Catalysis Science & Technology

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Catal. Sci. Technol.*, 2025, 15, 7067

Hydroxylation mechanism of lignin-derived aromatic substrates catalyzed by plant P450 cinnamate 4-hydroxylase

Sónia F. G. Santos,[ab] Paul James,[b] Rajesh Reddy Bommareddy,[b] Yunhong Jiang,[b] Jun Li,[c] Chun Li,[d] Warispreet Singh*[b] and Meilan Huang*[a]

Cytochrome P450 cinnamate 4-hydroxylase (C4H) is a pivotal enzyme in the phenylpropanoid pathway, playing a critical role in regulating lignin biosynthesis in plants. In contrast to the hydroxylation reactions catalyzed by human P450 enzymes, which have been extensively studied, the mechanistic understanding of plant P450-mediated hydroxylation of aromatic substrates remains limited. In this study, using comprehensive atomistic molecular dynamics (MD) simulations, we elucidated the binding pose of the native substrate *trans*-cinnamic acid and identified key residues contributing to the substrate specificity of the enzyme, which include Arg213 and a conserved hydrophobic pocket comprising Val118, Phe119, Val301, Ala302, Ile367 and Phe484. Additionally, we investigated the catalytic mechanism using hybrid quantum mechanics/molecular mechanics (QM/MM) calculations, evaluating all plausible C4H-catalysed pathways for aromatic hydroxylation. Our results reveal that among all investigated mechanisms, the most favourable pathway involves direct hydroxylation *via* electrophilic attack coupled with a proton shuttle. These findings provide valuable insights into the catalytic mechanism of C4H, which would pave the way for modifying lignin biosynthesis to regulate various lignin contents in plants, unlocking its potential applications in sustainable bioremediation and biomanufacturing.

Received 25th April 2025,
Accepted 18th September 2025

DOI: 10.1039/d5cy00502g

rsc.li/catalysis

Introduction

The valorization of lignin, the most abundant aromatic biopolymer in plant biomass, represents a promising strategy for advancing sustainable biorefineries and reducing dependence on fossil-based resources.[1] By converting lignin into high-value chemicals and renewable fuels, lignin bioconversion supports both environmental sustainability and economic viability. However, the structural complexity and inherent resistance of lignin to enzymatic degradation pose significant challenges[2–4] to its efficient processing at an industrial scale. A key regulatory step in lignin biosynthesis is catalyzed by cinnamate 4-hydroxylase (C4H), a cytochrome P450 monooxygenase of the CYP73A family, which catalyses an early and essential step in the phenylpropanoid pathway.[5] Beyond its role in lignin monomer biosynthesis,[5–11] C4H also contributes to the production of diverse secondary metabolites critical for plant development, defence and environmental adaptation. Given its central role in metabolism and lignin deposition, C4H has emerged as an attractive target for engineering lignin composition to enhance biomass utilization.[12–16]

Despite its importance, the industrial exploitation of C4H remains limited. The enzyme exhibits strict substrate specificity,[17] often excluding closely related analogues or displaying significantly reduced catalytic efficiency toward them. This narrow specificity constrains its utility in lignin depolymerization, where a broad array of chemically diverse intermediates is present. Overcome these limitations requires a comprehensive understanding of its active site architecture, substrate recognition and hydroxylation mechanism. Such insights are essential for regulating the lignin contents for industrial lignin valorisation, enabling the selective transformation of lignin-derived compounds while preserving vital agronomic traits such as plant robustness and resistance.

The catalytic mechanism of C(sp^3)–H bond hydroxylation by human P450 enzymes (*e.g.* CYP1A1,[18] CYP2C9 (ref. 19) and CYP3A4 (ref. 20)), which are crucial for drug metabolism and display broad substrate spectra has been extensively characterized. Such hydroxylation typically proceeds *via* hydrogen atom abstraction (HAA) by the iron(IV)-oxo porphyrin cation radical complex (Scheme S1). By contrast,

[a] *Department of Chemistry & Chemical Engineering, Queen's University, Belfast, BT9 5AG, UK. E-mail: m.huang@qub.ac.uk*
[b] *Department of Applied Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK. E-mail: w.singh@northumbria.ac.uk*
[c] *School of Chemistry and Chemical Engineering, Beijing Institute of Technology, Beijing, 100081, China*
[d] *Department of Chemical Engineering, Tsinghua University, Beijing 100084, China*

plant P450s are generally involved in specialized metabolic pathways and often exhibit narrower substrate specificity. Unlike the well-studied C(sp^3)–H bond hydroxylation mechanism in human P450 enzymes, detailed mechanistic and structural insights into plant P450-catalyzed hydroxylation of aromatic C(sp^2)–H bonds, especially in lignin-related biosynthesis pathways by C4H, which catalyzes the regioselective hydroxylation of aromatic substrates, remain comparatively limited.

In this work, we investigated the hydroxylation of aromatic substrate *trans*-cinnamic acid (TCA), the natural substrate of plant lignin biosynthesis enzyme P450 C4H, which plays a central role in lignin biosynthesis. Using QM/MM calculations, we examined four possible pathways for the C4H-catalyzed hydroxylation mechanism. This study provides molecular-level insights into the catalytic structure and mechanism of C4H, laying the foundation for regulating carbon flux in lignin biosynthesis, and enabling engineering of plants with optimized lignin content for lignin valorization.

Methodology

Protein preparation

The initial structure of cinnamate 4-hydroxylase (C4H, CYP73A) was derived from the crystal structure of the enzyme from *Sorghum bicolor* (PDB ID: 6VBY).[5] The protonation states of protein residues were determined using H++ server[21] at pH 7. The high-valent iron (IV)-oxo porphyrin-π-cation radical species, commonly known as compound I (Cpd I), was constructed and employed in the subsequent molecular docking, molecular dynamics (MD) simulations and QM/MM study. The oxidation reaction of the substrate was initiated from the hexacoordinated-Cpd I, with oxo and Cys443 serving as the axial ligands. The force field parameters of the Cpd I was taken from a previous study.[22] The Amber FF14SB[23] and GAFF force fields[24] were used in MD simulations. The parameters for the *trans*-cinnamic acid ligand were prepared using GAFF implemented in Antechamber[25] and the atomic charges for each ligand were calculated using the RESP method at the HF/6-31G* level of theory using Gaussian 16.[26]

Molecular docking

The natural substrate *trans*-cinnamic acid was docked to the MD simulated structure of the C4H enzyme in the close vicinity of Cpd I. Molecular docking was performed using the AutoDock 4.2 (ref. 27) suite with the LGA and the standard free energy scoring function.[28] The grid centre was set to x = 51.643 Å, y = 44.393 Å, z = 49.992 Å. A total of 300 LGA runs were carried out for each enzyme–substrate complex. The population was 300, the maximum number of generations and energy evaluations were 27 000 and 2 500 000, respectively.

MD simulations

The C4H enzyme was solvated in a truncated octahedral box with TIP3P[29] water with a cut-off distance of 15 Å. The periodic boundary conditions were employed throughout all the simulations. Each system was subjected to five successive steps of energy minimization. The first four steps consisted of 2500 cycles of steepest descent followed by 7500 cycles of the conjugate gradient minimization. These minimizations were performed as follows: first all the atoms of the protein and ligand were restrained with a harmonic force constant of 100 kcal mol^{-1} $Å^{-2}$. Then, only the protein backbone atoms and the ligand heavy atoms were restrained with a force constant gradually reduced to 100, 50, 25 and 10 kcal mol^{-1} $Å^{-2}$, respectively.

The systems were then gradually heated from 0 to 298.15 K over 20 ps under constant volume, with the protein and ligand atoms restrained using a harmonic force constant of 10 kcal mol^{-1} $Å^{-2}$. This was followed by three 100 ps equilibration steps at constant pressure (1 bar) and constant temperature, during which restraints on the protein backbone atoms and the ligand heavy atoms were gradually reduced from 25, 10, 5 and 1 kcal mol^{-1} $Å^{-2}$. Finally, 100 ps unrestrained MD simulation was performed.

Following the equilibration, four 500 ns MD replica runs with a time step of 2 fs were performed for each complex using the GPU-accelerated PMEMD in Amber20.[30] Periodic boundary conditions were applied and long-range interactions were treated using the particle mesh Ewald method.[31] The SHAKE[32] algorithm was employed to constrain the bonds involving the hydrogen atoms. Temperature control was maintained using the Langevin[33] thermostat with a collision frequency of 1.0 ps^{-1}. For each MD simulation, RMSD (Fig. S1), the distances between the *para*-carbon of the substrate and the iron oxo oxygen (Fig. S2), and cluster analysis were evaluated using cpptraj,[34] Jupyter notebook[35] and VMD.[36]

QM/MM calculations

All the QM/MM calculations started from the Cpd I, the main oxidant in P450 mediated reactions. The snapshots for the QM/MM calculations were obtained from the equilibrated MD trajectory from the cluster analysis. The most populated structure from the cluster analysis were then subjected to 1250 steps of steepest descend and 1250 steps of conjugate gradient minimization using Amber20.[30] The reaction profiles of C4H in oxidizing the substrate *trans*-cinnamic acid was investigated using QM/MM calculations implemented in ChemShell.[37] The QM region was calculated with ORCA 4.2.0 (ref. 38) while the MM region was treated using the AMBER force field within DL_POLY.[39] The effect of the protein environment on the polarization of the QM wavefunction was considered using electronic embedding scheme.[40] The treatment of cysteine in the QM region follows the approach adopted in previous studies.[41–45] The QM region consists of the Cpd I molecule with truncated propionate side chains,

substrate and the cysteine residue truncated at the Cβ atom (Scheme S2).

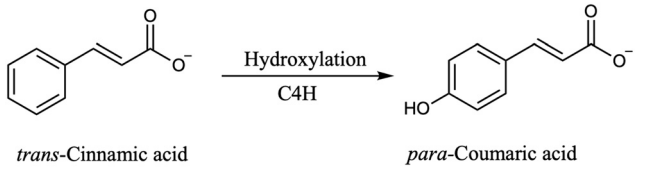
The residues within 10 Å of Cpd I, including *trans*-cinnamic acid and water molecules, were allowed to move freely during the geometry optimization, while the rest of the system was kept frozen. The water shell within 4 Å of the protein or within 20 Å of the QM atoms was retained. The hydrogen link atoms were used to cap the dangling bonds at the QM/MM boundary. The reaction coordinate was defined as the distance between the oxygen atom of the Fe(IV)=O in Cpd I and the *para*-carbon of the *trans*-cinnamic acid. The TS structures were obtained by performing a relaxed PES scan along the reaction coordinate with an increment of 0.1 Å.

All the QM calculations were performed with DFT using UB3LYP functional,[46,47] with D3 dispersion correction and BJ damping,[48] and def2-SVP basis set with iron in doublet spin state. The RIJCOSX[48] approximation was used to speed up the QM calculations. The highest energy point on the PES was fully optimized as TS using the dimer method implemented in the DL-FIND module of Chemshell. TS structures were validated by frequency calculations. The final reported energies include the electronic energies with single-point energy corrections using def2-TZVP basis set, combined with zero-point energy (ZPE) corrections (Tables S1–S3).

Results and discussion

Regioselectivity with the aromatic substrates

C4H catalyzes the hydroxylation of its natural substrate *trans*-cinnamic acid to *para*-coumaric acid (Scheme 1). The 3D structure of C4H in complex with *trans*-cinnamic acid was attained by molecular docking using the crystal structure of *Sorghum bicolor* C4H and further refined by MD and QM/MM simulations. MD simulations and cluster analysis of the complex showed that the phenylpropene group of *trans*-cinnamic acid is nested in a hydrophobic pocket constructed by conserved residues including Phe119, Val301, Ala302 and Phe484 (Fig. 1), which position the substrate near Cpd I. The carboxylic acid group of *trans*-cinnamic acid is stabilized by electrostatic interactions with Arg213 through ionic interactions. MD simulations show that the carbon atom at the *para*-position of *trans*-cinnamic acid is located in close proximity (3.1 Å) to the iron-oxo group. This orientation is accordance with the preferred regioselectivity of the C4H enzyme (Fig. 1).



trans-Cinnamic acid *para*-Coumaric acid

Scheme 1 Hydroxylation of *trans*-cinnamic acid to *para*-coumaric acid catalyzed by the P450 enzyme C4H.

In our previous studies on $GcoA_{P450}$ (ref. 43) and $AgcA_{P450}$ (ref. 42) enzymes, we demonstrated that water is crucial for the catalysis as it lowers the activation energy barrier by stabilizing the transition state through electrostatic interactions. Interestingly, during the MD simulations, a water molecule consistently remained near Cpd I in C4H (Fig. S3), forming hydrogen bonds with both the iron-oxo species and the substrate (Fig. 1 and S4 and S5).

The distribution of the mean distances between Cpd I and the *para* or *meta*-carbon atoms of the substrate during the MD simulations were measured (Fig. 2). The most frequent distance between iron-oxo species and the *para*-carbon (~3.76 Å) is shorter than those with the *meta*-carbons (4.41 Å and 4.29 Å, respectively), indicating a more stable and favourable interaction of the *para*-carbon maintains with the iron-oxo species and is consistent with the regioselectivity of the C4H enzyme.

Aromatic hydroxylation mechanism

Given the critical role of C4H in the phenylpropanoid metabolism pathway and its influence on lignin content in plants, it is essential to understand the aromatic hydroxylation mechanism. Iron-oxo of Cpd I can form a covalent bond with the aromatic substrate yielding an σ-complex intermediate (**IM1**). Following the formation of σ-complex, the hydroxylation of the aromatic substrate may proceed *via* different pathways: electrophilic attack (pathway A), hydroxylation through ketone (pathway B) and epoxidation (pathway C), or hydrogen atom abstraction directly from the substrate to Cpd I (pathway D) (Scheme 2).

In direct electrophilic attack pathway (pathway A), a proton-shuttle mechanism transfers the *ipso*-proton from the substrate to a nitrogen atom of the porphyrin ring, followed by re-shuttle of the proton within the intermediate, yielding the phenol.[16,49,50] Alternatively, the reaction may proceed through the formation of a ketone as the initial intermediate (pathway B) or occur *via* an epoxide intermediate (pathway C).[49–51] The ring subsequently opens to form a ketone, which then undergoes tautomerization to yield the hydroxylated phenol product. The fourth possible mechanism (pathway D) involves a hydrogen atom abstraction process, in which the iron-oxo species in Cpd I directly abstracts a hydrogen from the substrate, followed by an OH rebound to the substrate.[42,43,52]

During MD simulations, the *para*-carbon is consistently positioned toward the iron oxo, indicating the preference for oxidation of the *para*-carbon (Fig. 2). The starting configurations for the QM/MM calculations were obtained from MD simulations; four representative structures were selected from cluster analysis based on the substrate's position in relation to the iron-oxo porphyrin centre. The selection criteria included a distance between the *para*-carbon and the oxygen (of iron-oxo species) of less than 4.0 Å, and a Fe–O–C angle between 110° and 130°, in accordance with the guidelines proposed by Lonsdale *et al.*[53]

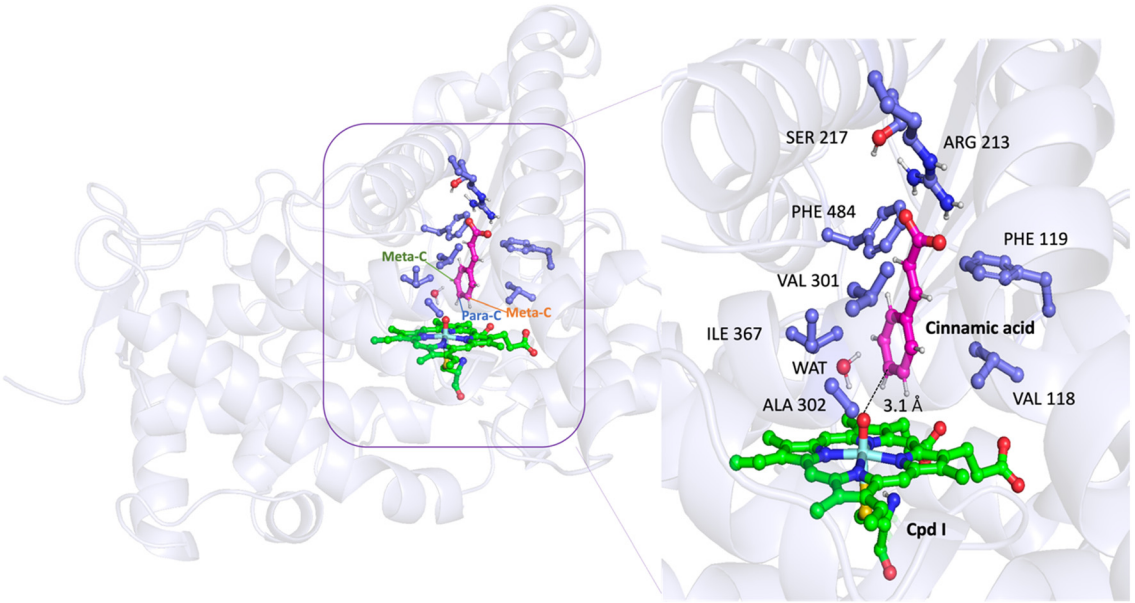


Fig. 1 The substrate (shown in coral) interacts with surrounding residues (shown in purple). The phenylpropene group of *trans*-cinnamic is surrounded by Val118, Phe119, Val301, Ala302, Ile367 and Phe484, while its carboxy group is stabilized by Arg213. The *meta* and *para*-carbon atoms of the substrate are labelled in the left panel. The protein is shown in light purple using the New Cartoon representation.

In P450 enzymes, the formation of the σ-adduct *via* the initial electrophilic attack on the π-system of the aromatic substrates[45] is the rate-limiting step in the hydroxylation reaction and is typically endothermic.[12,13,50,54] The energy barriers reported by QM/MM calculations generally range from 12 to 22 kcal mol^{-1},[13,14,45,54–56] although these values may vary depending on the substrate, reaction conditions and the computational methodology used.

We studied the formation of the σ-complex using QM/MM calculations. The reaction coordinate was defined as the distance between the oxygen atom of iron oxo in Cpd I and the *para*-carbon of the substrate. Relaxed PES scans along the reaction coordinate using QM/MM revealed reaction profiles with activation barriers for the electrophilic attack by the iron oxo species on the *para*-carbon ranging from 15.2 kcal mol^{-1} to 22.5 kcal mol^{-1}, based on the representative structures retrieved from four replica runs of MD simulations (Table 1). Since Arg213 stabilizes the substrate carboxylate, we also examined the effect of including this residue in the QM region and found the reaction profile would need to overcome a comparable barrier (Fig. S5), so in the remaining calculations, Arg213 is excluded from the QM region.

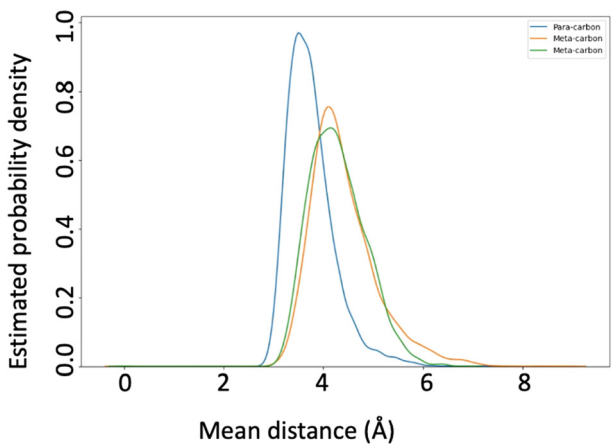
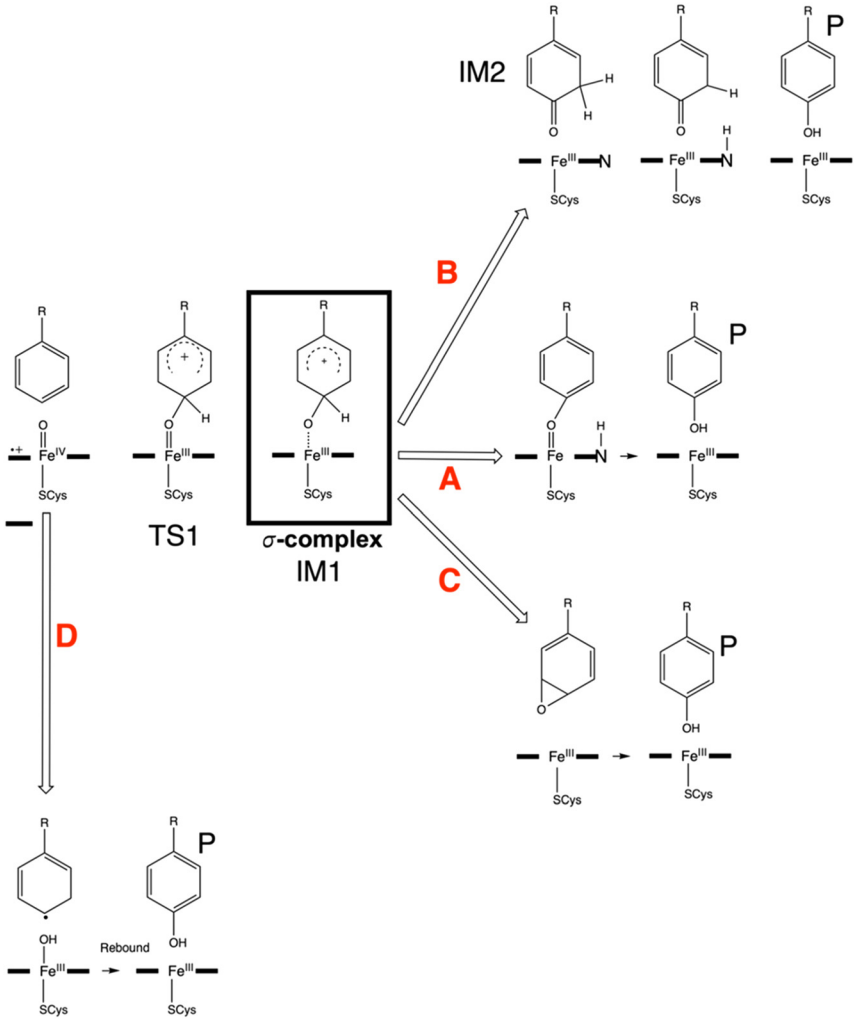


Fig. 2 Distributions of distances between the iron-oxo species and the aromatic carbons of *trans*-cinnamic acid: *para*-carbon (blue) and the two *meta*-carbons (orange and green).

The Boltzmann-weighted average barrier obtained from the QM/MM calculations is 16.03 kcal mol^{-1}, which closely aligns with the experimentally determined activation energy barrier of 18.9 kcal mol^{-1}.[5] Additionally, our previous study on P450$_{\mathrm{GcoA}}$ demonstrated the presence of water molecules in the enzyme's active site can influence these energy barriers, underscoring the crucial role of water in these biochemical reactions.[43] Notably, analysis of the geometry of the water molecule in the four starting structures revealed a water molecule positioned near the reactant, forming a H-bond with the iron-oxo oxygen in the transition state (TS). This interaction contributes to a lower energy barrier by stabilizing the substrate within the enzyme complex (Tables 1 and S4 and Fig. S6).

Direct hydroxylation *via* proton shuttle (pathway A)

Following the electrophilic attack and the formation of the σ-complex intermediate (**IM1**), our calculations disclosed a proton shuttle mechanism involving the nitrogen atom in the porphyrin ring. The *ipso*-proton is transferred from the *para*-carbon of the substrate to a nitrogen atom in the porphyrin ring, forming a transient



Scheme 2 Possible reaction pathways for the hydroxylation of aromatic substrate leading to the phenol product (**P**).

intermediate that subsequently facilitates the proton shuttling, yielding *p*-coumaric acid, the final product of this reaction (Fig. 3) (pathway A). This observation is consistent with previous studies by Shaik *et al.*[56] and Bathelt *et al.*,[12,13,54] which suggested a N-protonated porphyrin intermediate is involved in the direct formation of ketone and phenol, and the reaction proceeds *via* a proton relay from the site of electrophilic attack on the aromatic carbon to either the oxygen atom or the adjacent carbon.[12,13,54,56]

Table 1 QM/MM calculated activation barrier (kcal mol^{-1}) for the electrophilic attack by the iron-oxo species on the *para*-carbon. Also listed are the O–C distances [Å] and angle ϕ [degree] between the porphyrin plane and the substrate in the transition state (TS)

	d(O–C) Å	(Fe–O–C) ϕ (°)	Energy barrier (kcal mol^{-1})
Replica 1	1.92	122.46	15.2
Replica 2	1.85	123.8	22.5
Replica 3	1.93	121.58	22.1
Replica 4	1.92	123.54	20.1
Boltzmann-weighted average			16.03

The energy profile shows an activation barrier of 15.2 kcal mol^{-1}, corresponding to a transition state **TS1**, which leads to the formation of σ-complex intermediate I (**IM1**) at 7.96 kcal mol^{-1}. This is followed by a second activation barrier (**TS2**) of 6.73 kcal mol^{-1}, yielding the final product (**P**) with a significantly exergonic energy of −60.23 kcal mol^{-1}. These results indicate the reaction pathway is highly feasible and thermodynamically favourable (Fig. 3B).

These calculated activation barriers and reaction energies are comparable to those reported for aromatic hydroxylation reactions in previous studies. For example, the QM/MM calculated activation barriers for benzene hydroxylation in CYP2C9 for the addition of compound I to benzene range from 18.1 to 21.7 kcal mol^{-1}.[13] The activation barrier for the transition state in our aromatic hydroxylation is 15.2 kcal mol^{-1}, consistent with those reported in earlier from previous DFT studies using UB3LYP functional for human P450 enzymes, where the activation barriers for C–H hydroxylation reactions ranged from 13.8 to 19.9 kcal mol^{-1}.[14,19,20,54,57,58] The energy of the intermediate **IM1** is 7.96 kcal mol^{-1}, also aligning with the

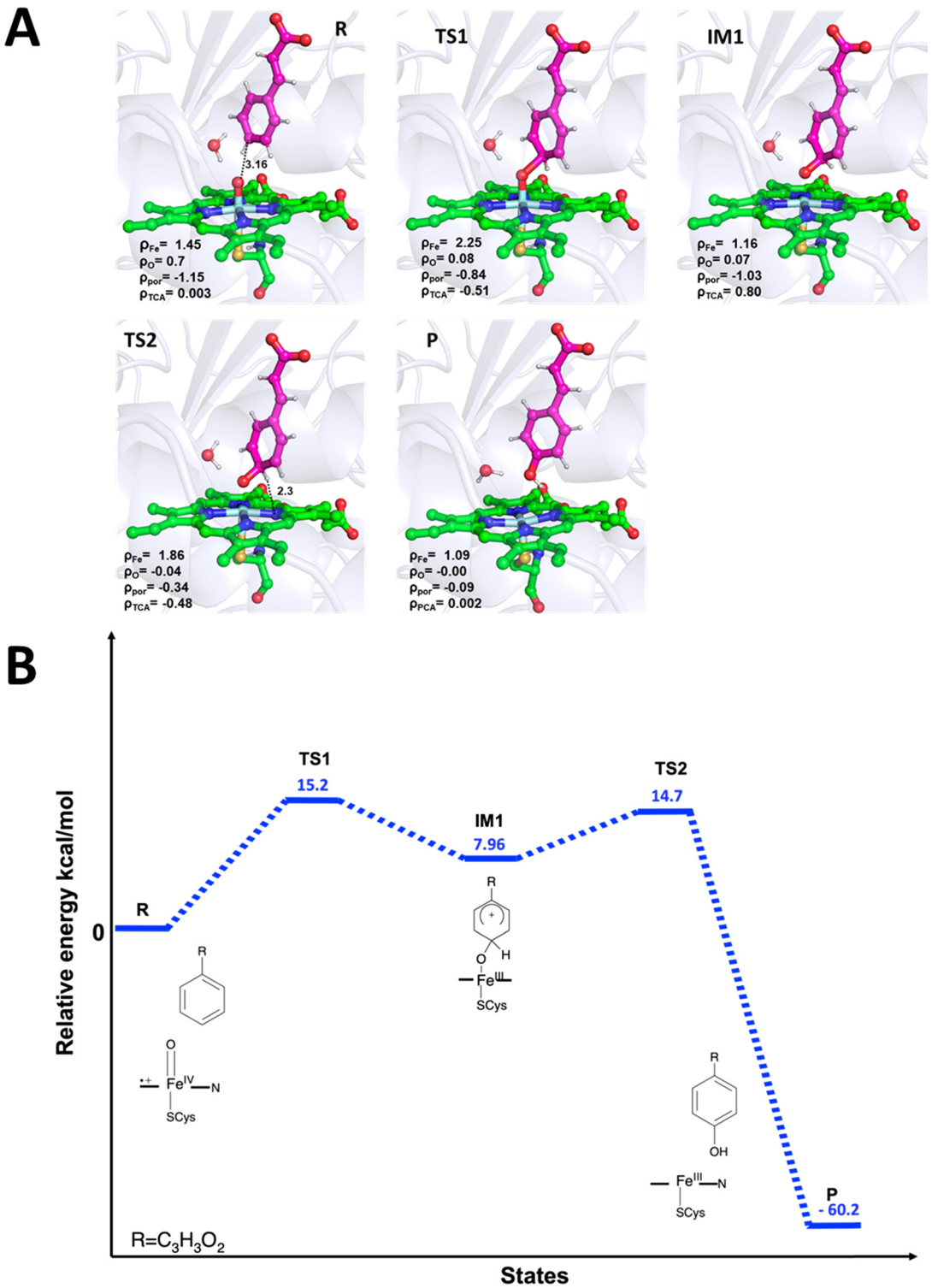


Fig. 3 (A) QM/MM optimized structures of the reaction species along pathway A, which involves the electrophilic attack of the substrate by compound I and proton shuttle mechanism. The C4H enzyme is shown in light blue using the New Cartoon presentation, while Cpd I, the substrate and the nearby water are shown in CPK mode. (B) QM/MM reaction profile corresponding to pathway A. All the calculations were performed at doublet spin $S = \frac{1}{2}$.

previously reported intermediate energies, which range from 4.9 to 17.8 kcal mol^{-1} depending on the specific reaction and computational model.[14,19,20,54,58] The second activation barrier associated with hydride transfer is 6.73 kcal mol^{-1}, which is in line with the previous reported values ranging from barrierless to 17.8 kcal mol^{-1}.[54] The energy of the final

phenol product is −60.23 kcal mol^{-1}, reflecting the highly exothermic nature of aromatic hydroxylation reactions catalysed by human P450 enzymes that yield phenol products.[14,20,54,58] Overall, our computational results show good agreement with the experimentally measured activation energy for the C4H enzyme, which is 18.9 kcal mol^{-1}.[5]

Hydroxylation *via* a ketone intermediate (pathway B)

In addition to the direct hydroxylation involving proton shuttle *via* the N-protonated porphyrin, the reaction may alternatively proceed *via* a spontaneous proton transfer from the *para*-position carbon to the *meta*-position carbon,

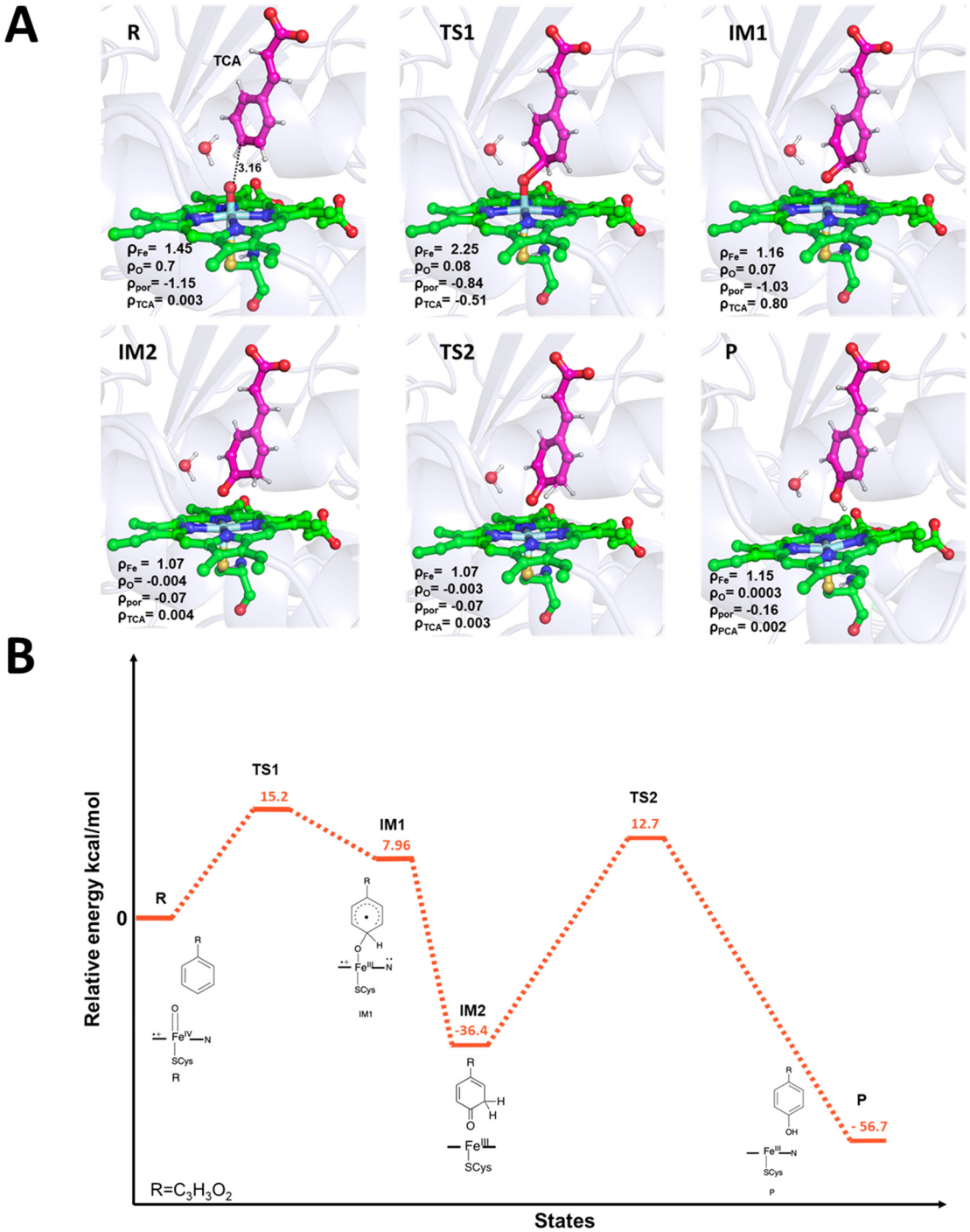


Fig. 4 (A) QM/MM optimized structures along the reaction pathway for the electrophilic attack by compound I (Cpd I), proceeding through the formation of a ketone intermediate (pathway B). The C4H enzyme is shown in light blue using New Cartoon representation, while Cpd I, the substrate and a nearby water are represented in CPK mode. (B) QM/MM reaction energy profile for hydroxylation *via* pathway B. All the calculations were performed at doublet spin ($S = \frac{1}{2}$).

forming a ketone intermediate II. Subsequently, the proton at the *meta*-position is transferred to the porphyrin ring nitrogen and ultimately shuttled to the oxygen at the *para*-position, giving the phenol product (pathway B, Fig. 4).

The first step in pathway B corresponding to the formation of the σ-complex intermediate I (**IM1**), is identical to that in pathway A, featuring an activation barrier of 15.2 kcal mol^{-1} for the transition state (**TS1**). The **IM1** at an energy of 7.96 kcal mol^{-1} then spontaneously transformed into the ketone intermediate II (**IM2**) at an energy of −36.4 kcal mol^{-1}. The reaction then proceeds through a second transition state (**TS2**) with a high activation barrier of 49.1 kcal mol^{-1} to yield the final product (**P**) at −56.7 kcal mol^{-1}. The reaction profile indicates that the formation of ketone intermediate **IM2** is facile and consistent with previous DFT studies.[14,20,54,58] However, once the ketone is formed, the reaction becomes energetically prohibitive due to the steep barrier of 49.1 kcal mol^{-1}

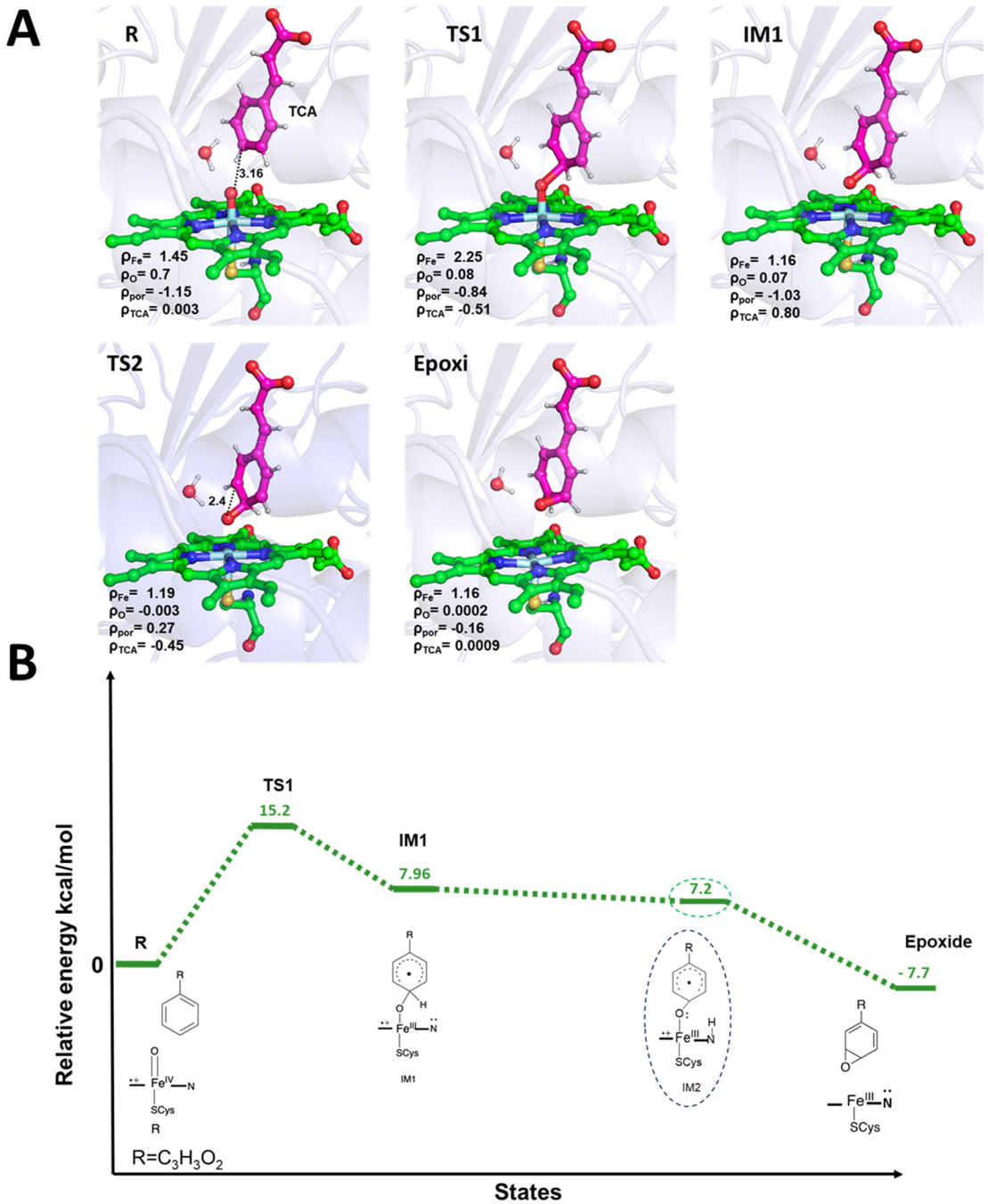


Fig. 5 (A) QM/MM optimized structures along the reaction pathway for hydroxylation through the epoxide intermediate (pathway C). The C4H enzyme is shown in light purple using the New Cartoon representation, while compound I, substrate and a nearby water are shown in CPK mode. (B) QM/MM reaction energy profile for hydroxylation through pathway C. All the calculations were performed at doublet spin ($S = \frac{1}{2}$).

required to reach the final product, rendering pathway B (Fig. 4B) unlikely for enzymatic catalysis. Interestingly, the conversion from the ketone to the phenol has been previously reported *via* a facile nonenzymatic process under physiological conditions.[20,59]

Hydroxylation through epoxide (pathway C)

Phenol formation was initially proposed to occur *via* the formation of an arene oxide (epoxide), either through a concerted single-step mechanism or *via* a stepwise process.[14,51,54] In the stepwise process, an σ-complex intermediate (**IM1**) is first formed, followed by intramolecular ring closure to generate the arene oxide intermediate. The epoxide then undergoes ring opening through nucleophilic attack, eventually giving the phenol product (pathway C, Fig. 5).

The initial step of pathway C mirrors those of pathways A and B, proceeding *via* the transition state (**TS1**) with an activation barrier of 15.2 kcal mol^{-1}. Following the formation of intermediate I (**IM1**) at 7.96 kcal mol^{-1}, spontaneously yielding the epoxide intermediate at −7.7 kcal mol^{-1} (Fig. 5B) *via* intramolecular ring closure without an additional barrier. An **IM2** may be formed by the N-ring protonation from **IM1**, although this state is not stable and only observed in few replica structures. This energetically favourable profile indicates that pathway C is both feasible and highly plausible. The low activation barrier for ring closure and barrierless formation of the epoxide are consistent with previous studies on aromatic hydroxylation.[13,14,20,54,58] However, additional steps are required to complete the reaction, including a non-enzymatic nucleophilic attack that opens the ring to yield the final product phenol.[20]

These findings are consistent with earlier density functional theory (DFT) studies,[12–14,45,54,56] which suggest that aromatic hydroxylation is initiated by an electrophilic attack. This mechanism, often involving carbocation intermediates, aligns with the dehydrogenase–oxidase activity commonly observed in cytochrome P450 enzymes.[14,54] Moreover, previous studies reported relatively low energy barriers for the subsequent rearrangement of these intermediates into epoxides, ketones or phenols, underpinning the pivotal role of electrophilic attack in P450-catalyzed aromatic hydroxylation.[45]

Herein, we investigated three possible pathways for the aromatic hydroxylation catalysed by P450 C4H, which proceed through the formation of the common σ-complex intermediate **IM1** (Fig. 6). Pathway A–C diverges following the formation of **IM1**. Among these, pathway A exhibits an overall feasible reaction profile for direct enzymatic conversion to the final phenol product. Pathway B, which proceeds *via* a ketone intermediate, shows favourable formation of the ketone intermediate, however, it requires a non-enzymatic tautomerization in solution to produce the phenol. Similarly,

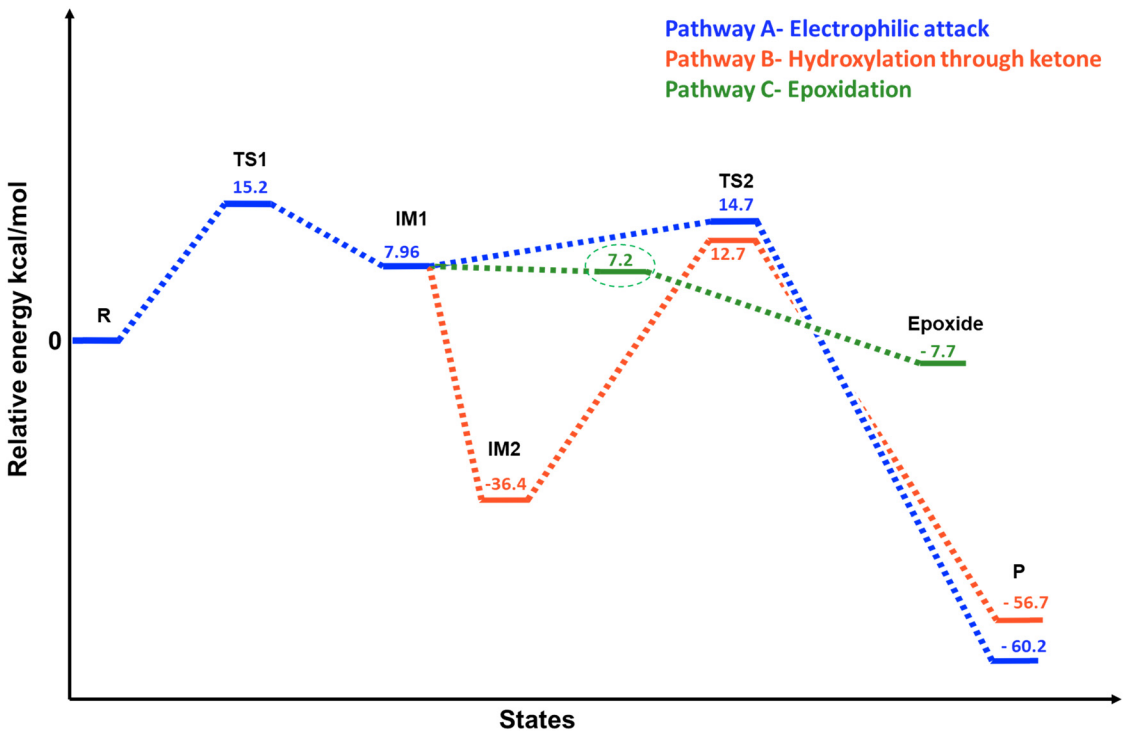


Fig. 6 Overall QM/MM energies profiles of pathways A–C reveal a shared initial step involving the formation of the σ-complex intermediate I (**IM1**), after which the reaction diverges into distinct rearrangement pathways. Pathway A proceeds through a proton shuttle mechanism, leading directly to the phenol product. Pathway B involves the formation of a ketone intermediate II (**IM2**) but requires overcoming a second high barrier of 49 kcal mol^{-1} to reach the final product. In pathway C, the reaction proceeds *via* spontaneous formation of an epoxide intermediate, however, as with pathway B, additional non-enzymatic steps are required to complete the conversion to the final phenol product.

pathway C, which involves an epoxide intermediate, appears energetically feasible and may represent the most prevalent pathway. Nonetheless, like pathway B, it also depends on additional non-enzymatic steps, specifically, nucleophilic ring opening to complete the transformation to the final phenol product.

Hydrogen atom abstraction (pathway D)

An alternative mechanism initiating directly from the iron-oxo species was also examined (pathway D). In this pathway, a hydrogen is abstracted from the *para*-position carbon of the aromatic substrate by the iron-oxo complex, followed by rebound of the hydroxyl group to the same carbon, yielding the phenol product. Although the hydrogen atom abstraction (HAA) mechanism is commonly observed in aliphatic hydroxylation by P450 enzymes,[12,13,18,42,43,45,50,54,56] it is not the case for aromatic hydroxylation: our calculations revealed that the activation barrier for the initial C–H bond cleavage in the HAA step exceeds 30 kcal mol^{-1} (Fig. S7), indicating this pathway is energetically unfavourable. As a result, further investigation into this mechanism was not pursued.

Conclusion

Cytochrome P450 cinnamate-4-hydroxylase (CYP450 C4H) is a key enzyme involved in regulating lignin content in plants. Compared to the human cytochrome systems, the mechanism of the plant P450 C4H enzyme is relatively less studied. Through comprehensive molecular dynamics (MD) simulations, we identified the binding mode of aromatic *trans*-cinnamic acid substrate within the active site of CYP450 C4H. The enzyme's high substrate specificity is attributed to several critical residues, including Arg213, which forms electrostatic interactions with the substrate's carboxylic group and a set of conserved hydrophobic residues (including Val118, Phe119, Val301, Ala302, Ile367 and Phe484) that stabilize the substrate in close proximity to the catalytic species compound I.

To elucidate the catalytic mechanism, we evaluated all plausible pathways for the aromatic hydroxylation catalyzed by CYP450 C4H. While the mechanism is inherently complex, our results indicate that direct hydroxylation *via* electrophilic attack followed by proton shuttle, is the most favourable pathway. Notably, it is the only fully enzymatic route among those examined.

The computational study sheds light on the substrate binding and catalytic mechanism of C4H, an important yet underexploited biocatalyst crucial for lignin biosynthesis in plants. These insights would set basis for tuning its activity and specificity to redirect the metabolic flux toward value-added products such as biofuels.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the supplementary information (SI).

Supplementary information: the SI includes electronic energies for the species along the reaction pathways, with ZPE corrections obtained from QM/MM calculations, the imaginary frequencies of the transition states, Cartesian coordinates from the optimized geometries, and the reaction profile on different iron spin state (double and quartet). See DOI: **https://doi.org/10.1039/d5cy00502g**.

Acknowledgements

The authors are grateful for the computing resources from the Kelvin 2 high-performance computing resource on NI-HPC funded by EPSRC (EP/T022175). YJ acknowledge the support for Leverhulme Trust (grant: RPG-2022-177). RRB would like to acknowledge EPSRC (grant: EP/Y002482/1) for the support.

References

1 F. Brienza, D. Cannella, D. Montesdeoca, I. Cybulska and D. P. Debecker, *RSC Sustainability*, 2024, **2**, 37–90.
2 C. Li, C. Chen, X. Wu, C. W. Tsang, J. Mou, J. Yan, Y. Liu and C. S. K. Lin, *Bioresour. Technol.*, 2019, **291**, 121898.
3 J. C. Chan, M. Paice and X. Zhang, *ChemCatChem*, 2020, **12**, 401–425.
4 T. D. H. Bugg, *Chem. Commun.*, 2024, **60**, 804–814.
5 B. Zhang, K. M. Lewis, A. Abril, D. R. Davydov, W. Vermerris, S. E. Sattler and C. Kang, *Plant Physiol.*, 2020, **183**, 957–973.
6 V. H. Salvador, R. B. Lima, W. D. dos Santos, A. R. Soares, P. A. F. Böhm, R. Marchiosi, M. de L. L. Ferrarese and O. Ferrarese-Filho, *PLoS One*, 2013, **8**, 69105.
7 Y. Hou, Y. Wang, X. Liu, N. Ahmad, N. Wang, L. Jin, N. Yao and X. Liu, *Int. J. Mol. Sci.*, 2023, **11**, 5393.
8 V. J. H. Sewalt, W. Ni, J. W. Blount, H. G. Jung, S. A. Masoud, P. A. Howles, C. Lamb and R. A. Dixon, *Plant Physiol.*, 1997, **115**, 41.
9 A. O'Connell, K. Holt, J. Piquemal, J. Grima-Pettenati, A. Boudet, B. Pollet, C. Lapierre, M. Petit-Conil, W. Schuch and C. Halpin, *Transgenic Res.*, 2002, **11**, 495–503.
10 R. W. Sykes, E. L. Gjersing, K. Foutz, W. H. Rottmann, S. A. Kuhn, C. E. Foster, A. Ziebell, G. B. Turner, S. R. Decker, M. A. W. Hinchee and M. F. Davis, *Biotechnol. Biofuels*, 2015, **8**, 128.
11 R. Vanholme, K. Morreel, C. Darrah, P. Oyarce, J. H. Grabber, J. Ralph and W. Boerjan, *New Phytol.*, 2012, **196**, 978–1000.
12 C. M. Bathelt, L. Ridder, A. J. Mulholland and J. N. Harvey, *J. Am. Chem. Soc.*, 2003, **125**, 15004–15005.
13 C. M. Bathelt, A. J. Mulholland and J. N. Harvey, *J. Phys. Chem. A*, 2008, **112**, 13149–13156.
14 S. Shaik, D. Kumar, S. P. de Visser, A. Altun and W. Thiel, *Chem. Rev.*, 2005, **105**, 2279–2328.
15 K. H. Mitchell, C. E. Rogge, T. Gierahn and B. G. Fox, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 3784–3789.
16 D. Kumar, G. N. Sastry and S. P. De Visser, *J. Phys. Chem. B*, 2012, **116**, 718–730.

17 L. Pollegioni, F. Tonin and E. Rosini, *FEBS J.*, 2015, **282**, 1190–1213.
18 T. Mokkawes, Z. Q. Lim and S. P. De Visser, *J. Phys. Chem. B*, 2022, **126**, 9591–9606.
19 A. Hermano Sampaio Dias, R. Yadav, T. Mokkawes, A. Kumar, M. S. Skaf, C. V. Sastri, D. Kumar and S. P. de Visser, *Inorg. Chem.*, 2023, **62**, 2244–2256.
20 D. Yue and H. Hirao, *J. Chem. Inf. Model.*, 2023, **63**, 7826–7836.
21 J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath and A. Onufriev, *Nucleic Acids Res.*, 2005, **33**, W368.
22 K. Shahrokh, A. Orendt, G. S. Yost and T. E. Cheatham, *J. Comput. Chem.*, 2012, **33**, 119–133.
23 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
24 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
25 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.*, 2006, **25**, 247–260.
26 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, Revision C.01*, Gaussian, Inc., Wallingford CT, 2016.
27 G. M. Morris, H. Ruth, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785–2791.
28 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, *J. Comput. Chem.*, 1998, **19**, 1639–1662.
29 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
30 D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, L. Wilson, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York and P. A. Kollman, *AMBER 2020*, University of California, San Francisco, 2020.
31 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
32 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.
33 J. A. Lzaguirre, D. P. Catarello, J. M. Wozniak and R. D. Skeel, *J. Chem. Phys.*, 2001, **114**, 2090–2098.
34 D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
35 T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla and C. Willing, *Positioning and Power in Academic Publishing: Players, Agents and Agendas - Proceedings of the 20th International Conference on Electronic Publishing, ELPUB 2016*, 2016, pp. 87–90.
36 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
37 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, *J. Mol. Struct.: THEOCHEM*, 2003, **632**, 1–28.
38 F. Neese, *WIREs Comput. Mol. Sci.*, 2012, **2**, 73–78.
39 W. Smith and T. R. Forester, *J. Mol. Graphics*, 1996, **14**, 136–141.
40 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1986, **7**, 718–730.
41 S. F. G. Santos, R. R. Bommareddy, G. W. Black and W. Singh, *Front. Chem.*, 2024, **11**, 1327398.
42 S. F. G. Santos, R. R. Bommareddy, G. W. Black, W. Singh and M. Huang, *Catal. Sci. Technol.*, 2023, **13**, 2070–2079.
43 W. Singh, S. F. G. Santos, P. James, G. W. Black, M. Huang and K. D. Dubey, *ACS Omega*, 2022, **7**, 21109–21118.
44 C. Hui, W. Singh, D. Quinn, C. Li, T. S. Moody and M. Huang, *Phys. Chem. Chem. Phys.*, 2020, **22**, 21696–21706.
45 S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar and W. Thiel, *Chem. Rev.*, 2010, **110**, 949–1017.
46 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter*, 1988, **37**, 785–789.
47 B. Miehlich, A. Savin, H. Stoll and H. Preuss, *Chem. Phys. Lett.*, 1989, **157**, 200–206.
48 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.
49 J. Oláh, A. J. Mulholland and J. N. Harvey, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 6050–6055.
50 S. P. De Visser, *Adv. Inorg. Chem.*, 2012, **64**, 1–31.
51 W. F. Trager, *Comprehensive Medicinal Chemistry II*, 2007, vol. 5, pp. 87–132.
52 J. Paulo, T. Zaragoza, D. C. Cummins, M. Qadri, E. Mubarak, M. A. Siegler, S. P. De Visser and D. P. Goldberg, *Inorg. Chem.*, 2019, **58**, 16761.
53 R. Lonsdale, K. T. Houghton, J. Zurek, C. M. Bathelt, N. Foloppe, M. J. De Groot, J. N. Harvey and A. J. Mulholland, *J. Am. Chem. Soc.*, 2013, **135**, 8001–8015.
54 C. M. Bathelt, L. Ridder, A. J. Mulholland and J. N. Harvey, *Org. Biomol. Chem.*, 2004, **2**, 2998–3005.

55 Z. Hou, Y. Li, M. Zheng, X. Liu, Q. Zhang and W. Wang, *Ecotoxicol. Environ. Saf.*, 2023, **258**, 114964.
56 S. P. de Visser and S. Shaik, *J. Am. Chem. Soc.*, 2003, **125**, 7413–7424.
57 C. Yuan, Q. Ouyang, X. Wang, X. Li, H. Tan and G. Chen, *Inorg. Chem.*, 2021, **60**, 6433–6445.
58 Z. Gan, J. Feng, J. Yin, J. Huang, B. Wang and J. Z. H. Zhang, *ACS Catal.*, 2024, **14**, 16277–16286.
59 P. R. Ortiz de Montellano, in *Cytochrome P450: Structure, Mechanism, and Biochemistry*, ed. P. R. Ortiz de Montellano, Springer International Publishing, Cham, 2015, pp. 111–176.