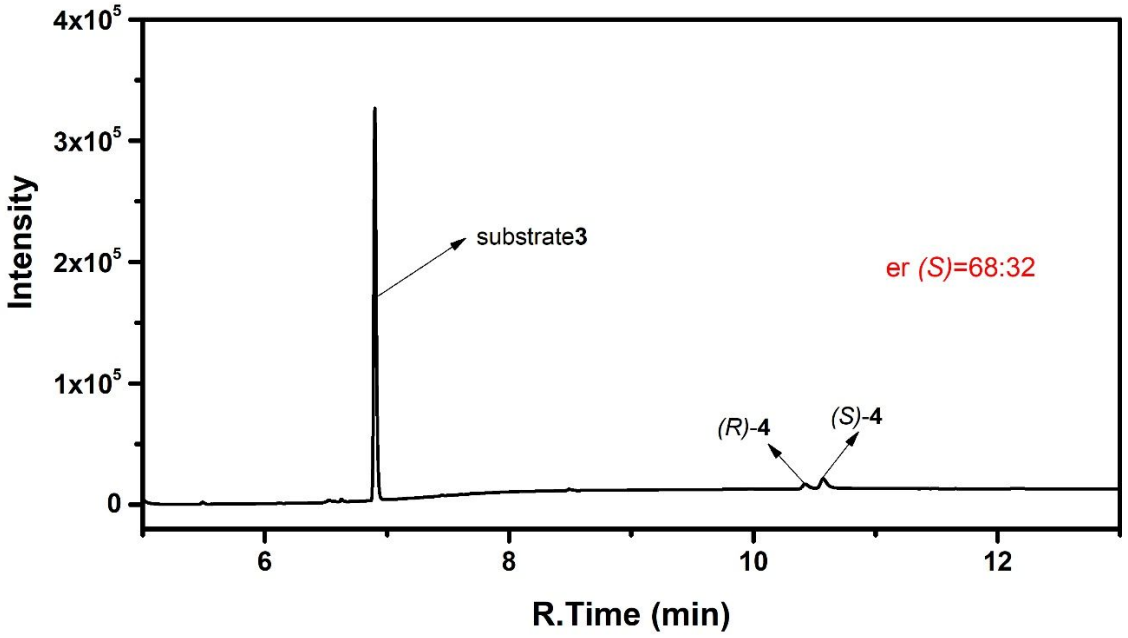

Figure S4. GC graph of WT-catalyzed reaction of substrate **3**.

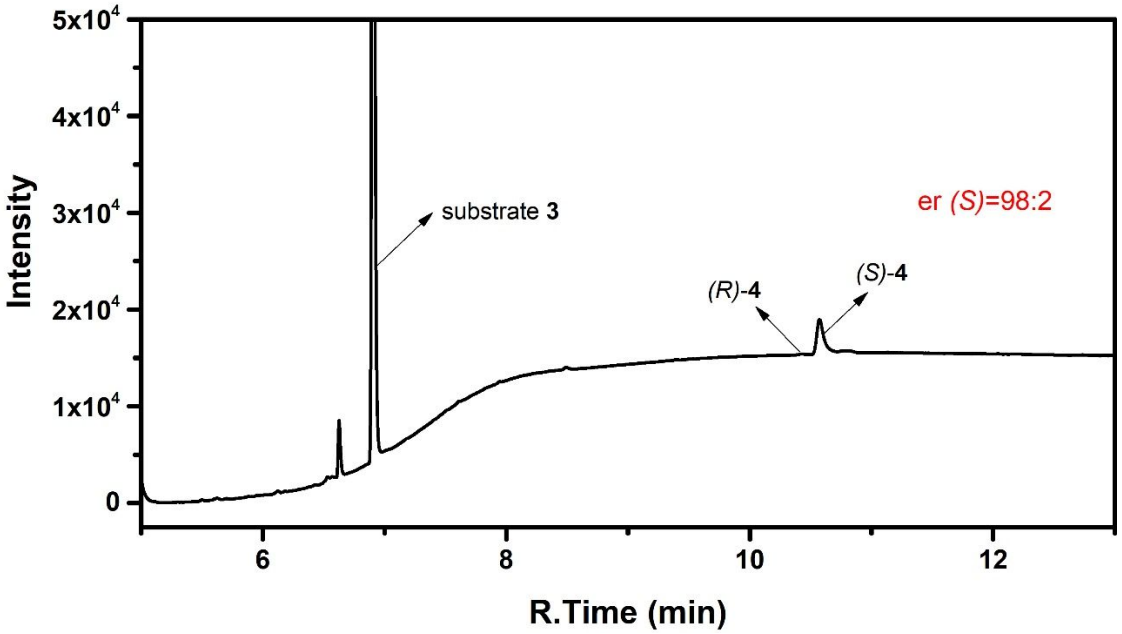

Figure S5. GC graph of WAJ-9-catalyzed reaction of substrate **3**.

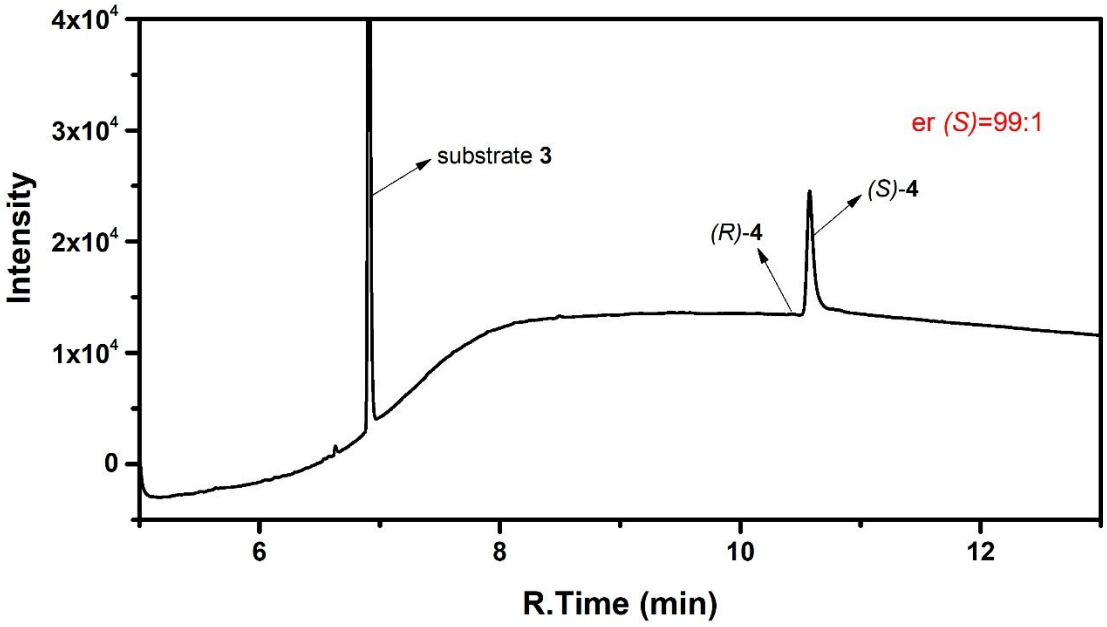

Figure S6. GC graph of A328F-catalyzed reaction of substrate **3**.

System setup and MD simulations

The initial structures of P450-BM3 were taken from PDB code of 2J1M[4] and 1JPZ.[5] In particular, PDB of 2JIM was used for substrate **1**, while 1JPZ was used for substrate **3**. In our initial setup, PDB of 2JIM is uniformly used for MD simulations of both substrate **1** and substrate **3**. However, we found substrate **3** is too flexible and tends to escape from the binding pocket after long time MD simulations. As such, we turned to PDB of 1JPZ for the MD simulations of substrate **3**, which yields more satisfactory results compared to PDB of 2JIM. Substrates **1** and **3** were docked into the active site of P450-BM3 using AutoDock Vina tool[6] in Chimera,[7] respectively. The docked poses with the highest docking scores were used for the subsequent MD simulations. Missing hydrogen atoms were added by module leap of Amber 18.[8] The force field for the Cpd I state was taken from the literature.[9] The general AMBER force field (GAFF)[10] was used for substrates **1** and **3**, while the partial atomic charges were obtained from the RESP method,[11] using HF/6-31G* level of theory. The parmchk utility from AMBERTools was used to generate the missing parameters for the ligands.14 Na^+ ions were added into the protein surface to neutralize the total charges of the systems. Finally, the resulting systems were solvated in a rectangular box of TIP3P[12] waters extending up to minimum cutoff of 15 Å from the protein boundary. The Amber ff14SB force field[13] was employed for the protein in all of the molecular dynamics (MD) simulations. The initial structures were fully minimized using combined steepest descent and conjugate gradient method. The systems were then gently annealed from 10 to 300 K under canonical ensemble for 0.05 ns with a weak restraint of 15 kcal/mol/Å. 1 ns of density equilibration was performed under isothermal-isobaric ensemble at target temperature of 300K and the target pressure of 1.0 atm using Langevin-thermostat[14] and Berendsen barostat[15] with collision frequency of 0.002 ns and pressure-relaxation time of 0.001 ns. Further equilibration of the systems was allowed for 4 ns to get well settled temperature and pressure. After proper minimizations and equilibrations, a productive MD run of 50 ns was performed for all the complex systems. For comparison, an extended MD run of 200 ns was also performed for substrate **1**. The MD simulations were performed with GPU version of Amber 18 package. [8]

QM/MM Calculations with Chemshell

All QM/MM calculations were performed using ChemShell, [16] combining Turbomole[17] for the QM region and DL_POLY[18] for the MM region. The propionate-truncated heme and the entire substrate were included in the QM region, whereas all of the protein residues and water within 14Å of Fe atom were included in the active region. The QM region was treated by the hybrid UB3LYP functional with the all electron basis set of def2-SVP (labeled as B1), whereas the remaining MM part of the system

was modeled at the classical level using the same parameters as in the classical MD simulations. The energies are further corrected with the large all-electron basis-set Def2-TZVP, labeled as B2. The dispersion energies were included in both optimizations and single-point energy calculations. The electronic embedding scheme[19] was used to account for the polarizing effect of the enzyme environment on the QM region. All of the reactions of Cpd I were studied in the S=1/2 state, since the S=3/2 and S=1/2 states of Cpd I show generally similar reactivities. [20]

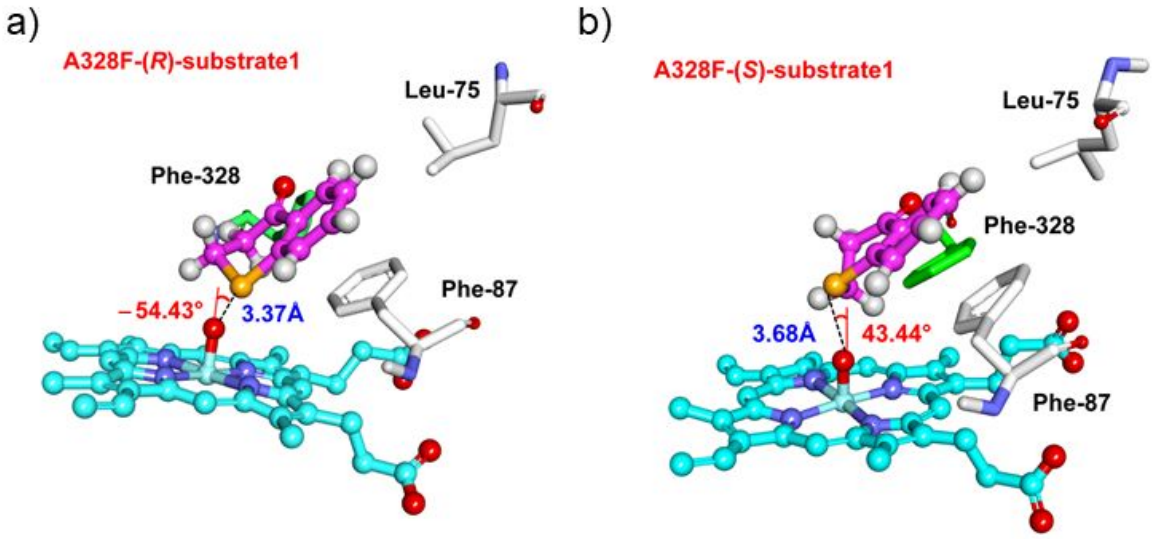

Figure S7. The representative binding modes of (*R*)-substrate **1** (a) and (*S*)-substrate **1** (b) in A328F system. (*R*)/(*S*)-substrate **1** and iron porphyrin are displayed with ball and stick model, residues Leu-75, Phe-87 and Phe-328 are displayed with stick model.

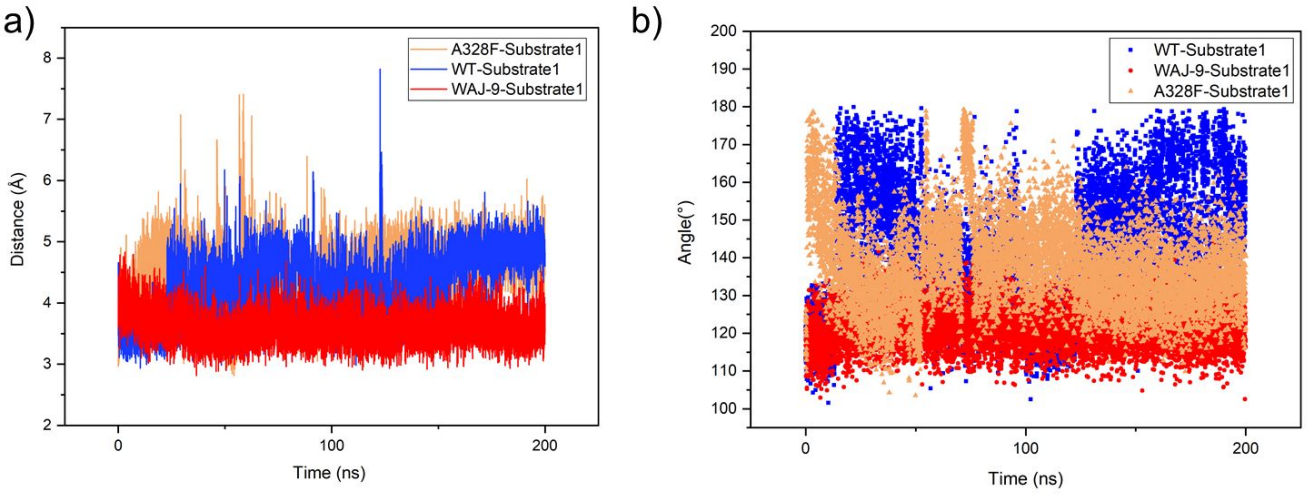

Figure S8. The fluctuation of the distance of Ox-S (a) and the ∠Fe-O-S angle (b) in WT, WAJ-9 and A328F of P450 BM3 during the MD simulations.

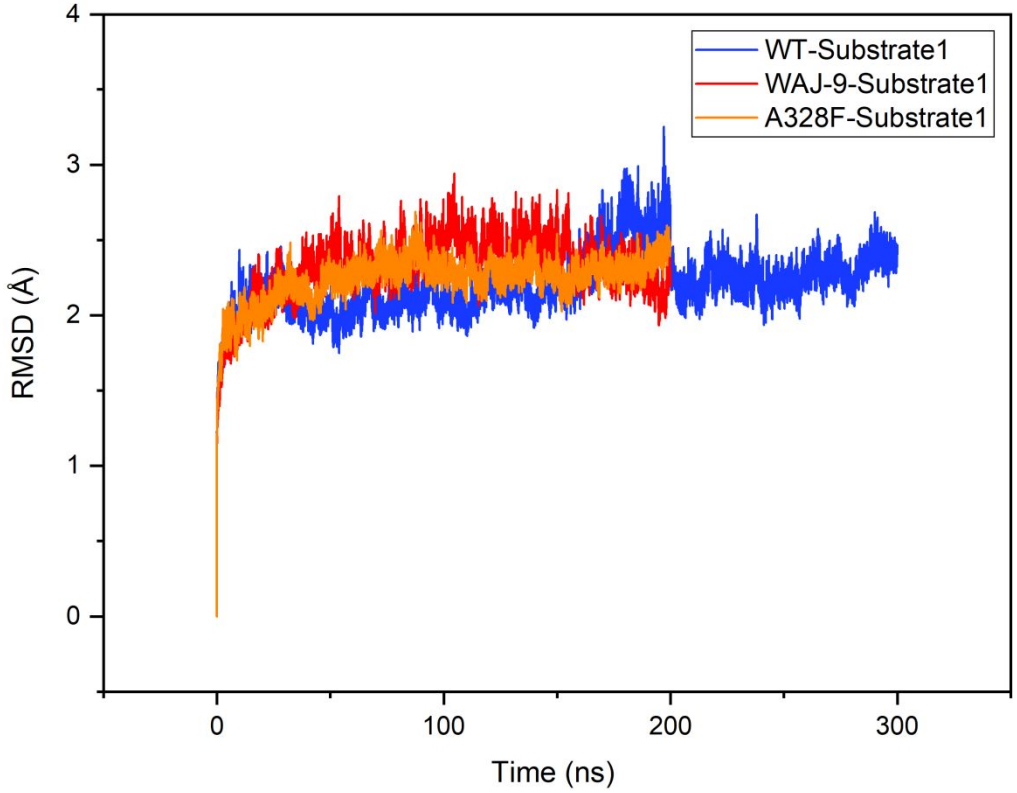

Figure S9. Time evolution of the root mean square deviations (RMSD) for the protein backbone and substrate during the MD of WT, WAJ-9 and A328F of P450 BM3. Due to some fluctuations around 200 ns in WT-sub1, this MD run has extended to 300ns, and it is seen that the RMSD keeps constant and thus converged.

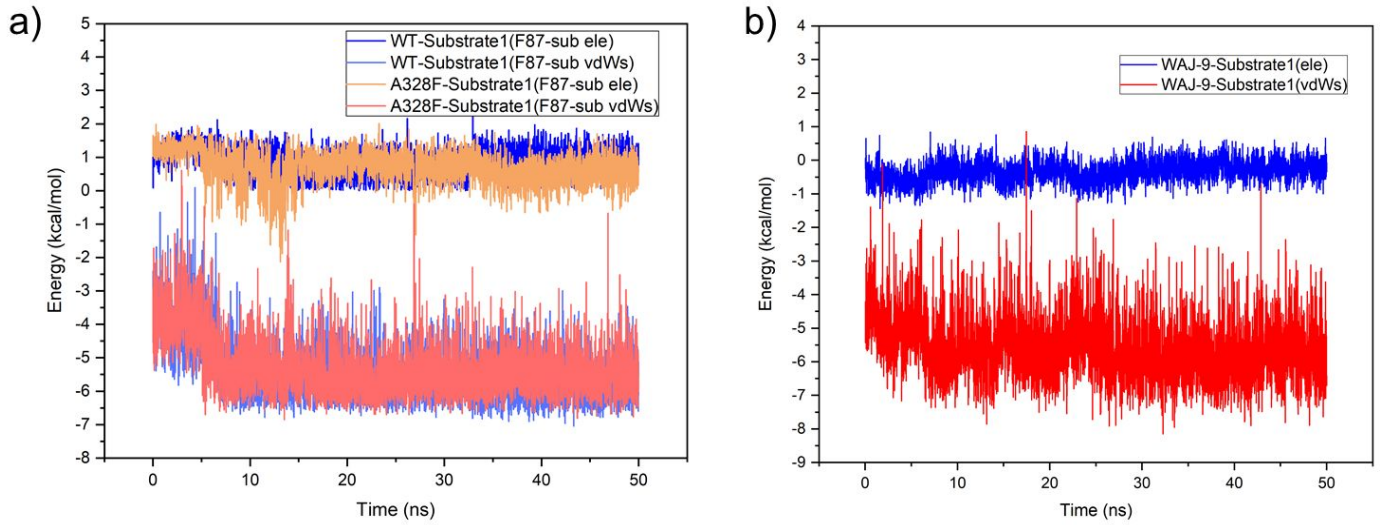

Figure S10. The fluctuation of electrostatic and van der Waals energies between substrate **1** and surrounding key residues in (a) WT and A328F; (b) WAJ-9; during MD simulations.

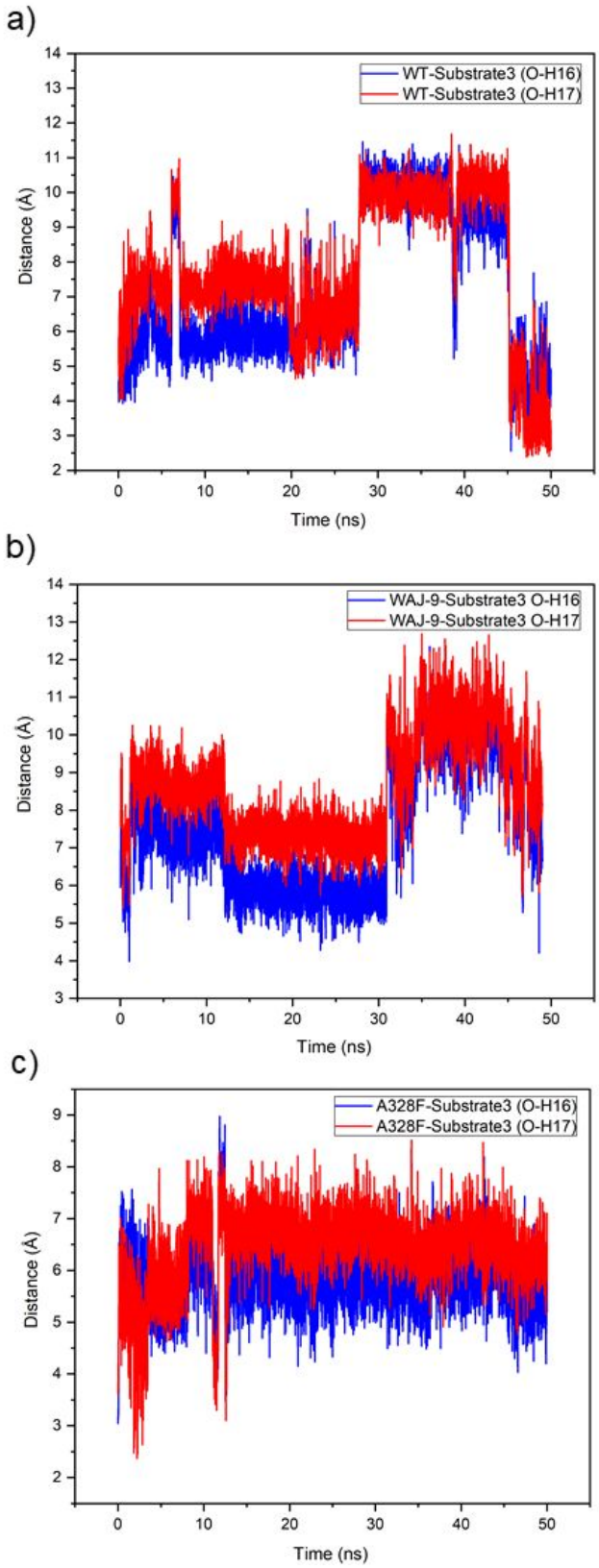

Figure S11. The fluctuation of the distance of Fe-Ox-H16/H17(substrate **3)** in (a): WT; (b): WAJ-9; (c) A328F during the MD simulations.

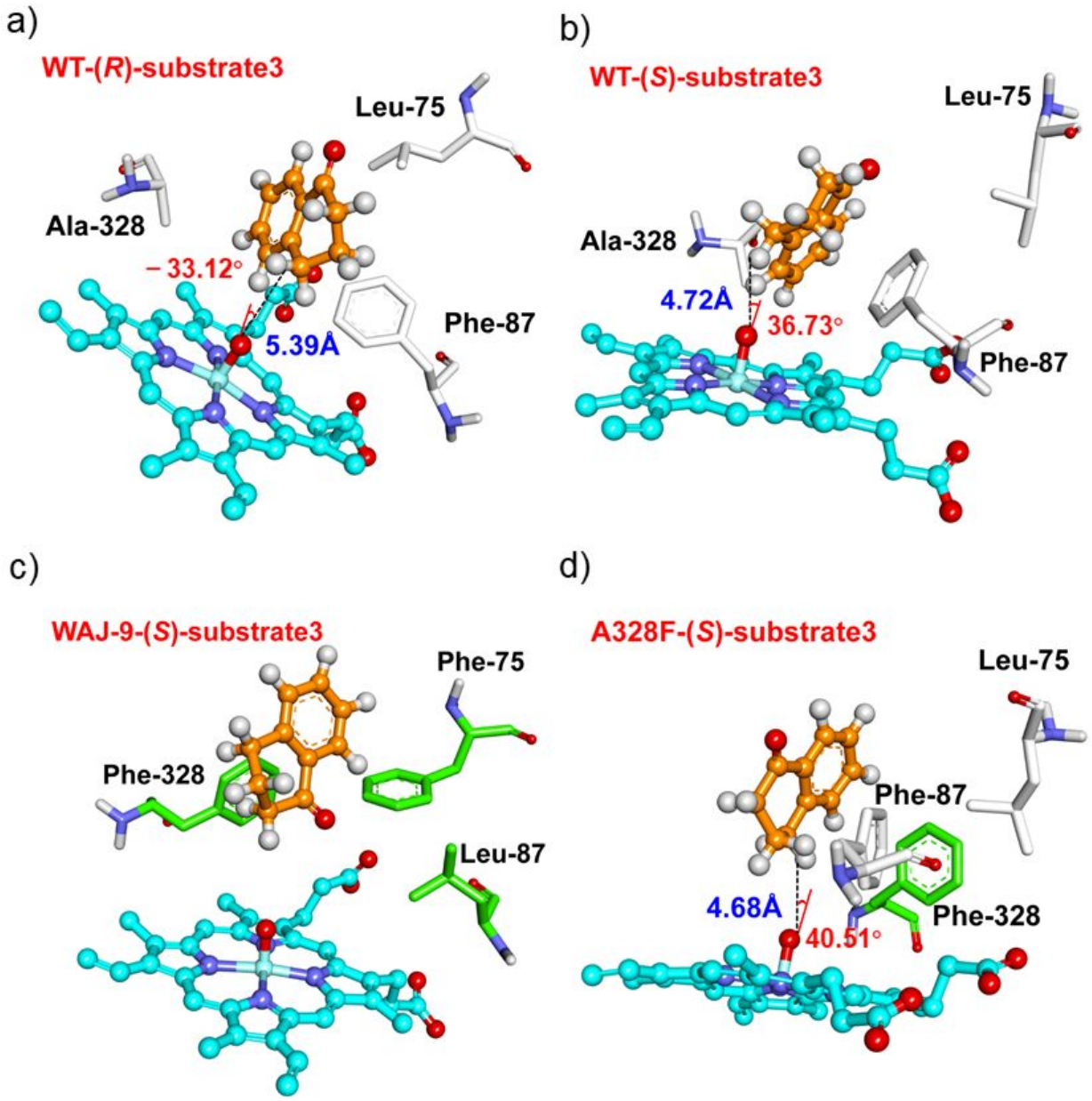

Figure S12. The representative binding modes of (a): (*R*)-substrate **3** in WT; (b): (*S*)-substrate **3** in WT; (c): (*S*)-substrate **3** in WAJ-9; (d): (S)-substrate **3** in A328F. Substrate **3** and iron porphyrin are demonstrated with ball and stick model, while residues Leu-75, Phe-87 and Phe-328 are demonstrated with stick model.

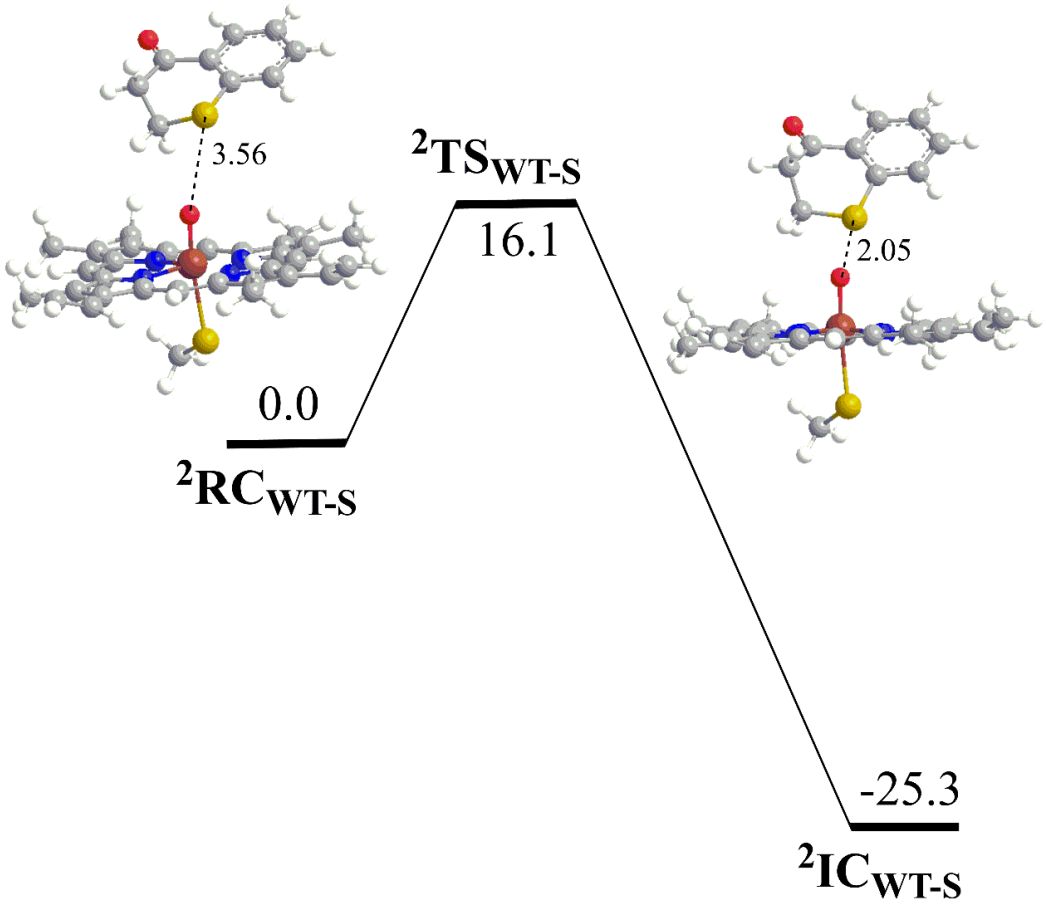

Figure S13. (a) QM(B2)/MM calculated energy profile (kcal/mol) for the Cpd I-mediated sulfoxidation of (*S*)-substrate **1** in WT. Alongside the QM(B1)/MM optimized structures of reactants and transistion states of QM region; key distances are given in angstroms.

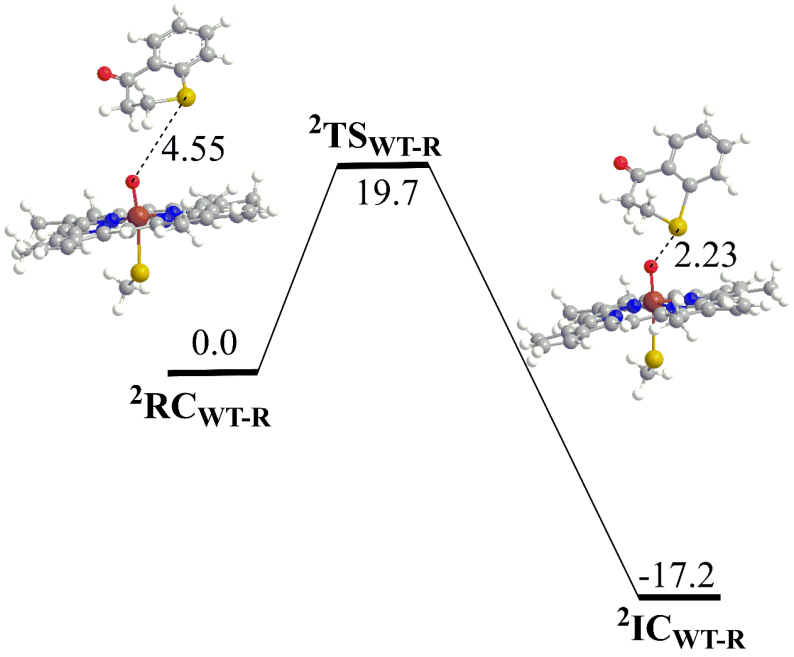

Figure S14. (a) QM(B2)/MM calculated energy profile (kcal/mol) for the Cpd I-mediated sulfoxidation of (*R*)- substrate **1** in WT. Alongside the QM(B1)/MM optimized structures of reactants and transistion states of QM region; key distances are given in angstroms.

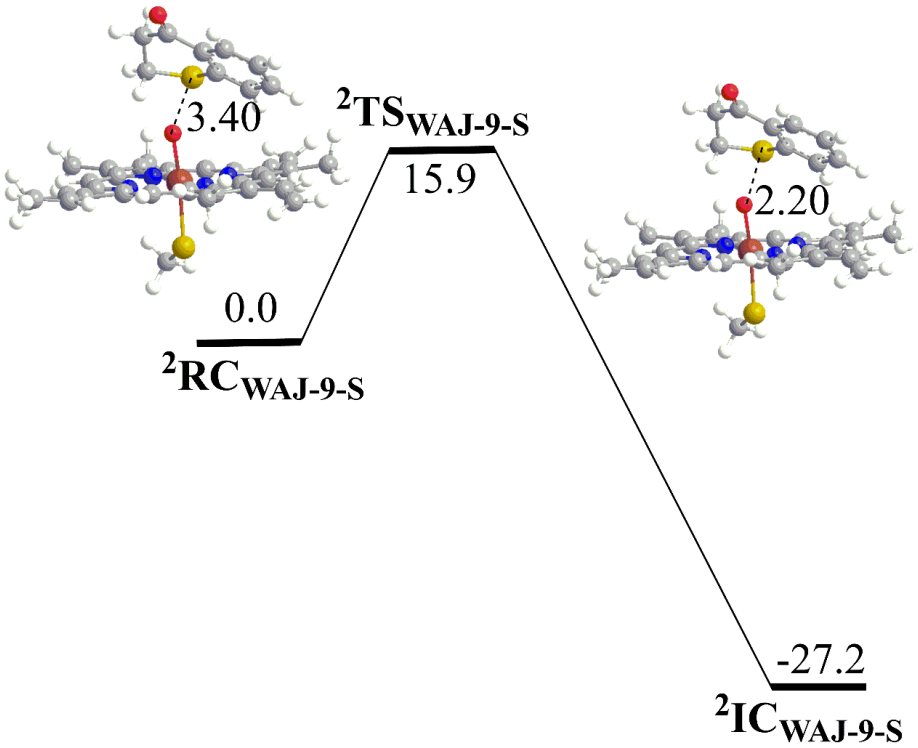

Figure S15. (a) QM(B2)/MM calculated energy profile (kcal/mol) for the Cpd I-mediated sulfoxidation of (*S*)- substrate **1** in WAJ-9 Alongside the QM(B1)/MM optimized structures of reactants and transistion states of QM region; key distances are given in angstroms.

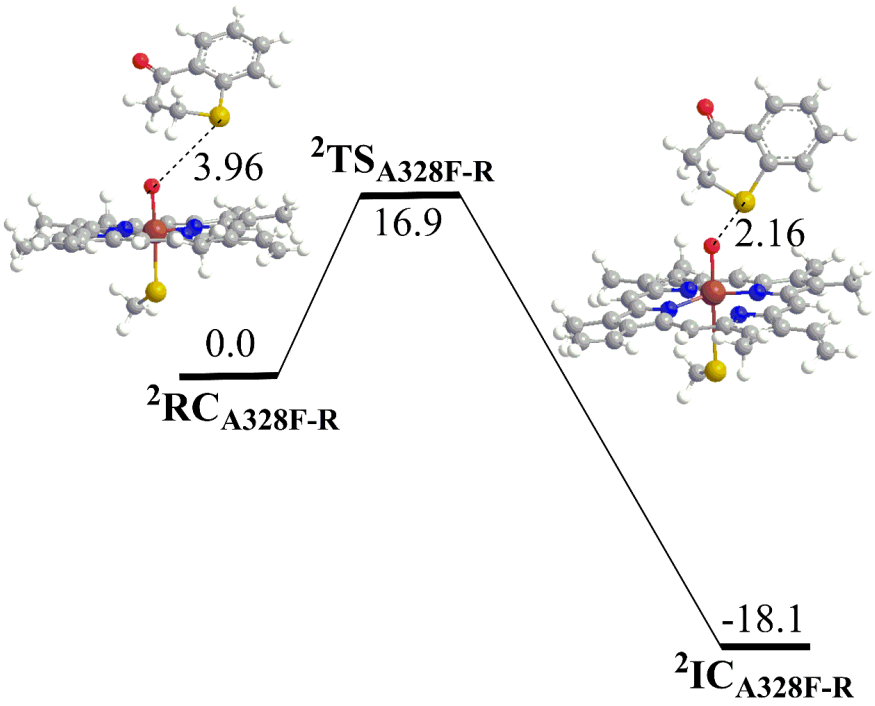

Figure S16. (a) QM(B2)/MM calculated energy profile (kcal/mol) for the Cpd I-mediated sulfoxidation of (*R*)- substrate **1** in A328F. Alongside the QM(B1)/MM optimized structures of reactants and transistion states of QM region; key distances are given in angstroms.

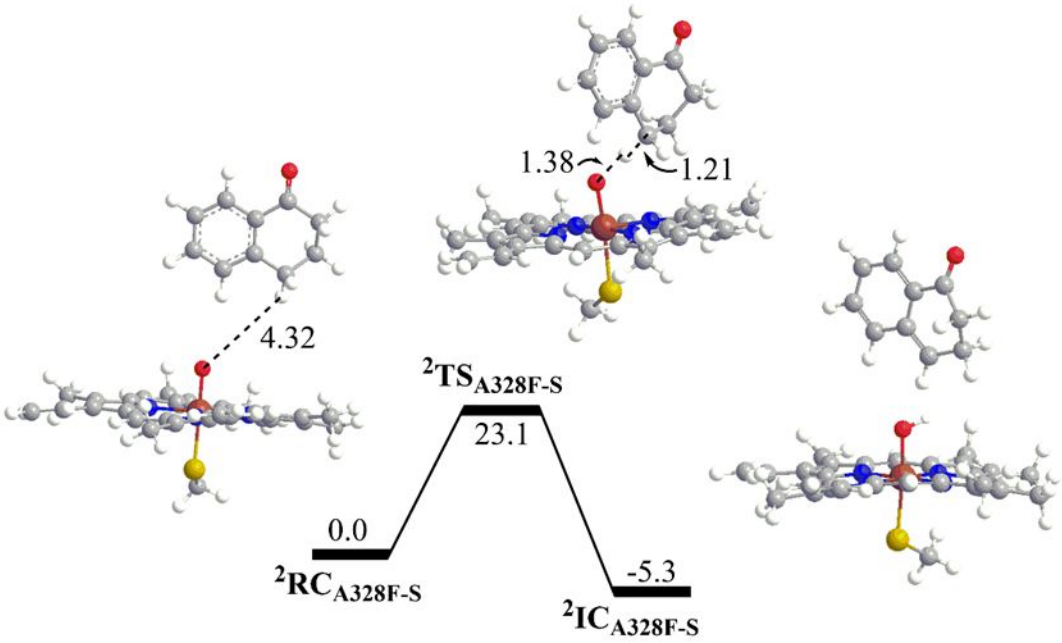

Figure S17. (a) QM(B2)/MM calculated energy profile (kcal/mol) for the Cpd I-mediated hydroxylation of (*S*)-substrate **3** in A328F. The ZPE energies are included. Alongside the QM(B1)/MM optimized structures of reactants and transistion states of QM region; key distances are given in angstroms.

Table S3. Calculated averaged interaction energies (electrostatic, van der Waals and the sum energy) between substrate **1** and the surrounding residues.

P450-Substrate 1		Electrostatic (kcal/mol)	Van der Waals (kcal/mol)	ele + vdWs (kcal/mol)
WT	F87---substrate **1**	0.846	- 4.745	- 3.900
WAJ-9	(F75-L87-F328)---substrate **1**	- 0.292	- 5.335	- 5.627
A328F	F87---substrate **1**	0.755	- 5.443	- 4.688

Table S4. Enantioselectivity of hydroxylation of substrate **3** derived from experiment and MD simulations, and the computed average Fe-Ox---H distance for the Cpd I of WT, WAJ-9 and A328F in the presence substrate **3**.

P450-Substrate 3	(*R*) : (*S*)	Fe-Ox---H Distance (Å)	
	Experiment	Fe-Ox---H16	Fe-Ox---H17
WT	(*R*) < (*S*)	7.213	7.822
WAJ-9	2 : 98	7.514	8.625
A328F	1 : 99	5.186	6.141

Table S5. Summary of the QM/MM calculated barriers, Fe-Ox---S Distance and the angle of ∠Fe-Ox-S obtained from MD simulations and QM/MM calculations in selected four snapshots.

P450-Substrate 1	Calculated barrier (kcal/mol)	Fe-Ox---S Distance (Å)		∠Fe-Ox-S (degree)	
		MD	QM/MM	MD	QM/MM
WT-*S*	16.1	3.3	3.6	167.4	162.7
WT-*R*	19.7	5.0	4.6	139.2	138.7
WAJ-9-*S*	15.9	3.4	3.4	121.8	117.9
A328F-*R*	16.9	4.1	4.0	124.0	131.8

References

(1) Li, A. ; Acevedo-Rocha C. G.; Reetz M. T.; Boosting the efficiency of site-saturation mutagenesis for a difficult-to-randomize gene by a two-step PCR strategy. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 6095-6103.

(2) Wang J.; Ilie A.; Reetz M. T. Chemo-and Stereoselective Cytochrome P450-BM3-Catalyzed Sulfoxidation of 1-Thiochroman-4-ones Enabled by Directed Evolution. *Adv. Synth. Catal.* **2017**, *359*, 2056-2060.

(3) Yu D.; Wang J.; Reetz M. T. Exploiting Designed Oxidase–Peroxygenase Mutual Benefit System for Asymmetric Cascade Reactions. *J. Am. Chem. Soc.* **2019**, *141*, 5655-5658.

(4) Kuper J.; Wong T. S.; Roccatano D.; Wilmanns M.; Schwaneberg U. Understanding a mechanism of organic cosolvent inactivation in heme monooxygenase P450 BM-3. *J. Am. Chem. Soc*. **2007**, *129*, 5786–5787.

[5] Haines D. C.; Tomchick D. R.; Machius M.; Peterson J. A. Pivotal role of water in the mechanism of P450BM-3. *Biochemistry*. **2001**, *40* , 13456–13465.

[6] Trott O., Olson A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem*. **2010**, *31*, 455–461.

[7] Pettersen E. F.; Goddard T. D.; Huang C. C.; Couch G. S.; Green-blatt D. M.; Meng E. C.; Ferrin T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[8] Case, D. A.; I.Y. Ben-Shalom, I. Y.; S.R. Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, III, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York D. M.; Kollman, P. A. **2018**, AMBER 2018, University of California, San Francisco.

(9) Oda, A.; Yamaotsu, N.; Hirono, S., New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. *J. Comput. Chem.* **2005**, *26*, 818–826.

(10) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(11) Bayly, C. I.; Cieplak, P.; Cornell, W. D; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(12) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions

for simulating liquid water. *J. Chem. Phys.*, **1983**, *79*, 926-935.

(13) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theor. Comput.* **2015**, *11*, 3696-3713.

(14) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D., Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **2001**, *114*, 2090−2098.

(15) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren,W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(16) (a) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjovoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 1–28. (b) Metz, S.; Kästner, J.; Sokol, A.; Keal, T.; Sherwood, P. ChemShell—a modular software package for QM/MM simulations. *WIREs Comput. Mol. Sci.* **2014**, *4*, 101−110.

(17) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(18) Smith, W.; Forester, T. R. DL_POLY_2. 0: A general-purpose parallel molecular dynamics simulation package. *J. Mol. Graph.* **1996**, *4*, 136–141.

(19) Bakowies, D.; Thiel, W. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

(20) Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. P450 Enzymes: Their structure, reactivity, and selectivity - modeled by QM/MM calculations. *Chem. Rev.* **2010**, *110*, 949-1017.