PAPER

Check for updates

Regio- and stereoselectivity in the CYP450$_{BM3}$-catalyzed hydroxylation of complex terpenoids: a QM/MM study†

Chenggong Hui,‡[a] Warispreet Singh,‡§[ab] Derek Quinn,[b] Chun Li,[cd] Thomas S. Moody[be] and Meilan Huang 🄳 *[a]

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 the hydroxylation of complex terpenoid substrates remains elusive, which has limited the rational engineering of P450 as a biocatalyst 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 a model compound. We found that the conformational change of the β1 sheet at the substrate entrance and the displacement of the β′ helix were 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 hydroxylation. There is a distinct linear correlation between activation barriers and reaction coordinates, indicating that reaction coordinates can be used as a facile descriptor for predicting the reactivity of P450$_{BM3}$. These findings would provide valuable guidance for predicting the selectivity and reactivity of P450$_{BM3}$ for the selective hydroxylation of non-native terpenoid substrates so as to prioritize the rationally designed enzymes for terpenoid biosynthesis.

1 Introduction

Cytochrome P450 is a superfamily of heme-containing monooxygenases 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 bonds 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

a *Department of Chemistry & Chemical Engineering, Queen's University, Belfast, BT9 5AG Northern Ireland, UK. E-mail: m.huang@qub.ac.uk*

b *Almac Sciences, Department of Biocatalysis and Isotope Chemistry, Almac House, 20 Seagoe Industrial Estate, Craigavon BT63 5QD, Northern Ireland, UK*

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

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

e *Arran Chemical Company Limited, Unit 1 Monksland Industrial Estate, Athlone, Co., Roscommon, Ireland*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cp03083j

‡ These authors contributed equally to the paper.

§ Current address: Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, England, UK.
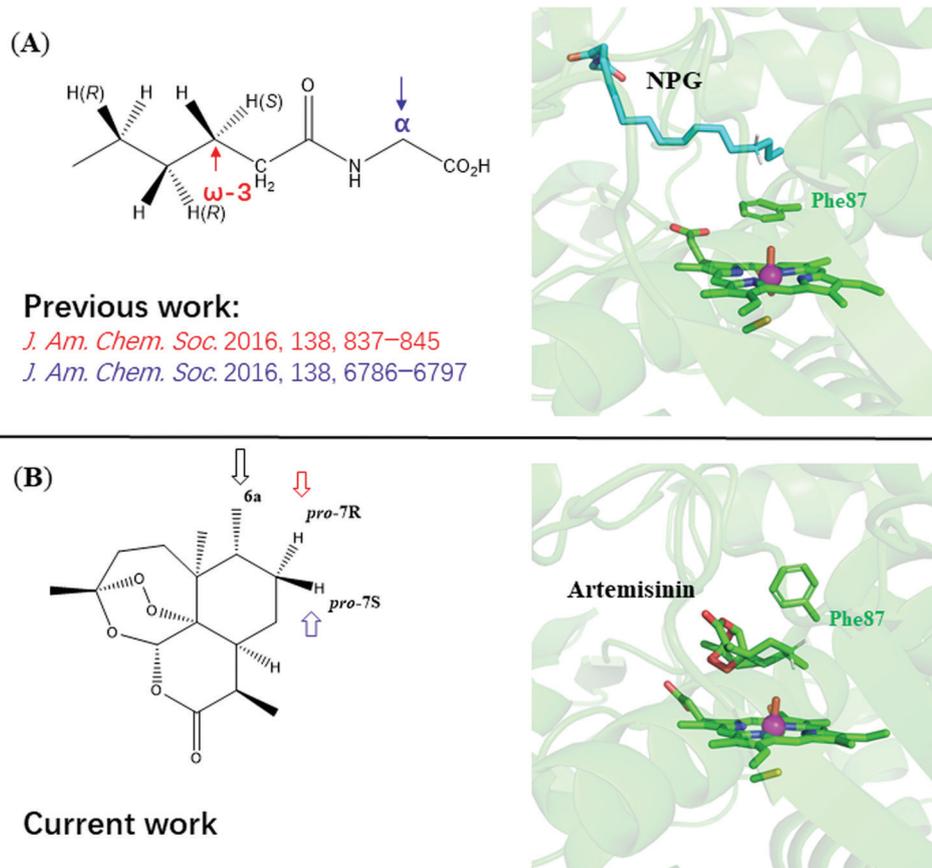
bonds in the molecules.[3–7] Of particular importance is the *Bacillus megaterium* P450$_{BM3}$ (CYP102A1), a catalytically self-sufficient fatty acid monooxygenase that exhibits the highest reported monooxygenase activity among the P450 family enzymes.

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 the artemisinin molecule.[10] Despite the outstanding performance of the P450$_{BM3}$ variants, there has been little study on their action mechanisms due to the 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 sites 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

21696 | *Phys. Chem. Chem. Phys.*, 2020, **22**, 21696–21706

This journal is © the Owner Societies 2020

Fig. 1 (A) Previous selectivity study on the fatty acid substrate NPG that mimics the native substrate of P450$_{BM3}$. (B) The current study is focused on the regio- and enantioselectivity of P450$_{BM3}$ for cyclic terpene compounds, using artemisinin as a model molecule.

particularly on the very short-lived reaction intermediates and transition states involved in the catalytic cycle.[11] For example, Dubey *et al.* studied the origin of the regio- and stereoselectivity in the P450$_{BM3}$ enzyme for the hydroxylation of fatty acid substrate *N*-palmitoylglycine, which mimics the native substrate, by applying the aforementioned computational methods (Fig. 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$_{SP\alpha}$ (Fig. 1A).[13]

However, to the best of 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 engineer the enzyme for the 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 a 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 β1 sheet at the substrate entrance and the displacement of the β′ helix which reshapes 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 the directed oxidative functionalization of C–H bonds in non-native terpenoid substrates.

2 Methods

Protein preparation

The initial structure of 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 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 the 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

This journal is © the Owner Societies 2020

Phys. Chem. Chem. Phys., 2020, **22**, 21696–21706 | **21697**

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 the restrained electrostatic potential (RESP) method at the HF/6-31G* level of theory using Gaussian 16.[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 the TIP3P[18] water molecule such that no protein atom was within 10 Å at any edge of the box. The periodic boundary conditions were used 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 ued. The systems were subjected to two successive steps of energy minimization first using steepest descent (2500 steps) and conjugate gradient algorithms (2500 steps) with proteins and ligands restrained with a constant of 50 kcal mol^{-1} A^{-2} followed by the steepest descent (2500 steps) and conjugate gradient algorithms (2500 steps) without restraint. The systems were gradually heated from 0 to 300 K for 250 ps under the NVT ensemble by restraining the solute molecules using a harmonic potential of 50 kcal mol^{-1} Å^{-2} using the Langevin thermostat[19] with a collision frequency of 1 ps^{-1}. The systems were then subjected to 50 ps of equilibration at 300 K 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 barostat[20] 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 the 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 algorithm[21] was used to constrain the bonds of all hydrogen atoms. Long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) method[22] involving a direct space and vdW cut-off of 10 Å. The trajectories were analysed using pytraj[23] and VMD.[24]

QM/MM calculations

The reaction profiles of P450$_{BM3}$ variants in the catalysis of artemisinin were studied using QM/MM calculations implemented

in Chemshell.[25] The QM calculations were performed using ORCA 4.2.0[26] 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 using 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 (1250) and conjugate gradient (1250) algorithms using Amber16.[16] The water shell within 4 Å of the protein or within 20 Å of the QM atoms was retained. The QM region consists of the whole Cpd I molecule, artemisinin and the C400 residue truncated at Cβ 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 the QM region was −2. The residues that are within 10 Å of Cpd I and artemisinin were allowed to move freely and the rest of the system was frozen during geometry optimization. The hydrogen linked atoms[29] 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 through a potential energy scan followed by a dimer TS search implemented in DL-FIND,[30] and were subsequently validated by frequency calculation. Only one imaginary frequency corresponding to the hydrogen transfer was shown. All QM calculations were performed with DFT using UB3LYP[31,32] with D3 dispersion correction and BJ damping,[33] which is widely used for the 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 the 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 was calculated with the same level of DFT functional and added to all the stationary points. The final energies were corrected at the UB3LYP-D3BJ/def2-TZVPP level. RIJCOSX approximation[37] was used in all the QM calculations. In the QM/MM calculation in each mutant, several snapshots were obtained, and the averaged barrier was calculated using eqn (1).

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

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 electrons from NADPH, while FMN reduces P450. The generation

21698 | Phys. Chem. Chem. Phys., 2020, 22, 21696–21706

This journal is © the Owner Societies 2020

of the active oxidant Cpd I in P450 requires the 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).
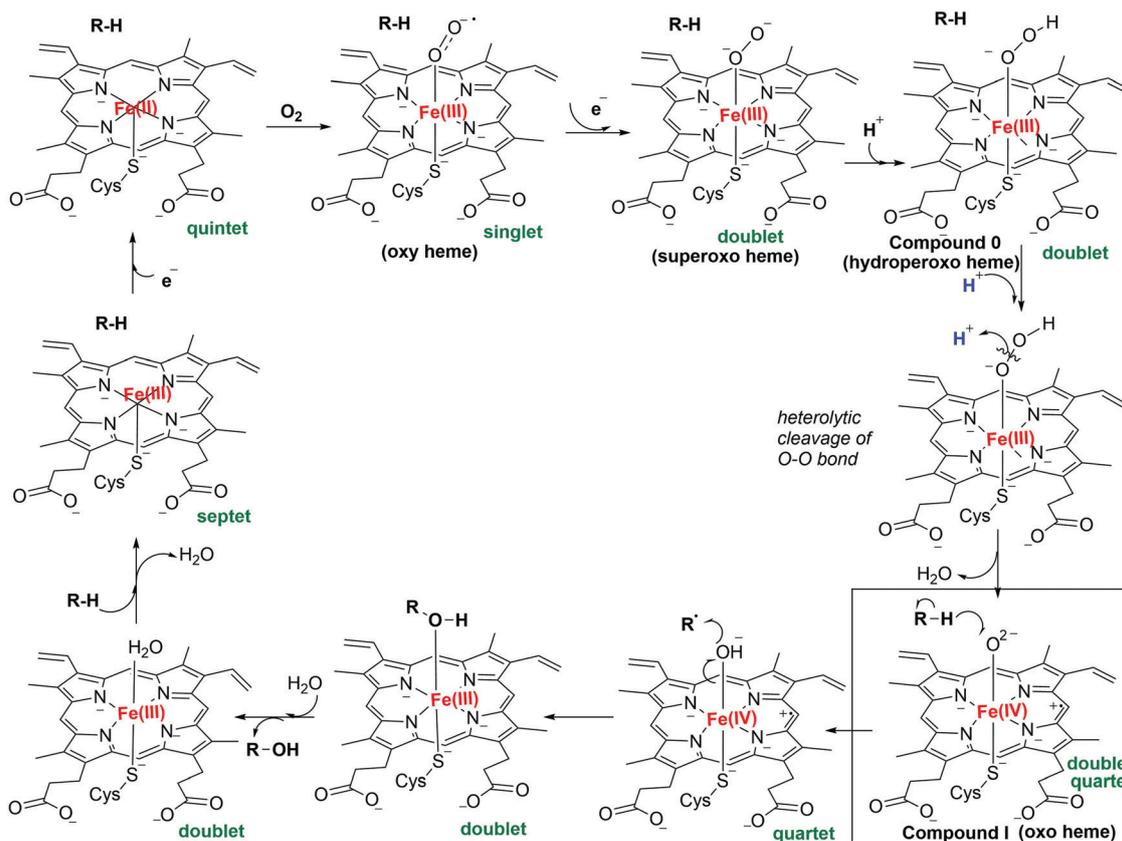
P450$_{BM3}$ (CYP102A1) is a self-sufficient fatty acid mono-oxygenase containing heme and FAD/FMN domains. In the crystal structure of apo P450$_{BM3}$ from *Bacillus megaterium* (PDB: 1BVY),[14] the Asp251 adjacent to the loop 240–250 that contacts with the FMN domain formed a salt bridge with Lys224. 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 β1 hairpin at the substrate entrance exhibits an open configuration and Arg47 on the hairpin formed a salt bridge with Glu352 (ESI,† Fig. 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 (Fig. 1) to give a mixture of hydroxylated products. The majority of the products were C7(S) enantiomers, 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 the C6a oxidized product and 100% enantio-selectivity 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; an additional four mutations resulted in unprecedented regioselectivity at position C6a.

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 CH$_2$ that gives a C7(S) enantiomer product (pro-C7(S)) (16 out of 50 docked poses), whereas few pro-C7(R) poses were also observed (2 out of 50 docked poses). In addition, the predicted binding energy was −5.31 kcal mol^{-1} for pro-C7(S) and −4.55 kcal mol^{-1} for pro-C7(R), indicating that the pro-C7(S) is more stable than pro-C7(R). These are in agreement with the experimental results that the C7(S) product is dominated by FL#62.[10] MD simulations of FL#62 in complex with the pro-C7(S) or pro-C7(R) ligand (Fig. S2–S5, ESI†) disclosed that Asp251 adjacent to the FMN contacting loop formed a salt bridge with Lys224 (Fig. S6A and B, ESI†). The hairpin Ala44–Arg47 located at the substrate entrance exhibited a closed conformation in both complexes (Fig. 2). In contrast to the crystal structure of apo P450$_{BM3}$ where a salt bridge is observed between Arg47 and Glu352 (Fig. S1, ESI†),[14] such a salt bridge was absent in the FL#62-pro-C7(S)/(R) complex (Fig. 2).

It should be noted that in the FL#62-pro-C7(S) complex, both propanoic acid carboxylates 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* the C10

Scheme 1 Proposed mechanism for the P450-catalyzed hydroxylation of substrate R−H.
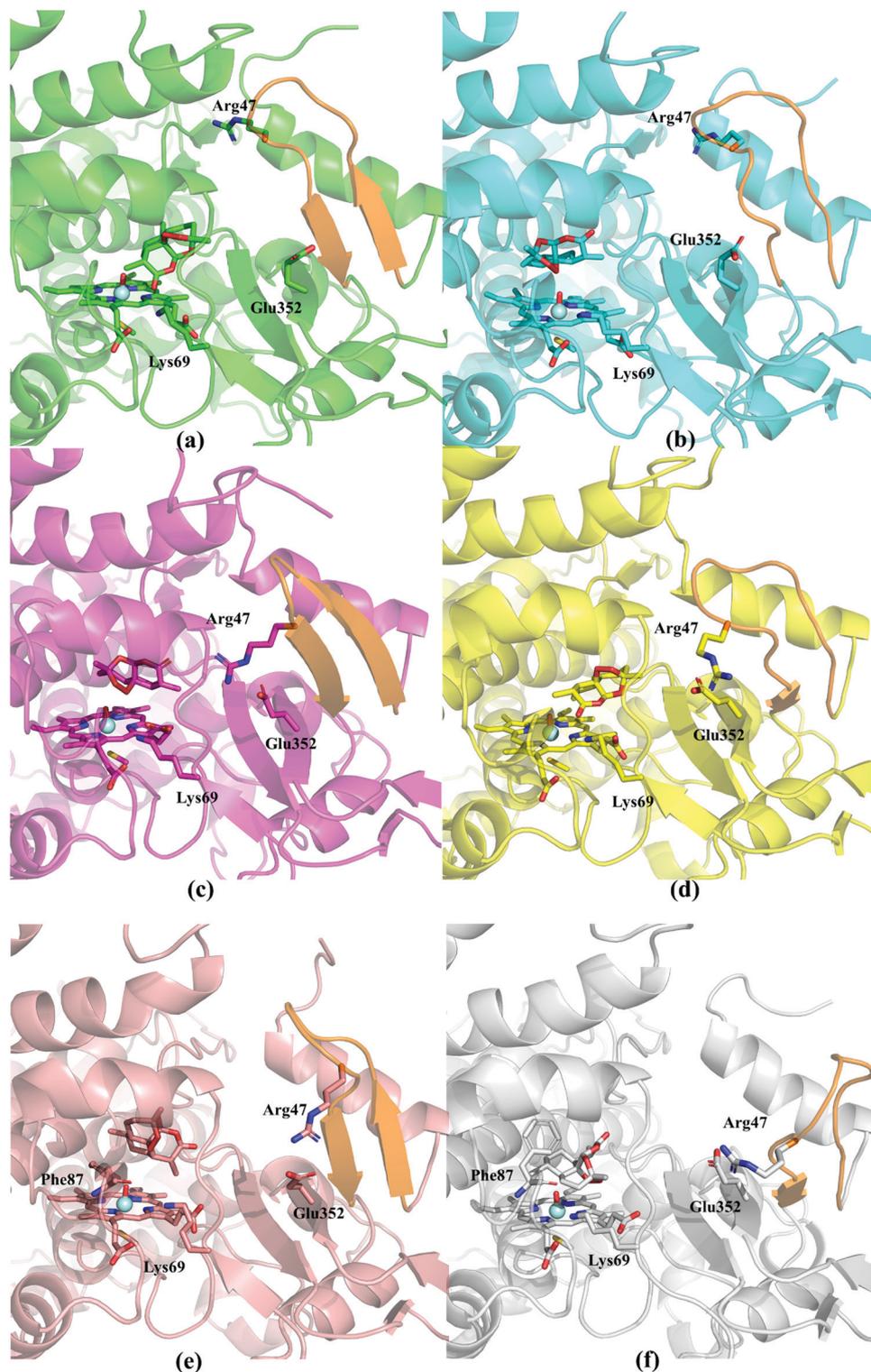
This journal is © the Owner Societies 2020

Phys. Chem. Chem. Phys., 2020, **22**, 21696–21706 | **21699**

Fig. 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), and (f) II-H10 with pro-C7(R) (Phe87-out). The hairpin near the substrate entrance is shown in orange.

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 bonds with one of the propanoic acid

carboxylates; however, unlike the enzyme in the presence of pro-C7(S), Lys69 in the FL#62-pro-C7(R) complex did not form a H-bond with the C10 carbonyl oxygen, which may account for the

21700 | *Phys. Chem. Chem. Phys.*, 2020, **22**, 21696–21706

This journal is © the Owner Societies 2020

low conversion toward the C7(R) product compared to the C7(S) product.

3.1.2 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 (Fig. S9, ESI†) showed that the pro-C7(S) hydrogen was consistently closer to Fe(IV)=O than pro-C7(R) hydrogen during the entire length of the MD trajectory (Fig. S10, ESI†), 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 the IV-H4 mutant in complex with pro-C7(S) artemisinin showed that Lys 224 formed a salt bridge with Asp251 (Fig. S6, ESI†). In addition, a salt bridge was formed between Glu352 and Arg47 located at the hairpin Ala44–Arg47, leading to a closed substrate entrance (Fig. 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 the substrate, locking the lactone of the substrate in the catalytic site such that the prochiral (S) hydrogen at the 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 (Fig. 3). The mutations at positions 78, 81 and 82 may synergistically introduce additional flexibility to 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 C7-(S) selectivity.

3.1.3 C7(R) selectivity. Previous QM/MM studies suggested that Phe87 plays an important role in controlling the (R) stereoselectivity of oxidation reactions for the 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, and V184T, which gave a hydroxylated product at the C7 position with 100% R selectivity. The MD simulations of the II-H10-pro-C7(R) complex (Fig. S11, ESI†) showed that the distance
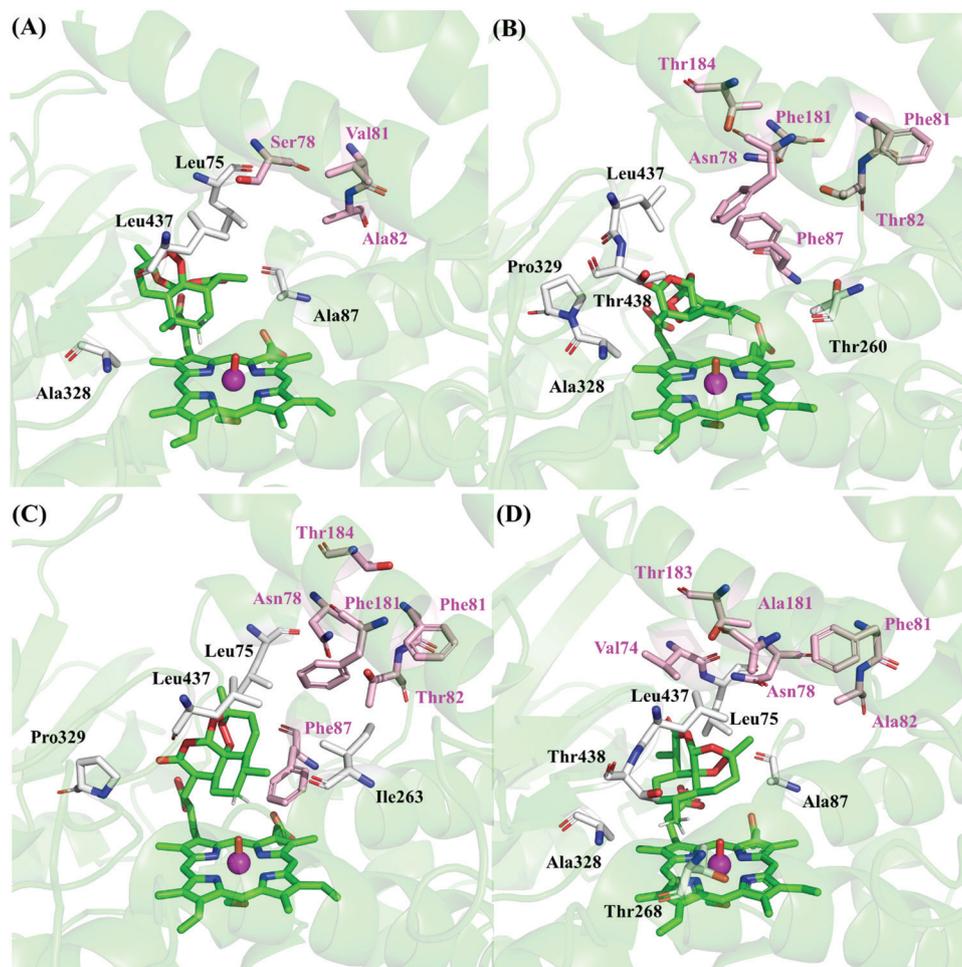

Fig. 3 MD simulated structures of P450$_{BM3}$ variants in complex with artemisinin. (A) IV-H4 with pro-C7(S). (B) II-H10 with pro-C7(R) (Phe87-out). (C) II-H10 with pro-C7(R) (Phe87-in). (D) X-E12 with preference to the abstraction of the methyl H from C6a. The mutated residues in relation to the FL#62 template are shown in pink.

between the Fe=O and pro-C7(R) hydrogen was generally shorter than with C7(S) or C6a (Fig. S12, ESI†), 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 towards 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, and A87F located on the flexible loop adjacent to the β′ helix may help 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 β1 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 the 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 (Fig. 3B). In contrast, in the Phe87-in conformation, the β1 hairpin is closed. Leu437 pointed away from the binding pocket so that the hydrophobic pocket composed of Leu75, Ile263, Pro329 and Leu437 is enlarged (Fig. 3C). Closing of the hairpin pushed the β′ helix along with the 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 (Fig. 2).

Since both conformations were observed in the extensive MD simulations of II-H10 in complex with pro-C7(R) artemisinin, both were considered in the subsequent QM/MM calculations in the study of C7(R) enantioselectivity. It was found that the Phe87-out conformation displayed a lower energy barrier than th eePhe87-in conformation in the hydrogen abstraction from the pro-(R) hydrogen by heme Fe(IV)=O (see Section 3.2.3).

3.1.4 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 was retained largely around 2.5 Å (Fig. S6, ESI†). 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 (Fig. 2c). It should be noted that the salt bridge was also present in IV-H4-pro7(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 the 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. Furthermore, Thr268 on the distal H-helix, which was previously suggested to be important for the O$_2$ binding and activation for the catalysis of the 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 at the β′ helix (Fig. 3D). Hence, altering the loop region 69–73 and the adjacent β′-helix, as well as the 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 (Fig. 4).

3.2 Origin of reactivity through 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.

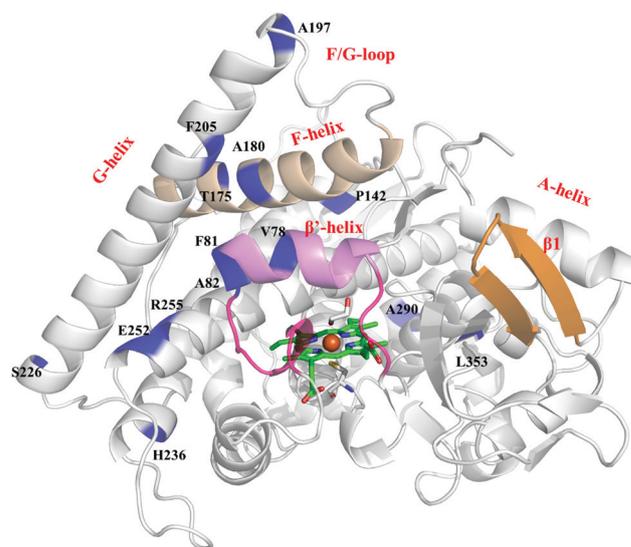Among the four reaction profiles, the activation barrier from cluster 0 showed the lowest barrier of 22.1 kcal mol^{-1} (ESI,†

Fig. 4 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 are shown in purple.
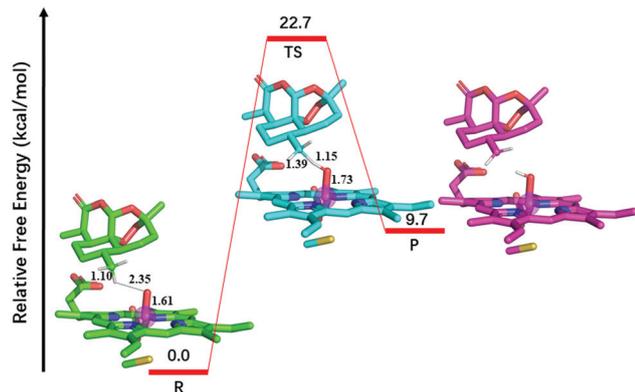
21702 | *Phys. Chem. Chem. Phys.*, 2020, **22**, 21696–21706

This journal is © the Owner Societies 2020

Fig. 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 eqn (1). The distance in this graph was from cluster 0.

Table S1 and Fig. S16). An average barrier of 22.7 kcal mol^{-1} was obtained according to eqn (1) (Fig. 5).

A basis set benchmark study was performed for cluster 0, the representative structure of the X-E12 variant from MD simulations, which has the lowest barrier for hydrogen abstraction. In addition to the def2-SVP basis set (B1), geometry optimizations were also conducted using the def2-TZVP basis set (B2). The difference in the energy barrier using the two basis sets was less than 2 kcal mol^{-1} (Fig. S16, ESI†). The error was mainly attributed by the potential energy calculated at the two basis sets, while the ZPEs at these basis sets were almost the same (Table S2A and B, ESI†). 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 the 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, ESI†). Our calculations showed that the two spin states were highly degenerated, with a barrier of 22.7 kcal mol^{-1} for $S = 1/2$ and 22.8 kcal mol^{-1} for $S = 3/2$ based on the energies of the geometries optimized with the def2-TZVP (B2) basis set (Fig. S17, ESI†). Since the two spin states of iron in Cpd I are highly degenerated, all of the remaining calculations were performed with doublet spin state.

In the optimized reactant of X-E12-artemisinin (Fig. 5), the distance between the Fe(ɪᴠ)═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 (ɪᴠ)═O oxo in the Cpd I oxidant. The angle of the hydrogen atom of the C6a methyl group approaching the Fe (ɪᴠ)═O (C6a–O–Fe) was 120.3° (Table S3, ESI†), in agreement with the previous 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 that the 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 that to Fe═O oxygen is decreased from 2.35 Å to 1.15 Å (Table S3, ESI†).

The hydrogen abstraction is followed by the rebound step, where the Fe(ɪᴠ)–OH intermediate generated from hydrogen abstraction is rebounded to the substrate radical. The hydrogen abstraction has been shown to be a rate-limiting step of the P450-catalyzed hydroxylation reaction.[42–46] The rebound step calculations were performed on two representative clusters C1 and C2 (Fig. S18, ESI†). The relative free energy profile shows that the barrier associated with the rebound step is lower than that of the hydrogen abstraction step, indicating that hydrogen abstraction is the rate limiting step. This is in agreement with previous literature;[43–46] so the later discussion is focused on hydrogen abstraction, the rate limiting step of 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 was studied on three representative cluster structures derived from the MD simulations of IV-H4 in complex with artemisinin. Among them, cluster 3 showed the lowest barrier of 18.7 kcal mol^{-1} (Fig. 6). The reaction profile of abstraction of the prochiral hydrogen atom that gave a C7(S) hydroxylate was also investigated for the three representative cluster structures. The reaction barriers associated with hydrogen abstraction leading to a C7(R) hydroxylated product were consistently higher than those of the C7(S) by 5–10 kcal mol^{-1} (Fig. 6 and Fig. S19, ESI†). These results are in good agreement with the experimental findings that IV-H4 showed the absolute (S) selectivity C7 position.[10] The difference in the energy barriers associated with the abstraction of the prochiral hydrogen leading to R/S hydroxylated products is similar to the previously reported barrier difference for the hydroxylation reaction at the ω or α CH$_2$ site of the fatty acid substrate catalyzed by P450$_{BM3}$ or P450$_{SPα}$, respectively.[12,13]

In the QM/MM optimized reactant structure (Fig. 7), C7 is 3.07 Å away from Fe═O oxo and the angle formed by C7–O–Fe is 124.2° (Table S3, ESI†). 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 5undergoes a longer path to the reactant than TS-S (Fig. S20, ESI†).

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 (Fig. S11 and S12, ESI†), whereas in the Phe87-in complex (Fig. S13 and S14, ESI†), Phe87 was nested between Cpd I and the ligand. Comparison of the switching between Phe87-in/Phe87-out conformations (Fig. S14, ESI†) and
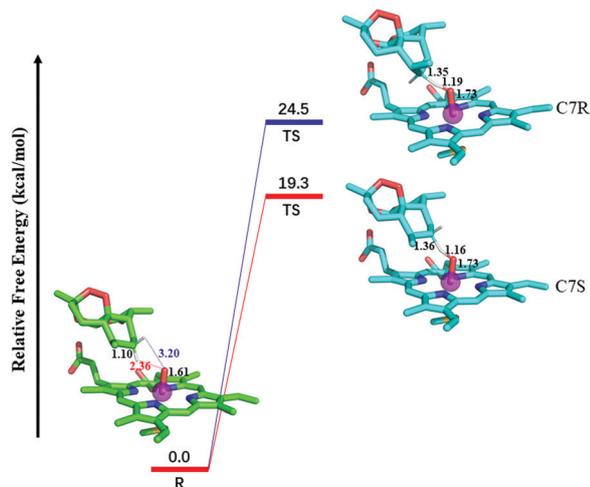
This journal is © the Owner Societies 2020

Phys. Chem. Chem. Phys., 2020, **22**, 21696–21706 | **21703**

Fig. 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 eqn (1). The distance in this graph was obtained from cluster 3.

the overall visiting frequency (Fig. S15, ESI†) from multiple MD replica runs starting from these two poses showed that the Phe87-in conformation is slightly more stable than the Phe87-out conformation. The relative visiting frequency corresponds to an energy difference of 0.7 kcal mol^{-1} (Fig. 7). First, the reaction profiles for both Phe87-out orientations were studied based on two representative structures from the cluster analysis of MD simulations. For the Phe87-out complex, the average activation barrier calculated using eqn (1) for hydrogen abstraction was around 18.9 kcal mol^{-1} (Fig. 7). The distance between C7 and Fe=O (3.12 Å) is comparable to the other P450$_{BM3}$ variant complexes (Table S3, ESI†), 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, the angles formed by O–H and C–H were 1.17 Å and 1.34 Å, respectively.

QM/MM calculations were then conducted for Phe87-in, also starting from two representative structures from the cluster analysis of the MD simulations. Cluster 1 showed a slightly lower activation barrier of 24.6 kcal mol^{-1} than cluster 2 with a reaction energy of 8.0 kcal mol^{-1} (Fig. S21, ESI†). Since Phe87 is located in the proximity of Cpd I, the effect of including the
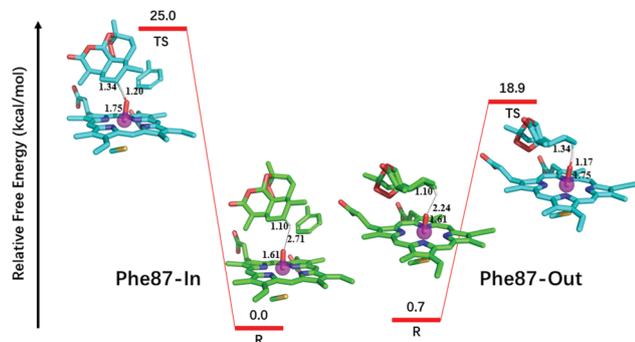
residue in the QM region was also investigated. Little difference in the reaction barrier was observed when Phe87 was included in the QM region (denoted as L-QM). In the reactant, the distance between C7 and Fe=O oxo was 3.44 Å (Table S3, ESI†), a distance which is larger than X-E12-artemisinin with 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 a 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 (Fig. 7). The former showed a lower activation barrier (18.9 kcal mol^{-1}) than the latter (25.0 kcal mol^{-1}). 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 a similar orientation as Phe87-in, also nested between 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 the bulky cyclic terpene substrate such as artemisinin. The flexible loop (Phe81–Glu93) adjacent to the β′ helix needs to adopt an appropriate conformation so as to position Phe87 in favour of substrate binding and subsequent selective hydroxylation.

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 a correlation between the O–H distances and the reaction barriers for hydrogen abstraction at different
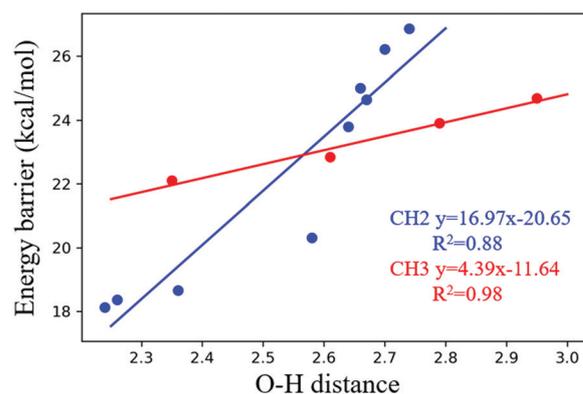

Fig. 7 Free energy profile of hydrogen abstraction by II-H10 from two different poses Phe87-in and Phe87-out.

Fig. 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 CH$_3$ in the XE-12 variant is shown in red. The hydrogen abstraction from the prochiral CH$_2$ of C7 in IV-H4 and II-H10 variants is shown in blue.

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, ESI†). Strikingly, a positive correlation is observed between the activation barriers and O–H distances, for the hydrogen abstraction at C6a by the X-E12 variant, or C7(*S*) by IV-H4 and C7(*R*) by II-H10 (Fig. 8). This indicates that the representative structure with the shortest O–H distance is associated with a 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 for the enzyme engineering of P450$_{BM3}$, since it would be possible to predict the reactivity of the designed variants for hydroxylation at the specific C–H site to be activated, based on the QM/MM optimized reaction coordinates of the selected representative structures derived from MD simulations.

4 Conclusions

Elucidating the selectivity of C–H activation at various positions in drug molecules by P450 is crucial for understanding the drug metabolism mechanism as well as rationally engineering the enzyme as a biocatalyst for the synthesis of high-value products that are difficult to access *via* chemical synthesis routes. The site selectivity of the native fatty acid substrate by P450$_{BM3}$ has been studied by QM/MM studies, however, little is known regarding the binding and selectivity of 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 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. A comparison of the MD simulated structures of these variants and their template FL#62 disclosed a remarkable displacement of the β′ helix with a subtle difference in 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 a 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 β1 hairpin at the substrate entrance displayed an open conformation; however, it was closed in the 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 the 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 the C6a and C7 positions of artemisinin was evaluated through QM/MM calculations, by estimating the activation energy barriers associated with the hydrogen abstraction from the C6a methyl group or from the prochiral CH$_2$ hydrogen atoms at C7 to Fe=O. Different representative structures obtained from the 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 that 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 provide valuable insight for rationally engineering P450 for the selective hydroxylation of complex cyclic molecules.

Author contributions

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

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgements

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. Hoebenreich, 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.

This journal is © the Owner Societies 2020

Phys. Chem. Chem. Phys., 2020, **22**, 21696–21706 | **21705**

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. Sevrioukova, H. Li, H. Zhang, J. A. Peterson and T. L. Poulos, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 1863.

15 K. Shahrokh, A. Orendt, G. S. Yost and T. E. Cheatham III, *J. Comput. Chem.*, 2012, **33**, 119–133.

16 D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao and P. A. Kollman, *Amber2018*, University of California, San Francisco, 2018.

17 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16*, 2016.

18 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

19 J. A. Izaguirre, D. P. Catarello, J. M. Wozniak and R. D. Skeel, *J. Chem. Phys.*, 2001, **114**, 2090–2098.

20 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

21 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

22 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

23 D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.

24 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

25 P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer and C. Lennartz, *THEOCHEM*, 2003, **632**, 1–28.

26 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 73–78.

27 W. Smith and T. R. Forester, *J. Mol. Graphics*, 1996, **14**, 136–141.

28 D. Bakowies and W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580–10594.

29 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1986, **7**, 718–730.

30 J. Kästner, J. M. Carr, T. W. Keal, W. Thiel, A. Wander and P. Sherwood, *J. Phys. Chem. A*, 2009, **113**, 11856–11865.

31 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.

32 B. Miehlich, A. Savin, H. Stoll and H. Preuss, *Chem. Phys. Lett.*, 1989, **157**, 200–206.

33 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

34 A. Altun, D. Kumar, F. Neese and W. Thiel, *J. Phys. Chem. A*, 2008, **112**, 12904–12910.

35 H. Chen, J. Song, W. Lai, W. Wu and S. Shaik, *J. Chem. Theory Comput.*, 2010, **6**, 940–953.

36 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

37 F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.*, 2009, **356**, 98–109.

38 D. C. Haines, D. R. Tomchick, M. Machius and J. A. Peterson, *Biochemistry*, 2001, **40**, 13456–13465.

39 K. Zhang, S. El Damaty and R. Fasan, *J. Am. Chem. Soc.*, 2011, **133**, 3242–3245.

40 C. J. C. Whitehouse, S. G. Bell and L.-L. Wong, *Chem. Soc. Rev.*, 2012, **41**, 1218–1260.

41 H. Yeom, S. G. Sligar, H. Li, T. L. Poulos and A. J. Fulco, *Biochemistry*, 1995, **34**, 14733–14740.

42 S. Shaik, D. Kumar, S. P. de Visser, A. Altun and W. Thiel, *Chem. Rev.*, 2005, **105**, 2279–2328.

43 T. Kamachi and K. Yoshizawa, *J. Am. Chem. Soc.*, 2003, **125**, 4652–4661.

44 D. Kumar, S. P. de Visser, P. K. Sharma, S. Cohen and S. Shaik, *J. Am. Chem. Soc.*, 2004, **126**, 1907–1920.

45 I. Viciano and S. Martí, *J. Phys. Chem. B*, 2016, **120**, 3331–3343.

46 W. Lai, H. Chen, S. Cohen and S. Shaik, *J. Phys. Chem. Lett.*, 2011, **2**, 2229–2235.

47 S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar and W. Thiel, *Chem. Rev.*, 2010, **110**, 949–1017.

21706 | *Phys. Chem. Chem. Phys.*, 2020, **22**, 21696–21706

This journal is © the Owner Societies 2020