Supporting information

Regio- and Stereoselectivity of $CYP450_{BM3}$-Catalyzed Hydroxylation of Complex Terpenoids: A QM/MM Study

Chenggong Hui[1≠], Warispreet Singh [1, 2≠,§], Derek Quinn [2], Chun Li, [3,4] Thomas S. Moody[2,5], Meilan Huang[1]*

[1] Department of Chemistry & Chemical Engineering, Queen's University, Belfast, BT9 5AG Northern Ireland, United Kingdom

[2] Almac Sciences, Department of Biocatalysis and Isotope Chemistry, Almac House, 20 Seagoe Industrial Estate, Craigavon BT63 5QD, Northern Ireland, United Kingdom

[3] Institute for Synthetic Biosystem/Department of Biochemical Engineering, School of Chemistry and Chemical Engineering, Beijing Institute of Technology, Beijing 100081, P.R. China

[4] Key Lab of Industrial Biocatalysis Ministry of Education, Institute of Biochemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing, 100084, P. R. China

[5] Arran Chemical Company Limited, Unit 1 Monksland Industrial Estate, Athlone, Co. Roscommon, Ireland

[≠]*These authors contribute equally to the paper*

[§]*Current address: Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, England, United Kingdom*

Table S1 The activation energy (E + ZPE) barrier (kcal/mol) for hydrogen abstraction starting from representative cluster structures of $P450_{BM3}$ variants. Geometry optimizations were conducted using UB3LYP/def2-SVP(B1) with D3BJ and single point energy was corrected using def2-TZVPP(B2')

	Cluster	O-H (Å)	EE_{B1}	$E_{B1} + ZPE_{B1}$	$E_{B2'} + ZPE_{B1}$	Average
X-E12 C6a-CH_3	C0	2.35	25.5	20.7	22.1	**22.7**
	C1	2.95	27.8	23.3	24.7	
	C2	2.79	26.6	22.3	23.9	
	C3	2.61	26.6	21.9	22.8	
IV-H4 *pro*-C7(S)	C1	2.58	24.7	20.5	20.3	**19.3**
	C2	2.66	29.0	24.5	25.0	
	C3	2.36	21.6	17.0	18.7	
IV-H4 *pro*-C7(R)	C1	3.09	33.4	29.2	30.3	**24.5**
	C2	3.01	35.5	30.8	31.3	
	C3	3.20	25.9	21.3	23.8	
II-H10 *pro*-C7(R)	F87-in1_L-QM	2.64	27.4	22.4	23.8	**24.2**
	F87-in2_L-QM	2.74	29.6	24.9	26.9	
	F87-in1_S-QM	2.67	28.9	23.9	24.6	**25.0**
	F87-in2_S-QM	2.7	30.2	25.2	26.2	
	F87-out1	2.26	21.4	17.0	18.4	**18.2**
	F87-out2	2.24	22.1	17.0	18.1	

Table S2 Benchmark on basis set and multiplicity for the activation energy barrier (kcal/mol) of X-E12. Geometry optimizations were conducted using UB3LYP with D3BJ. Def2-SVP (B1) and def2-TZVP (B2) basis sets and S=1/2 and S=2/3 multiplicities were calculated. Frequency calculations were conducted for validating the transition states obtained.

(A) Electric energy difference in the two basis sets and multiplicities.

Multiplicities	**S=1/2**		**S=3/2**	
Basis set	B1	B2	B1	B2
$\Delta E^{\neq}$ (E_{TS}-E_R)	25.5	27.4	26.1	27.5
ΔE (E_P-E_R)	19.6	19.4	19.8	19.5
$\Delta E^{\neq}_{B2}$ ($E_{TS} - E_R$) – $\Delta E^{\neq}_{B1}$ ($E_{TS} - E_R$)	**1.8**		**1.4**	
ΔE_{B2} ($E_P - E_R$) – ΔE_{B1} ($E_P - E_R$)	-0.2		-0.2	

(B) ZPE and frequency difference in the two basis sets B1 and B2.

Multiplicities	**S=1/2**		**S=3/2**	
ΔE_R ($E_{B2} - E_{B1}$)	-0.7		-0.7	
ΔE_{TS} ($E_{B2} - E_{B1}$)	-0.5		-0.4	
ΔE_P ($E_{B2} - E_{B1}$)	-0.8		-0.8	
$\Delta E^{\neq}_{B2}$ ($E_{TS} - E_R$) – $\Delta E^{\neq}_{B1}$ ($E_{TS} - E_R$)	**0.2**		**0.3**	
ΔE_{B2}($E_P - E_R$) – ΔE_{B1} ($E_P - E_R$)	-0.0		-0.1	
TS Imaginary Freq	-1847.773	-1952.050	-1913.878	-1955.459

(C) E + ZPE difference in the two basis sets and multiplicities.

Multiplicities	**S=1/2**		**S=3/2**	
Basis set	B1	B2	B1	B2
$\Delta E^{\neq}$(E_{TS}-E_R)	20.7	22.7	21.1	22. 8
ΔE(E_P-E_R)	17.0	16.8	17.21	16.9

Table S3 Geometry parameters in the QM/MM optimized geometries. Basis set and multiplicity benchmark for the activation energy (E + ZPE) barrier. Geometry optimizations were conducted using UB3LYP with D3BJ and two basis sets def2-SVP (B1) and def2-TZVP (B2) for X-E12. For the other variants IV-H4 and II-H10 variants, same DFT method with B1 basis set was used in the geometry optimizations. The unit of distance is Å and that of bond angle is °.

Stationary points	**C selectivity**	X-E12-C6a				IV-H4 *pro*-C7(S)	II-H10 *pro*-C7(R)-Phe87 in		II-H10 *pro*-C7(R)-Phe87 out
	Cluster	(Cluster 0)				(Cluster 3)	**(S-QM)** (Cluster 1)	**(L-QM)** (Cluster 1)	**(S-QM)** (Cluster 2)
	Multiplicity	S=1/2		S=3/2		S=1/2	S=1/2	S=1/2	S=1/2
	Basis set	B1	B2	B1	B2	B1	B1	B1	B1
Reactant	**Fe=O**	1.61	1.62	1.61	1.62	1.61	1.61	1.61	1.61
TS		**1.73**	**1.73**	**1.73**	**1.73**	**1.73 (7S)**	**1.74(7R)**	**1.75(7R)**	**1.75(7R)**
Reactant	**O-H**	**2.35**	**2.45**	**2.35**	**2.40**	**2.36**	**2.69**	**2.71**	**2.24**
TS		**1.15**	**1.15**	**1.15**	**1.15**	**1.16 (7S)**	**1.20(7R)**	**1.20(7R)**	**1.17(7R)**
Reactant	**C-H**	1.10	1.09	1.10	1.09	1.10	1.10	1.10	1.10
TS		**1.39**	**1.39**	**1.40**	**1.40**	**1.36 (7S)**	**1.34(7R)**	**1.34(7R)**	**1.34(7R)**
Reactant	**Fe=O…C**	**3.13**	**3.18**	**3.13**	**3.15**	**3.07**	**3.44**	**3.45**	**3.12**
Reactant	***angle***	**120.3**	**121.5**	**120.3**	**122.0**	**124.2**	**154.5**	**153.5**	**163.3**

Table S4 The relative free energies (kcal/mol) of each stationary points for hydrogen abstraction and rebound steps in XE12. Re-TS-IM1 is the hydrogen abstraction step, and IM1-TS2-P is the rebound step.

		R	**TS1**	**IM1**	**TS2**	**P**
C1	$\mathbf{E_{B1} + ZPE_{B1}}$	0.0	23.4	13.5	18.0	-44.2
	$\mathbf{E_{B2'} + ZPE_{B1}}$	**0.0**	**24.7**	**11.8**	**17.5**	**-40.5**
C2	$\mathbf{E_{B1} + ZPE_{B1}}$	0.0	22.3	9.6	14.8	-47.4
	$\mathbf{E_{B2'} + ZPE_{B1}}$	**0.0**	**23.9**	**8.8**	**15.2**	**-45.4**

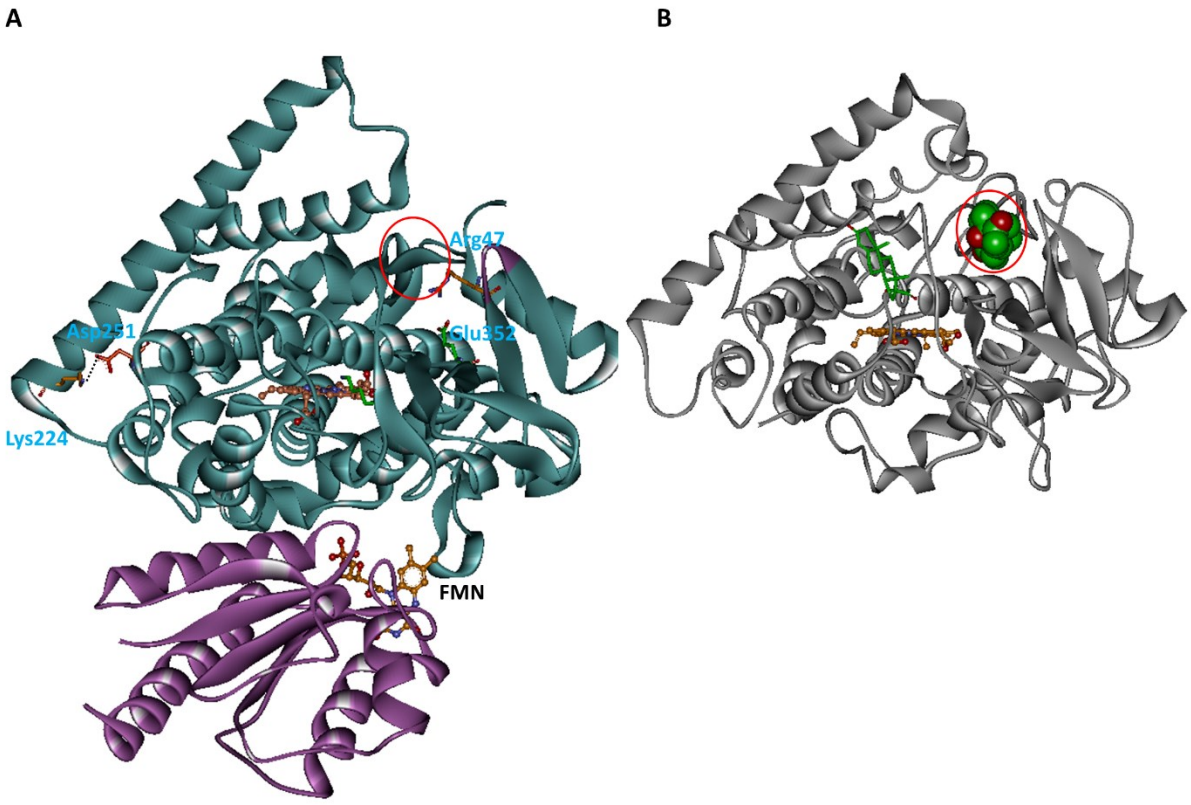


Figure S1. Comparison of crystal structures of $P450_{BM3}$ and P450c21 (A) the crystal structure of $P450_{BM3}$ composed of heme and FMN domains (PDB: 1BVY); **(B)** The crystal structure of bovine P450c21 in complex with two steroids (PDB: 3QZ1), [1] one steroid (shown in stick mode) located at the catalytic centre and the other (shown in CPK mode) located at the substrate entrance. The substrate entrance is highlighted in red circle.

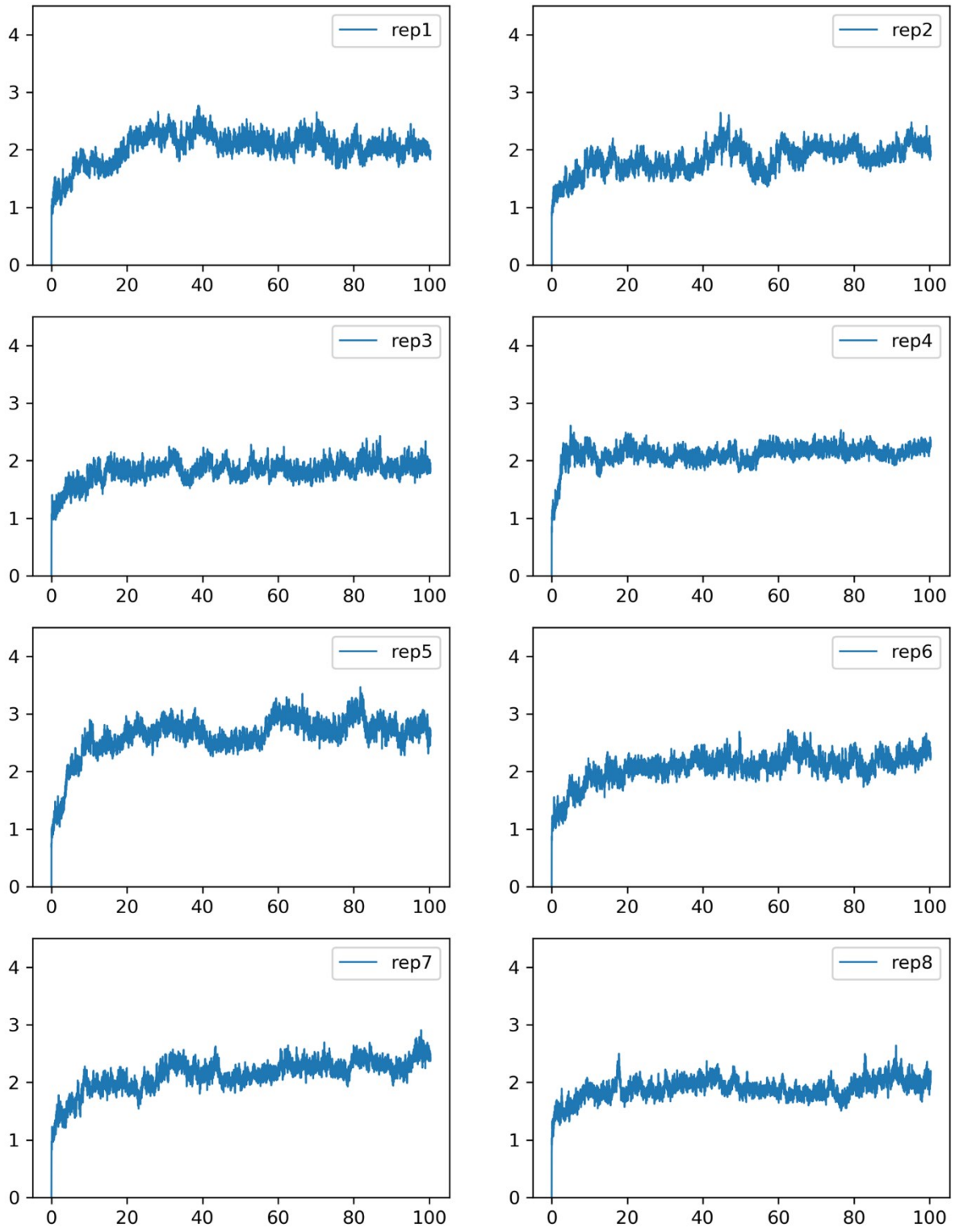
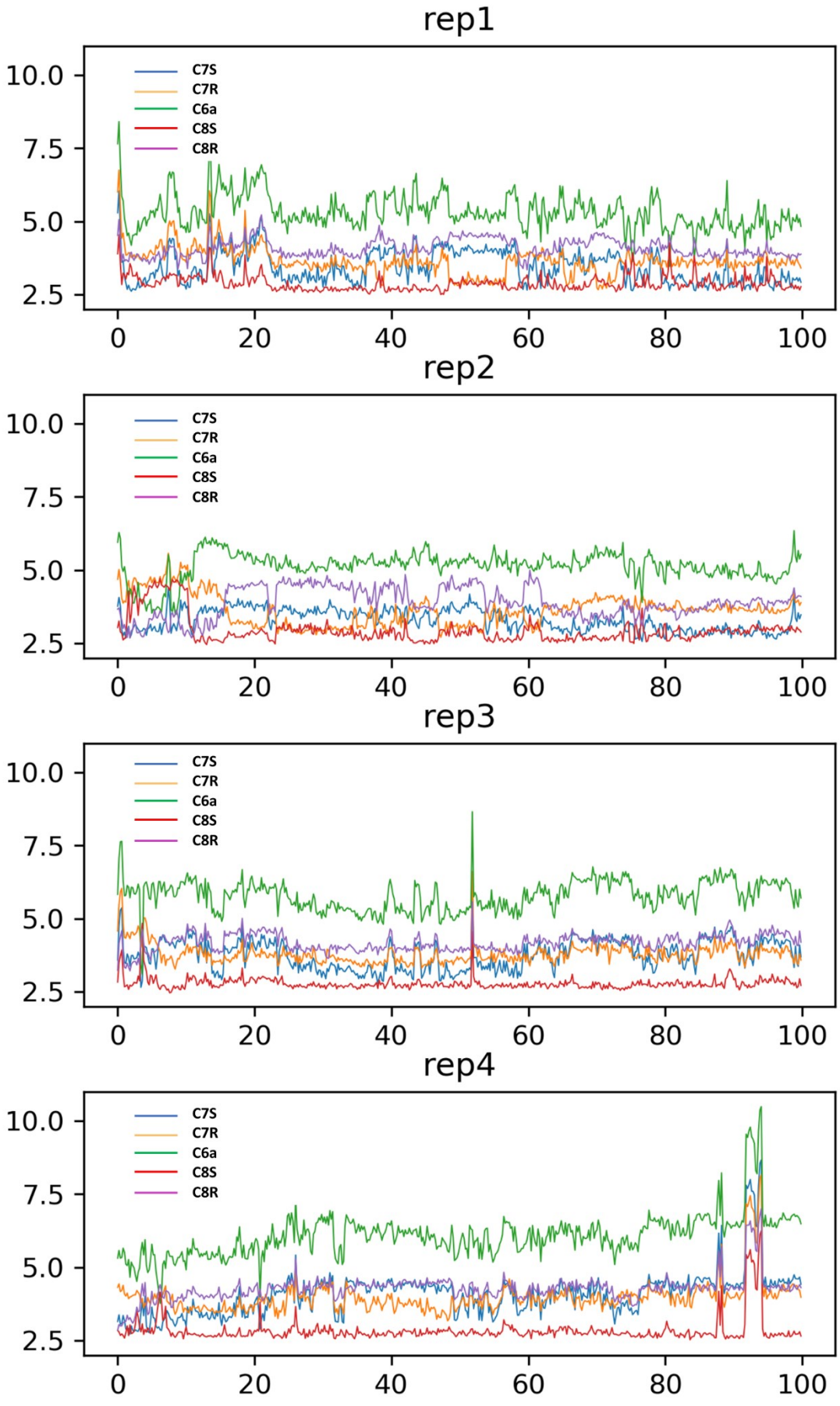


Figure S2. RMSD of Cα of FL≠62, the $P450_{BM3}$ variant template bound with *pro*-C7(S) from eight replicas of MD simulations


rep1
C7S
C7R
C6a
C8S
C8R
10.0
7.5
5.0
2.5
0
20
40
60
80
100
rep2
C7S
C7R
C6a
C8S
C8R
10.0
7.5
5.0
2.5
0
20
40
60
80
100
rep3
C7S
C7R
C6a
C8S
C8R
10.0
7.5
5.0
2.5
0
20
40
60
80
100
rep4
C7S
C7R
C6a
C8S
C8R
10.0
7.5
5.0
2.5
0
20
40
60
80
100

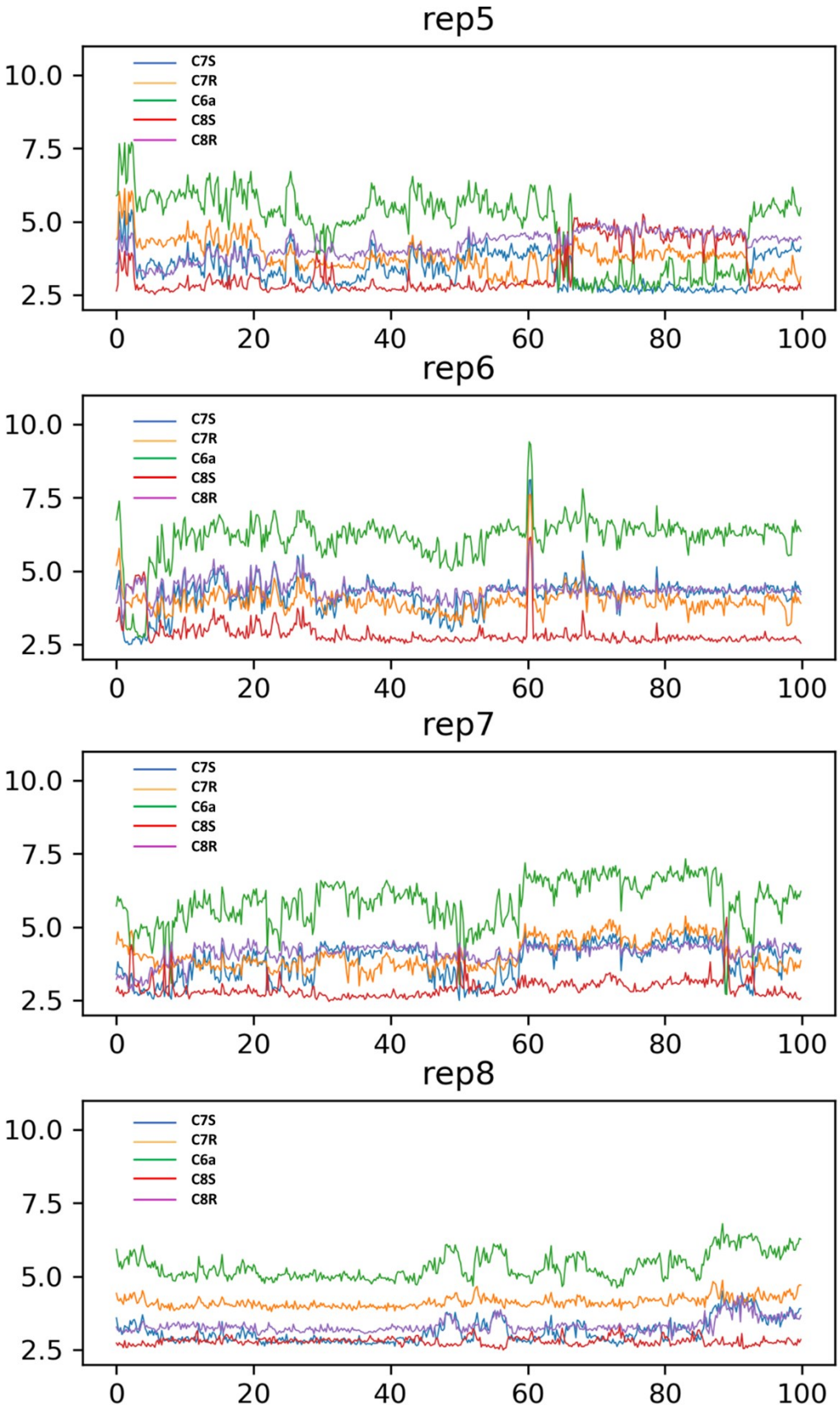


Figure S3. Distance between Fe=O oxo and different H in FL≠62, the $P450_{BM3}$ variant template bound with *pro*-C7(S) from eight replicas of MD simulations.

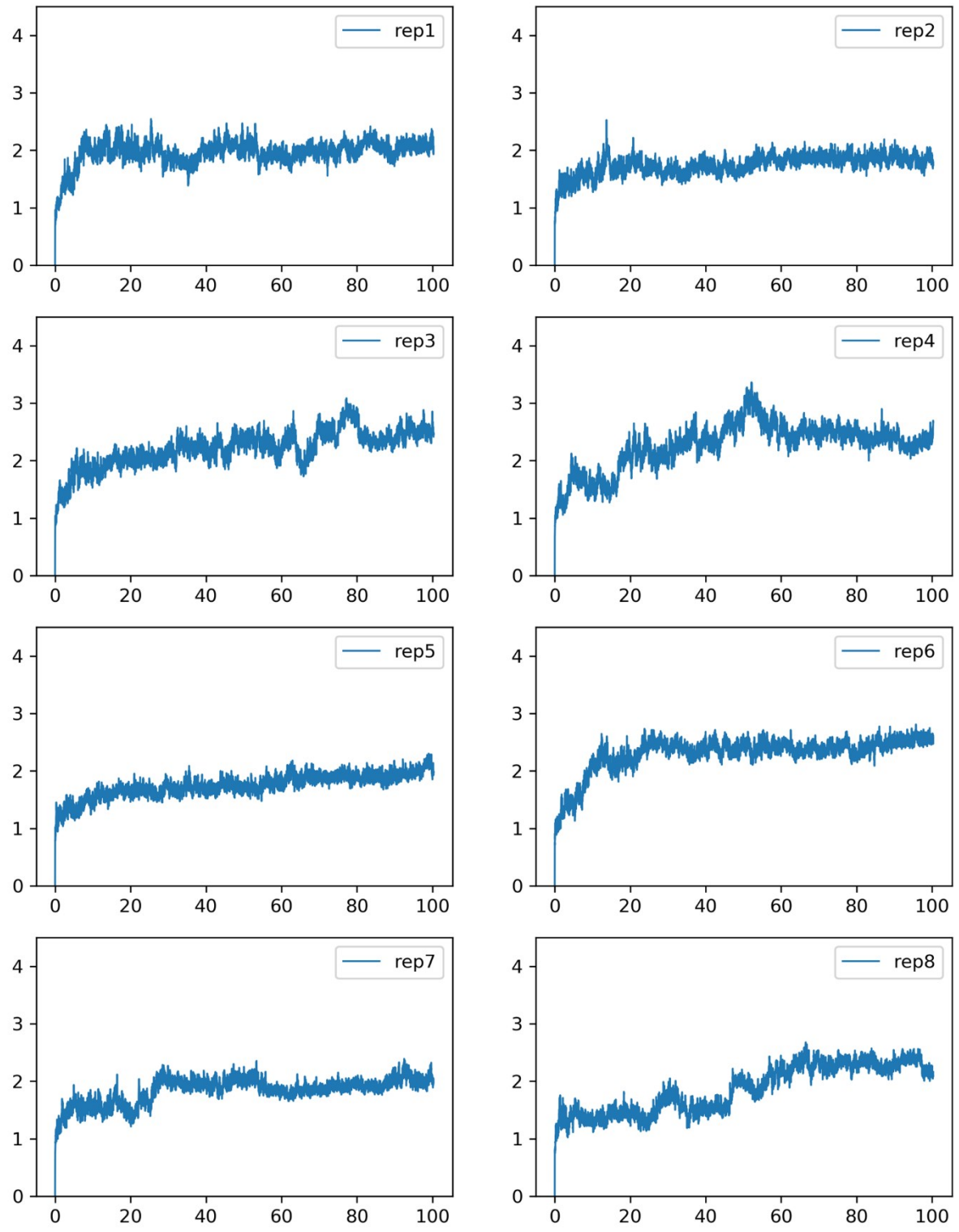
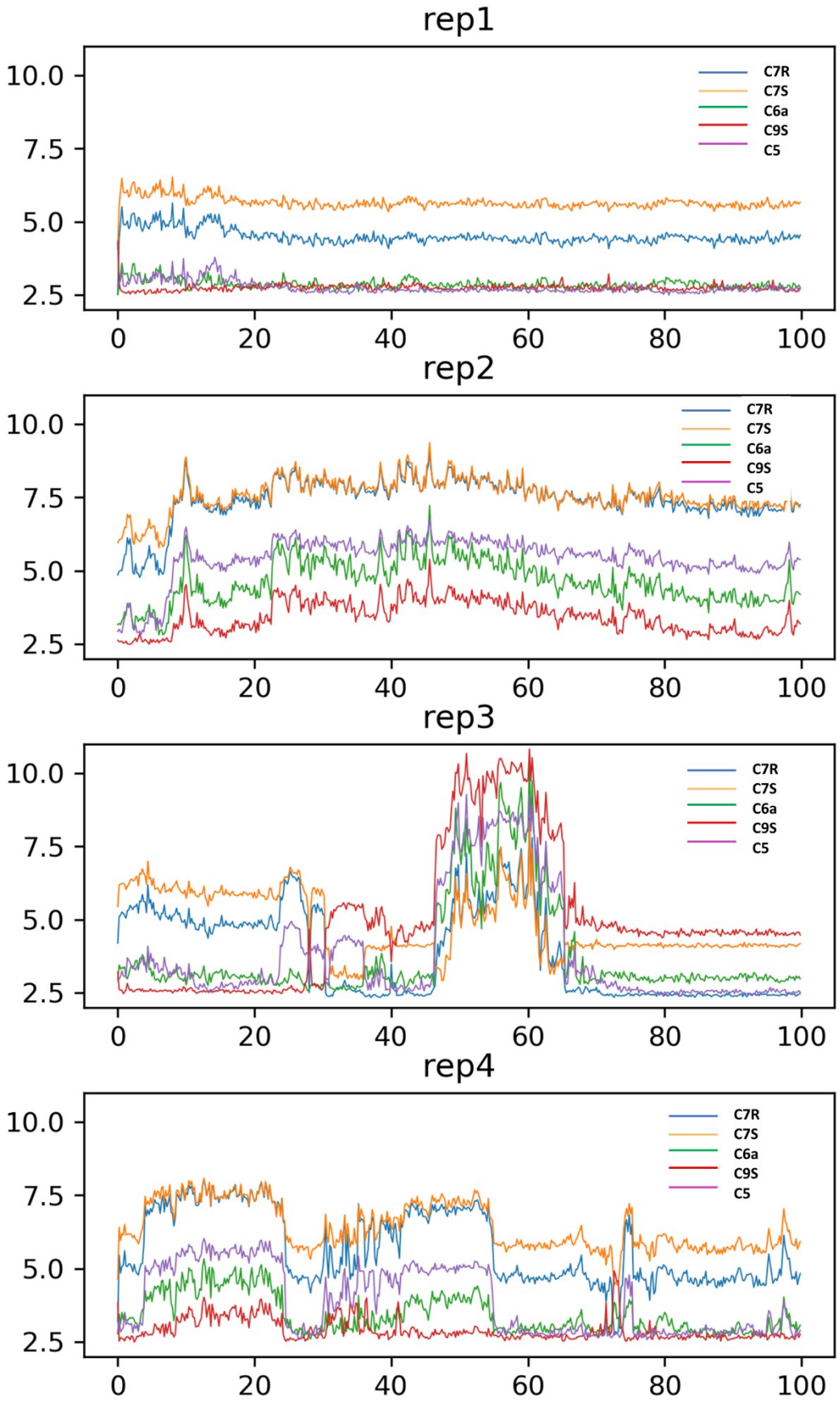


Figure S4. RMSD of Cα of FL≠62, the $P450_{BM3}$ variant template bound with *pro*-C7(R) from eight replicas of MD simulations.


rep1
rep2
rep3
rep4
C7R
C7S
C6a
C9S
C5
10.0
7.5
5.0
2.5
0
20
40
60
80
100

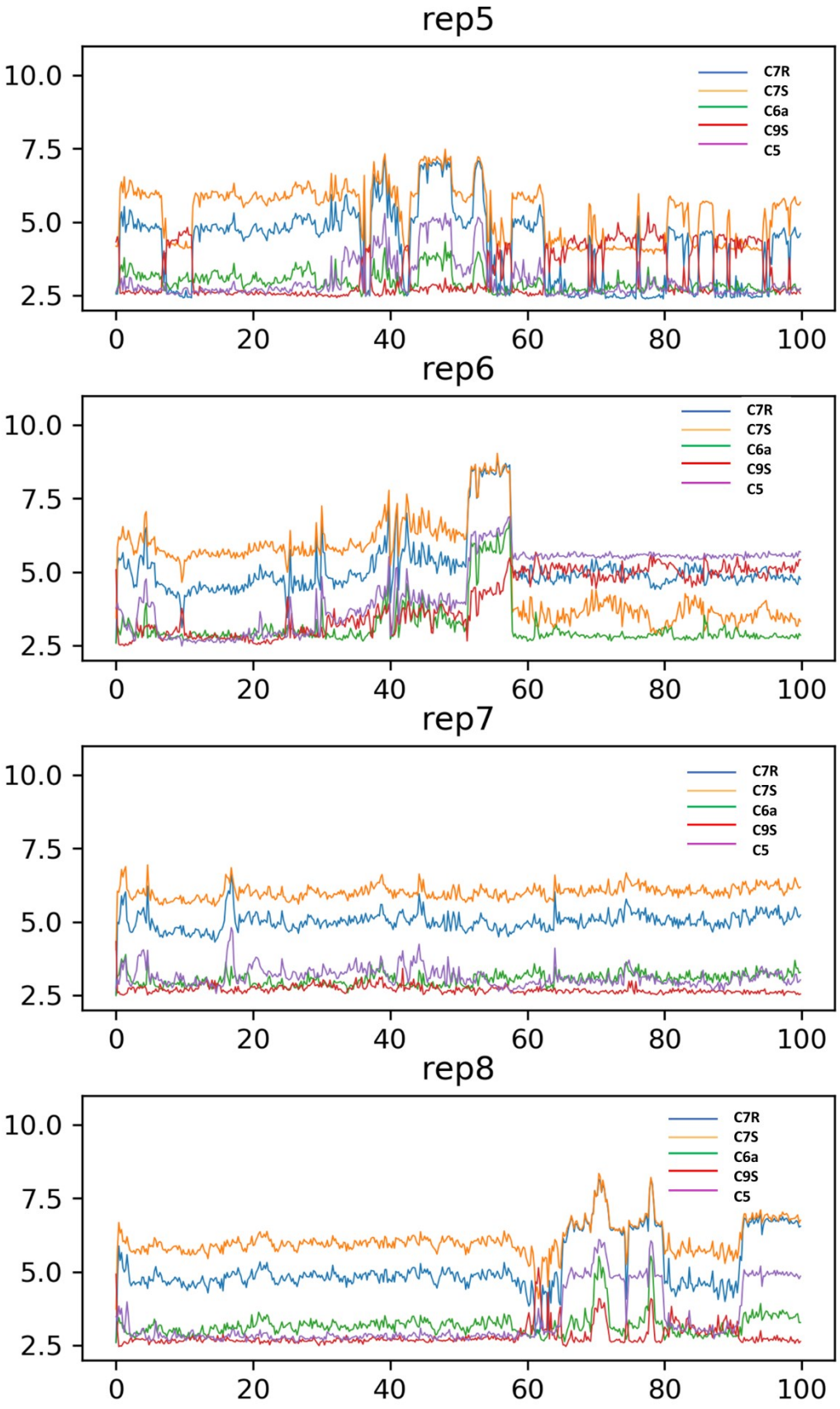


Figure S5. Distance between Fe=O and different H in FL≠62, the $P450_{BM3}$ variant template bound with *pro*-C7(R) from eight replicas of MD simulations.

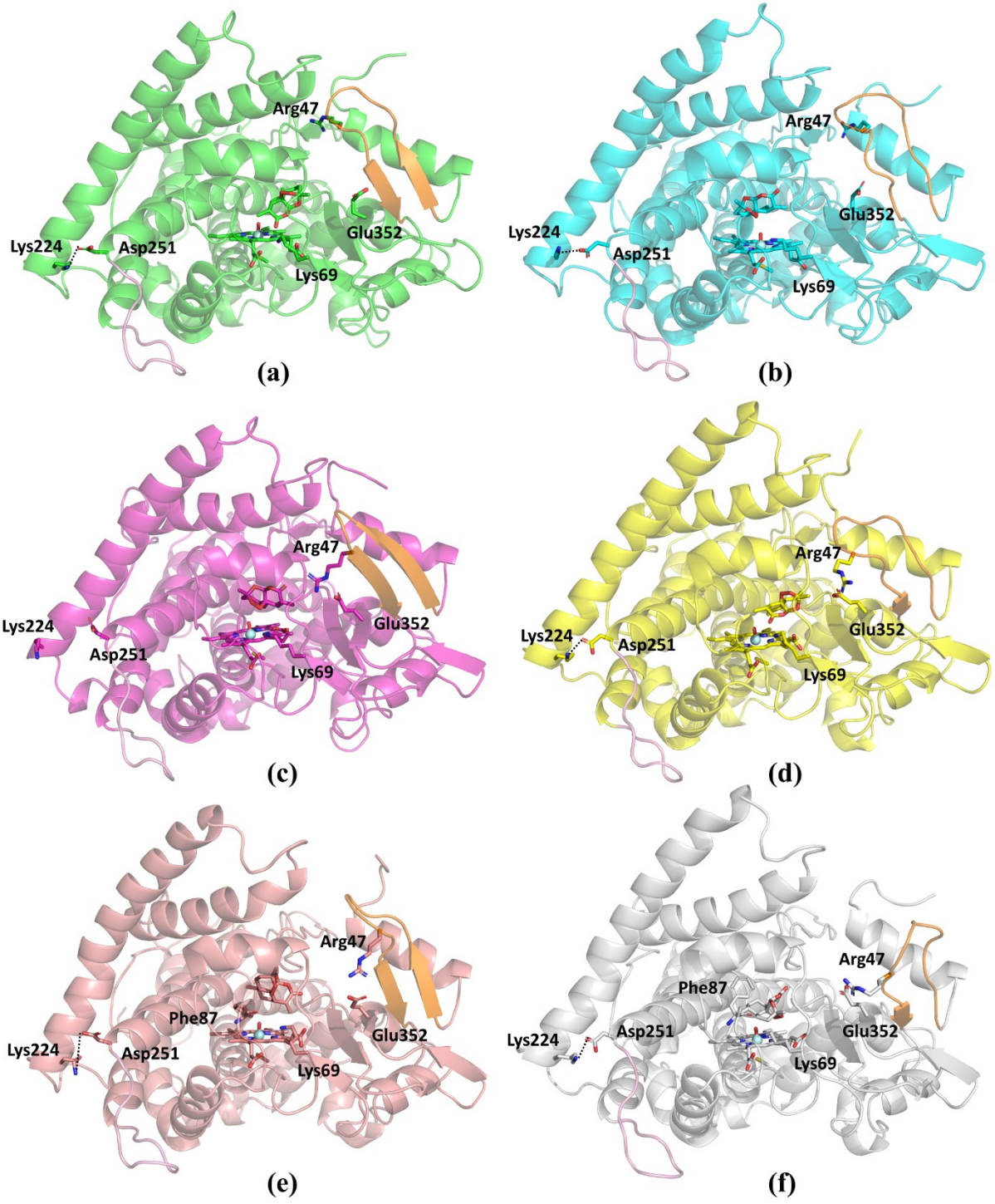


Figure S6. MD simulated structures of P450$_{BM3}$ variants in complex with arteminsinin. (a) FL#62 with *pro*-C7(S), (b) FL#62 with *pro*-C7(R), (c) X-E12 with preference for abstraction of H from C6a, (d) IV-H4 with *pro*-C7(S), (e) II-H10 with *pro*-C7(R) (Phe87-in), (f) II-H10 with *pro*-C7(R) (Phe87-out). The β1 hairpin at the substrate entrance is shown in orange and the loop contacting the FMN domain is shown in pink.

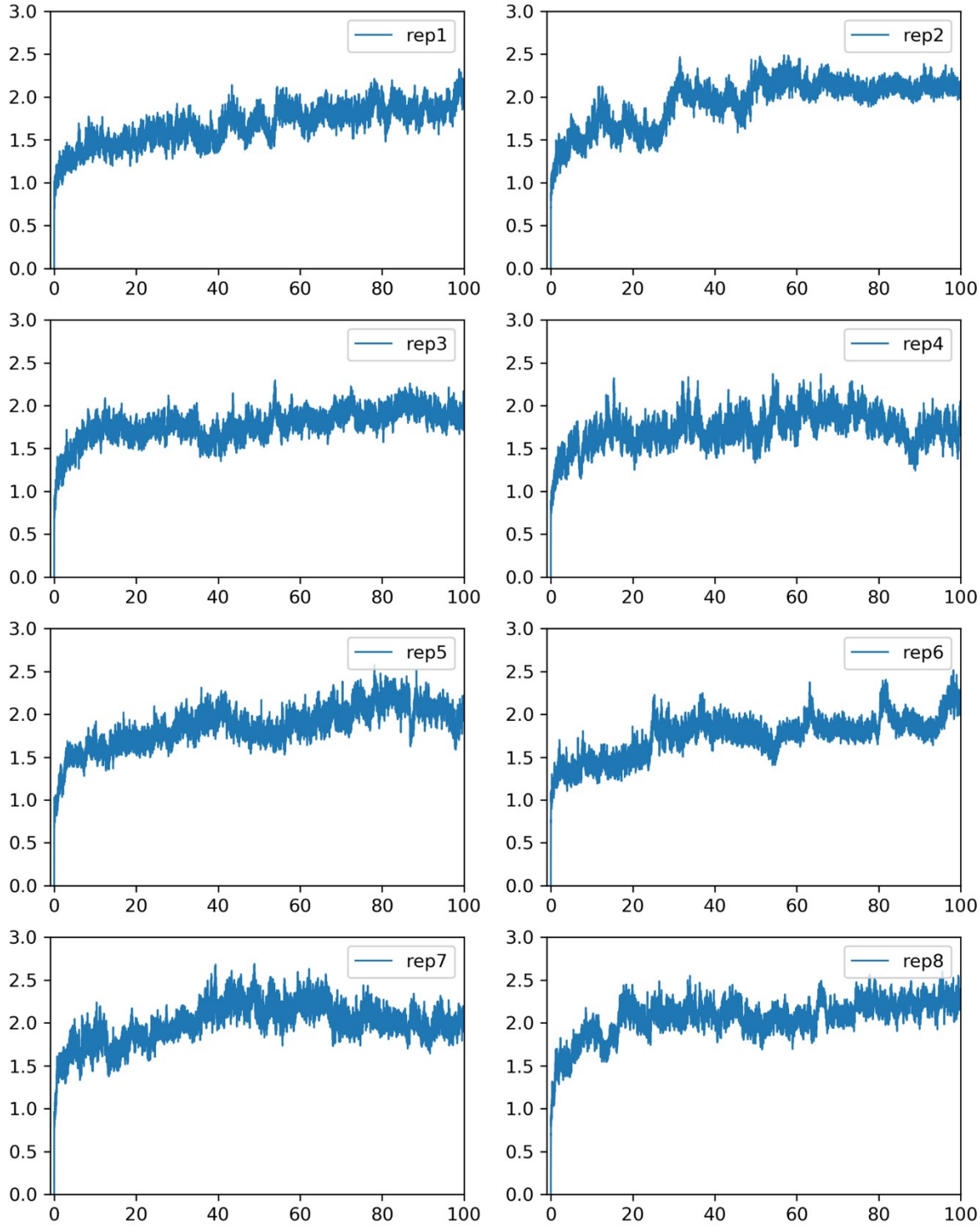
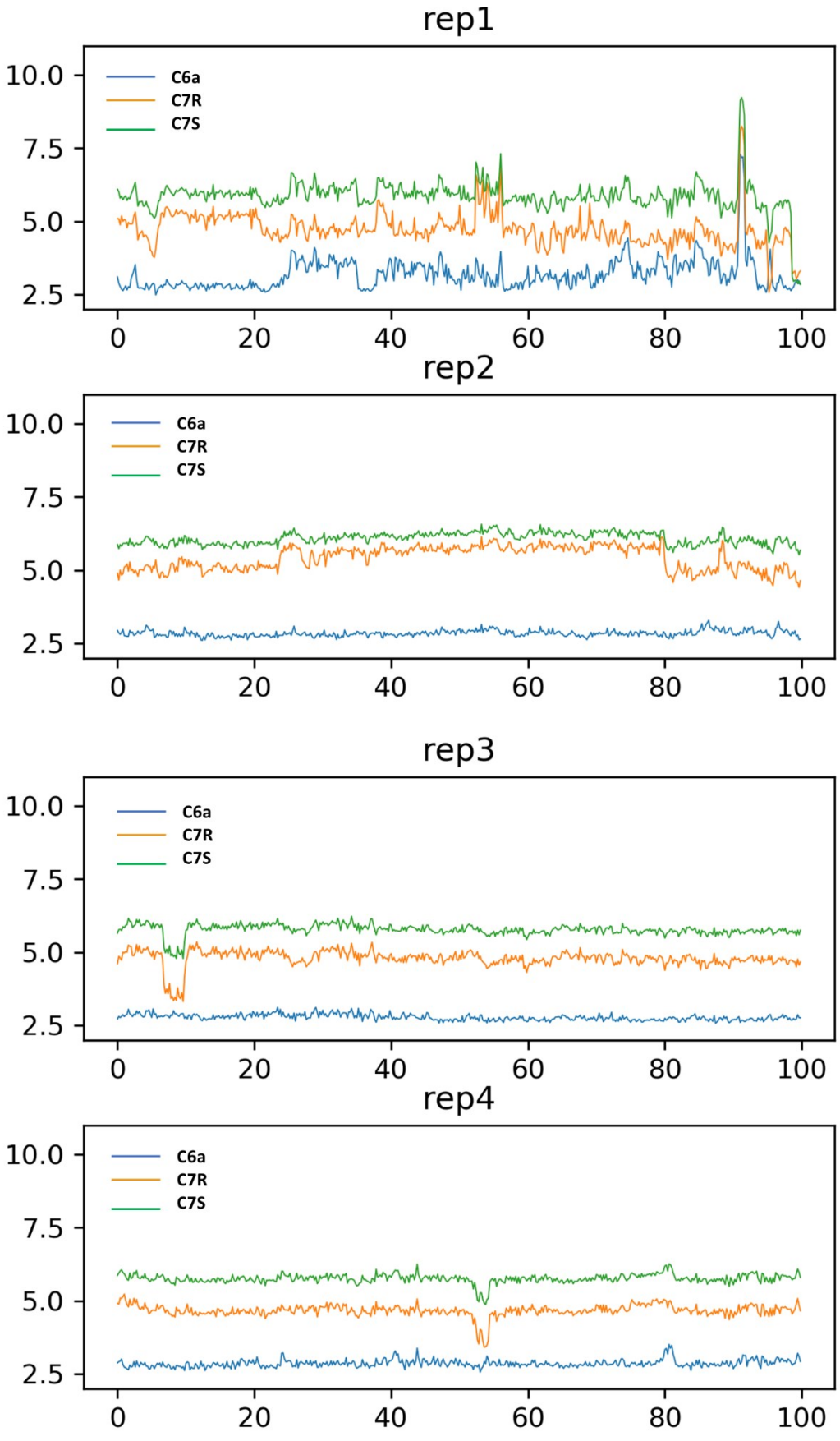


Figure S7. RMSD of Cα in MD simulations of X-E12. Eight replicas of MD simulations were run for X-E12, the $P450_{BM3}$ variant bound in complex with artemisinin with C6a preference.


rep1
C6a
C7R
C7S
10.0
7.5
5.0
2.5
0
20
40
60
80
100
rep2
C6a
C7R
C7S
10.0
7.5
5.0
2.5
0
20
40
60
80
100
rep3
C6a
C7R
C7S
10.0
7.5
5.0
2.5
0
20
40
60
80
100
rep4
C6a
C7R
C7S
10.0
7.5
5.0
2.5
0
20
40
60
80
100

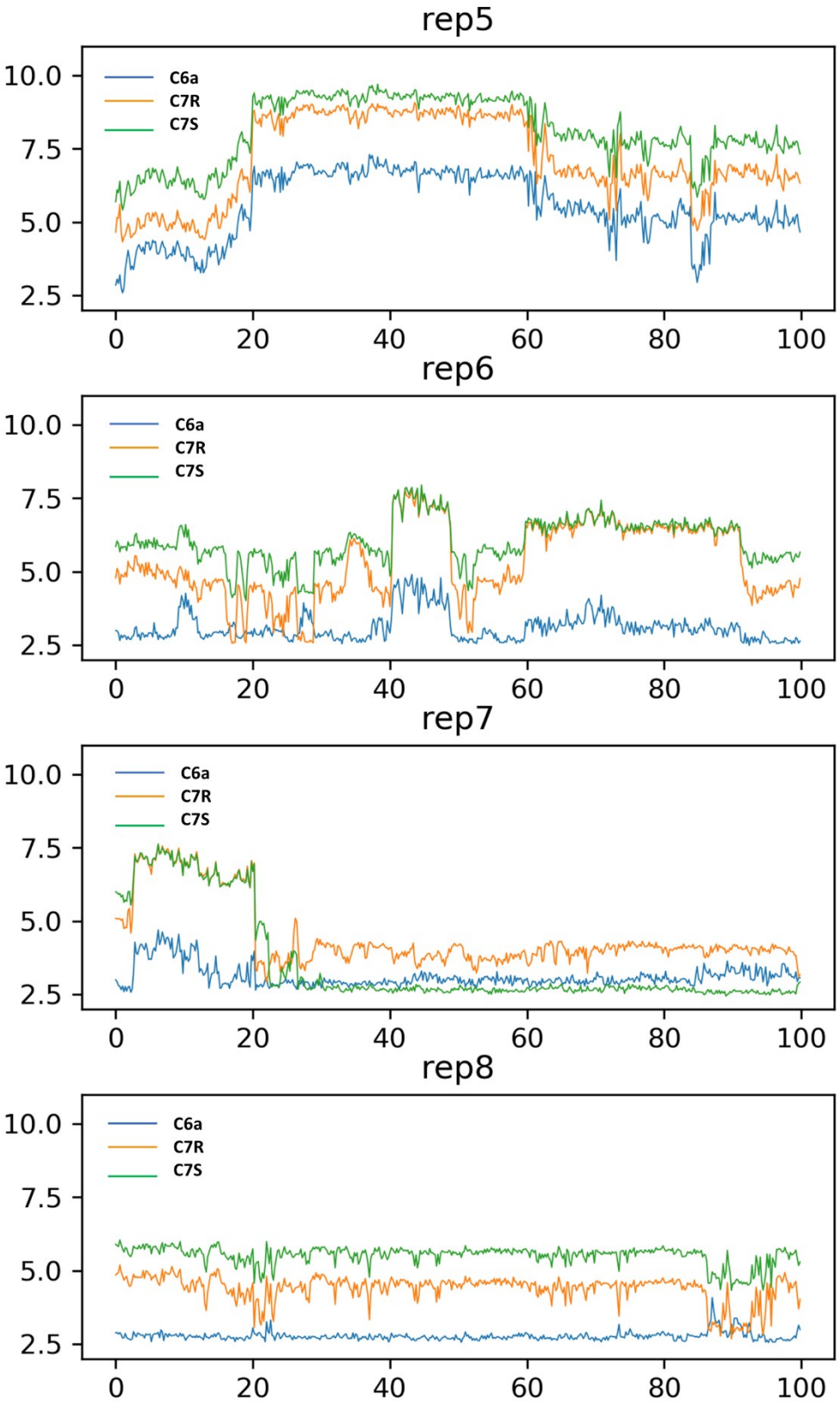


Figure S8. Distance between Fe=O oxo and different H in MD simulations of X-E12. Eight replicas of MD simulations were run for X-E12 start with the enzyme bound with C6a arteminsinin.

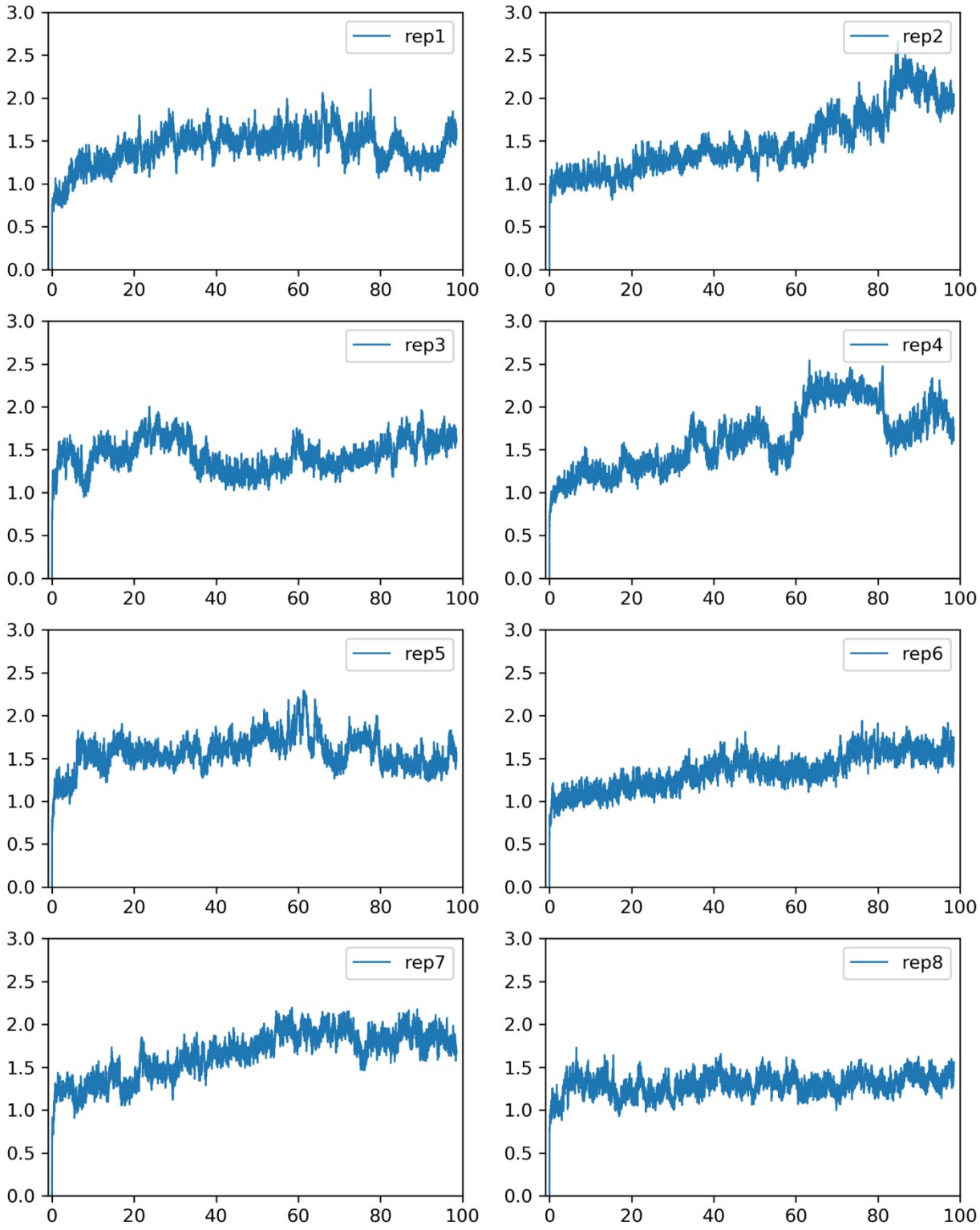
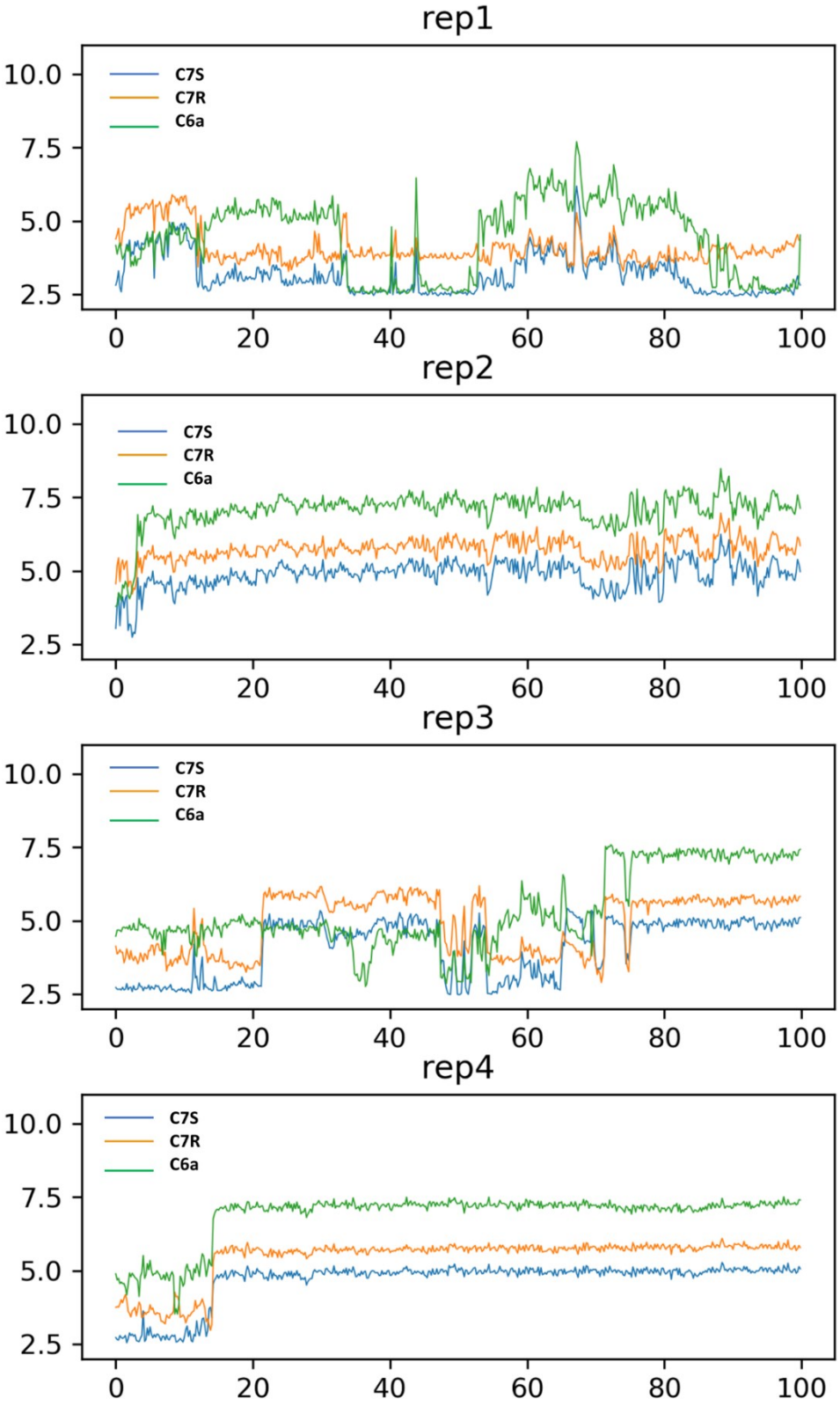


Figure S9. RMSD of Cα in MD simulations of IV-H4. Eight replicas of MD simulations of were run for IV-H4, the $P450_{BM3}$ variant bound with *pro*-C7(S) artemisinin.


rep1
C7S
C7R
C6a
10.0
7.5
5.0
2.5
0
20
40
60
80
100
rep2
C7S
C7R
C6a
10.0
7.5
5.0
2.5
0
20
40
60
80
100
rep3
C7S
C7R
C6a
10.0
7.5
5.0
2.5
0
20
40
60
80
100
rep4
C7S
C7R
C6a
10.0
7.5
5.0
2.5
0
20
40
60
80
100

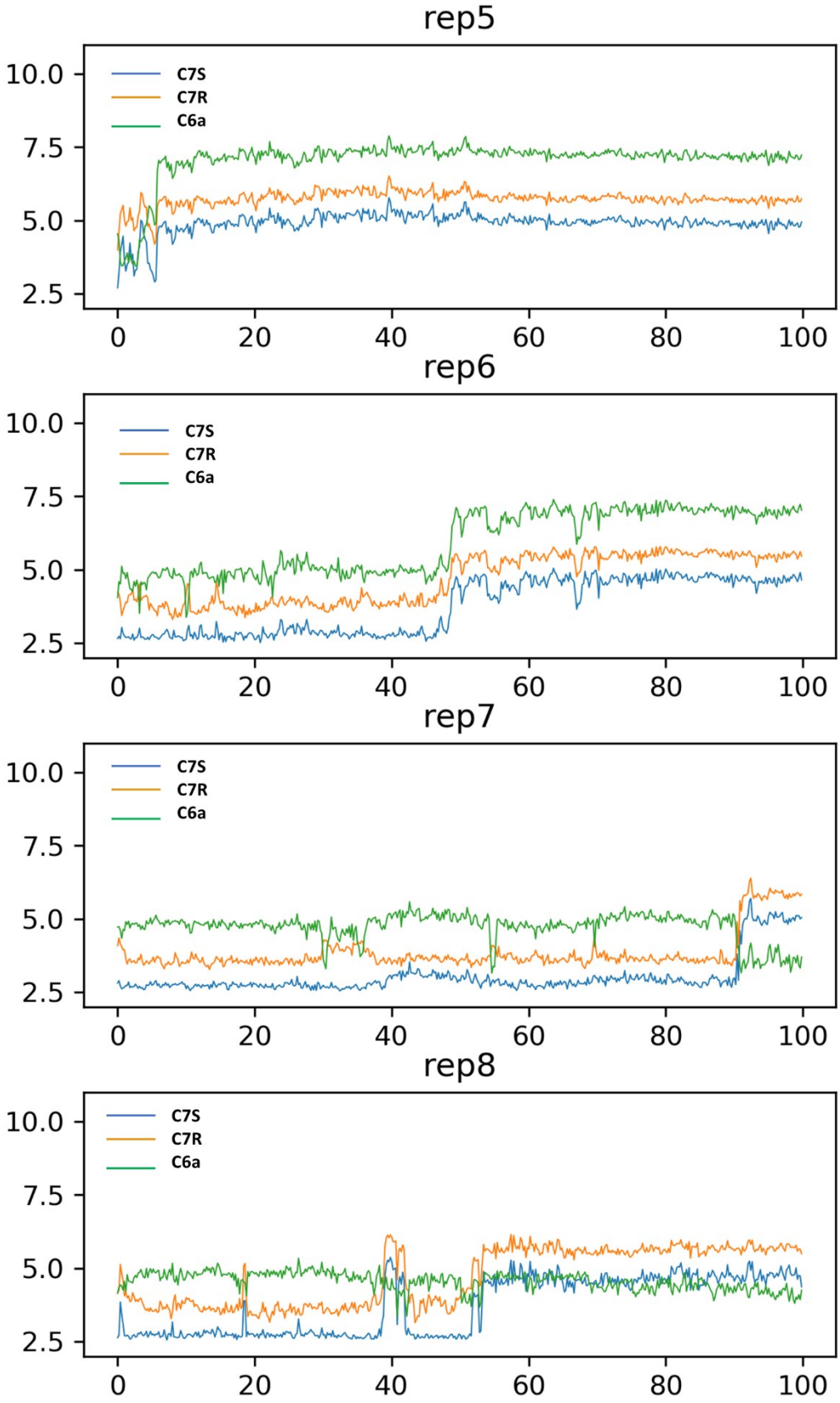


Figure S10. Distance between Fe=O oxo and different H in MD simulations of IV-H4. Eight replicas of MD simulations were run for IV-H4, the $P450_{BM3}$ variant bound with *pro*-C7(S) arteminsinin.

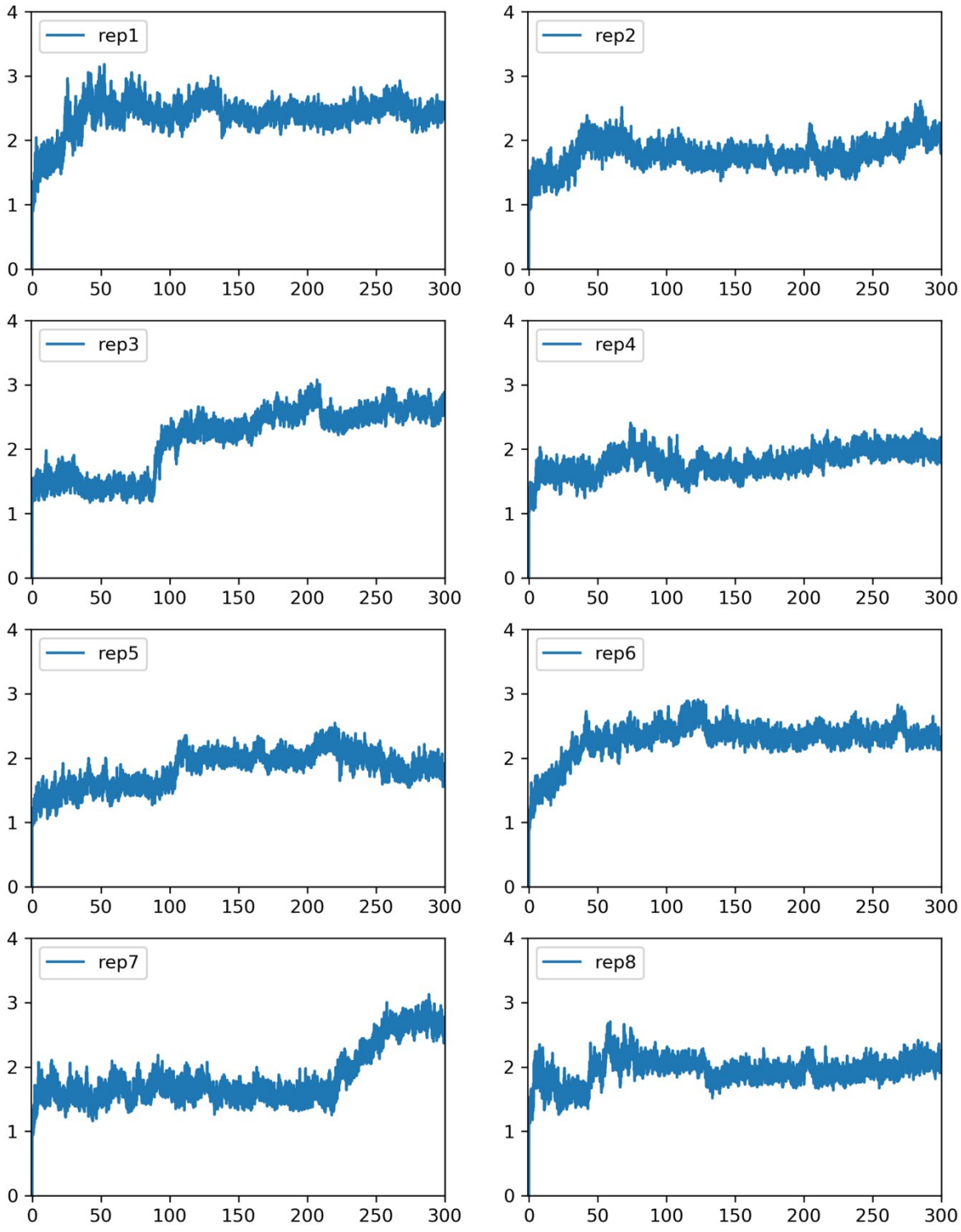
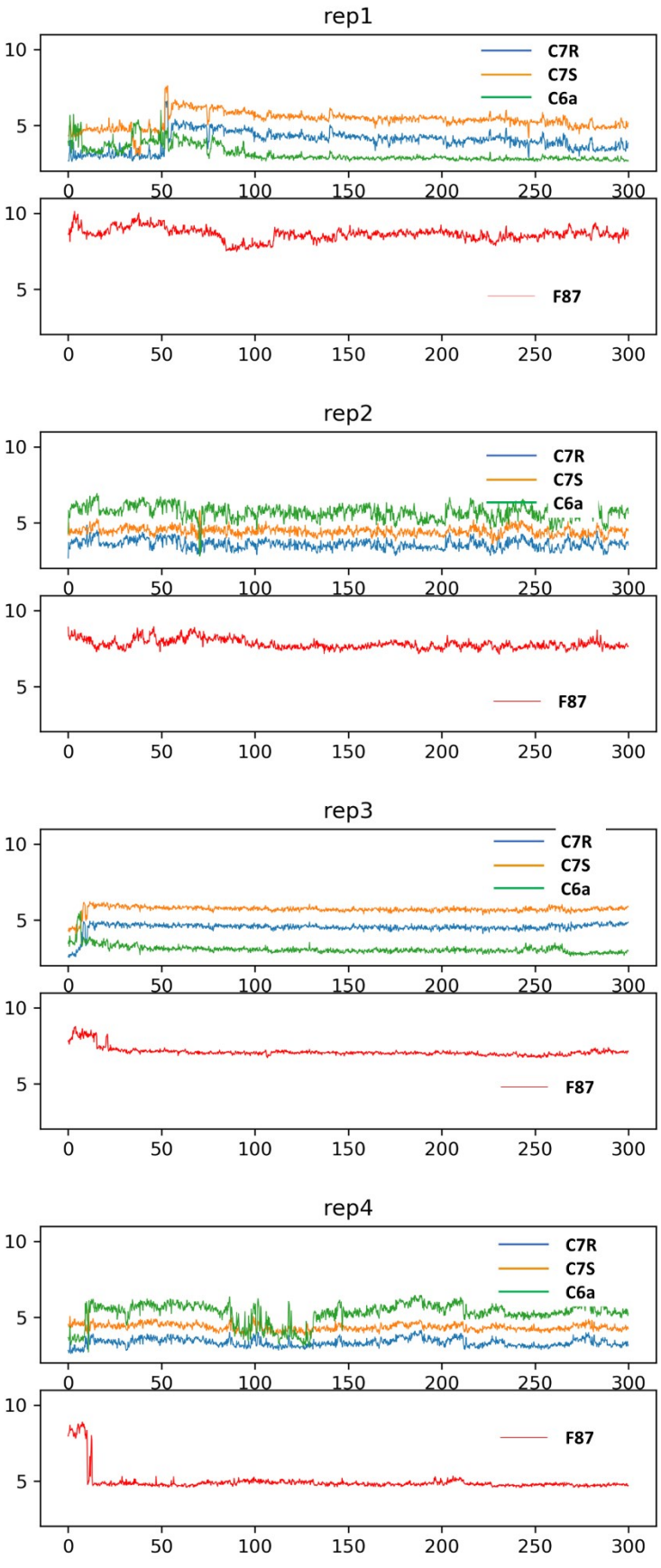


Figure S11. RMSD of Cα in MD simulations of II-H10 starting from Phe87-out conformation. Eight replicas of MD simulations were run for II-H10, the $P450_{BM3}$ variant bound with *pro*-C7(R) artemisinin. Phe87 points away from the catalytic site at the beginning of the simulations.


rep1
C7R
C7S
C6a
F87
rep2
C7R
C7S
C6a
F87
rep3
C7R
C7S
C6a
F87
rep4
C7R
C7S
C6a
F87

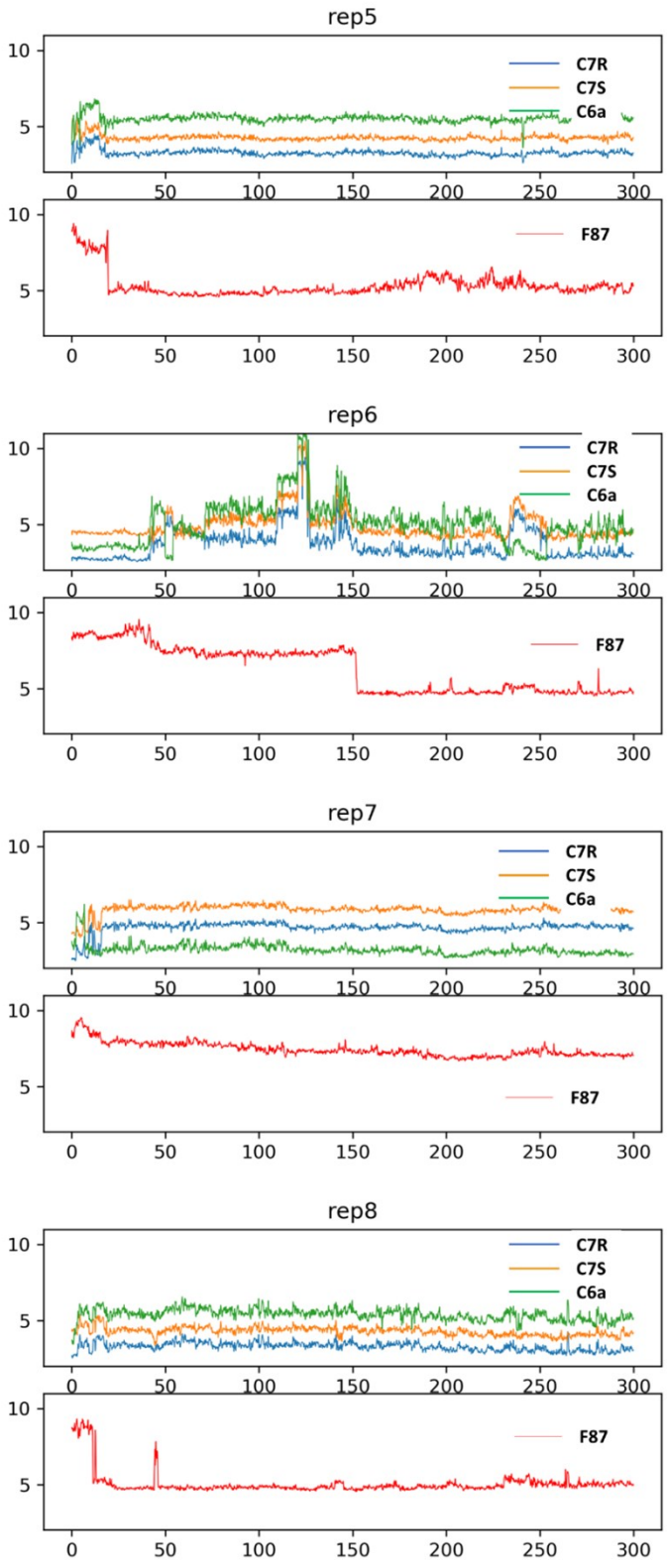


Figure S12. Distance between Fe=O oxo and different H in MD simulations of II-H10 starting from Phe87-out conformation. Eight replicas of MD simulations were run for II-H10, the $P450_{BM3}$ variant bound with *pro*-C7(R) artemisinin. Phe87 points outward the catalytic site at the beginning of the simulations. The distance between the mass centre of Phe87 benzene and the oxygen atom of Fe=O is shown in red plot.

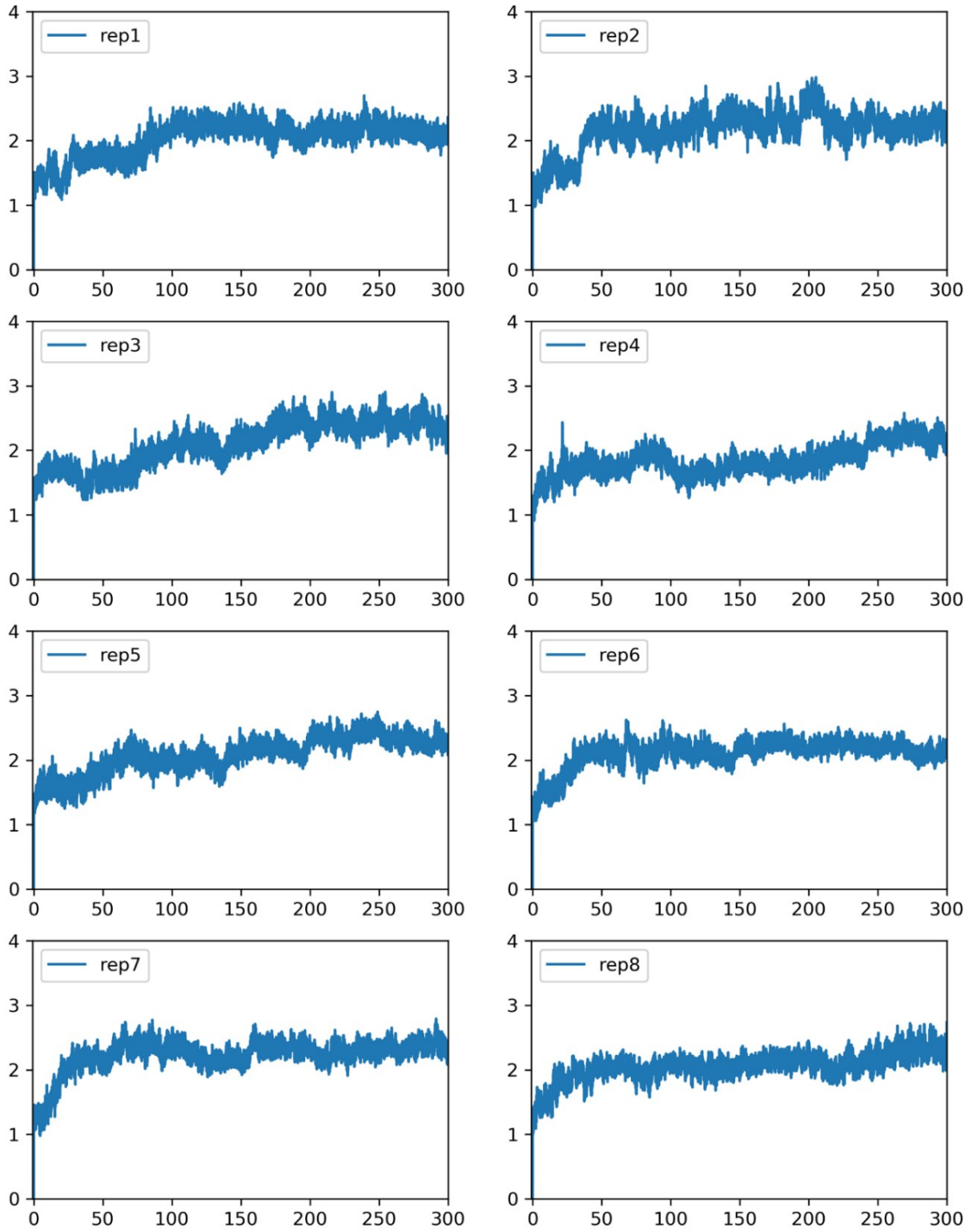
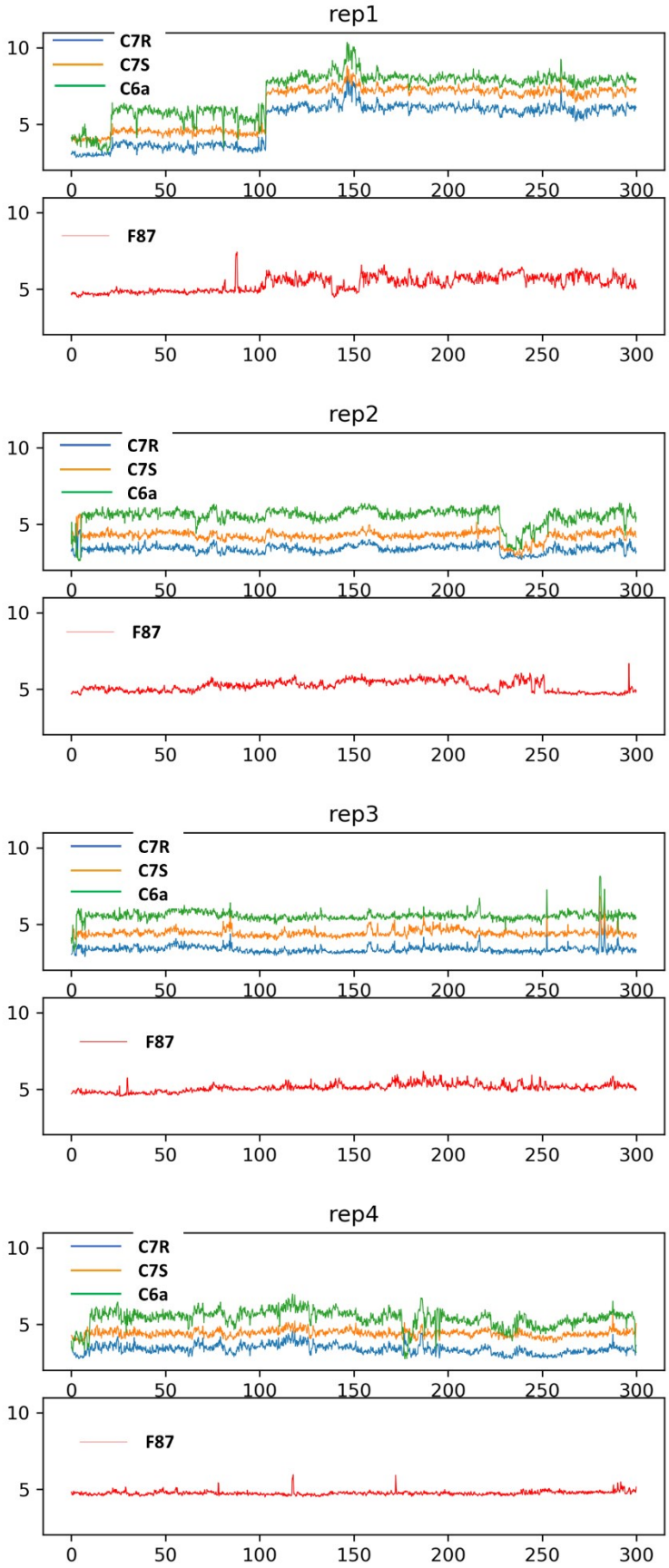


Figure S13. RMSD of Cα in MD simulations of II-H10 starting from Phe87-in conformation. Eight replicas of MD simulations were run for II-H10, the $P450_{BM3}$ variant bound with *pro*-C7(R) artemisinin. Phe87 points inward the catalytic site at the beginning of the simulations.


rep1
C7R
C7S
C6a
F87
rep2
C7R
C7S
C6a
F87
rep3
C7R
C7S
C6a
F87
rep4
C7R
C7S
C6a
F87
10
5
0
50
100
150
200
250
300

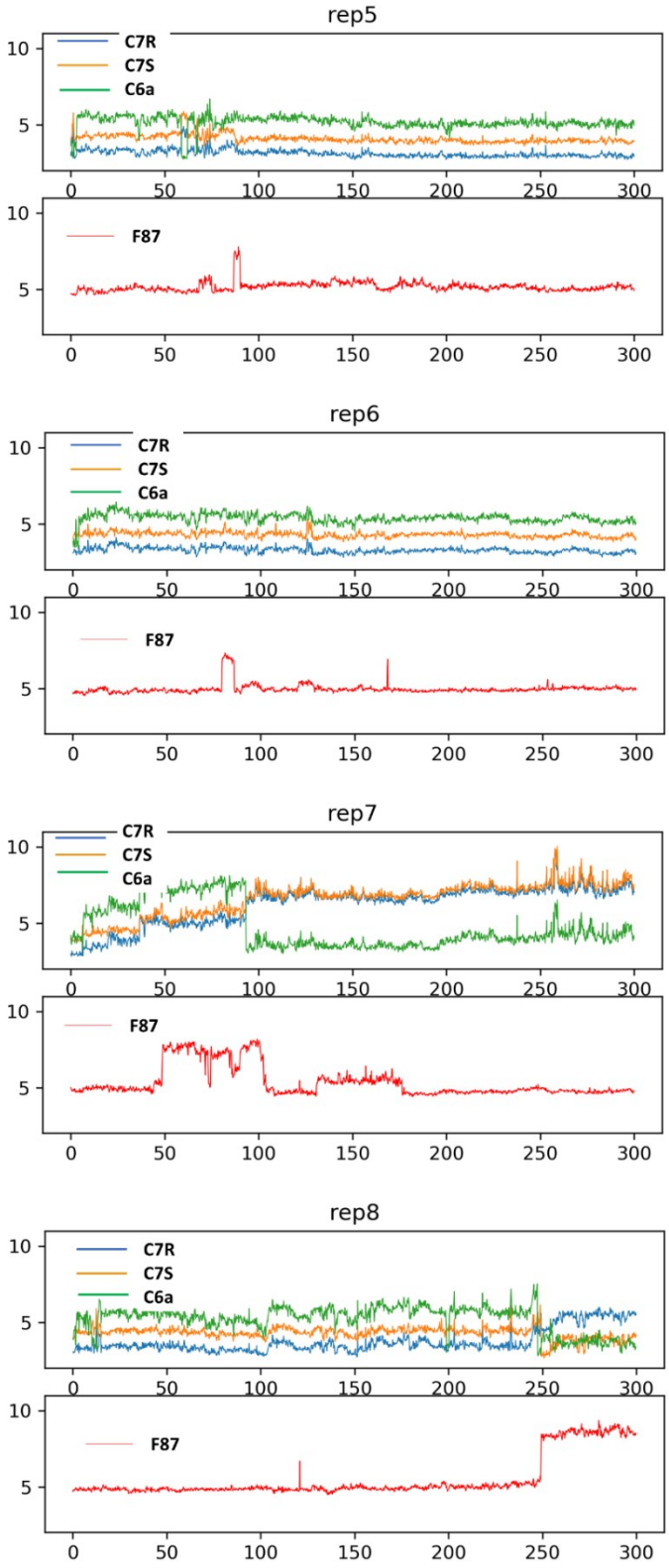


Figure S14. Distance between Fe=O oxo and different H in MD simulations of II-H10 starting from Phe87-in conformation. Eight replicas of MD simulations were run for II-H10, the $P450_{BM3}$ variant bound with *pro*-C7(R) artemisinin. Phe87 points inward the catalytic site at the beginning of the simulations. The distance between the mass centre of Phe87 benzene and the oxygen atom of Fe=O is shown in red plot.

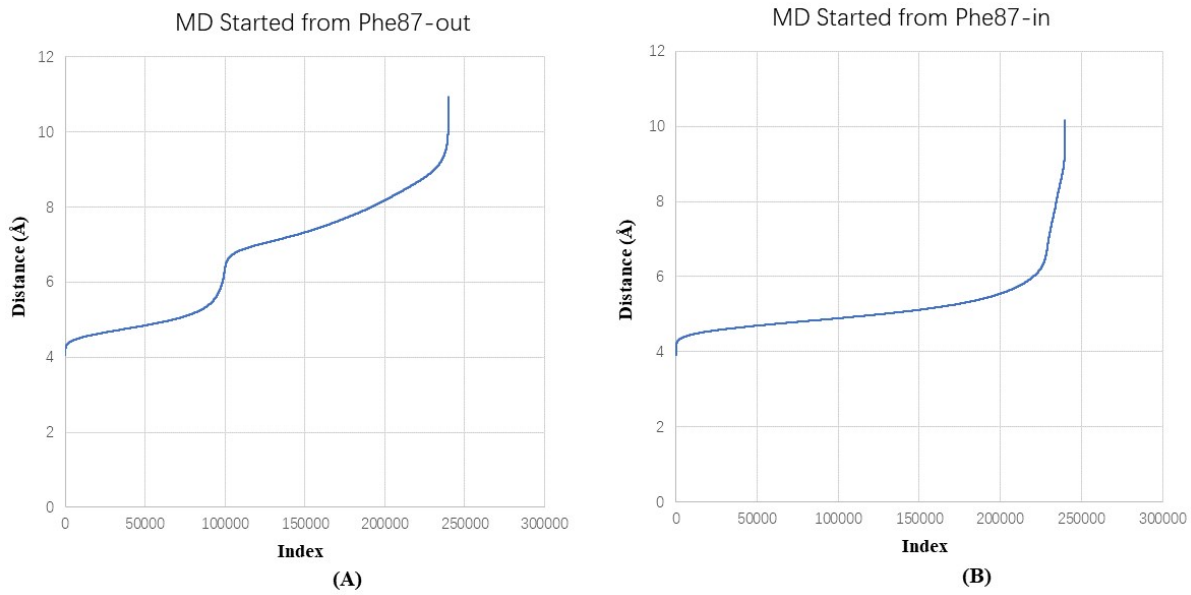


Figure S15. Distance from Fe oxo to the centre of mass of Phe87 benzene. The distance was sorted. The X axis is the index after sorted and Y axis is the distance. The cut-off number of 6.28 Å that distinguishes the Phe87-in and Phe87-out conformations was obtained based on the first derivative of distance-index. The number of Phe87-in/ Phe87-out was 99590/140410 when the MD was started from Phe87-out pose (A). The number of Phe87-in/ Phe87-out snapshot was 225807/14193 when the MD was started from Phe87-in pose (B). These two visiting frequencies correspond to an energy difference of -0.2 kcal/mol and 1.65 kcal/mol. In average, the energy difference between Phe87-out and Phe87-in ($\Delta G_{\text{Phe87-out- Phe87-in}}$) is 0.73 ± 0.92 kcal/mol.

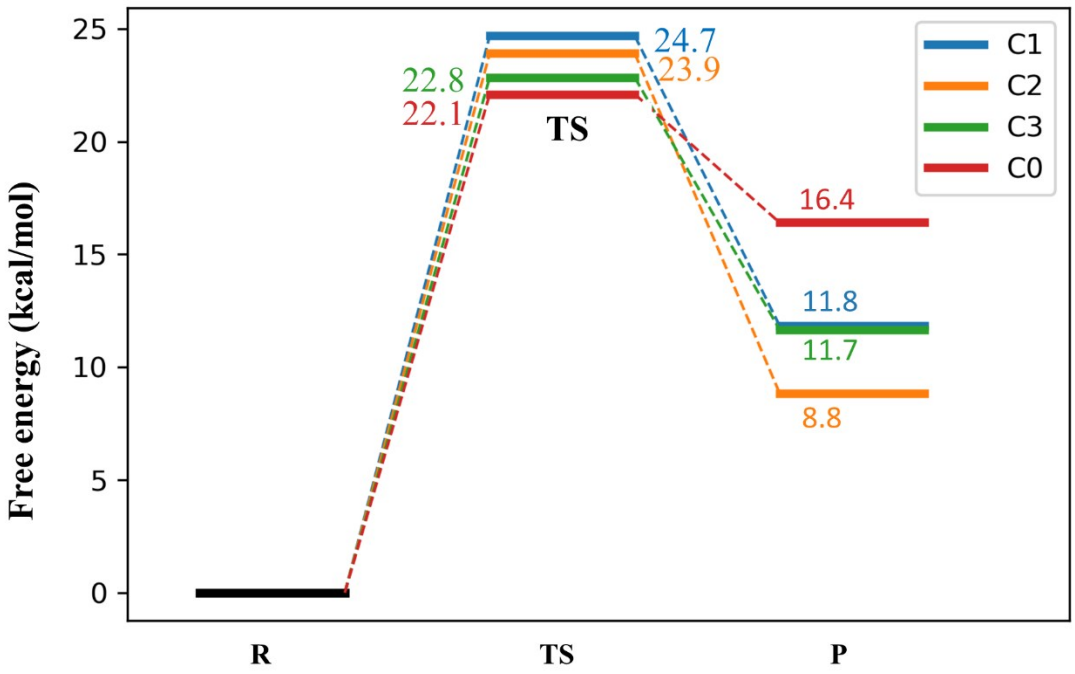


Figure S16. Free Energy profiles for the H abstraction in XE-12. Four representative cluster structures were selected. The reaction profiles corresponding to cluster 0, 1, 2, 3 are shown in red, blue, orange and green, respectively. Geometry optimization and frequency calculations were performed using UB3LYP D3BJ with def2-SVP basis set. The electronic energy was corrected by single point energy calculations with UB3LYP/def2-TZVPP +ZPE.

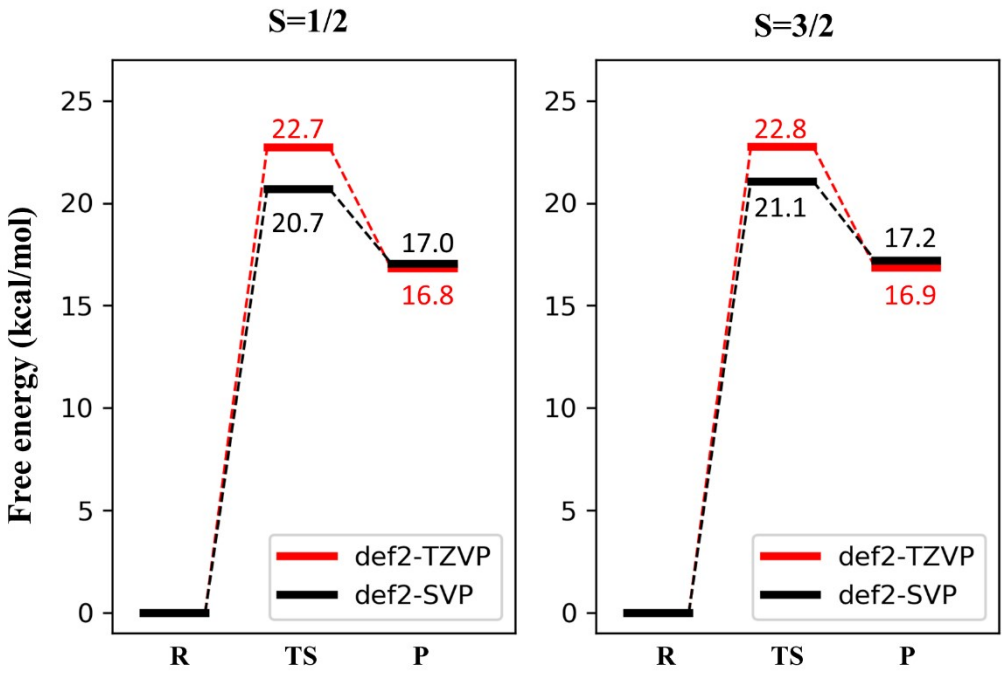


Figure S17 Benchmark on basis set and multiplicity for relative energy (E + ZPE) of XE-12. Geometry optimizations were conducted using B3LYP D3BJ with def2-SVP and def2-TZVP basis sets. S=1/2 and S=2/3 multiplicities were calculated at same level.

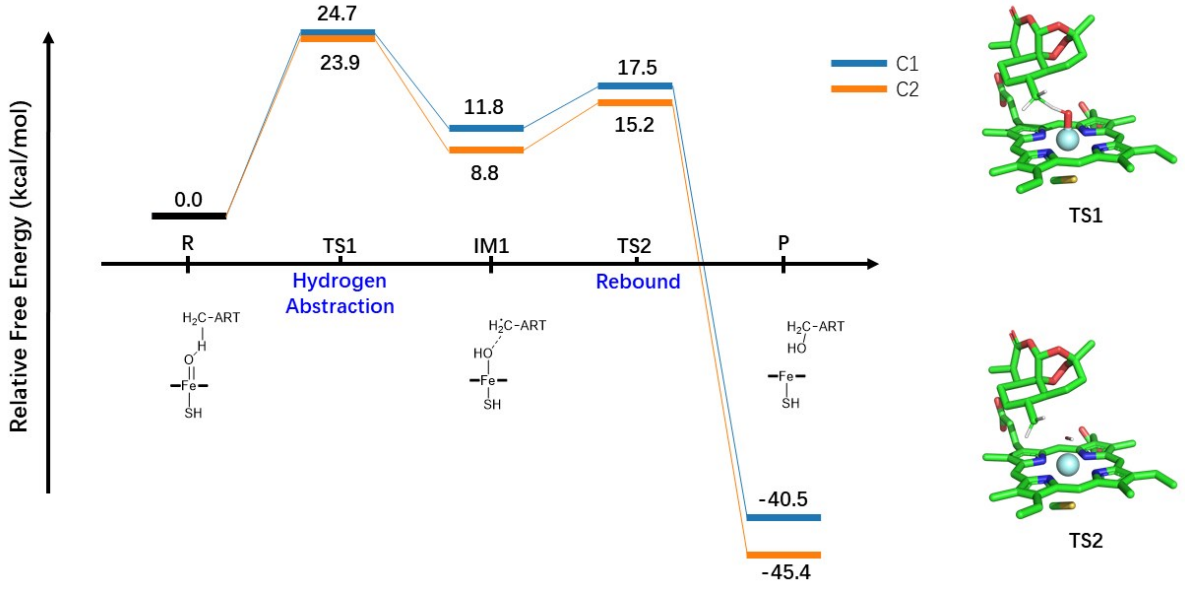


Figure S18 Relative free energy profile of Hydrogen abstraction and Rebound in XE12. The QM/MM reaction profile started from two representative MD cluster structures C1 and C2, which are shown in blue and orange plots, respectively. The optimized structures of the transition states for the two potential energy scans are shown in stick mode.

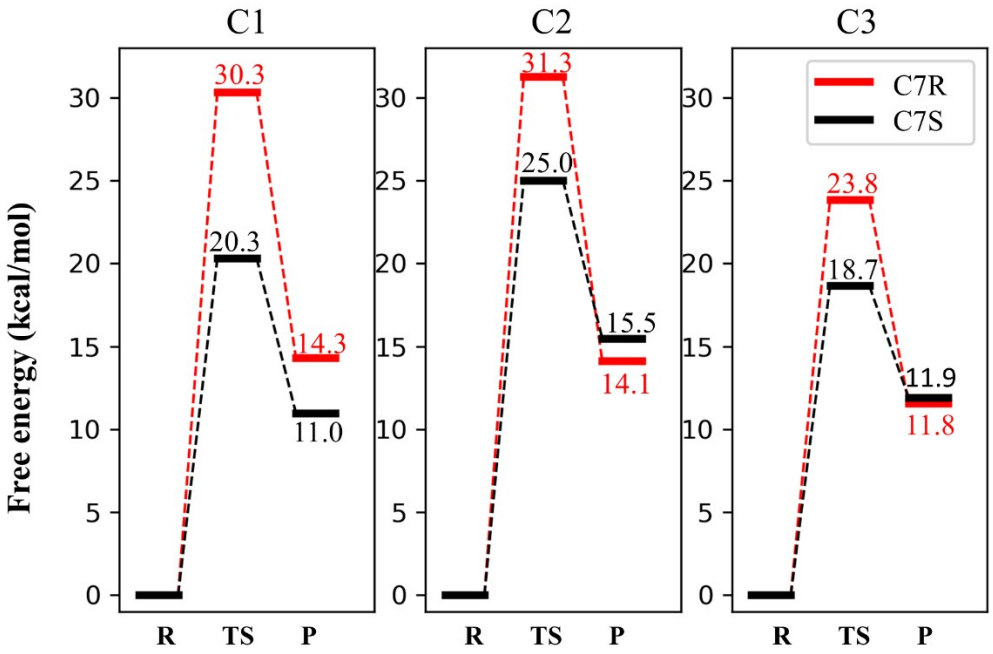
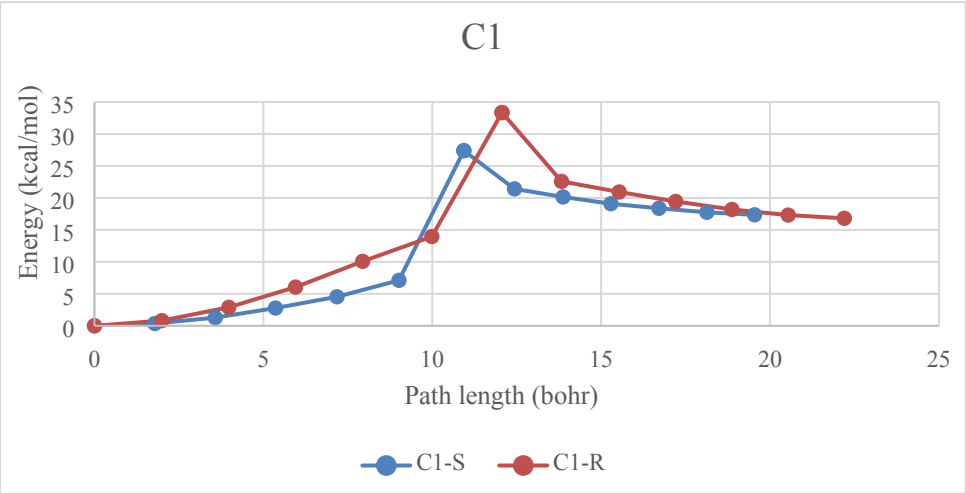
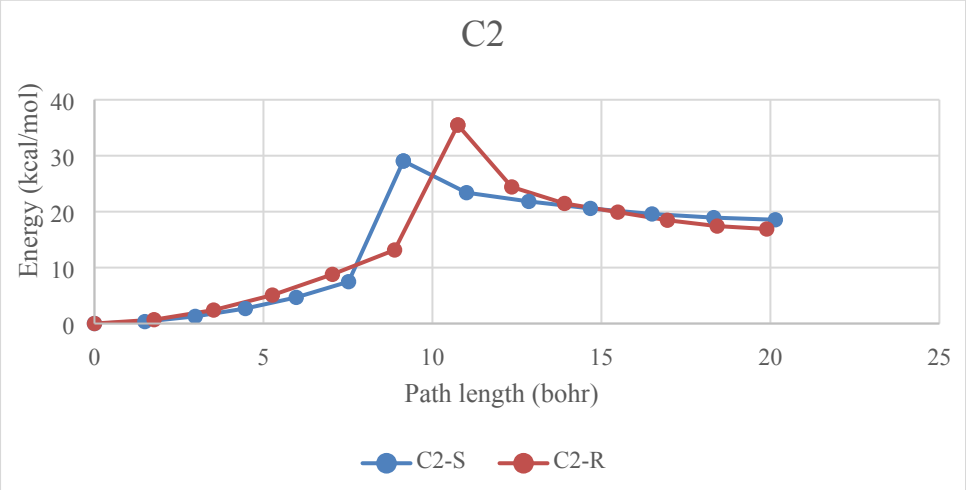


Figure S19 Energy profiles for hydrogen abstraction of arteminsinin by Fe=O in IV-H4. Abstraction of the two pro-7(R) and pro-C7(S) hydrogen atoms of arteminsinin was calculated. Three representative cluster structures (Cluster 1, 2, 3) were selected as the starting structures of the reactant.





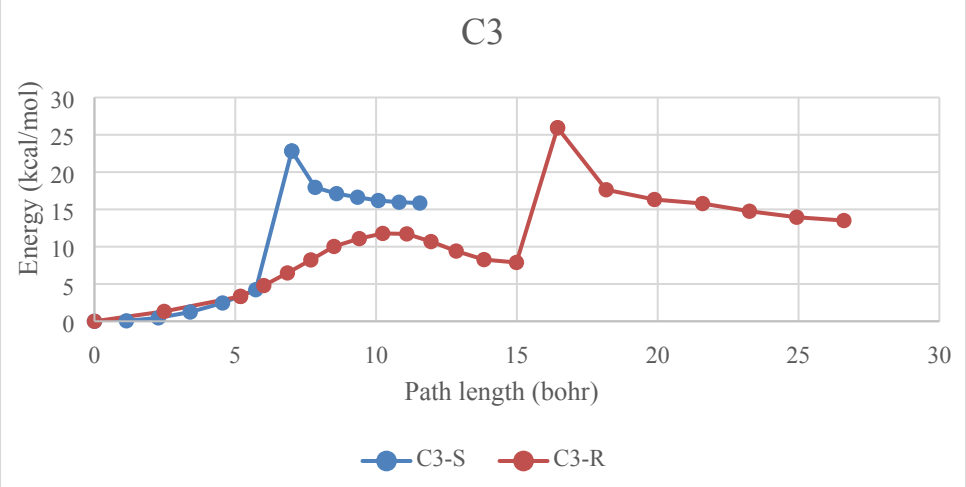
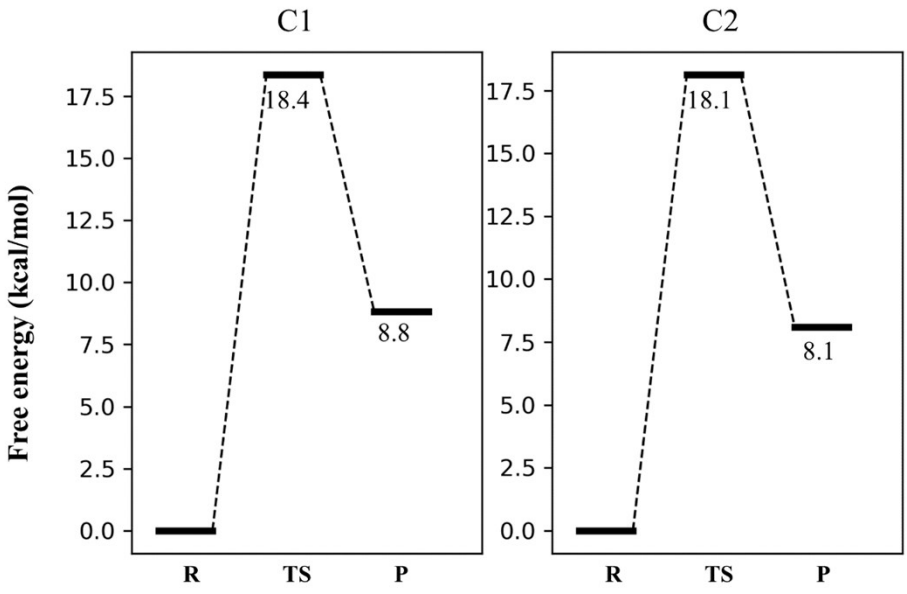
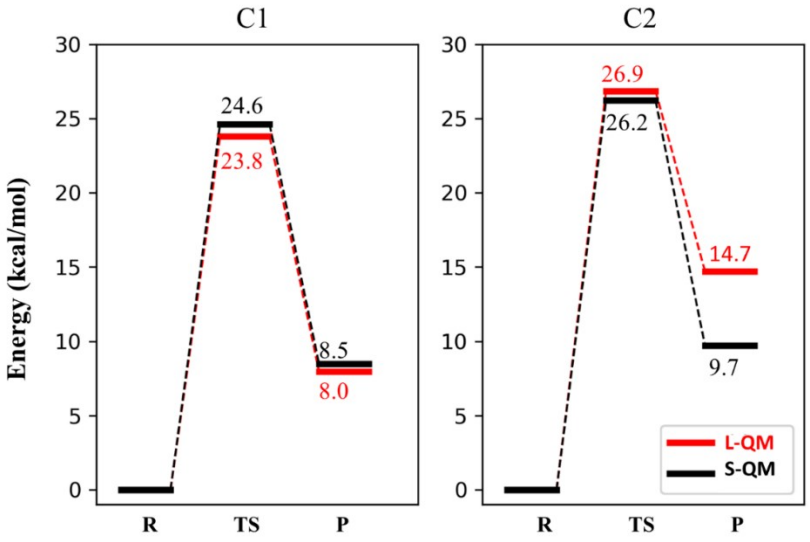


Figure S20. NEB Energy path in 3 clusters in IV-H4. Y axis is Energy in kcal/mol. X axis is path length in bohr. The peak in C3-R at 10 bohr corresponds to chair to boat flip on the six-membered ring.



(A)



(B)

Figure S21. Free Energy profiles for hydrogen abstraction of pro-C7(R) of arteminsinin by Fe(IV)=O in II-H10. (A)Two representative cluster structures from MD simulations (Cluster 1, 2) were selected as the starting structures of the reactant. **(B)** Free Energy profiles for hydrogen abstraction of arteminsinin by II-H10, starting from Phe87-in conformation. Two representative cluster structures from MD simulations (Cluster 1, 2) were selected as the starting structures of the reactant. L-QM denotes the Phe87 was included in the QM region and S-QM denotes Phe87 was excluded from the QM region.

Reference:

1. Zhao, B.; Lei, L.; Kagawa, N.; Sundaramoorthy, M.; Banerjee, S.; Nagy, L. D.; Guengerich, F. P.; Waterman, M. R., Three-dimensional Structure of Steroid 21-Hydroxylase (Cytochrome P450 21A2) with Two Substrates Reveals Locations of Disease-associated Variants. *J. Biol. Chem.* **2012,** *287* (13), 10613-10622.