Supporting Information

Quantum Chemical Modeling of Enantioselective Sulfoxidation and Epoxidation Reactions by Indole Monooxygenase *Vp*IndA1

Qinrou Li[1,2], Shiqing Zhang[2,3], Fufeng Liu[1], Hao Su[2,3], Xiang Sheng[2,3]*

[1] *College of Biotechnology, Tianjin University of Science and Technology, Tianjin, 300457 P. R. China*
[2] *Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, P. R. China*
[3] *National Center of Technology Innovation for Synthetic Biology, National Engineering Research Center of Industrial Enzymes and Key Laboratory of Engineering Biology for Low-Carbon Manufacturing, Tianjin 300308, P.R. China*

* Corresponding author: shengx@tib.cas.cn

Contents

1. Selected residues in the computational model

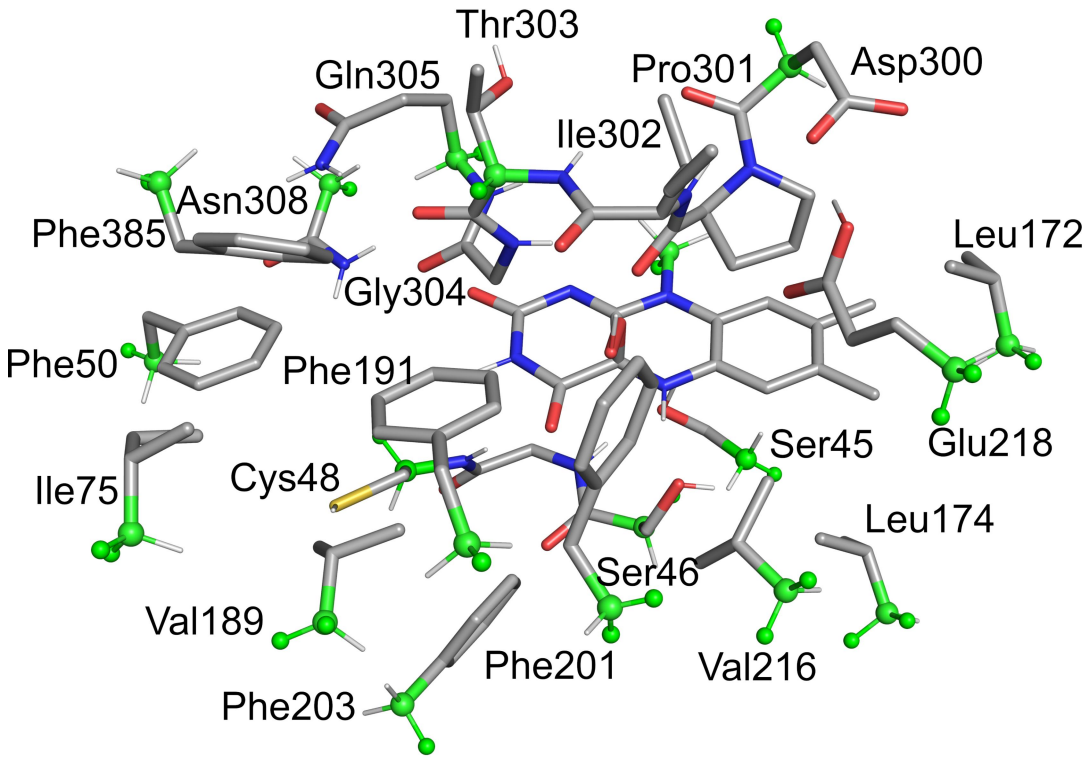

Figure S1. The selected residues in the computational cluster model used in the present study, in which Asp300 is in the deprotonated state and Glu218 is in the protonated state. The FAD_{OOH} is also shown in the figure. The model was designed based on the crystal structure of the wide-type *Vp*IndA1 (PDB:7Z4X). Atoms fixed during geometry optimization are shown in green color with balls and sticks. For clarity, most of the hydrogen atoms are omitted in the figure.

2. Prediction of the p*K*a values of Glu218 and Asp300

Table S1. Prediction of the p*K*a values of Glu218 and Asp300 by the constant pH molecular dynamics simulations (CpHMD) at the experimental pH (7.5) using wide-type *Vp*IndA1 in complex with FAD.

	Generalized Born implicit solvent model		Explicit solvent	
	Glu218	Asp300	Glu218	Asp300
p*K*a	10.36	1.50	13.20	1.80

Technical Details of the constant pH molecular dynamics (CpHMD) simulations

The AMBER 20 software was used for conducting CpHMD simulations.[1] The starting structure was constructed based on the crystal structure of wide-type *Vp*IndA1 crystallized with FAD (PDB ID:7Z4X). The AMBER ff99SB force field coupled with the necessary modifications for CpHMD simulations was used, and salt concentration was set to 0.1 M.[2-4] Both implicit and explicit solvents were considered for the prediction of the p*K*a values of Glu218 and Asp300. The general AMBER force field (gaff2) was applied to FAD.[5] The constructed model was solvated in the 12 Å TIP3P water box and neutralized by adding sodium ions in the explicit solvent model.[6] The GB model developed by Onufriev et al. (igb=2)[7] was used for the implicit solvent model. To maintain the micro-environment of the residues as that in the crystal structure, a harmonic potential of 5 kcal/mol/ $Å^2$ was applied to the protein backbones and FAD during the simulations. Simulation of each solvent model lasted for 100 ns in a production MD run with a time step of 2 fs. Protonation state changes were attempted every 25 steps in both solvent models, and in the explicit solvent model, solvent water relaxation occurred every 100 steps.

Table S2. Predicted p*K*a values of Glu218 and Asp300 by the PROPKA server[8, 9].

	Prediction by PROPKA	
	Glu218	Asp300
p*K*a	12.66	5.92

3. Results on the MPS sulfoxidation involving FAD_{OO-}

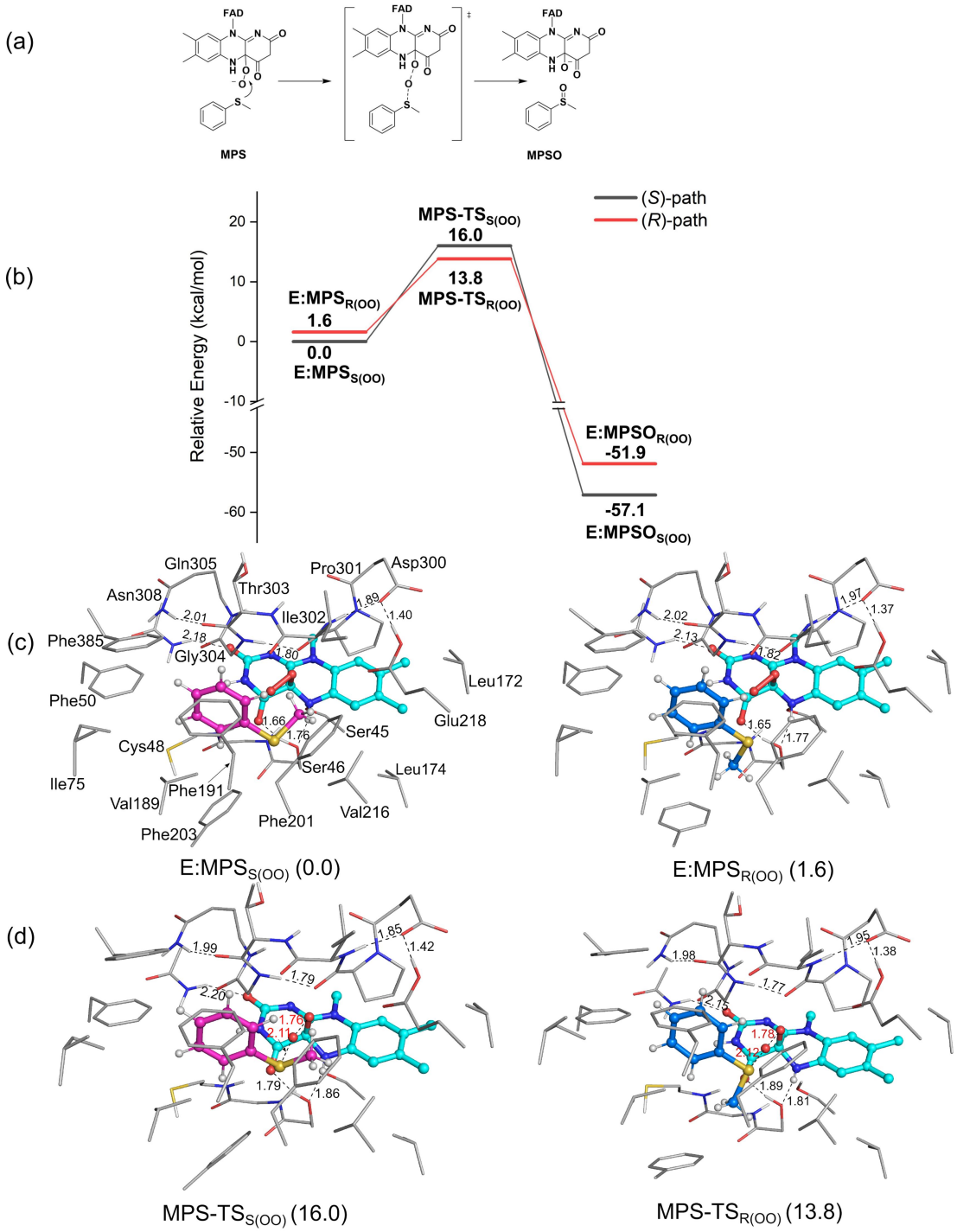

Figure S2. Calculation results on the MPS sulfoxidation involving FAD_{OO-} rather than FAD_{OOH} in the proposed mechanism: (a) Reaction mechanism, (b) Energy profile, (c) structures of the enzyme−substrate complexes and (d) structures of the transition states. The energies relative to **E:MPS$_{R(OO)}$** are provided in parentheses in kcal/mol. For clarity, most of the hydrogen atoms are omitted in the figure.

4. Results on the indene epoxidation involving FAD_{OO-}

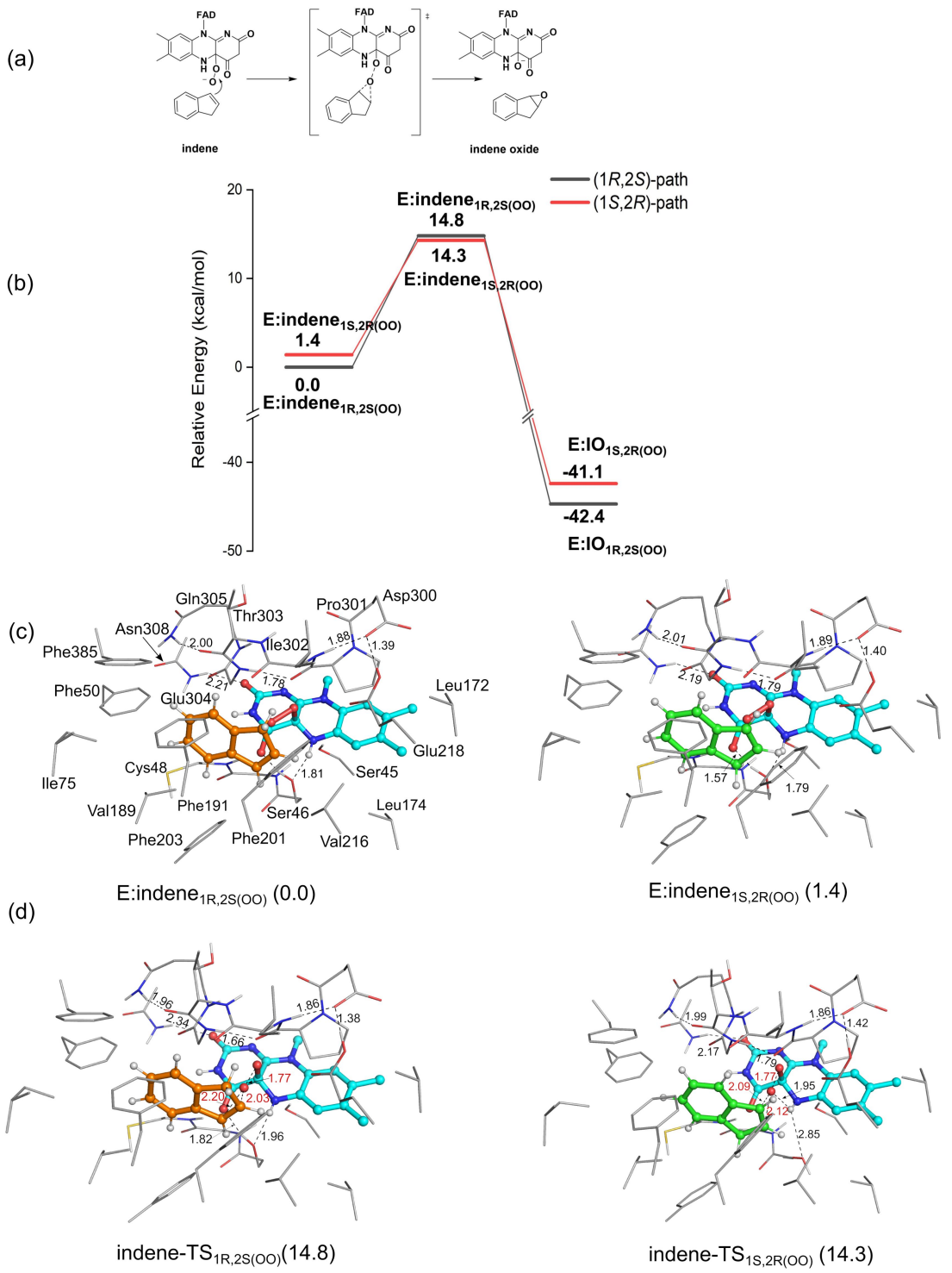

Figure S3. Calculation results on the indene epoxidation involving FAD_{OO-} rather than FAD_{OOH} in the proposed mechanism: (a) Reaction mechanism, (b) Energy profile, (c) structures of the enzyme−substrate complexes and (d) structures of the transition states. The energies relative to **E:indene$_{1R,2S(OO)}$** are provided in parentheses in kcal/mol. For clarity, most of the hydrogen atoms are omitted in the figure.

5. Other optimized structures of the E:MPS complexes

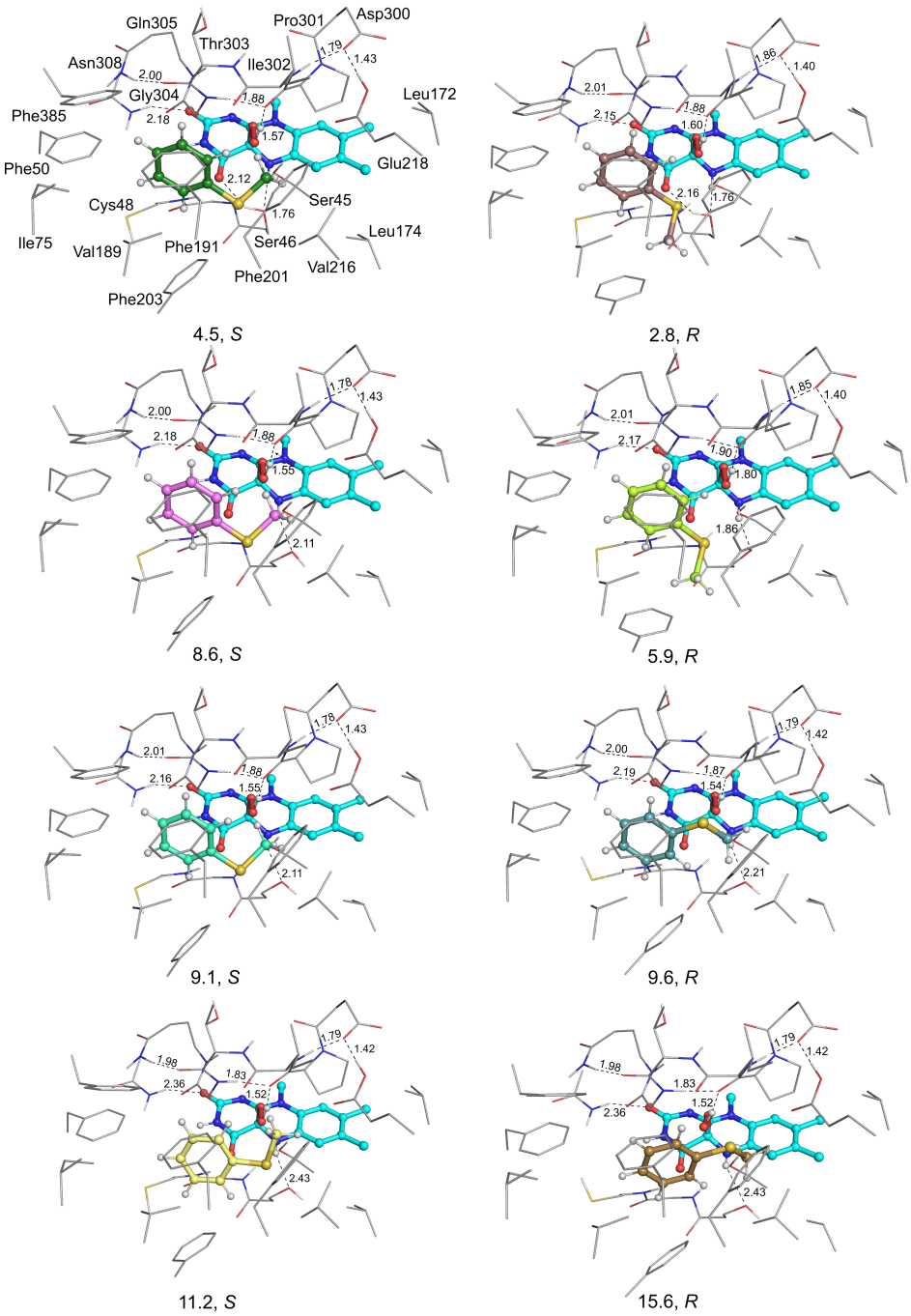

Figure S4. Other optimized structures of **E:MPS** with different conformation of the MPS substrate in addition to those shown in **Figure 3** in the main text. The energies are all relative to the energies of **E:MPS**$_S$ in the main text and are given in kcal/mol. For clarity, most of the hydrogen atoms are omitted in the figure. Selected distances are given in Å.

6. Transition states for the MPS sulfoxidation with the "phenyl-right" mode

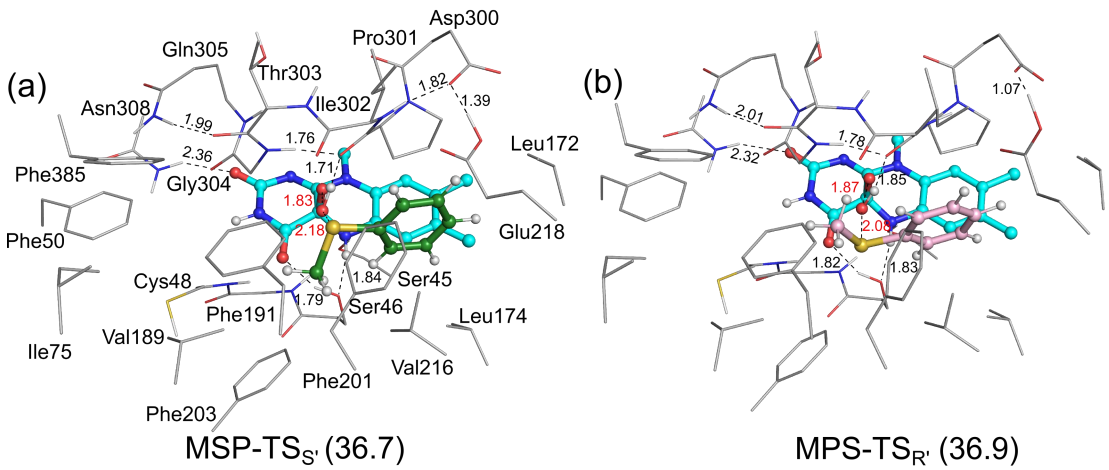

Figure S5. The optimized structures of TSs with the "Phenyl-right" mode are involved in (a) the *R*-pathway and (b) the *S*-pathway. The energies, provided in parentheses in kcal/mol, are all referenced relative to **E:MPS**$_S$ in the main text. For clarity, most of the hydrogen atoms are omitted in the figure. Selected distances are given in Å.

7. Other structures of the transition state for the MPS sulfoxidation

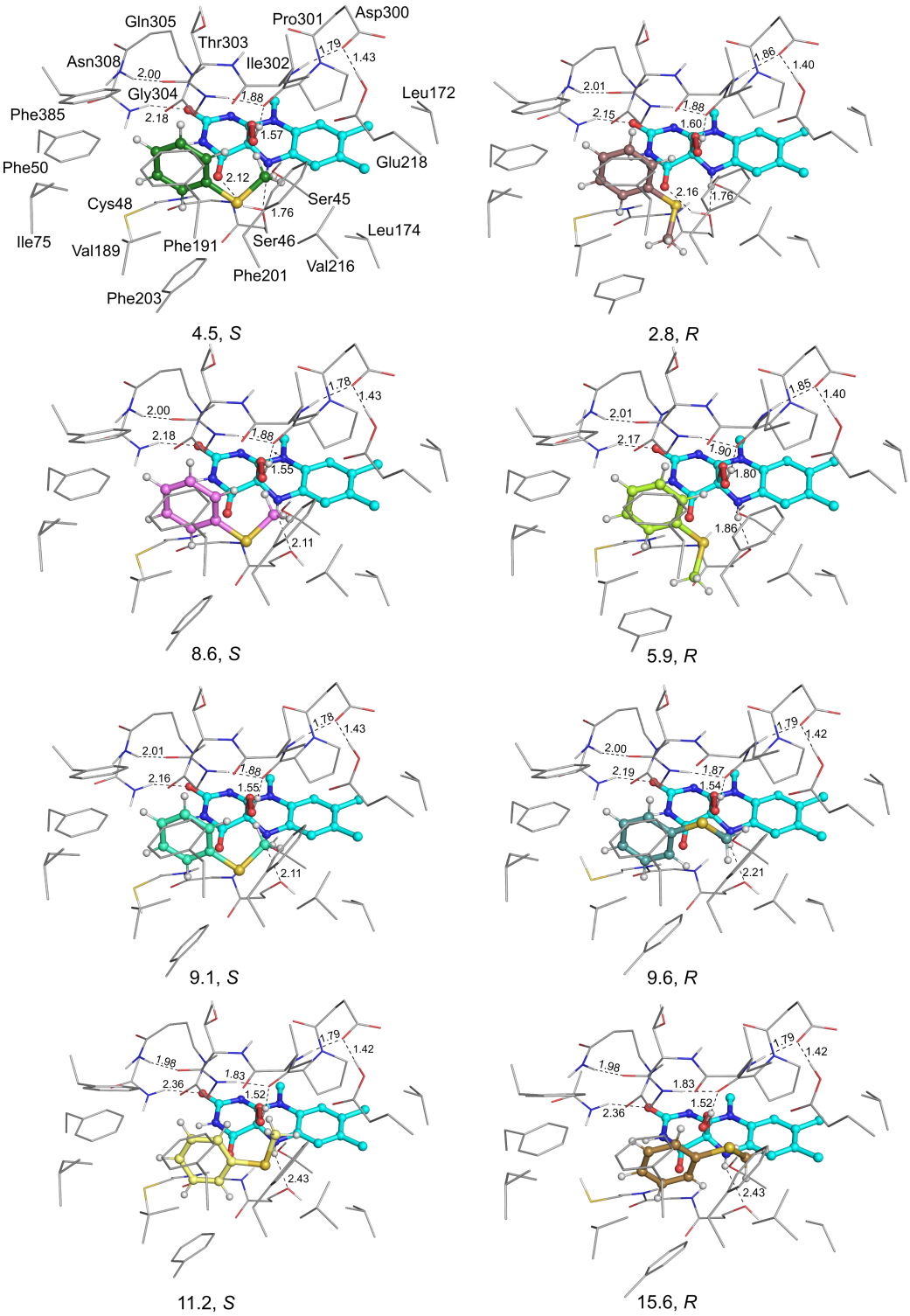

Figure S6. Other optimized structures of the TSs for the sulfoxidation reaction in addition to those shown in **Figure 4** in the main text. The energies are all relative to the energies of **E:MPSs** in the main text and are given in kcal/mol. For clarity, most of the hydrogen atoms are omitted in the figure. Selected distances are given in Å.

8. Optimized structures of the E:MPSO complexes

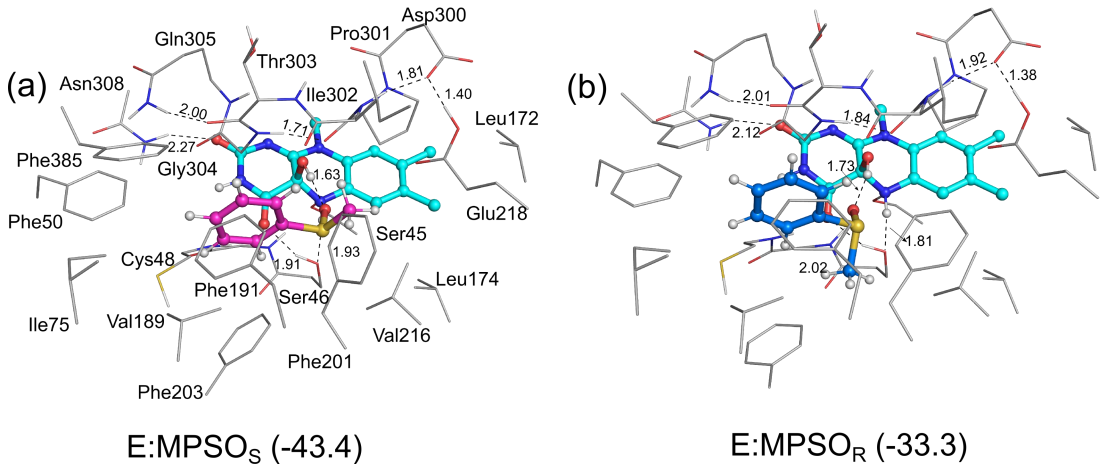

Figure S7. Optimized structures of **E:MPSO$_S$** and **E:MPSO$_R$**. The energies, provided in parentheses in kcal/mol, are all referenced relative to **E:MPS$_S$** in the main text. For clarity, most of the hydrogen atoms are omitted in the figure. Selected distances are given in Å.

9. Schematic representation of MPS-TS$_R$ and MPS-TS$_S$

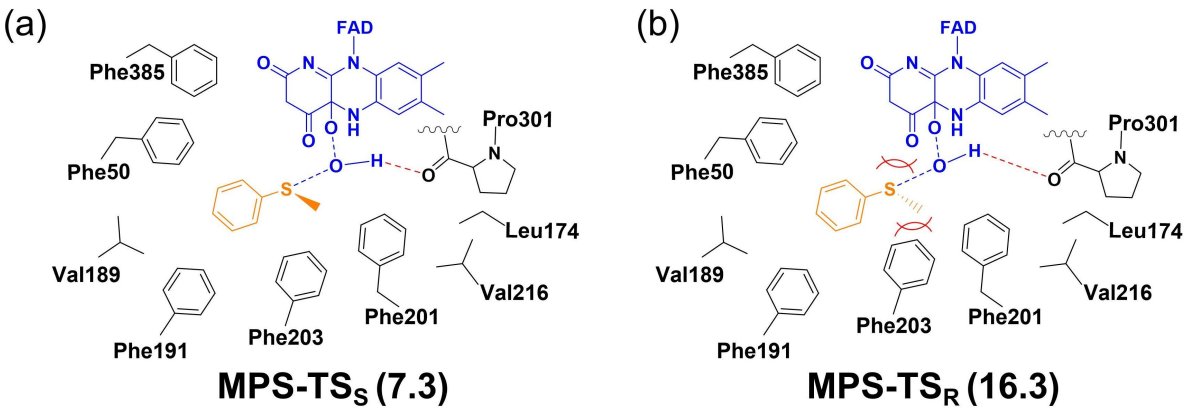

Figure S8. Schematic representation of **MPS-TS$_S$** and **MPS-TS$_R$** involved in the sulfoxidation reaction. The energies, provided in parentheses in kcal/mol, are all referenced relative to **E:MPS$_S$** in the main text.

10. Comparison of the active site structures of *Pp*StyA and *Vp*IndA1

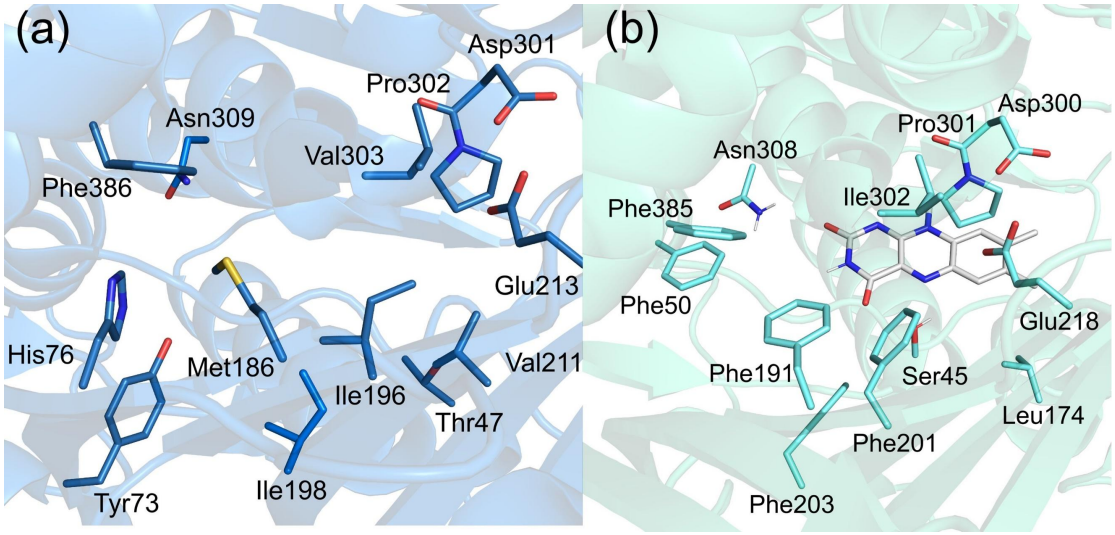

Figure S9. Active site structures of (a) the styrene monooxygenase from *Pseudomonas putida* (*Pp*StyA, PDB 3IHM)[10] and (b) *Vp*IndA1 (PDB 7Z4X)[11].

11. Optimized structures of the lowest-energy E:indene complexes

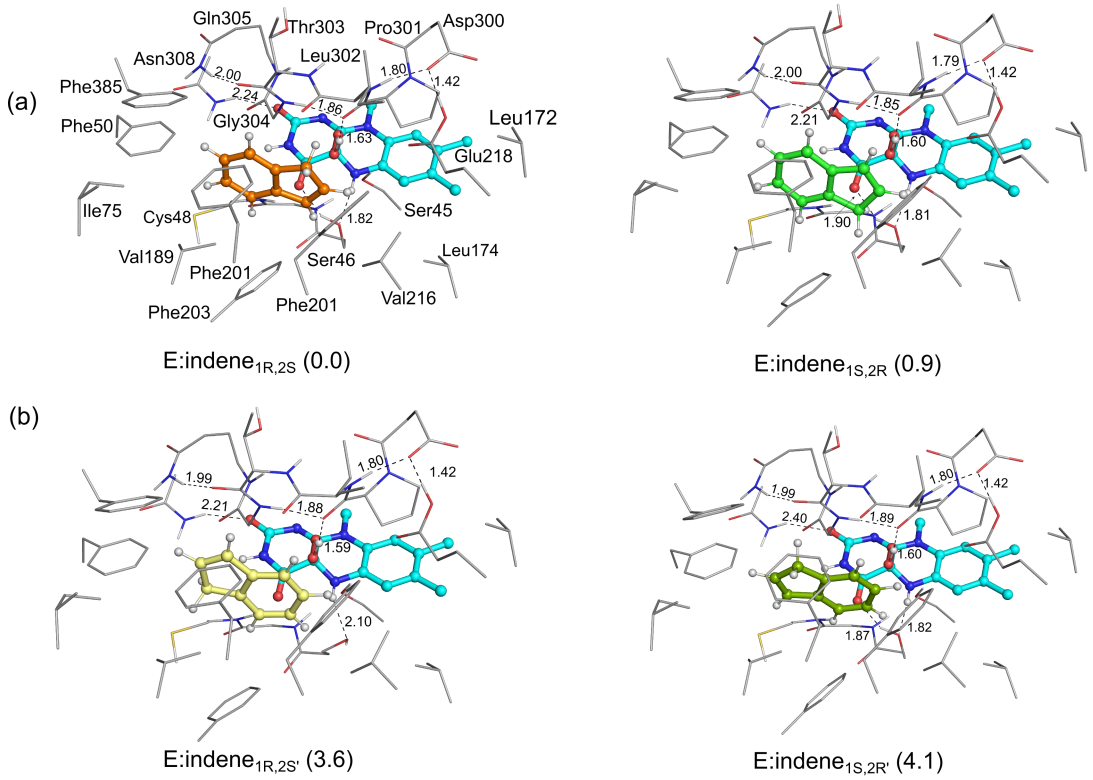

Figure S10. Optimized structures of the lowest-energy ES complexes with the "Phenyl-left" mode (a) and with the "Phenyl-right" mode (b). The subscript 1R,2S (in **E: indene $_{1R,2S}$** and **E:indene$_{1R,2S'}$**) and the 1S,2R (in **E: indene $_{1S,2R}$** and **E: indene $_{1S,2R'}$**) denote the configurations of the respective products originating from this ES complex. The energies, which are provided in parentheses in kcal/mol, are relative to **E:indene$_{1R,2S}$**. For clarity, most of the hydrogen atoms are omitted in the figure. Selected distances are given in Å.

12. Other optimized structures of the E:indene complexes

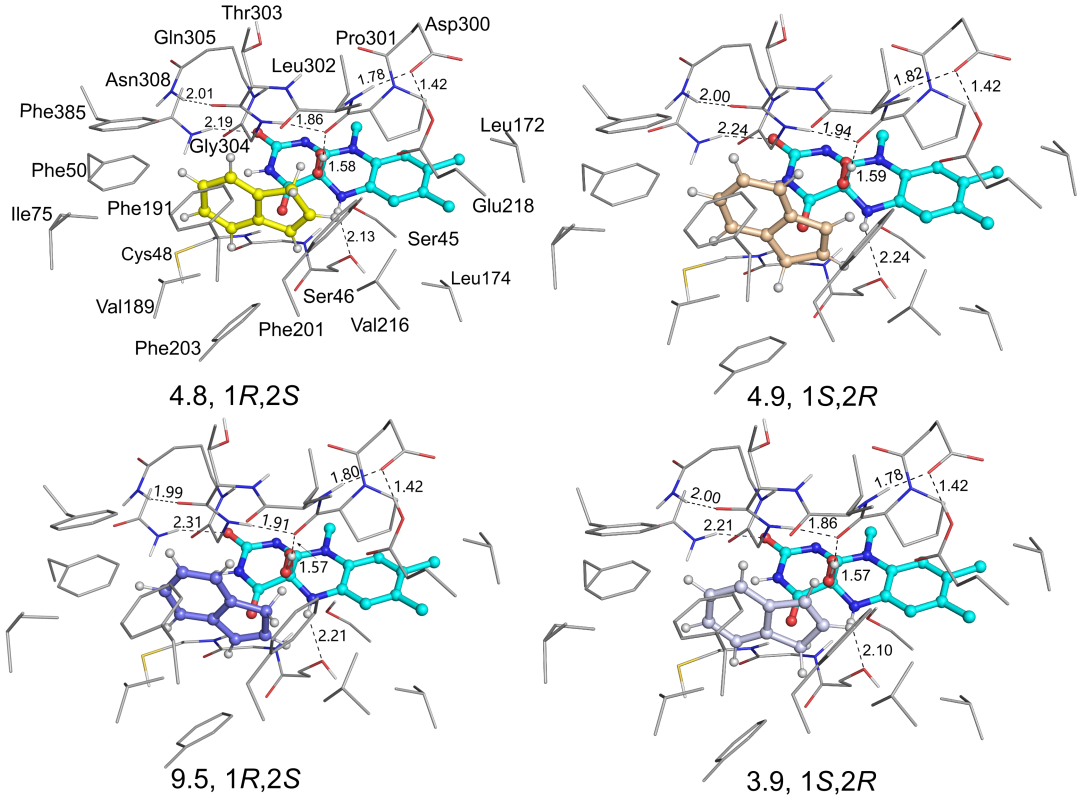

Figure S11. Other optimized structures of **E:indene** with different conformation of the indene substrate in addition to those shown in **Figure S10**. The energies are all relative to the energies of **E:indene$_{1R,2S}$** in **Figure S10** and are given in kcal/mol. For clarity, most of the hydrogen atoms are omitted in the figure. Selected distances are given in Å.

13. Transition states for the indene epoxidation with the "phenyl-right" mode

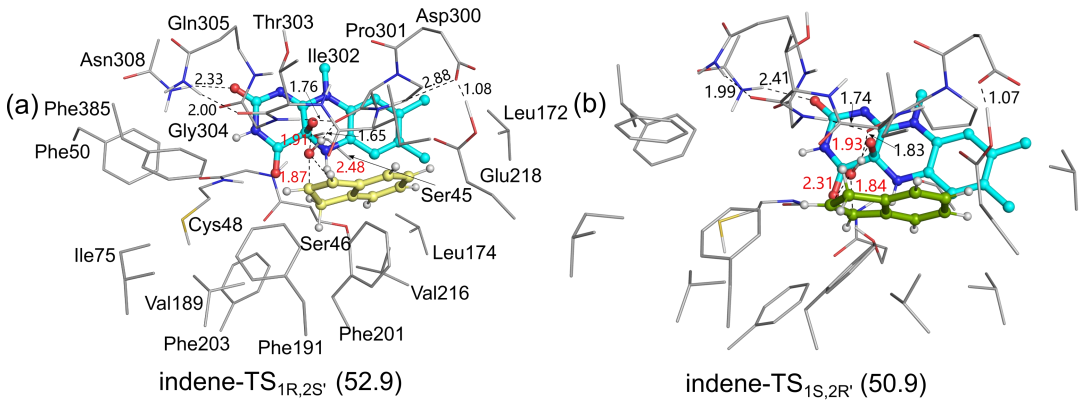

Figure S12. The optimized structures of TSs with the "Phenyl-right" mode involved in (a) the (1*R*,2*S*)-pathway and (b) the (1*S*,2*R*)-pathway. The energies, provided in parentheses in kcal/mol, are all referenced relative to **E:indene$_{1R,2S}$** in **Figure S10**. For clarity, most of the hydrogen atoms are omitted in the figure. Selected distances are given in Å.

14. Other structures of the transition state for the indene epoxidation

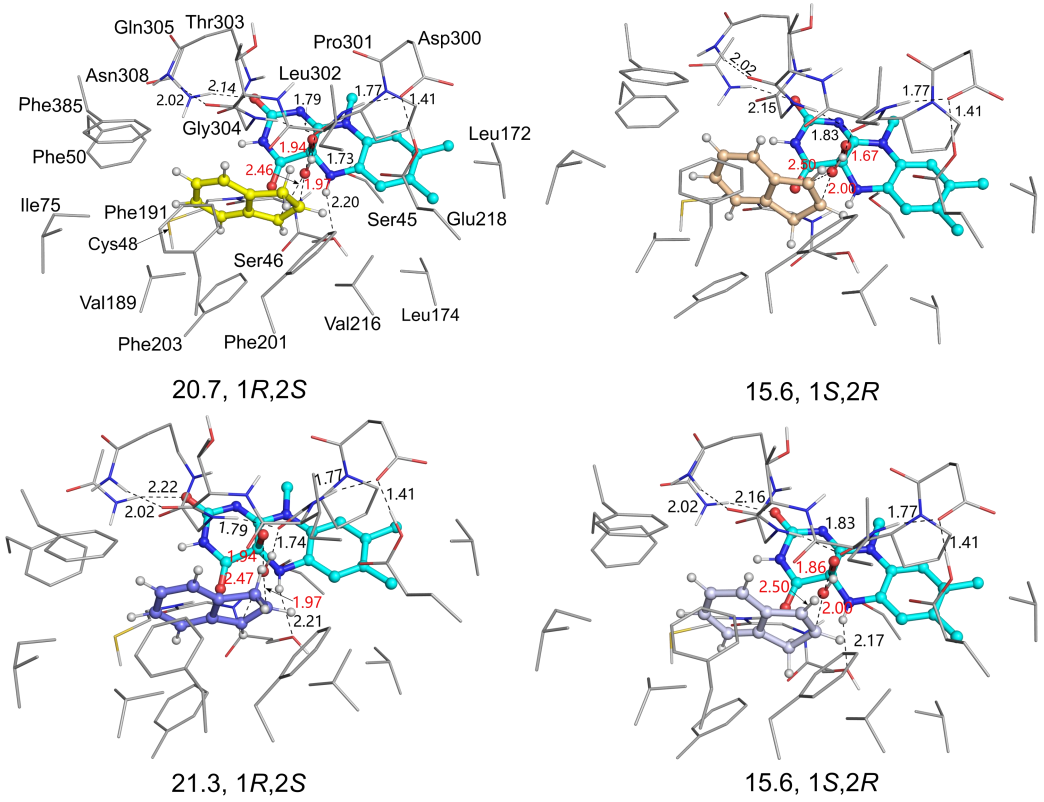

Figure S13. Other optimized structures of the TSs for the indene epoxidation in addition to those shown in **Figure 5** in the main text. The energies are all relative to the energies of **E:indene$_{1R,2S}$** in **Figure S10** and are given in kcal/mol. For clarity, most of the hydrogen atoms are omitted in the figure. Selected distances are given in Å.

15. Optimized structures of the indene-Int complexes

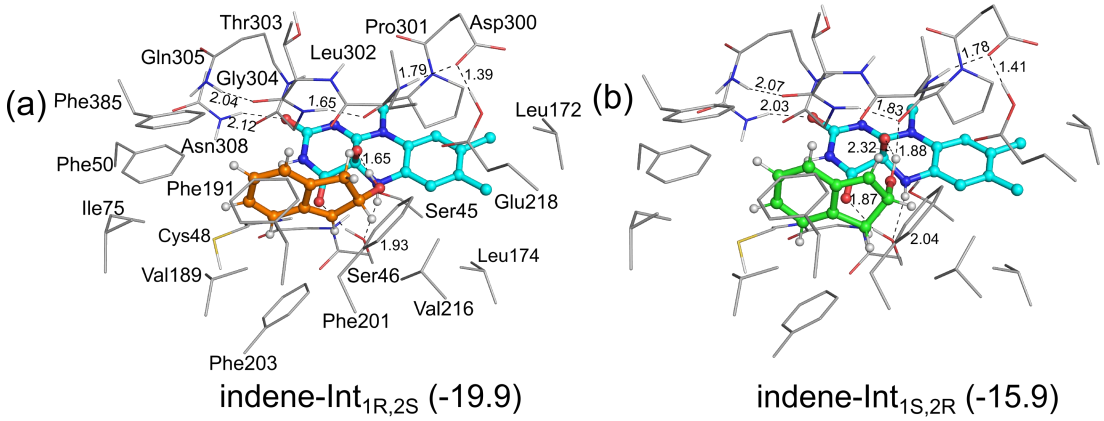

Figure S14. Optimized structures of **indene-Int$_{1R,2S}$** and **indene-Int$_{1S,2R}$**. The energies, provided in parentheses in kcal/mol, are all referenced relative to **E:indene$_{1R,2S}$** in **Figure S10**. For clarity, most of the hydrogen atoms are omitted in the figure. Selected distances are given in Å.

16. Optimized structures of the E:IO complexes

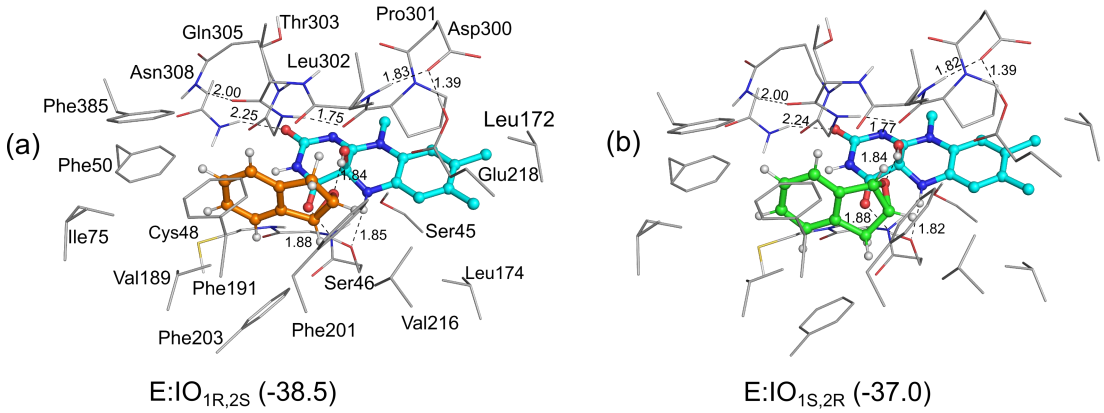

Figure S15. Optimized structures of **E:IO$_{1R,2S}$** and **E:IO$_{1S,2R}$**. The energies, provided in parentheses in kcal/mol, are all referenced relative to **E:indene$_{1R,2S}$** in **Figure S10**. For clarity, most of the hydrogen atoms are omitted in the figure. Selected distances are given in Å.

17. Optimized structures for the Phe191Met/Phe201Leu/Ile302Val reaction

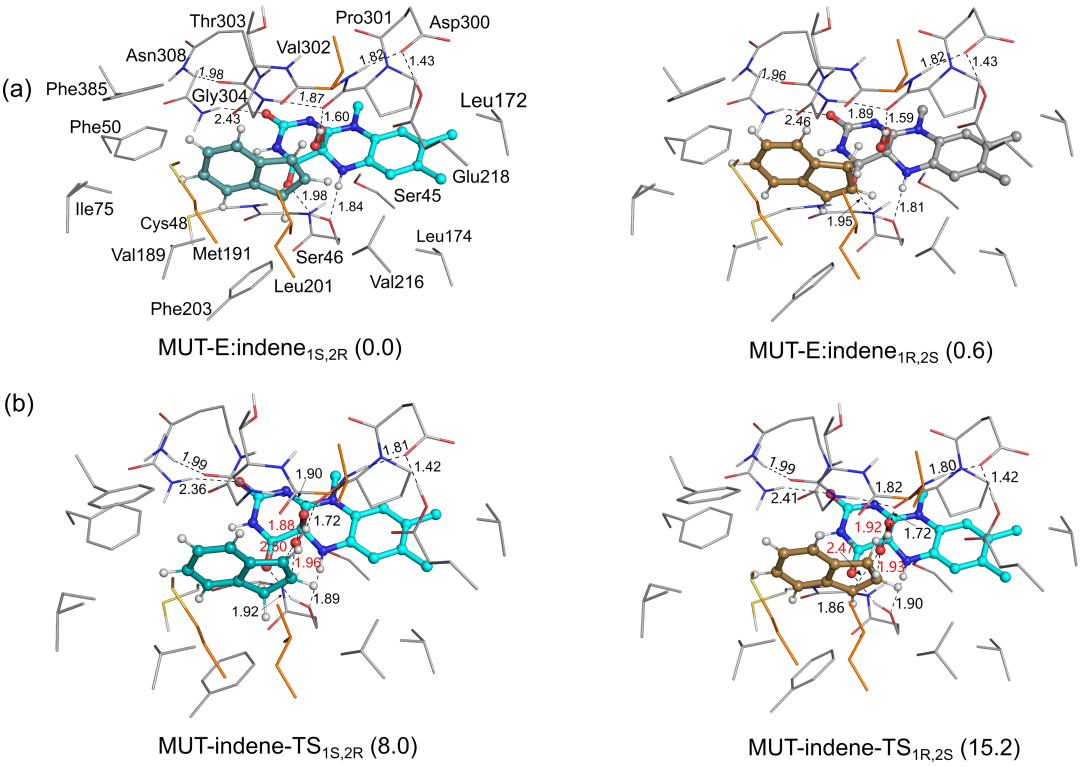

Figure S16. Optimized structures of (a) the active site model in complex with indene (**MUT-E:indene**) and (b) the transition states of indene epoxidation (**MUT-indene-TS**) for the reaction of the Phe191Met/Phe201Leu/Ile302Val mutant. For clarity, most of the hydrogen atoms are omitted in the figure. Selected distances are given in Å.

18. Results on the intrinsic reaction coordinate (IRC) analysis

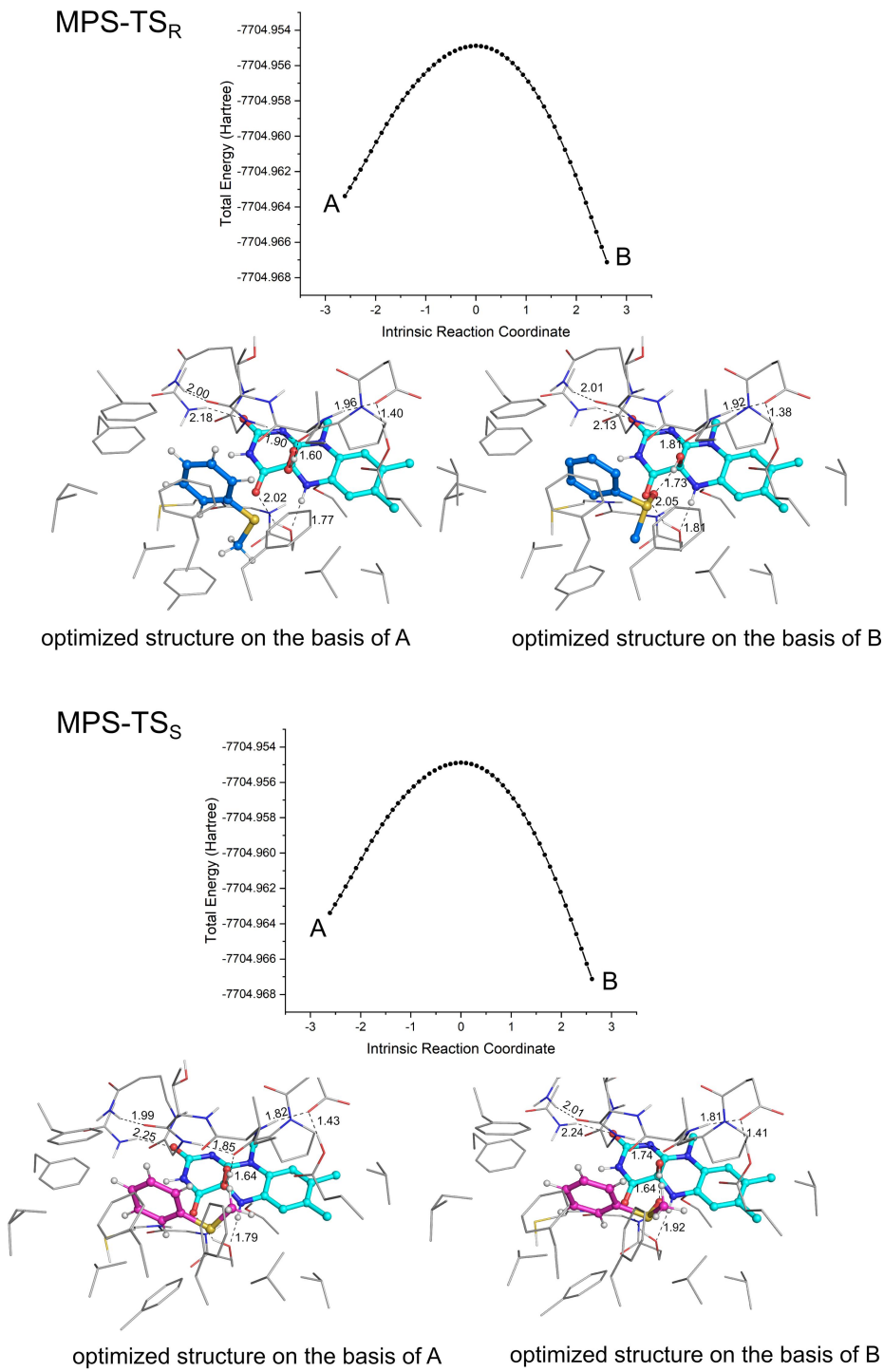

Figure S17. The IRC analysis results of **MPS-TS$_S$** and **MPS-TS$_R$**. For clarity, most of the hydrogen atoms are omitted in the figure for the structures. Selected distances are given in Å.

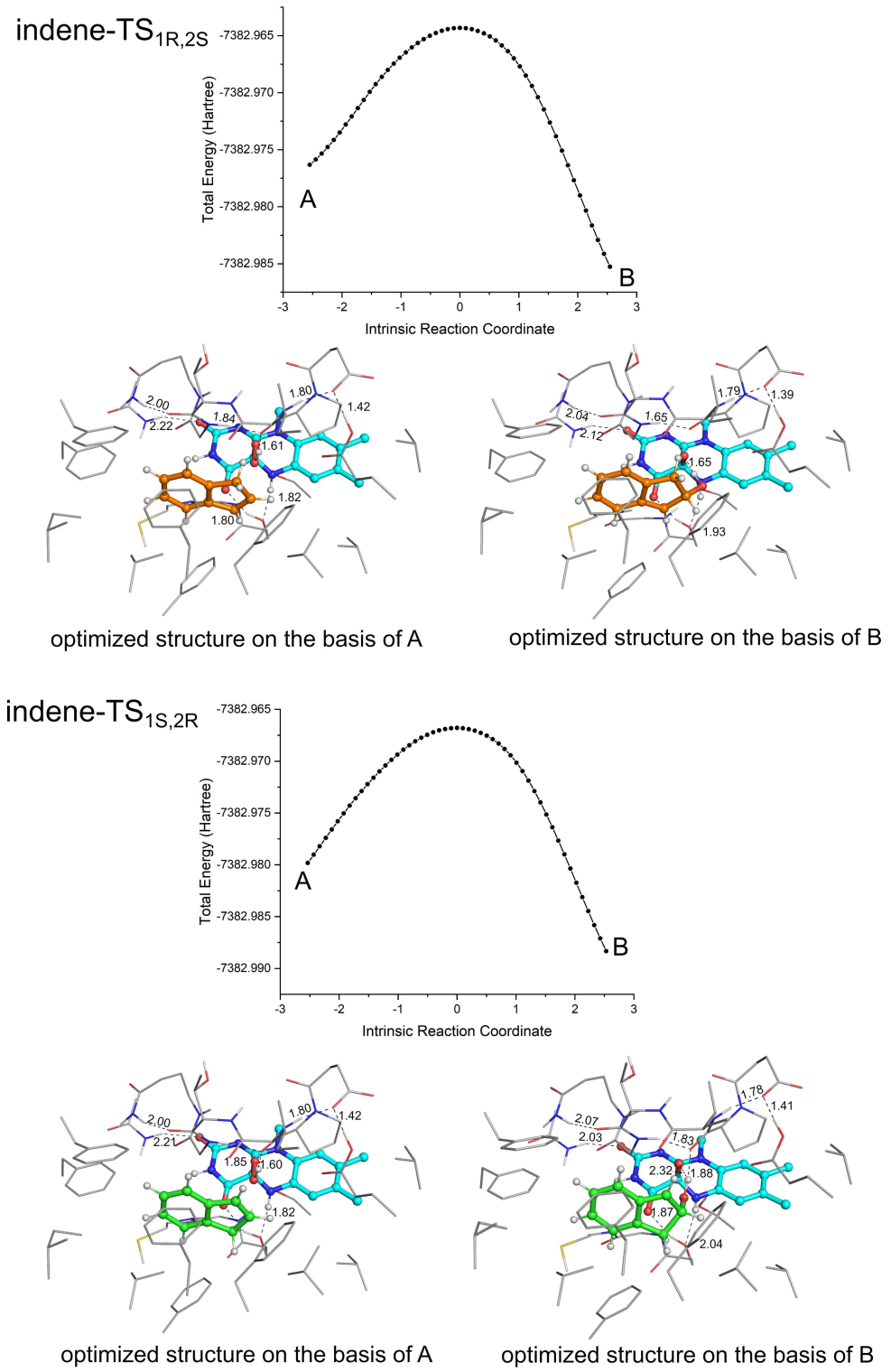

Figure S18. The IRC analysis results of **indene-TS$_{1S,2R}$** and **indene-TS$_{1R,2S}$**. For clarity, most of the hydrogen atoms are omitted in the figure for the structures. Selected distances are given in Å.

19. Absolute energies and energy corrections

Table S2. Calculated absolute energies and energy corrections.
BS1=6-31G(d,p) , BS2=6-311+G(2d,2p)

	E_{BS1} (au)	E_{BS2} (au)	$E_{solvation}$ (au)	E_{ZPE} (au)	E_{total} (au)	ΔE_{total} (kcal/mol)
The sulfoxidation of methyl phenyl sulfide (MPS)						
E:MPS$_S$	-7704.983634	-7706.949986	-7705.040302	2.695876	-7704.310778	0.0
E:MPS$_R$	-7704.979510	-7706.947040	-7705.036116	2.695626	-7704.308020	1.7
MPS-TS$_S$	-7704.969341	-7706.937525	-7705.024861	2.693870	-7704.299175	7.3
MPS-TS$_R$	-7704.954890	-7706.924170	-7705.010882	2.695328	-7704.284834	16.3
E:MPSO$_S$	-7705.031542	-7707.013247	-7705.094529	2.696283	-7704.379950	-43.4
E:MPSO$_R$	-7705.026328	-7707.002466	-7705.084753	2.697050	-7704.363841	-33.3
The epoxidation of indene						
E:indene$_{1R,2S}$	-7382.992234	-7384.944909	-7383.050802	2.706965	-7382.296512	0.0
E:indene$_{1S,2R}$	-7382.990541	-7384.944131	-7383.048161	2.706671	-7382.295080	0.9
indene-TS$_{1R,2S}$	-7382.966760	-7384.918165	-7383.024517	2.705800	-7382.270122	16.6
indene-TS$_{1S,2R}$	-7382.971463	-7384.923497	-7383.028809	2.705666	-7382.275177	13.4
indene-Int$_{1S,2R}$	-7383.009958	-7384.964166	-7383.073481	2.705842	-7382.321848	-15.9
indene-Int$_{1R,2S}$	-7383.021319	-7384.973160	-7383.084048	2.707652	-7382.328237	-19.9
E:IO$_{1R,2S}$	-7383.057193	-7385.007681	-7383.116910	2.709523	-7382.357875	-38.5
E:IO$_{1S,2R}$	-7383.053251	-7385.004669	-7383.113026	2.709040	-7382.355405	-37.0

20. References

1. D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, L. Wilson, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York and P. A. Kollman, *AMBER 2020*, University of California, San Francisco, CA, 2020.
2. J. Mongan, D. A. Case and J. A. McCammon, *J Comput Chem*, 2004, **25**, 2038-2048.
3. J. A. Wallace and J. K. Shen, *Journal of Chemical Theory and Computation*, 2011, **7**, 2617-2629.
4. F. Hofer, J. Kraml, U. Kahler, A. S. Kamenik and K. R. Liedl, *Journal of Chemical Information and Modeling*, 2020, **60**, 3030-3042.
5. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J Comput Chem*, 2004, **25**, 1157-1174.
6. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926-935.
7. A. Onufriev, D. Bashford and D. A. Case, *Proteins: Structure, Function, and Bioinformatics*, 2004, **55**, 383-394.
8. C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski and J. H. Jensen, *Journal of Chemical Theory and Computation*, 2011, **7**, 2284-2295.
9. E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.-W. Wei, M. J. Holst, J. A. McCammon and N. A. Baker, *Protein Science*, 2018, **27**, 112-128.
10. U. E. Ukaegbu, A. Kantz, M. Beaton, G. T. Gassner and A. C. Rosenzweig, *Biochemistry*, 2010, **49**, 1678-1688.
11. J. Kratky, D. Eggerichs, T. Heine, S. Hofmann, P. Sowa, R. H. Weiße, D. Tischler and N. Sträter, *Angewandte Chemie International Edition*, 2023, **62**, e202300657.