



**QUEEN'S
UNIVERSITY
BELFAST**

Regio- and Stereoselectivity of CYP450BM3-Catalyzed Hydroxylation of Complex Terpenoids: A QM/MM Study

Hui, C., Singh, W., Quinn, D., Li, C., Moody, T., & Huang, M. (2020). Regio- and Stereoselectivity of CYP450BM3-Catalyzed Hydroxylation of Complex Terpenoids: A QM/MM Study. *Physical Chemistry Chemical Physics*, 22(38), 21696-21706. <https://doi.org/10.1039/D0CP03083J>

Published in:

Physical Chemistry Chemical Physics

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright 2020 Royal Society of Chemistry. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

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

Chenggong Hui^{1,‡}, Warispreet Singh^{1, 2,‡,§}, Derek Quinn², 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

Corresponding author: Meilan Huang

m.huang@qub.ac.uk

[‡] 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*

ABSTRACT

The site-selective C-H oxidation of terpenoids by P450 attracts great attention because of their wide range of biological activities. However, the binding and catalytic mechanism of P450 for hydroxylation of complex terpenoid substrates remains elusive, which has limited the rational engineering of P450 as biocatalysts for terpenoid biosynthesis. Here, we studied the origin of the selectivity and reactivity of P450_{BM3} in the hydroxylation of terpenoids by combining molecular dynamics simulations and QM/MM calculations, using artemisinin as model compound. We found that the conformational change of the β 1 sheet at the substrate entrance and the displacement of the β' helix was critical for reshaping the binding pocket to modulate substrate entrance and positioning the C-H to be activated toward the oxidative species of P450 for the subsequent hydrogen abstraction, the rate-determining step of the hydroxylation. There is a distinct linear correlation between the activation barriers and the reaction coordinates, indicating the reaction coordinates can be used as a facile descriptor for predicting the reactivity of P450_{BM3}. These findings would provide a valuable guidance for predicting the selectivity and reactivity of P450_{BM3} for selective hydroxylation of non-native terpenoid substrates so as to prioritize the rationally designed enzymes for terpenoid biosynthesis.

1. INTRODUCTION

The cytochrome P450 is a superfamily of heme-containing mono-oxygenases that can catalyze a broad range of chemical reactivity such as hydroxylation, dealkylation, sulfoxidation and even unusual cyclization.^{1, 2} It can catalyze site-selective oxidation of C-H bond in terpenoids that are generally not accessible by traditional metal catalysis with high regio- and stereoselectivity, due to the different electronic and steric properties of the C-H bonds in the molecules.³⁻⁷ Of particular importance is the *Bacillus megaterium* P450_{BM3} (CYP102A1), a catalytically self-sufficient fatty acid monooxygenase that exhibits the highest reported mono-oxygenase activity among the P450 family enzymes.

Terpenoids are the most populated natural compounds and their biosynthesis has attracted great interest due to their potential biological activities and therapeutic applications.^{8, 9} For example, artemisinin, a sesquiterpene lactone compound, is among the most effective drugs used for the treatment of malaria. Several P450 variants from *Bacillus megaterium* obtained through directed evolution showed complete regio- and enantioselectivity in the oxidization of artemisinin at C6a and C7 positions located at the upper hemisphere of artemisinin molecule.¹⁰ Despite the outstanding performance of the P450_{BM3} variants, there is little study on their action mechanisms due to lack of the crystal structure of P450_{BM3} in complex with a bulky cyclic substrate.

Molecular docking, Molecular dynamics (MD) simulations and quantum mechanics/molecular mechanics (QM/MM) provide valuable insight into ligand binding and catalytic reactions in the vicinity of active site of enzymes, which can be utilized in structure-based rational protein engineering studies. P450 and its variants have been studied extensively using density functional theory to disclose the information on the catalytic mechanism particularly on the very short-lived reaction intermediates and transition states involved in the catalytic cycle.¹¹

e.g. Dubey et al studied the origin of the regio- and stereoselectivity in P450_{BM3} enzyme for the hydroxylation of fatty acid substrate N-palmitoylglycine, which mimics the native substrate, by applying the aforementioned computational methods (Figure 1A).¹² Ramanan et al successfully implemented MD and QM/MM calculations to understand the origin of fatty acid oxidation at the α -position with S-enantioselectivity in P450_{SP α} (Figure 1A).¹³ However, to our knowledge, the hydroxylation mechanism of P450 toward a non-native cyclic substrate has not been reported yet. It should be noted that the previously reported selective P450_{BM3} variants for terpenoids were usually derived from a template that already contains a number of mutations. In order to rationally and efficiently engineering the enzyme for selective hydroxylation of bulky natural products such as terpenoids, it is necessary to answer the following questions: (i) What is the critical region in the enzyme that would dictate its selectivity in the hydroxylation of non-native cyclic terpenoids? (ii) Is there a quantitative property that can be easily acquired for predicting the reactivity of the enzymes? Here we studied the hydroxylation of non-native terpenoid substrates using artemisinin as model compound, by performing molecular docking, molecular dynamics simulations and QM/MM calculations. We found that the substrate entrance and binding are dependent on the conformational change of the β 1 sheet at the substrate entrance and the displacement of the β' helix which reshaping the binding pocket to accommodate the cyclic substrate. There is a clear linear correlation between the free energy barriers required to activate the C-H bonds and reaction coordinates. The information obtained from the study would shed light for the design of selective biocatalysts for directed oxidative functionalization of C-H bonds in non-native terpenoid substrates.

2. METHODS

Protein preparation

The initial structure of the cytochrome P450_{BM3} was taken from the crystal structure of *Bacillus megaterium* P450_{BM3} (PDB: 1BVY).¹⁴ The ethylene glycol molecule in the crystal structure was removed and replaced with an oxygen atom. The resulting high-valent-iron (IV)-oxo porphyrin- π -cation-radical active species, commonly known as the compound I (Cpd I), was used in the subsequent docking, MD and QM/MM study. The oxidation reaction of the substrate is initiated from the hexacoordinated Cpd I. It is the main oxidant in the P450 enzyme and performs the rate determining step of the reaction by abstracting hydrogen atom from a C-H bond. The force field parameters of the Cpd I were taken from a recent study.¹⁵ The missing hydrogen atoms were added using the tleap module implemented in Amber 16.¹⁶ The parameters for Artemisinin were prepared using the general Amber force field (GAFF) implemented in Antechamber. The atomic charges of the ligand were calculated using restrained electrostatic potential (RESP) method at HF/6-31G* level of theory using Gaussian16.¹⁷

Molecular docking

The ligand was docked to the MD simulated structures of the P450_{BM3} enzyme or its variants in the close vicinity of Cpd I. Molecular docking was performed using the AutoDock 4.2 suite with the Lamarckian genetic algorithm (LGA) and the standard free energy scoring function. The partial charges for the Cpd I in the docking study were taken from a previous study.¹⁵ A total of 100 LGA runs were carried out for each ligand: protein complex. The population was 300, the maximum number of generations was 27,000 and the maximum number of energy evaluations was 2,500,000. Flexible docking was run for II-H10, where Phe87 was set as the flexible residue.

MD simulations

The P450_{BM3} variants were solvated into a truncated octahedral box with TIP3P¹⁸ water molecule such that no protein atom was within 10 Å to any edge of the box. The periodic boundary conditions were employed in all the simulations. The counter ions were added to the protein surface to neutralize the total charge of the system. The Amber FF14SB force field was employed. The systems were subjected to two successive steps of energy minimization first using steepest descent (2,500 steps) and conjugate gradient algorithms (2,500 steps) with protein and ligand restrained with constant of 50 kcal mol⁻¹ Å⁻² followed by steepest descent (2,500 steps) and conjugate gradient algorithms (2,500 steps) without restraint. The systems were gradually heated from 0 to 300K for 250 ps under the NVT ensemble by restraining the solute molecules using harmonic potential of 50 kcal mol⁻¹ Å⁻² using Langevin thermostat¹⁹ with a collision frequency of 1 ps⁻¹. The systems were then subjected to 50ps of equilibration at 300K in the NPT ensemble with a restraint of 5.0 kcal mol⁻¹ Å⁻². An additional 1 ns NPT with a restraint constant of 5.0 kcal mol⁻¹ Å⁻² was performed for the IV-H4 and II-H10 mutants prior to the final production MD. The Barendsen barostat²⁰ was used to maintain the pressure at 1 bar. A production MD run was performed using the GPU version of PMEMD with Amber 16 in a NPT ensemble with a target pressure of 1 bar and a pressure coupling constant of 2 ps. Eight replicas of MD runs were conducted for each complex, making a total simulation time of 0.8-4.8 μs. The SHAKE algorithm²¹ was used to constrain the bonds of all hydrogen atoms. Long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) method²² with a direct space and vdW cut-off of 10 Å. The trajectories were analysed using pytraj²³ and VMD²⁴.

QM/MM calculations

The reaction profiles of P450_{BM3} variants in the catalysis of artemisinin were studied by using QM/MM calculations implemented in Chemshell.²⁵ The QM calculations were performed using ORCA 4.2.0²⁶ and the MM part was defined using DL_POLY.²⁷ The effect of the protein

environment on the polarization of the QM wavefunction was described by the electronic embedding scheme.²⁸ The snapshots for the QM/MM calculations were obtained from the equilibrated MD trajectory using cluster analysis. The representative snapshots from cluster analysis were then subsequently subjected to the energy minimization by using steepest descend (1,250) and conjugate gradient (1,250) algorithms using Amber16.¹⁶ The water shell within 4 Å of the protein or within 20 Å of the QM atoms were retained. The QM region consists of the whole Cpd I molecule, Artemisinin and the C400 residue truncated at C β positions. For II-H10, in addition to the aforementioned atoms, a benchmark study was conducted with and without Phe87 in the QM region in view of the flexibility of Phe87 observed from the MD simulations. The overall charge of QM region was -2. The residues which are within 10 Å of Cpd I and artemisinin were allowed to move freely and the rest of the system was frozen during the geometry optimization. The hydrogen link atoms²⁹ were used to saturate the dangling bond at the QM/MM boundary.

The reaction coordinate was defined by the distance between the oxygen atom of the Fe(IV)=O in Cpd I and the hydrogen atom to be abstracted. The transition state (TS) structures were obtained by potential energy scan followed by dimer TS search implemented in DL-FIND³⁰, and were subsequently validated by frequency calculation. Only one imaginary frequency corresponding to the hydrogen transfer was showed. All QM calculations were performed with DFT using UB3LYP^{31, 32} with D3 dispersion correction and BJ damping,³³ which is widely used for iron containing heme and non-heme system.^{34, 35}

A basis set benchmark study was performed for one of the P450_{BM3} variants, X-E12 using two basis sets def2-SVP (denoted as B1) and def2-TZVP (denoted as B2),³⁶ and two well established spin states for the iron in the Cpd I of P450, doublet spin state (S=1/2) and quartet spin state (S=3/2).

Following the benchmark study, geometry optimizations were carried out for all the other variants using UB3LYP-D3BJ/def2-SVP. The ZPE were calculated with the same level of DFT functional and added to all the stationary points. The final energies were corrected at UB3LYP-D3BJ/def2-TZVPP level. RIJCOSX approximation³⁷ was used in all the QM calculations. In the QM/MM calculation in each mutant, several snapshots were calculated, and the averaged barrier was calculated using Equation 1.

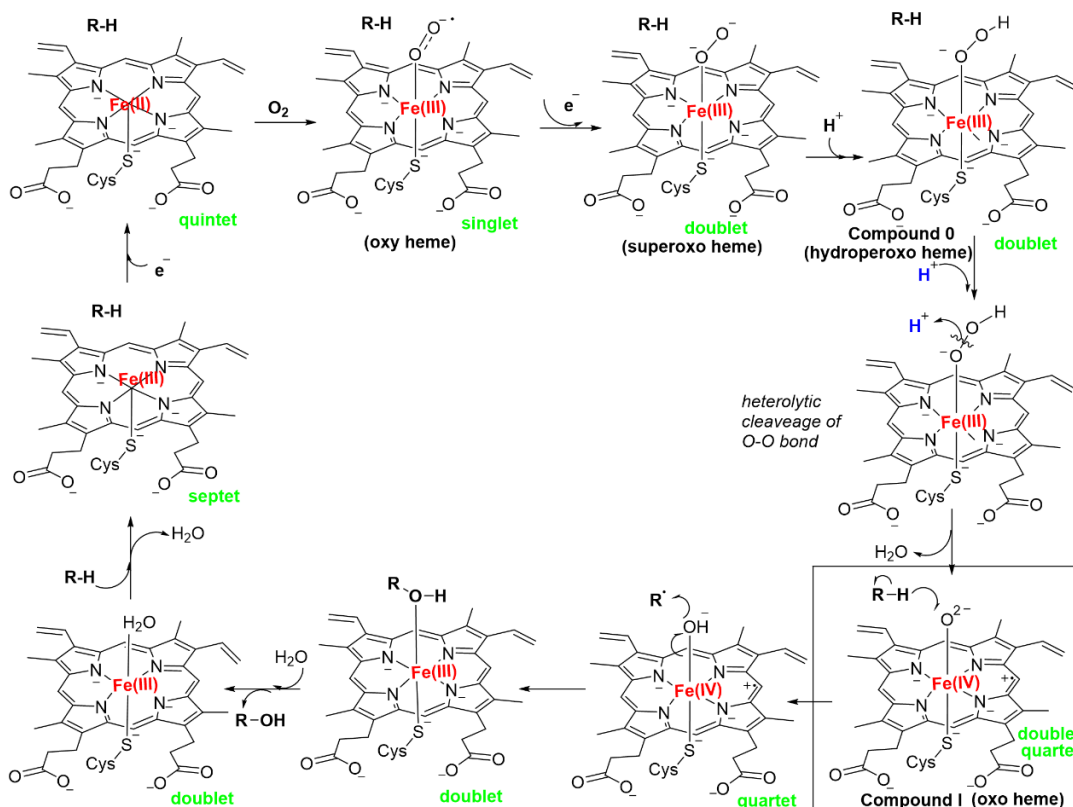
$$\Delta E_{av}^{\ddagger} = -RT \ln \left(\frac{1}{n} \sum_{i=1}^n \exp \left(\frac{-\Delta E_i^{\ddagger}}{RT} \right) \right) \quad (\text{Equation 1})$$

3 RESULTS AND DISCUSSION

3.1 Origin of regio- and stereoselectivity

3.1.1 Binding of Artemisinin in the P450_{BM3} variant template FL#62

In the catalytic cycle by eukaryotic P450, the electron transfer occurs from a FAD/FMN dependent NADPH-cytochrome P450 oxidoreductase (CPR) to P450 monooxygenase. FAD accepts electron from NADPH, while the FMN reduces P450. Generation of the active oxidant Cpd I in P450 requires activation of molecular oxygen by the reducing cofactor NADPH, followed by protonation at the distal oxygen of the superoxo intermediate, consecutive protonation of Compound 0 (Cpd 0), a ferric-hydroperoxo intermediate and leaving of a water molecule (Scheme 1).



Scheme 1. Proposed mechanism for the P450-catalyzed hydroxylation of substrate R-H.

P450_{BM3} (CYP102A1) is a self-sufficient fatty acid monooxygenase containing heme and FAD/FMN domains. In the crystal structure of apo P450_{BM3} from *Bacillus megaterium* (PDB:1BVY),¹⁴ Asp251 adjacent to the loop 240-250 that contacts with the FMN domain formed a salt bridge with Lys224. The highly flexible loop is missing in the crystal structure of P450_{BM3} in complex with a fatty acid substrate N-palmitoylglycine (PDB: 1JPZ).³⁸ The β 1 hairpin at the substrate entrance exhibits an open configuration and Arg47 on the hairpin formed a salt bridge with Glu352 (Supplementary information, Figure S1).

Starting from the F87A variant which showed minimal but detectable activity towards the probe molecules, an active FL#62 template was identified using a fingerprinting method.³⁹ FL#62 converted the module compound artemisinin (Figure 1) to give a mixture of hydroxylated products. Majority of the products was C7 (S) enantiomer, and C7 (R) and C6a oxidized products were also observed. Further evolution of FL#62 yielded three selective

variants, X-E12, IV-H4 and II-H10, which gave 94% selectivity for C6a oxidized product and 100% enantioselectivity for C7(R) and C7(S) products.¹⁰ Two P450 variants that are only different in six amino acid positions showed a complete opposite stereoselectivity at position C7; additional four mutations resulted in unprecedented regioselectivity at position C6a.

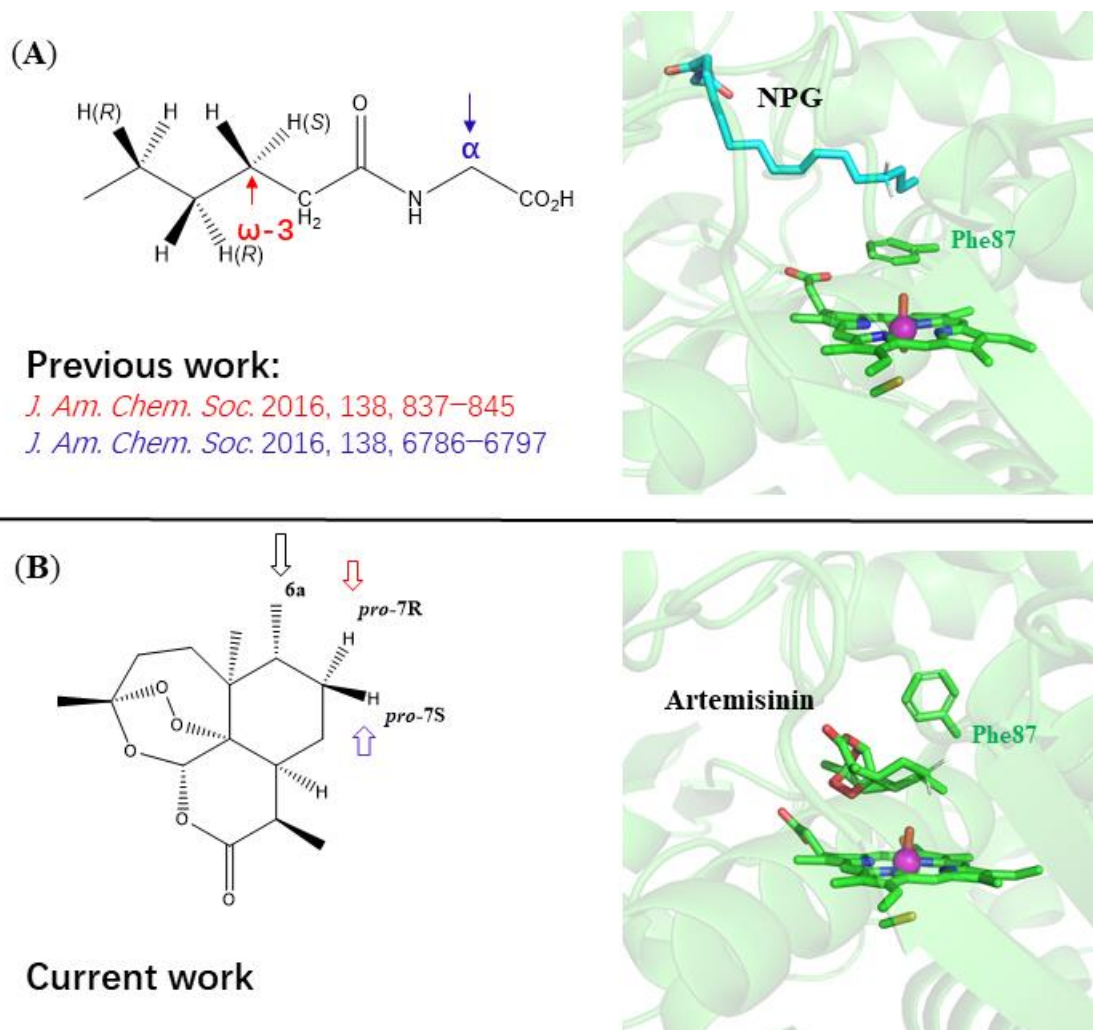
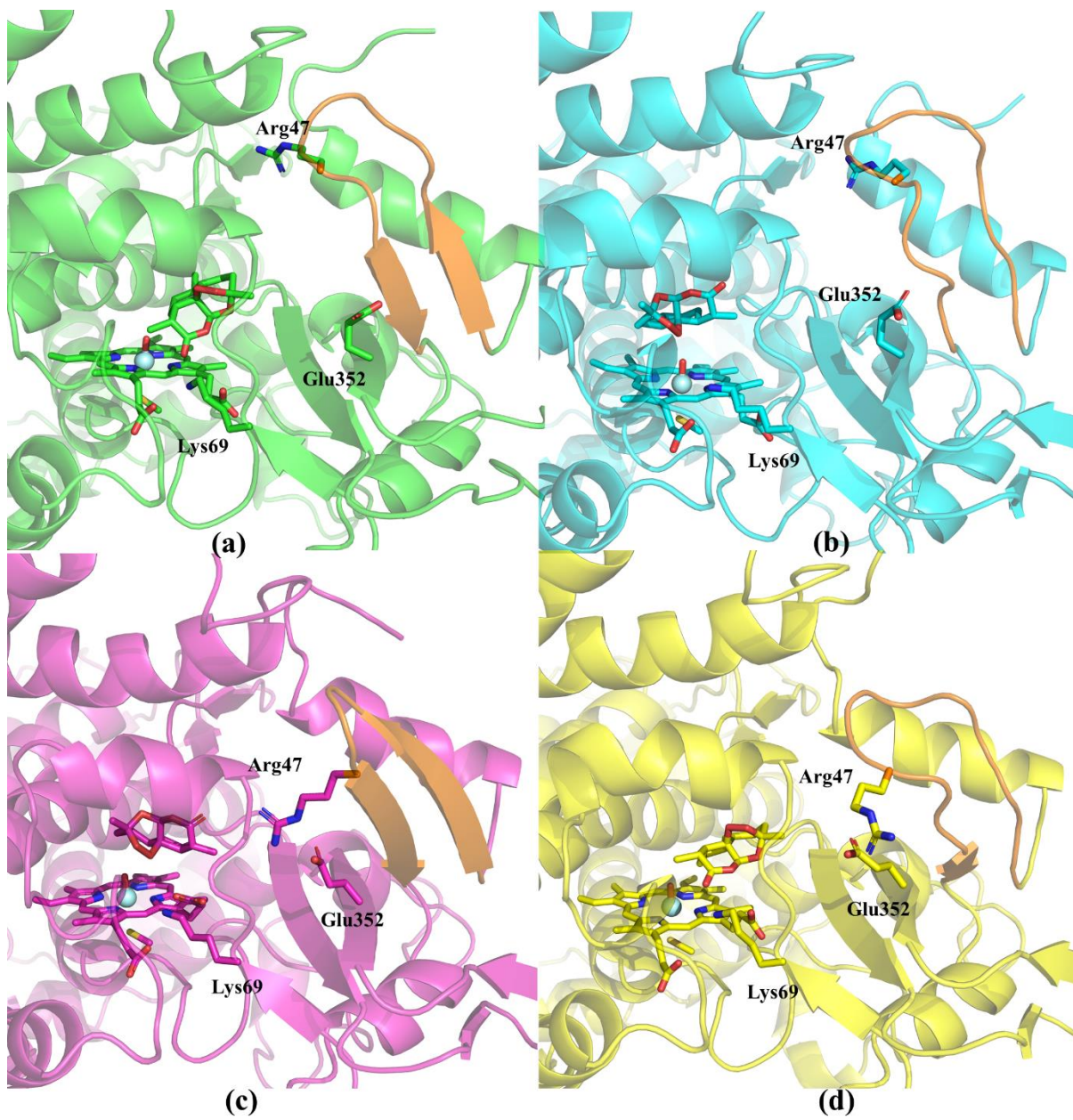


Figure 1. (A) Previous selectivity study on the fatty acid substrate NPG that mimics the native substrate of P450_{BM3}. (B) Current study is focused on the regio- and enantioselectivity of P450_{BM3} for cyclic terpene compounds, using artemisinin as a model molecule.

In order to understand the binding of artemisinin in FL#62, the ligand was first docked into the active site of the variant. The dominant docked poses showed preference to the prochiral CH₂ that gives C7 (S) enantiomer product (*pro*-C7(S)) (16 out of 50 docked poses), whereas

215 few *pro*-C7(R) poses were also observed (2 out of 50 docked poses). In addition, predict
216 binding energy was -5.31 kcal/mol for *pro*-C7(S) and -4.55 kcal/mol for *pro*-C7(R), indicating
217 the *pro*-C7(S) is more stable than *pro*-C7(R). These are in agreement with the experimental
218 results that the C7(S) product is dominant by FL#62.¹⁰ MD simulations of FL#62 in complex
219 with the *pro*-C7(S) or *pro*-C7(R) ligand (Figure S2-5) disclosed that Asp251 adjacent to the
220 FMN contacting loop formed a salt bridge with Lys224 (Figure S6A&B). The hairpin Ala44-
221 Arg47 located at the substrate entrance exhibited a closed conformation in both complexes
222 (Figure 2). In contrast to the crystal structure of apo P450_{BM3} where a salt bridge is observed
223 between Arg47 and Glu352 (Figure S1),¹⁴ such salt bridge was absent in the FL#62-*pro*-
224 C7(S)/(R) complex (Figure 2).



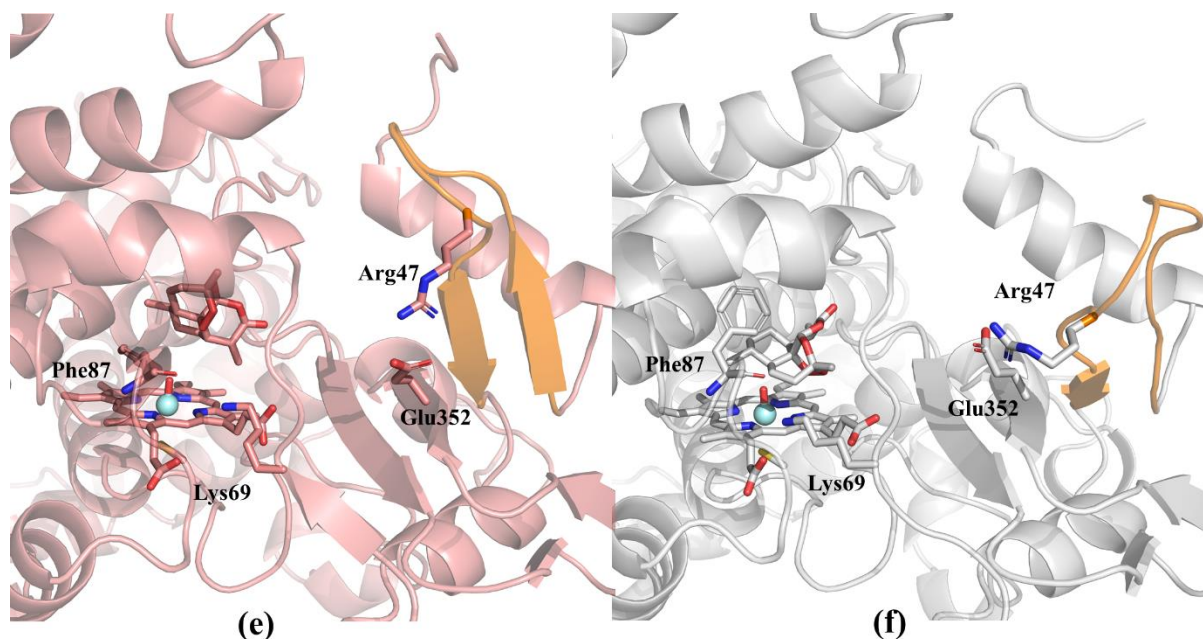


Figure 2. MD simulated structures P450_{BM3} variants in complex with artemisinin. (a) FL#62 with *pro*-C7(S), (b) FL#62 with *pro*-C7(R), (c) X-E12 with preference to the abstraction of the methyl H from C6a, (d) IV-H4 with *pro*-C7(S), (e) II-H10 with *pro*-C7(R) (Phe87 -in), (f) II-H10 with *pro*-C7(R) (Phe87 -out). The hairpin near the substrate entrance is shown in orange.

It should be noted that in FL#62-*pro*-C7(S) complex, both propanoic acid carboxylate interact with Lys69 that is located at loop 69-73 near the substrate entrance. Lys69 was involved in the hydrogen bond with the substrate via C10 carbonyl group of the lactone. Interestingly, in the FL#62-*pro*-C7(R) complex, Lys69 moved away from the substrate such that it only forms ionic bond with one of propanoic acid carboxylates, however, unlike the enzyme in presence of *pro*-C7(S), Lys69 in FL#62-*pro*-C7(R) complex did not form a H-bond with the C10 carbonyl oxygen, which may account for the low conversion toward the C7(R) product compared to the C7(S) product.

3.1. 3 C7(S) selectivity

In the IV-H4 P450_{BM3} variant, three additional mutations A78S, S81V and V82A were introduced in the active site of the FL#62 variant. Selective hydroxylation at C7 with complete S-stereoselectivity was achieved by IV-H4.¹⁰ MD simulations (Figure S9) showed that the *pro*-C7(S) hydrogen was consistently closer to the Fe(IV)=O than *pro*-C7(R) hydrogen during the entire length of MD trajectory (Figure S10), in agreement with the experimental result that 7(S)-hydroxylated artemisinin is the preferred product by the mutant.¹⁰ Lys224 was located on the FMN contacting loop 240-250, and it formed a salt bridge with Asp251 in the crystal structure of apo P450_{BM3}¹⁴. The MD simulated structure of IV-H4 mutant in complex with *pro*-C7(S) artemisinin showed that Lys 224 formed a salt bridge with Asp251 (Figure S6). In addition, a salt bridge was formed between Glu352 and Arg47 located at the hairpin Ala44-Arg47, leading to a closed substrate entrance (Figure 2). Interestingly, this salt bridge was absent in the FL#62-*pro*-C7(S) complex. Lys69 formed a hydrogen bond with the C10 carbonyl oxygen of substrate, locking the lactone of the substrate in the catalytic site such that the prochiral (S) hydrogen at C7 position was pushed toward the vicinity of Fe(IV)=O. The substrate was nested in a hydrophobic pocket composed of Leu75, Ala87, Ala328 and Leu437 (Figure 3). The mutations at positions 78, 81 and 82, may synergistically introduce additional flexibility of the loop adjacent to the β' helix, which caused Leu75 on the helix to form favourable hydrophobic interaction with the substrate, helping to correctly orientate the substrate to achieve the C7-(S) selectivity.

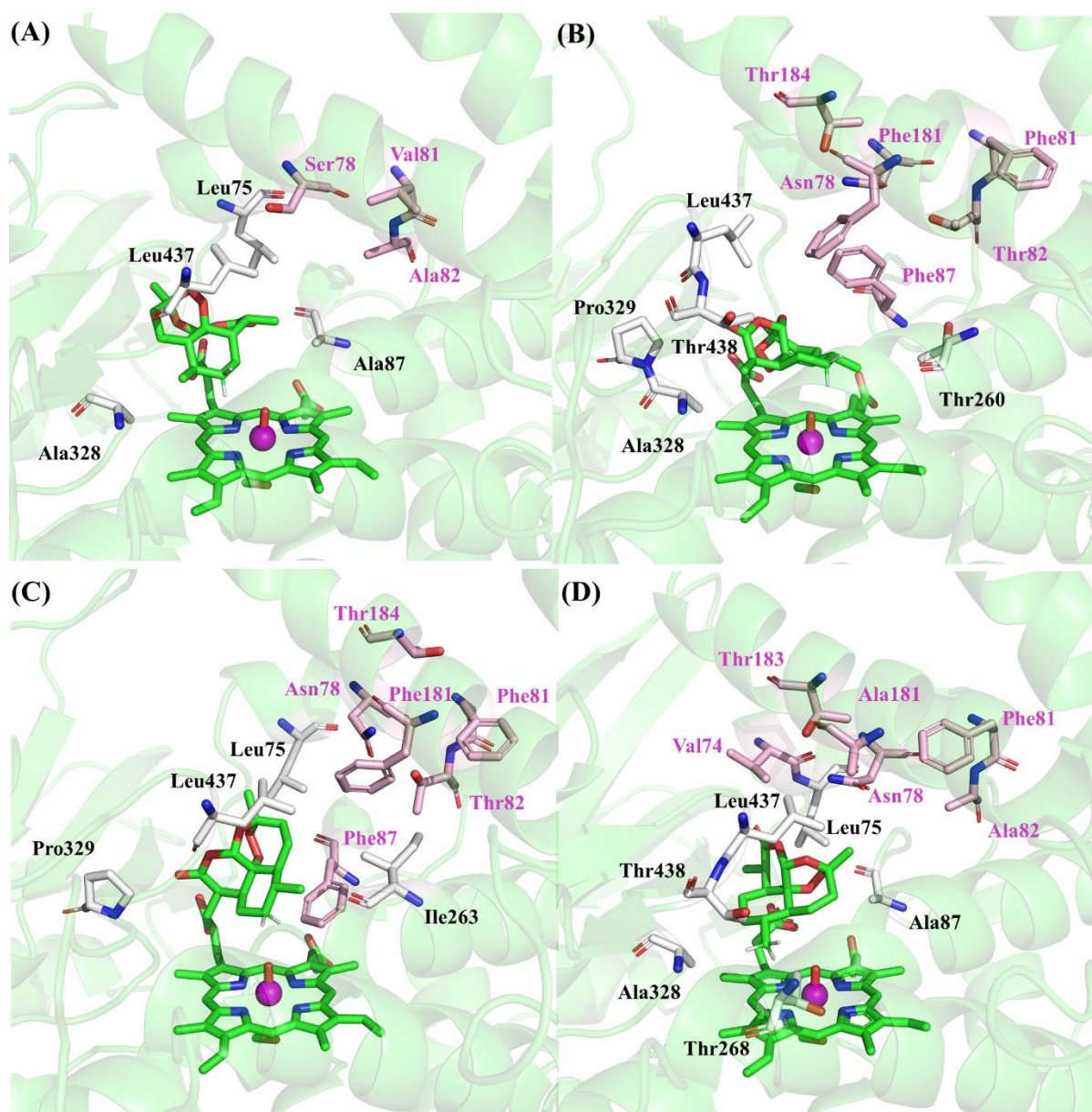


Figure 3. MD simulated structures of P450_{BM3} variants in complex with artemisinin. (A) IV-H4 with *pro*-C7(S), (C) II-H10 with *pro*-C7(R) (Phe87-out). (C) II-H10 with *pro*-C7(R) (Phe87-in). (D) X-E12 with preference to the abstraction of the methyl H from C6a. The mutated residues in relation to the FL#62 template are shown in pink.

3.1.4 C7(R) selectivity

Previous QM/MM studies suggested that Phe87 play an important role in controlling the (R) stereoselectivity of oxidation reactions for fatty acid substrate.¹² In the II-H10 variant, the small

alanine at position 87 was mutated back into phenylalanine. In addition to A87F, five mutations were introduced in the II-H10 variant compared to the FL#62 template, namely, A78N, S81F, V82T, L181F, V184T, which gave hydroxylated product at C7 position with 100 % R selectivity. The MD simulations of the II-H10-*pro*-C7(R) complex (Figure S11) showed that the distance between the Fe=O and *pro*-C7(R) hydrogen were generally shorter than with C7(S) or C6a (Figure S12), in agreement with the experimentally observed exclusive R-enantiomer product.¹⁰

MD simulations of II-H10 in complex with artemisinin disclosed that Phe87 exhibited two representative conformations. One pointed away from the catalytic site (hereafter called Phe87-out), and the other pointed inward the catalytic centre and resides between Cpd I and artemisinin (hereafter called Phe87-in). In both conformations, Glu352 and Arg47 formed an ionic bond. The four mutations A78N, S81F, V82T, A87F located on the flexible loop adjacent to the β' helix may helped to orientate Phe87 appropriately in relation to the substrate and catalytic centre, while the mutation of the second sphere residues L181F and V184T on the F helix further stabilized the substrate binding by favourable H-bond or hydrophobic interactions with the first-sphere residues.

The $\beta 1$ hairpin around the substrate entrance Ala44-Arg47 displayed an open conformation in the Phe87-out conformation, similar to the P450_{BM3} in complex with a native fatty acid substrate analogue.³⁸ Lys69 formed an ionic bond with the distal propionic carboxylate group of porphyrin ring and also formed a H-bond with the backbone carbonyl of Phe87. Artemisinin in the Phe87-out conformation is nested in a relatively compact hydrophobic pocket formed by Phe87, Thr260, Ala328, Pro329, Leu437 and Thr438 (Figure 3B). In contrast, in the Phe87-in conformation, the $\beta 1$ hairpin is closed. Leu437 pointed away from the binding pocket so that the hydrophobic pocket composed of Leu75, Ile263, Pro329 and Leu437 is enlarged (Figure 3C). Closing of the hairpin pushed the β' helix along with F/G helix away

from the catalytic centre, as a result, the substrate moved further away from the heme centre allowing Phe87 to move toward the heme centre and position itself between the heme porphyrin ring and the substrate. It is also worth noting that a salt bridge was observed between Lys224 and Asp251 near the FMN contacting loop in the Phe87-out conformation whereas this interaction is absent in the Phe87-in conformation (Figure 2).

Since both conformations was observed in the extensive MD simulations of II-H10 in complex with *pro*-C7(R) artemisinin, both was considered in the subsequent QM/MM calculations in the study of C7(R) enantioselectivity. It was found that Phe87-out conformation displayed a lower energy barrier than Phe87-in conformation in the hydrogen abstraction from the *pro*-(R) hydrogen by heme Fe(IV)=O (see section 3.2.3).

3.1.5 C6a selectivity

Similar to FL#62, X-E12 contains a F87A mutation, which was suggested to broaden the substrate scope of P450_{BM3}.⁴⁰ In the X-E12 mutant, six additional mutations were introduced compared to FL#62, namely, A74V, A78N, S81F, V82A, L181A and V184T and the selectivity toward C6a was increased significantly from 7 % to 94 %. During the MD simulations the distance between C6a and Fe=O retained largely around 2.5Å (Figure S6). Mutating Ala74 on the β' helix into Val enabled favourable hydrophobic interaction with Leu437. Meanwhile, the replacement of the residues on the β'-helix that was underpinned by the flexibility of the adjacent loop 69-73 reshaped a favourable hydrophobic pocket to accommodate the substrate. The β1 hairpin Ala44-Arg47 at the mouth of the substrate entrance was closed up and a salt bridge was formed between Arg47 and Glu352 (Figure 2C). It should be noted that the salt bridge was also present in IV-H4-*pro*7(S) and II-H10-*pro*-C7(R) complexes, whereas it was absent in the FL#62-*pro*-7(S)/(R) complex. Thus, maintaining the salt bridge between Arg47 and Glu352 caused β' helix to move closer to the substrate and would be important for achieving the selectivity at C6a or C7 of artemisinin. Closing the substrate entrance pushed the

substrate further in the active site with the C6a positioned closer to the Cpd I oxidant. Further, Thr268 on the distal H-helix, which was previously suggested to be important for the O₂ binding and activation for the catalysis of fatty acid substrate,⁴¹ formed a hydrophobic interaction with the C6 methyl group of tricyclic terpene, so as to lock it in the active site facilitating the subsequent oxidative reaction. As a result, artemisinin is nested in a hydrophobic site formed by Ala87, Thr268, Ala328, Leu437, Thr438 as well as Val74 and Leu75 located on at the β' helix (Figure 3D). Hence, altering the loop region 69-73 and the adjacent β'-helix, as well as F-helix (181-190) that is in direct contact with the β'-helix may modulate the substrate entrance and also reshape the catalytic pocket, and thus may further enhance the regioselectivity toward the C6a of artemisinin (Figure 4).

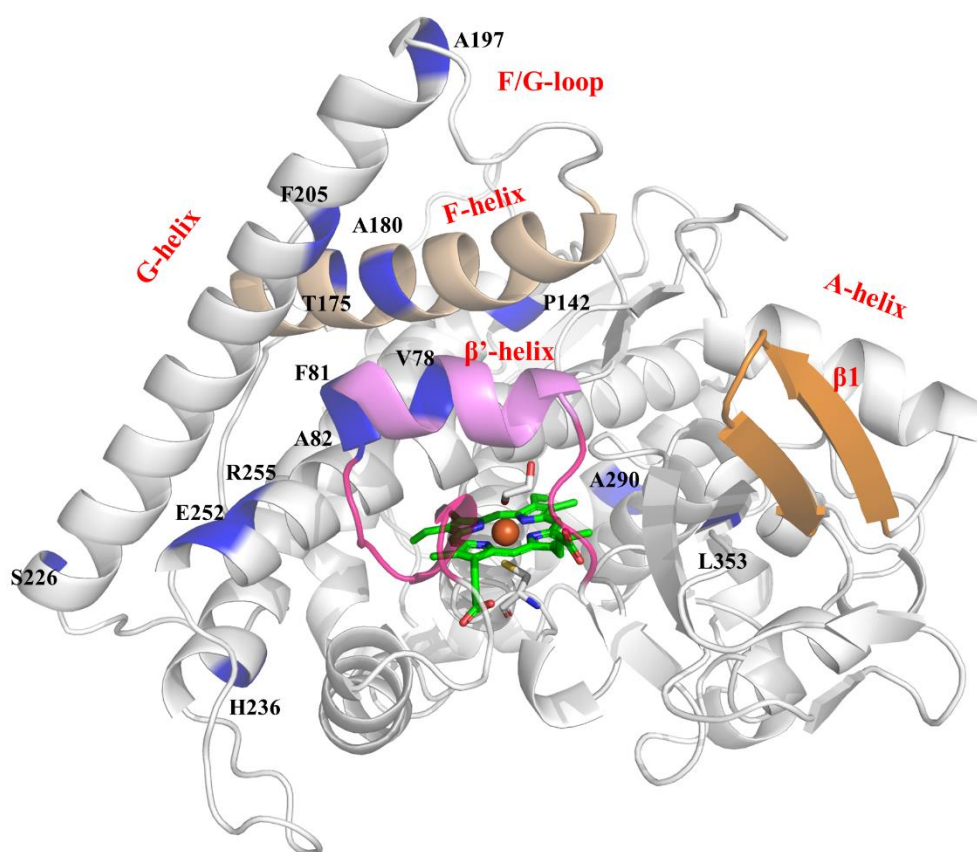


Figure 4. The crystal structure of P450_{BM3} (PDB: 1BVY).¹⁴ The heme group in the active site of P450_{BM3} is shown in stick representation. The positions mutated in the FL#62 template are shown in blue. The proposed hot spots from this research is shown in purple.

3.2 Origin of reactivity by QM/MM calculations

3.2.1 C-H activation at C6a of artemisinin

Four representative structures from MD simulations of X-E12 in complex with the ligand were selected for QM/MM studies in order to investigate the effect of conformational sampling on the barrier of the rate limiting step of the reaction.

Among the four reaction profiles, the activation barrier from cluster 0 showed the lowest barrier of 22.1 kcal/mol (Supplementary information Table S1, Figure S16). An average barrier of 22.7 kcal/mol was obtained according to equation 1 (Figure 5).

A basis set benchmark study was performed for cluster 0, the representative structure of X-E12 variant from MD simulations, which has the lowest barrier for the hydrogen abstraction. In addition to def2-SVP basis set (B1), geometry optimizations were also conducted using def2-TZVP basis set (B2). The difference in the energy barrier using the two basis sets was less than 2 kcal/mol (Figure S16). The error was mainly attributed by the potential energy calculated at the two basis sets, while the ZPE at these basis sets were almost same (Table S2A &B). Thus, energies obtained by geometry optimization using a double Zeta basis set (def2-SVP: B1) with a triple Zeta basis set (def2-TZVP: B2 or def2-TZVPP: B2') single point energy correction would give reaction energies as accurate as those calculated from full optimization using large def2-TZVP basis set (B2//B2).

Doublet ($S=1/2$) and quartet ($S=3/2$) spin states are the two plausible spin states of iron in Cpd I of heme enzymes and exhibit similar reactivity,^{42, 47} so these two spin states were calculated in the benchmark study (Table S2). Our calculations showed that the two spin states were highly degenerated, with a barrier of 22.7 kcal/mol for $S=1/2$ and 22.8 kcal/mol for $S=3/2$ based on the energies of the geometries optimized with def2-TZVP (B2) basis set (Figure S17). Since

the two spin states of iron in Cpd I are highly degenerated all of the remaining calculations were performed with doublet spin state.

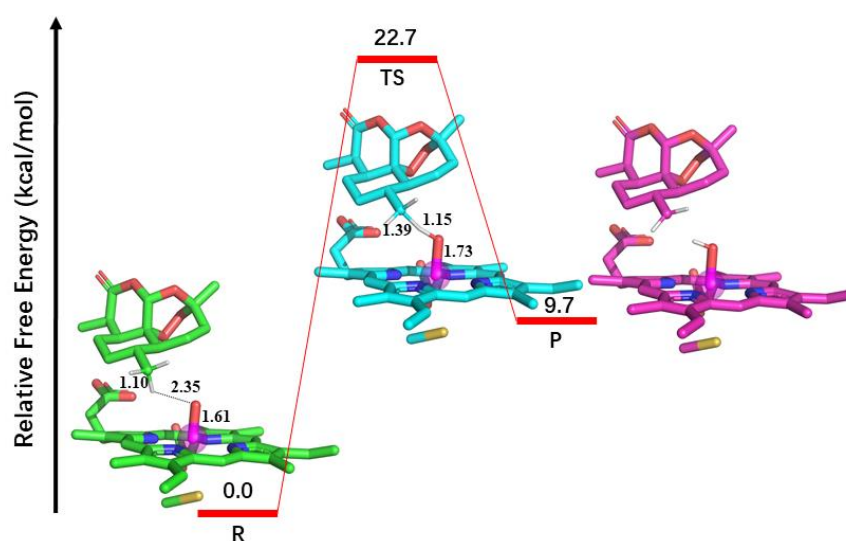


Figure 5 Free energy profile of hydrogen abstraction by X-E12. Four snapshots from cluster analysis were used here, and the energy here was averaged using Equation 1. The distance in this graph was from Cluster 0.

In the optimized reactant of X-E12-artemisinin (Figure 5), the distance between the Fe(IV)=O and C6a was 3.13 Å. The hydrogen atom to be abstracted from the C6a methyl was 2.35 Å away from the Fe (IV)=O oxo in the Cpd I oxidant. The angle of hydrogen atom of the C6a methyl group approaching the Fe (IV)=O (C6a-O-Fe) was 120.3°(Table S3), in agreement with previously research that suggested that the optimal angle for an approaching hydrogen atom to be abstracted for either aliphatic or aromatic hydroxylation is around 110-130°, which would provide a good orbital overlap and reduce the activation barrier.¹¹ The optimized transition state (TS) showed Fe=O bond increased from 1.61 Å in the reactant to 1.73 Å in the TS. The distance between the hydrogen to be transferred and C6a is increased from 1.10 Å to 1.39 Å and to Fe=O oxygen is decreased from 2.35 Å to 1.15 Å (Table S3).

The hydrogen abstraction is followed by the rebound step, where the Fe(IV)-OH intermediate generated from the hydrogen abstraction is rebounded to the substrate radical. The hydrogen

abstraction has been shown to be rate-limiting step of the P450-catalyzed hydroxylation reaction.⁴²⁻⁴⁶ The rebound step calculations were run on two representative clusters C1 and C2 (Figure S18). The relative free energy profile shows that the barrier associated with the rebound step is lower than that of the hydrogen abstraction step, indicating the hydrogen abstraction is the rate limiting step. This is in agreement with previous literature,⁴³⁻⁴⁶ so the later discussion is focused on the hydrogen abstraction, the rate limiting step of the P450-catalyzed hydroxylation.

3.2.2 C-H activation of artemisinin for C7 (S) selectivity

The IV-H4 P450_{BM3} variant showed an absolute regio- and enantioselectivity for C7 hydroxylation.¹⁰ The hydroxylation of C7 were studied on three representative cluster structures from the MD simulations of IV-H4 in complex with artemisinin. Among them, cluster 3 showed the lowest barrier of 18.7 kcal/mol (Figure 6). The reaction profiles of abstraction of the prochiral hydrogen atom that gave a C7(S) hydroxylated was also investigated for the three representative cluster structures. The reaction barriers associated with the hydrogen abstraction leading to a C7(R) hydroxylated product were consistently higher than those of the C7(S) by 5-10 kcal/mol (Figure 6, Figure S19). These results are in good agreement with the experimental findings that IV-H4 showed absolute (S) selectivity C7 position.¹⁰ The difference in the energy barriers associated with the abstraction of the prochiral hydrogen leading to the R/S hydroxylated products is similar to the previously reported barrier difference for the hydroxylation reaction at the ω or α CH₂ site of the fatty acid substrate catalyzed by P450_{BM3} or P450_{SP α} , respectively.^{12, 13}

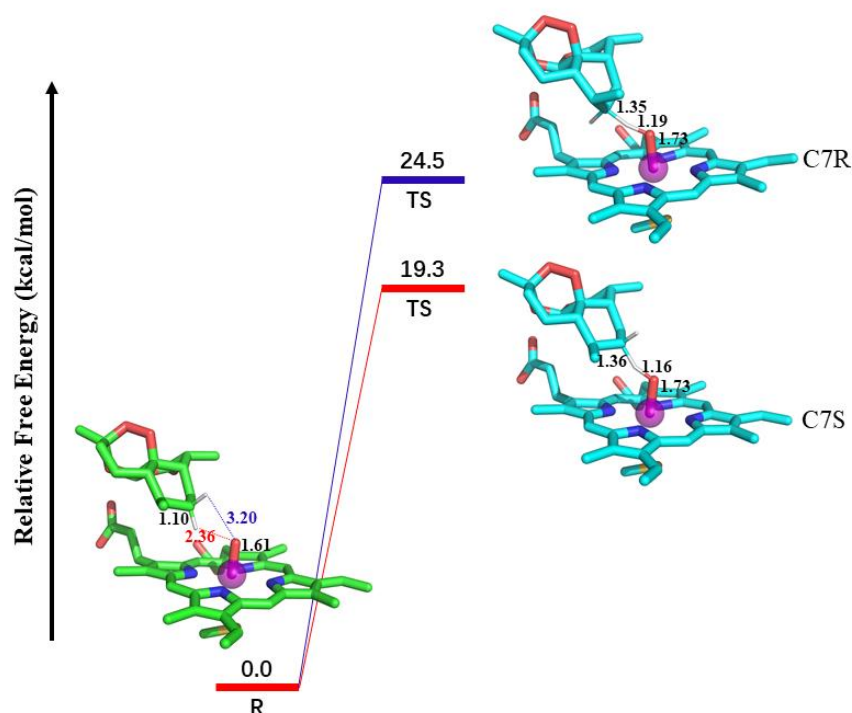


Figure 6 Free energy profile of hydrogen abstraction by IV-H4. 3 snapshots from cluster analysis were used here, and the energy here was averaged using Equation 1. The distance in this graph was from Cluster 3.

In the QM/MM optimized reactant structure (Figure 7), C7 is 3.07 Å away from Fe=O oxo and the angle formed by C7-O-Fe is 124.2° (Table S3). In the transition state, the bond length of O-H to be formed is 1.16 Å, and that of C7-H to be broken is 1.36 Å.

In order to validate the transition state, nudged elastic band (NEB) calculations were conducted. It is shown that TS has correct connectivity to link to the reactant and TS-R indeed undergoes a longer path to the reactant than TS-S (Figure S20).

3.2.3 C-H activation of artemisinin for C7 (R) selectivity

Phe87 was previously suggested to be crucial for the substrate scope.¹² Interestingly, in our MD simulations of II-H10 in complex with artemisinin, Phe87 adopted two distinct conformations. In the Phe87-out complex, Phe87 pointed away from the active site (Figure S11

&12), whereas in the Phe87-in complex (Figure S13 &14), Phe87 was nested between Cpd I and the ligand. Comparison of the switching between Phe87-in/Phe87-out conformations (Figure S14) and the overall visiting frequency (Figure S15) from multiple MD replicas run started from these two poses showed the Phe87-in is slightly more stable than the Phe87-out conformation. The relative visiting frequency corresponds to an energy difference of 0.7 kcal/mol (Figure 7). First the reaction profiles for both Phe87-out orientations were studied based on two representative structures from cluster analysis of MD simulations. For the Phe87-out complex, the average activation barrier calculated using Equation 1 for the hydrogen abstraction was around 18.9 kcal/mol (Figure 7). The distance between C7 and Fe=O (3.12 Å) is comparable to the other P450_{BM3} variants complexes (Table S3), however, the angle formed by Fe=O-C7 was 163.3°, much larger than those observed in the other two variants. In the transition state, O-H and C-H were 1.17 Å and 1.34 Å, respectively.

QM/MM calculation were then conducted for Phe87-in, also starting from two representative structures from the cluster analysis of the MD simulations. Cluster 1 showed slightly lower activation barrier 24.6 kcal/mol than cluster 2 with a reaction energy of 8.0 kcal/mol (Figure S21). Since Phe87 is located in the proximity of Cpd I, the effect of including the residue in QM region was also investigated. Little difference in the reaction barrier was observed when Phe87 was included in the QM region (denoted as L-QM). In the reactant, the distance between C7 and the Fe=O oxo was 3.44Å (Table S3) distance is larger than X-E12-artemisinin with C6a preference, or IV-H4-artemisinin with C7(S) preference. Interestingly, similar to the Phe87-out complex, the angle formed by Fe=O-C7 is also large being 154.5°, both notably larger than those observed in the other two variants that give the C6 or C7(R) selectivity. In addition, the transition state showed a longer OH distance of 1.20 Å and shorter CH distance of 1.34 Å compared to the other two variants.

The reaction profiles of II-H10 based on the two Phe87-out and Phe87-in conformations were compared (Figure 7). The former showed lower activation barrier (18.9 kcal/mol) than the latter (25.0 kcal/mol). It should be noted that Phe87 in the crystal structure of P450_{BM3} in complex with a fatty acid substrate N-Palmitoylglycine (NPG),³⁸ adopts an similar orientation as Phe87-in, also nested between the Cpd I and the fatty acid. Thus, the preference of Phe87-out over Phe87-in for the C7(R) enantioselectivity suggests the different binding pocket requirement for the native fatty acid substrate and bulky cyclic terpene substrate such as artemisinin. The flexible loop (Phe81-Glu93) adjacent to β' helix needs to adopt an appropriate conformation so as to position Phe87 in favour of the substrate binding and subsequent selective hydroxylation.

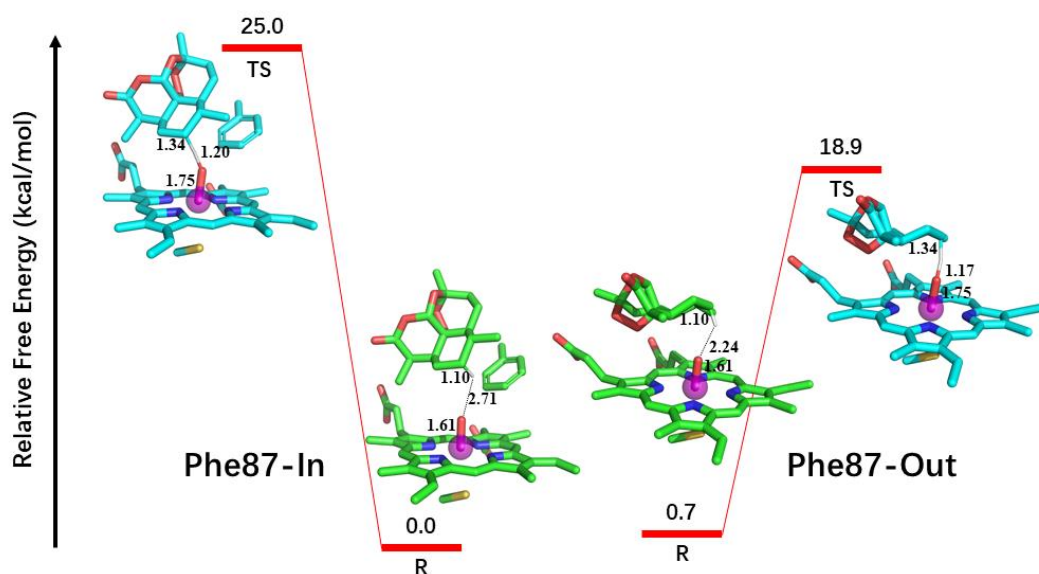


Figure 7 Free energy profile of hydrogen abstraction by II-H10 from two different poses Phe87-in and Phe87-out.

3.3 Correlation between OH bond lengths and reaction barriers

QM/MM calculations were conducted for the representative cluster structures from MD simulations for three P450_{BM3}-variants using QM/MM calculations. The O-H distance between

the H to be abstracted from artemisinin and the iron Oxo of Compound I) in the QM/MM optimized reactants from different clusters spanned a good range, therefore, we were interested to investigate if there is correlation between the O-H distances and the reaction barriers for hydrogen abstraction at different positions of artemisinin. The activation energies were plotted as a function of the O-H distances in the QM/MM optimized reactant structures (Table S1). Strikingly, a positive correlation is observed between the activation barriers and O-H distances, for the hydrogen abstraction at C6a by X-E12 variant, or C7(S) by IV-H4 and C7(R) by II-H10 (Figure 8). This indicates that the representative structure with shortest O-H distance is associated with lower reaction barrier, and therefore can be selected for the subsequent reaction mechanism study without the need to conduct complete free energy profile calculations for many representative cluster structures. This would be valuable in enzyme engineering of P450_{BM3}, since it would be possible to predict the reactivity of the designed variants for hydroxylation at specific C-H site to be activated, based on the QM/MM optimized reaction coordinates of the selected representative structures from MD simulations.

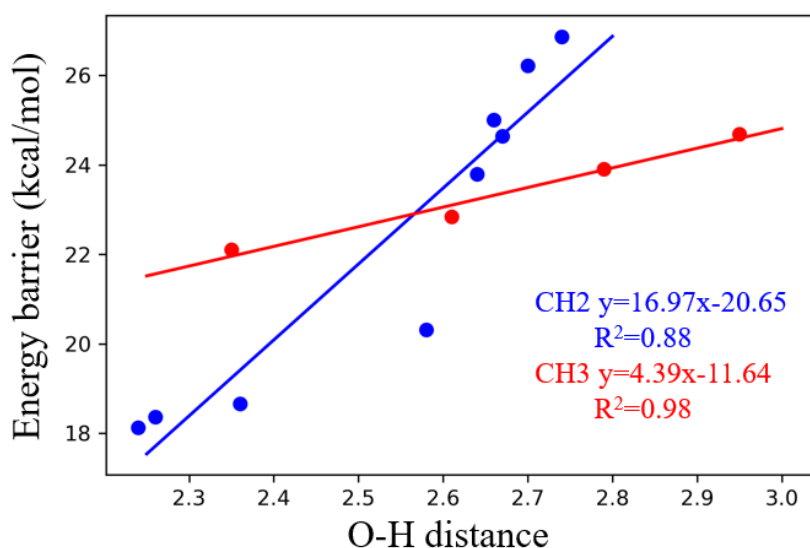


Figure 8. Correlation of free energy barriers with the reaction coordinate (i.e. the OH distance corresponding to the abstraction of H from the terpenoid to the iron Oxo of Compound I) in

QM/MM optimized reactant structures. H abstraction from C6a CH₃ in the XE-12 variant are shown in red. The hydrogen abstraction from the prochiral CH₂ of C7 in IV-H4 and II-H10 variants are shown in blue.

4 CONCLUSIONS

Elucidating the selectivity of C-H activation at various positions in drug molecules by P450 is crucial for understanding drug metabolism mechanism as well as rationally engineering the enzyme as biocatalysts for synthesis of high-value products that are difficult to be accessible via chemical synthesis routes. The site selectivity of native fatty acid substrate by P450_{BM3} has been studied by QM/MM studies, however, little is known for the binding and selectivity for non-native complex bulky natural products such as terpenoids. Here, molecular docking, molecular dynamics simulations and QM/MM calculations were conducted for P450_{BM3}-catalyzed hydroxylation, using artemisinin as a model compound, in order to understand the selectivity and the C-H activation at the C6a and C7 positions of the substrate.

Three P450_{BM3} variants, X-E12, IV-H4 and II-H10 that were reported to convert artemisinin with high regio- or enantioselectivities at C6a or C7 positions were studied. Comparison of the MD simulated structures of these variants and their template FL#62 disclosed remarkable displacement of the β' helix with the subtle difference in the mutations among them. Different interactions around the substrate binding pocket and substrate entrance were observed. The selectivity of the variants was attributed by the few different mutations in the three variants that reshaped the binding pocket, leading to favourable hydrophobic pocket to accommodate the substrate. Lys69 at the substrate entrance formed a favourable hydrogen bond with C10 of the lactone in IV-H4, whereas this additional hydrogen bond was absent in other two variants. The β 1 hairpin at the substrate entrance displayed an open conformation, however, it was closed in other X-E12 and II-H10 variants. Therefore, we propose altering the β' helix and its

adjacent secondary structures including the two adjacent loops and F-helix would modulate the substrate entrance and binding.

The activation barrier can be used to estimate the catalytic efficiency of the variants. The reactivity of the C-H activation at C6a and C7 positions of artemisinin was evaluated by QM/MM calculations, by estimating the activation energy barriers associated with the hydrogen abstraction from the C6a methyl group or from the prochiral CH₂ hydrogen atoms at C7 to Fe=O. Different representative structures obtained from MD simulations were used for the subsequent QM/MM calculation. A clear correlation was observed between the activation barriers and the distances between the Fe=O and the hydrogen to be abstracted, indicating the OH distance in the QM/MM optimized reactant would serve as a facile indicator on the reactivity of the enzyme for C-H activation. The results reported in this research would provide valuable insight for rationally engineering of P450 for selective hydroxylation of complex cyclic molecules.

ASSOCIATED CONTENT

Supplementary Information

The supplementary Information is attached separately in PDF format and is available free of charge.

AUTHOR INFORMATION

Corresponding Author

Correspondence and request for materials should be addressed to M.H. m.huang@qub.ac.uk

Author Contributions

CH and WS conducted the experiments. MH supervised the project. All authors read and approved the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors acknowledge the financial support from INVEST NI Research and Development Programme, part financed by the European Regional Development Fund under the Investment for Growth and Jobs programme 2014-2020. This work was supported by the Natural Science Foundation of China (No. 21736002, and the National Key Research and Development Program of China (2018YFA0901800). We are grateful for the computing resources from QUB high performance computing Tier2 computing resource funded by EPSRC (EP/T022175).

REFERENCES

1. P. R. Ortiz de Montellano, *Chem Rev*, 2010, **110**, 932-948.
2. J. A. McIntosh, C. C. Farwell and F. H. Arnold, *Curr Opin Chem Biol*, 2014, **19**, 126-134.
3. W. Sun, H. Xue, H. Liu, B. Lv, Y. Yu, Y. Wang, M. Huang and C. Li, *Acs Catal*, 2020, **10**, 4253-4260.
4. S. Kille, F. E. Zilly, J. P. Acevedo and M. T. Reetz, *Nat. Chem.*, 2011, **3**, 738-743.
5. A. Li, C. G. Acevedo-Rocha, L. D'Amore, J. Chen, Y. Peng, M. Garcia-Borràs, C. Gao, J. Zhu, H. Rickerby, S. Osuna, J. Zhou and M. T. Reetz, *Angew. Chem., Int. Ed.*, 2020, **59**, 12499-12505.
6. C. G. Acevedo-Rocha, C. G. Gamble, R. Lonsdale, A. Li, N. Nett, S. Hoebeinreich, J. B. Lingnau, C. Wirtz, C. Fares, H. Hinrichs, A. Deege, A. J. Mulholland, Y. Nov, D. Leys, K. J. McLean, A. W. Munro and M. T. Reetz, *Acs Catal*, 2018, **8**, 3395-3410.
7. J. Wang, Y. Zhang, H. Liu, Y. Shang, L. Zhou, P. Wei, W.-B. Yin, Z. Deng, X. Qu and Q. Zhou, *Nat. Commun.*, 2019, **10**, 3378.
8. H. Xiao, Y. Zhang and M. Wang, *Trends Biotechnol.*, 2019, **37**, 618-631.
9. X. Zhang, Y. Peng, J. Zhao, Q. Li, X. Yu, C. G. Acevedo-Rocha and A. Li, *Bioresour. Bioprocess.*, 2020, **7**, 2.
10. K. Zhang, B. M. Shafer, M. D. Demars, H. A. Stern and R. Fasan, *J. Am. Chem. Soc.*, 2012, **134**, 18695-18704.
11. R. Lonsdale, K. T. Houghton, J. Żurek, C. M. Bathelt, N. Foloppe, M. J. de Groot, J. N. Harvey and A. J. Mulholland, *J. Am. Chem. Soc.*, 2013, **135**, 8001-8015.
12. K. D. Dubey, B. Wang and S. Shaik, *J. Am. Chem. Soc.*, 2016, **138**, 837-845.
13. R. Ramanan, K. D. Dubey, B. Wang, D. Mandal and S. Shaik, *J. Am. Chem. Soc.*, 2016, **138**, 6786-6797.
14. I. F. Sevioukova, H. Li, H. Zhang, J. A. Peterson and T. L. Poulos, *PNAS*, 1999, **96**, 1863.

- 557 15. K. Shahrokh, A. Orendt, G. S. Yost and T. E. Cheatham III, *J. Comput. Chem.*, 2012, **33**, 119-
558 133.
- 559 16. D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H.
560 Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S.
561 LeGrand, P. Li, C., Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H.
562 Nguyen, H. T. Nguyen, I., Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L.
563 Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao
564 and P. A. Kollman, 2016, **University of California, San Francisco**.
- 565 17. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G.
566 Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J.
567 Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J.
568 L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T.
569 Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada,
570 M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H.
571 Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J.
572 J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K.
573 Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M.
574 Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B.
575 Foresman and D. J. Fox, *Journal*, 2016.
- 576 18. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*,
577 1983, **79**, 926-935.
- 578 19. J. A. Izaguirre, D. P. Catarello, J. M. Wozniak and R. D. Skeel, *J. Chem. Phys.*, 2001, **114**, 2090-
579 2098.
- 580 20. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem.*
581 *Phys.*, 1984, **81**, 3684-3690.
- 582 21. J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327-341.
- 583 22. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.
- 584 23. D. R. Roe and T. E. Cheatham, *J Chem Theory Comput*, 2013, **9**, 3084-3095.
- 585 24. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33-38.
- 586 25. P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A.
587 A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S.
588 C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer and C.
589 Lennartz, *J. Mol. Struct. THEOCHEM*, 2003, **632**, 1-28.
- 590 26. F. Neese, *WIREs Comput Mol Sci*, 2012, **2**, 73-78.
- 591 27. W. Smith and T. R. Forester, *J. Mol. Graph.*, 1996, **14**, 136-141.
- 592 28. D. Bakowies and W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580-10594.
- 593 29. U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1986, **7**, 718-730.
- 594 30. J. Kästner, J. M. Carr, T. W. Keal, W. Thiel, A. Wander and P. Sherwood, *J. Phys. Chem. A*,
595 2009, **113**, 11856-11865.
- 596 31. C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785-789.
- 597 32. B. Miehlisch, A. Savin, H. Stoll and H. Preuss, *Chem. Phys. Lett.*, 1989, **157**, 200-206.
- 598 33. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456-1465.
- 599 34. A. Altun, D. Kumar, F. Neese and W. Thiel, *J. Phys. Chem. A*, 2008, **112**, 12904-12910.
- 600 35. H. Chen, J. Song, W. Lai, W. Wu and S. Shaik, *J Chem Theory Comput*, 2010, **6**, 940-953.
- 601 36. F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297-3305.
- 602 37. F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.*, 2009, **356**, 98-109.
- 603 38. D. C. Haines, D. R. Tomchick, M. Machius and J. A. Peterson, *Biochemistry*, 2001, **40**, 13456-
604 13465.
- 605 39. K. Zhang, S. El Damaty and R. Fasan, *J. Am. Chem. Soc.*, 2011, **133**, 3242-3245.
- 606 40. C. J. C. Whitehouse, S. G. Bell and L.-L. Wong, *Chem. Soc. Rev.*, 2012, **41**, 1218-1260.
- 607 41. H. Yeom, S. G. Sligar, H. Li, T. L. Poulos and A. J. Fulco, *Biochemistry*, 1995, **34**, 14733-14740.

- 608 42. S. Shaik, D. Kumar, S. P. de Visser, A. Altun and W. Thiel, *Chem Rev*, 2005, **105**, 2279-2328.
609 43. T. Kamachi and K. Yoshizawa, *J. Am. Chem. Soc.*, 2003, **125**, 4652-4661.
610 44. D. Kumar, S. P. de Visser, P. K. Sharma, S. Cohen and S. Shaik, *J. Am. Chem. Soc.*, 2004, **126**,
611 1907-1920.
612 45. I. Viciano and S. Martí, *J. Phys. Chem. B*, 2016, **120**, 3331-3343.
613 46. W. Lai, H. Chen, S. Cohen and S. Shaik, *J. Phys. Chem. Lett.*, 2011, **2**, 2229-2235.
614 47. S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar and W. Thiel, *Chem Rev*, 2010, **110**, 949-1017.

615