

Cite this: *Chem. Sci.*, 2023, 14, 14316

All publication charges for this article have been paid for by the Royal Society of Chemistry

# Computationally guided bioengineering of the active site, substrate access pathway, and water channels of thermostable cytochrome P450, CYP175A1, for catalyzing the alkane hydroxylation reaction†

Mohd Taher,<sup>1</sup> Kshatresh Dutta Dubey<sup>2</sup> and Shyamalava Mazumdar<sup>1</sup>

Understanding structure–function relationships in proteins is pivotal in their development as industrial biocatalysts. In this regard, rational engineering of protein active site access pathways and various tunnels and channels plays a central role in designing competent enzymes with high stability and enhanced efficiency. Here, we report the rational evolution of a thermostable cytochrome P450, CYP175A1, to catalyze the C–H activation reaction of longer-chain alkanes. A strategy combining computational tools with experiments has shown that the substrate scope and enzymatic activity can be enhanced by rational engineering of certain important channels such as the substrate entry and water channels along with the active site of the enzyme. The evolved enzymes showed an improved catalytic rate for hexadecane hydroxylation with high regioselectivity. The Q67L/Y68F mutation showed binding of the substrate in the active site, water channel mutation L80F/V220T showed improved catalytic activity through the peroxide shunt pathway and substrate entry channel mutation W269F/I270A showed better substrate accessibility to the active pocket. All-atom MD simulations provided the rationale for the inactivity of the wild-type CYP175A1 for hexadecane hydroxylation and predicted the above hot-spot residues to enhance the activity. The reaction mechanism was studied by QM/MM calculations for enzyme–substrate complexes and reaction intermediates. Detailed thermal and thermodynamic stability of all the mutants were analyzed and the results showed that the evolved enzymes were thermally stable. The present strategy showed promising results, and insights gained from this work can be applied to the general enzymatic system to expand substrate scope and improve catalytic activity.

Received 5th June 2023  
Accepted 10th November 2023

DOI: 10.1039/d3sc02857g

rsc.li/chemical-science

## 1 Introduction

Engineering and development of enzymes with new or improved catalytic activities are crucial for the advancement of biocatalysts for chemical and pharmaceutical industries.<sup>1,2</sup> Protein engineering provides an eco-friendly and sustainable way to design new biocatalysts for promoting green chemistry. In recent years, advances in protein engineering strategies such as directed evolution by random mutagenesis and site-directed mutagenesis have paved the way to achieve this goal through the repurposing of natural enzymes. The directed evolution

technique is the most commonly used approach, which involves screening of the active variants from a large library of mutants.<sup>3</sup> Extensive studies are reported to design a ‘smart library’ of mutants to improve the efficiency of the directed evolution process.<sup>4,5</sup> A structure-guided rational protein design approach could provide an efficient method to produce engineered enzymes with a minimal screening procedure and provides a better understanding of the structure–function relationship in the protein.<sup>6</sup> Cytochrome P450s are amongst the most extensively studied enzymes and are considered as potential candidates for biotechnological applications.<sup>7,8</sup> The engineered cytochrome P450 enzymes were shown to be more catalytically active and were also capable of catalyzing ‘new to nature’ chemical reactions such as C–C coupling<sup>9</sup> and nitrogen insertion into an unactivated C–H bond.<sup>10</sup> The rational evolution of the enzymes, has mostly been focused on active site engineering using site-directed and saturation mutagenesis,<sup>11</sup> and also by using novel decoy molecules to expand the substrate scope.<sup>12,13</sup> Modification of the metal cofactor<sup>14,15</sup> along with active site mutation<sup>16,17</sup> has also been investigated to catalyze abiological

<sup>1</sup>Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400005, India. E-mail: chiraltaher@gmail.com; shyamal@tifr.res.in

<sup>2</sup>Department of Chemistry, School of Natural Science, Shiv Nadar Institution of Eminence Delhi-NCR, NH91, Tehsil Dadri, Greater Noida, Uttar Pradesh 201314, India. E-mail: kshatresh.dubey@snu.edu.in

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3sc02857g>



reactions by using this enzyme. In cytochrome P450s, the active site is deeply buried inside the core of the protein matrix and surrounded by a complex network of tunnels and channels which connect the active site with the protein surface. Tunnels in the cytochrome P450s serve as pathways for the substrate<sup>18</sup> and other essential entities such as H<sub>2</sub>O, protons, and products to enter or exit the active site<sup>19</sup> of the enzyme. These tunnels are lined with amino acid residues that are critical for substrate recognition and binding, as well as for controlling the flow of substrates and products through the enzyme.<sup>20</sup> Hence, a study of the tunnels would provide better insight into the protein structure, and bioengineering of these tunnels may offer an efficient approach for the design of engineered enzymes with improved catalytic activity.<sup>21</sup> The tunnels in cytochrome P450 enzymes are often highly specific and can discriminate between different substrates based on their size, shape, and chemical properties.<sup>21,22</sup> The tunnels often contain aromatic amino acids, which act as gatekeepers for the incoming substrate.<sup>23</sup> Improvement of the catalytic efficiency of cytochrome P450s and other enzymes by engineering some of these tunnels has been recently reported.<sup>24–26</sup> However, the combined effects of rational design of the substrate binding site along with modification of important tunnels to expand the substrate scope and to increase the catalytic efficiency have not yet been reported. In the present work, we have used computational tools such as molecular docking and molecular dynamics simulation tools to complement the experiments for the development of engineered enzymes for a very challenging C<sup>(sp<sup>3</sup>)</sup>–H activation reaction. Due to its importance at the industrial scale, several researchers have attempted to develop enzymes for alkane hydroxylation.<sup>27–30</sup> Arnold and co-workers developed cytochrome P450-based alkane hydroxylating enzymes with a remarkably high catalytic activity using the directed evolution technique.<sup>31</sup> Using the decoy molecule strategy,<sup>32</sup> Reetz and co-workers showed that CYP102A1 from *Bacillus megaterium*, also known as P450<sub>BM3</sub> can be converted into an alkane hydroxylating enzyme.<sup>33</sup> These findings were very promising, giving rise to high regio- and enantioselectivity, but still had several limitations such as low thermal stability of the enzyme, low tolerance towards organic solvents, and the use of the expensive electron donor NADPH. In a report, Cirino *et al.*<sup>34</sup> engineered the P450<sub>BM3</sub> enzyme to use cheaper and greener oxidants such as H<sub>2</sub>O<sub>2</sub> for catalyzing the monooxygenation reaction as an alternate source of expensive NADPH, but the catalytic activity was very poor due to low coupling efficiency. To address these limitations, we have engineered the active site and two important tunnels *i.e.* the substrate entry channel and water channel in a thermostable cytochrome P450 from *Thermus thermophilus*, CYP175A1, to catalyze the alkane hydroxylation reaction through the peroxide shunt pathway using H<sub>2</sub>O<sub>2</sub>. CYP175A1 has a close sequence similarity with P450<sub>BM3</sub>. Like P450<sub>BM3</sub>, CYP175A1 has been shown to catalyze the fatty acid hydroxylation reaction<sup>35</sup> indicating that both enzymes have similar active pocket sizes and environments. Hence, we anticipate that CYP175A1 could be a thermostable analog for P450<sub>BM3</sub>. Wild-type CYP175A1 does not show any detectable catalytic activity for the alkane hydroxylation reaction. For proof of the concept,

we selected hexadecane as the alkane substrate to investigate the alkane hydroxylation activity in the variants of the enzyme. Hexadecane has a close structural resemblance with palmitoleic acid (hexadecane-9-enoic acid), which was previously shown to efficiently bind inside the active pocket of the wild-type enzyme with the long hydrophobic tail residing inside the active site.<sup>35</sup> We employed a simple molecular docking tool to understand the probable substrate orientations inside the active pocket of the enzyme and selected the lowest energy binding pose of the enzyme–substrate complex to identify the important residues (hotspots) in the substrate binding site for mutagenesis. This helped us create a virtual library of mutants, which was screened using molecular docking to identify the most suitable mutant. We have selected the most suitable mutant from the analysis of the docking results, and performed all-atom MD simulations on the substrate-bound structure of the mutant. The MD simulation data analysis helped us to further assess the stability of the enzyme–substrate complex in the presence of segmental dynamics. The MD simulations further suggested that there is a bottleneck at the substrate entry channel, and the water channel is also not properly formed at the active site of this mutant, which could result in lower activity. Tunnel analyses using CAVER<sup>36</sup> have also suggested the presence of a bottleneck at the substrate entry channel. Guided by the MD simulation data we made mutations to activate the ‘water channel’ inside the protein matrix and also modified the residues at the bottleneck of substrate entry. The design of the mutant enzymes based on the MD simulations and docking analysis was validated experimentally. The experimental results show good agreement with the computational results. Furthermore, QM/MM calculations were carried out on the mutant enzyme to understand the reaction mechanism and energetics. The results showed that the developed mutant follows the typical cytochrome P450 reaction mechanism for C–H activation.

## 2 Results and discussion

In contrast to P450<sub>BM3</sub>, which is widely used as a scaffold for various bioengineering studies,<sup>37–39</sup> the active site architecture of CYP175A1 is less elucidated. We, therefore, started our study to understand the active site topology of CYP175A1 vis-à-vis CYP102A1 (*i.e.*, P450<sub>BM3</sub>). In doing so, we first superimposed the crystal structures of CYP175A1 (PDB ID: 1N97)<sup>40</sup> with those of CYP102A1 (PDB ID: 1FAG).<sup>41</sup> Superimposition shows that both structures have almost identical tertiary structural folds with an RMSD of 1.07 Å (Fig. S1 in the ESI†). Cavity analysis using Pymol showed that the active site of CYP175A1 is reasonably large and is of U shape (Fig. S2 in the ESI†) geometry. Similar to cytochrome P450<sub>BM3</sub>, the active site of CYP175A1 contains mostly hydrophobic amino acids (Table S1 in the ESI†). These data suggested that longer-chain alkanes can easily be accommodated inside the active pocket of CYP175A1. However, screening of medium to long-chain alkanes for hydroxylation by wild-type CYP175A1 shows no detectable activity, and therefore, we aim to bioengineer the CYP175A1 enzyme for C–H hydroxylation using computational guided rational designing. We have taken

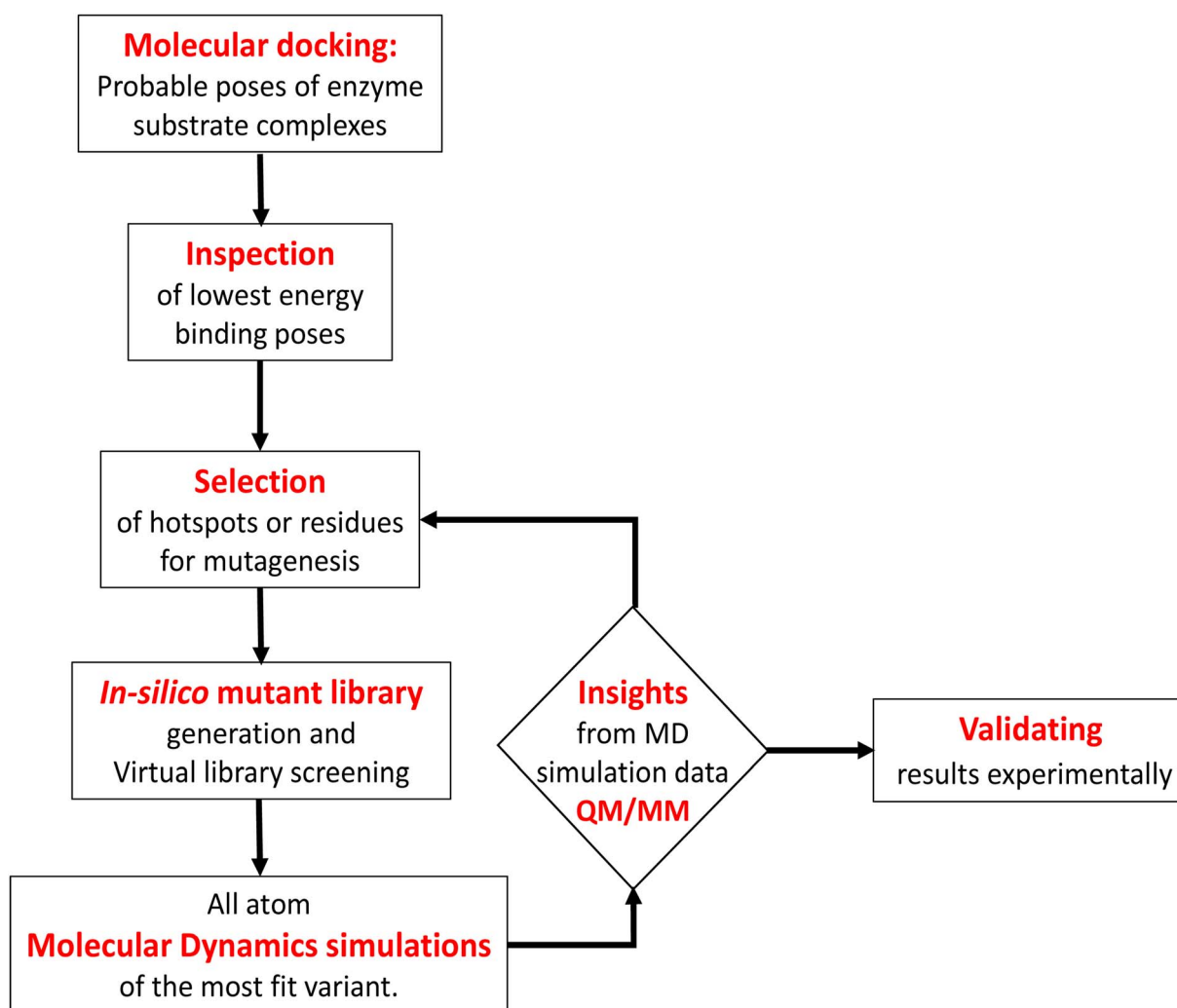


hexadecane as a substrate to start rational design using a strategy highlighted in Scheme 1. Initially, hexadecane was docked into the active pocket of wild-type CYP175A1 (see Fig. S3 and S4 and section 6 in the ESI† for details), and the best pose based on the lowest energy score was chosen. Fig. 1a shows the lowest energy pose of hexadecane docked into the active pocket of the wild-type (WT) enzyme, while Fig. 1b shows some strategic residues close to the active site. Based on the docked structure of the substrate (Fig. 1b) we adopted three key strategies during bioengineering:

### 2.1 Substrate binding engineering

To understand the rationale behind the absence of any detectable catalytic activity of the WT enzyme, we performed the MD simulation on the lowest energy-docked structure of the substrate–enzyme (WT) complex. To our surprise, we found that the substrate was ejected out of the binding pocket of the enzyme during the course of the simulation (see supplementary video number VS1†) signifying that the active pocket of the WT enzyme is not suitable for substrate binding. A closer inspection

of the MD trajectory reveals that active site residues Q67 and Y68 provide steric repulsion on the substrate that leads to unbinding of the substrate. The presence of Q67 residue hindering substrate (palmitoleic acid) binding has also been proposed earlier by Poulos and co-workers.<sup>40</sup> In addition, the polar group of these residues interacts with the propionate of heme, which provides stiffness to the active site, and therefore, we anticipate that the replacement of these polar residues with appropriate non-polar residues could increase the binding of hexadecane. With these insights, we prepared a library of double mutants of Q67 and Y68 by replacing these two residues with non-polar residues and docked the substrate in the active site. Interestingly, the docking/virtual screening data (Tables S2, S3, and Fig. S5 in the ESI†) of the substrate into the double-mutant variants showed an enhanced docking score relative to that of the WT complex, which indicated that replacement of Q → L and Y → F (*i.e.*, Q67L/Y68F termed an LF mutant) could be promising for hexadecane binding. To further substantiate the docking result, we performed the MD simulation of the best-docked pose of the substrate in complex with the LF variant of



Scheme 1 Flowchart of computationally guided design of an alkane hydroxylating enzyme.



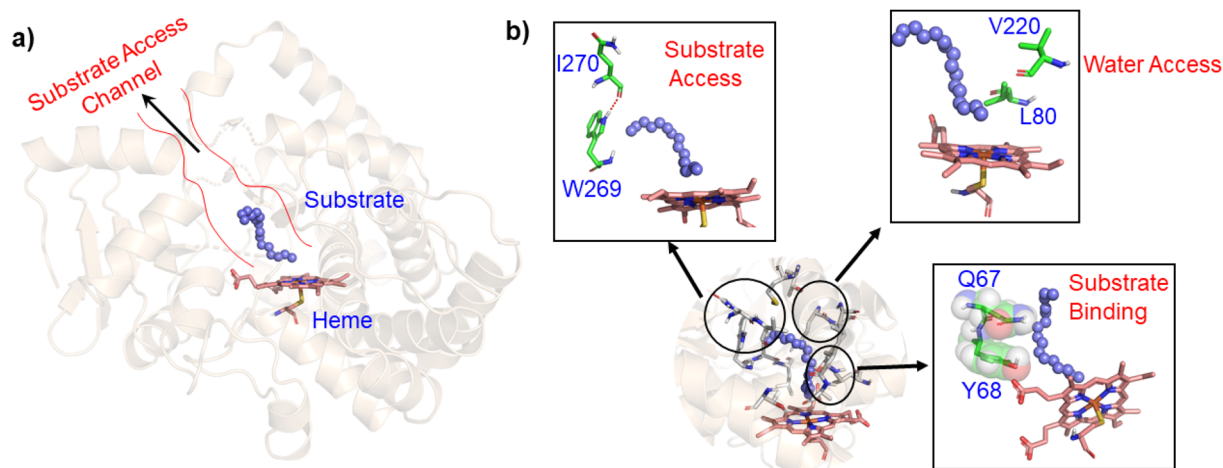


Fig. 1 (a) Lowest energy docking pose of wild-type CYP175A1 with hexadecane (blue balls), and the substrate access channel is highlighted by red lines. (b) Strategic residues which may affect catalytic activity. Note the blockage of the substrate access channel by H-bonds between I270 and W269, and steric blockage by Q67-Y68 for substrate binding.

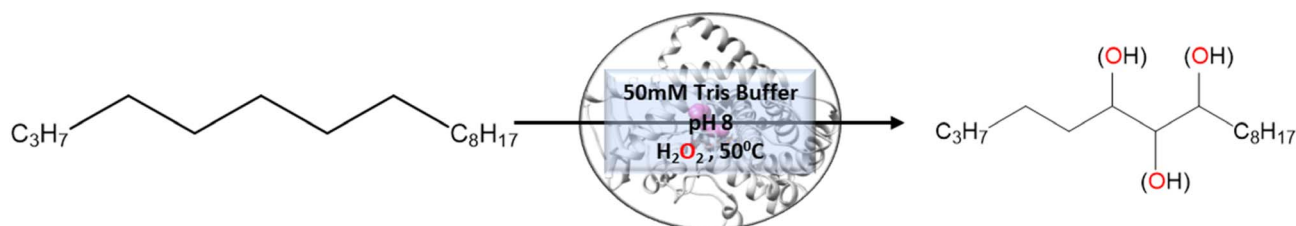
CYP175A1. Surprisingly, in contrast with the WT simulation, where the substrate came out of the binding pocket; the substrate shows stable binding with the LF mutant, and it was not ejected out of the binding pocket (supplementary video VS2†) during MD simulation. It is important to note here that the formed hydroxylated product will be polar and hence will have lower binding affinity compared to the substrate due to which another incoming substrate molecule can easily replace the product from the active pocket during turnover.

To validate the above computational prediction, we performed LF mutation using site-directed mutagenesis by PCR and measured the experimental catalytic efficiency. The mutant protein was expressed and purified according to the previously reported protocols<sup>42</sup> with minor modifications (see the ESI† for details). The purified protein showed a characteristic peak at 450 nm in a reduced CO-bound state (Fig. S9 in the ESI†). Purified mutant proteins were used to catalyze the hexadecane hydroxylation reaction as shown in Scheme 2. The mutant protein showed moderate catalytic activity (TON = 52) towards hexadecane through the reductive pathway but only a trace level of activity through the peroxide shunt pathway using H<sub>2</sub>O<sub>2</sub> (Tables 1 and 2 in the ESI† for more details). It is important to note here that the wild-type enzyme was not showing any detectable catalytic activity for hexadecane. Our designed double mutant LF showed catalytic activity for hexadecane hydroxylation, which reflects that the substrate is indeed

binding inside the active pocket of the LF mutant and also that the environment inside the active pocket of the wild-type enzyme was not favorable for highly non-polar substrates such as hexadecane. In the LF mutant we have incorporated small hydrophobic amino acids, which possibly increase the hydrophobicity and the accessibility of the active pocket, and hence, a bulky substrate like hexadecane showed binding inside the active pocket. Lower activity through the shunt pathway might be either due to irreversible inactivation of the protein by H<sub>2</sub>O<sub>2</sub> or poor coupling of H<sub>2</sub>O<sub>2</sub> with the protein active pocket, as has also been observed in other reports.<sup>43</sup>

## 2.2 Water channel engineering

As mentioned above, the double mutant LF has enhanced the catalytic activity compared to the WT enzyme. However, its catalytic activity was 'moderate' and insufficient to be a practical, robust catalyst through the peroxide shunt mechanism. A recent study by Shaik and co-workers<sup>44</sup> highlighted the importance of water molecules close to the heme site in the C-H activation reaction. We anticipated that as water and H<sub>2</sub>O<sub>2</sub> are very much similar and completely miscible, activating the water channel may enhance the abundance of the H<sub>2</sub>O<sub>2</sub> near the heme active site, and thus may increase the reaction rate of the peroxide-mediated C-H activation pathway. We, therefore, focused on the water channel inside the active pocket of



Scheme 2 Hexadecane reaction catalysed by mutants of CYP175A1 through the peroxide shunt pathway.





Table 1 Details of sequential mutations on CYP175A1 by the multi generation approach

| Generation | Mutant description                            | Mutation sites and corresponding mutations | Abbreviation |
|------------|---|--|--------------|
| 1st        | Active site mutation                          | Q67L/Y68F                                  | LF           |
| 2nd        | Water channel activation mutations            | L80F/V220T                                 | FT           |
| 3rd        | Substrate entry channel opening mutations     | W269F/I270A                                | FA           |
| 4th        | Water channel + substrate entry               | L80F/V220T/W269F/I270A                     | FTFA         |
| 5th        | Active site + substrate entry                 | Q67L/Y68F/W269F/I270A                      | LFFA         |
| 6th        | Active site + water channel                   | Q67L/Y68F/L80F/V220T                       | LFFT         |
| 7th        | Active site + substrate entry + water channel | Q67L/Y68F/V220T/W269F/I270A                | LFTFA        |
| 8th        | Combining all                                 | Q67L/Y68F/L80F/V220T/W269F/I270A           | LFFTFA       |

Table 2 Evaluation of catalytic activity: product distribution by various evolved enzymes for hexadecane hydroxylation<sup>a</sup>

| Evolved enzyme | TON w.r.t the FA<br>(W269F/I270A) mutant | Product distribution <sup>b</sup> |      |      |                     |
|----------------|--|-----------------------------------|------|------|---------------------|
|                |  | 6-OH                              | 7-OH | 8-OH | Others <sup>c</sup> |
| Wild-type      | N.D.                                     | —                                 | —    | —    | —                   |
| LF             | 2  | 8                                 | 92   | —    | —                   |
| FA             | 1  | 13                                | 83   | 4    | —                   |
| FT             | N.D.                                     | —                                 | —    | —    | —                   |
| FTFA           | 5  | 61                                | 37   | 2    | —                   |
| LFFT           | 8  | 10                                | 82   | 8    | —                   |
| LFFA           | 12                                       | 17                                | 80   | 2    | <1                  |
| LFTFA          | 15                                       | 2                                 | 92   | 4    | <3                  |
| LFFTFA         | 19                                       | 2                                 | 95   | 1    | <2                  |

<sup>a</sup> Catalytic activities are reported through the peroxide shunt pathway. The reaction was performed for 24 hours. <sup>b</sup> Product distribution (in percentage) was quantified by GC-FID at least in triplicate. The error was less than 5%. <sup>c</sup> Unidentified product peaks. N.D. stands for not determined.

CYP175A1 and hypothesized that enhancing hydration close to the heme site may increase the catalytic activity through the peroxide shunt pathway. The analysis of MD trajectories of the WT and LF mutant shows very low water occupancy throughout the simulations as shown in Fig. 2a (see Fig. S7 and supporting video VS2†). Through a closer inspection of the protein topology and MD trajectory of the WT enzyme, we found that L80 and V220 are two residues that may block the water access to the heme site. Therefore, we mutated V220 by using a threonine (Thr) residue since Thr is known for its role in oxygen activation and water channel formation in cytochrome P450 chemistry.<sup>45–47</sup> Earlier studies on cytochrome P450<sub>BM3</sub> showed that a phenylalanine residue close to the heme site plays the role of gatekeeper and can regulate the water aqueduct and substrate orientations.<sup>48,49</sup> Inspired by these findings, we mutated L80 → F80 by assuming that Phe can play a gatekeeper role in substrate/water access since L80 in CYP175A1 occupies an equivalent position to that of Phe87 in cytochrome P450<sub>BM3</sub>. Interestingly, the MD simulation of this new tetra mutant incorporating mutations in two key areas *i.e.*, the substrate binding site and water channel site (Q67L/Y68F/L80F/V220T, termed the LFFT mutant) showed a significant enhancement of the water occupancy in the first hydration shell of the heme center as shown in Fig. 2b. The computationally predicted mutation of L80F/V220T was further verified by an experimental study, and we found that L80F and V220T mutations on the LF mutant (resulting in the LFFT mutant) were able to increase the

catalytic activity through the peroxide shunt pathway, as shown in Fig. 4. A recent study by Zhao *et al.* showed that the peroxxygenase activity of CYP199A4 and CYP153<sub>AM.aq</sub> can be increased by rational engineering of the water channel,<sup>50</sup> and closely agrees with our results.

### 2.3 Substrate access engineering

Since easy access of the substrate to the active pocket and efficient product exit are crucial for catalysis, theoretical as well as experimental studies on understanding the substrate access channels in various cytochrome P450s have been performed previously, and it has been established that the catalytic activity can be influenced by how the substrate approaches the active site through the entry channel.<sup>26,51</sup> We envisioned that catalytic activity could be improved by increasing the substrate accessibility to the active pocket. We therefore, focused on the engineering of the substrate access channel. We used the CAVER software tool to analyze the tunnel inside the crystal structure of the WT CYP175A1. CAVER tunnel analysis showed that the substrate entry tunnel has a bottleneck mainly in the vicinity of the W269 and I270 residues, as shown in Fig. 3. We substituted these residues with smaller residues of a similar type and analyzed the tunnel size using CAVER followed by docking of hexadecane into the mutant enzymes. Based on the good docking score and wider neck of the channel, we identified W269F and I270A as the best mutants as shown in Fig. 3. The binding of the substrate in the W269F/I270A mutant was further



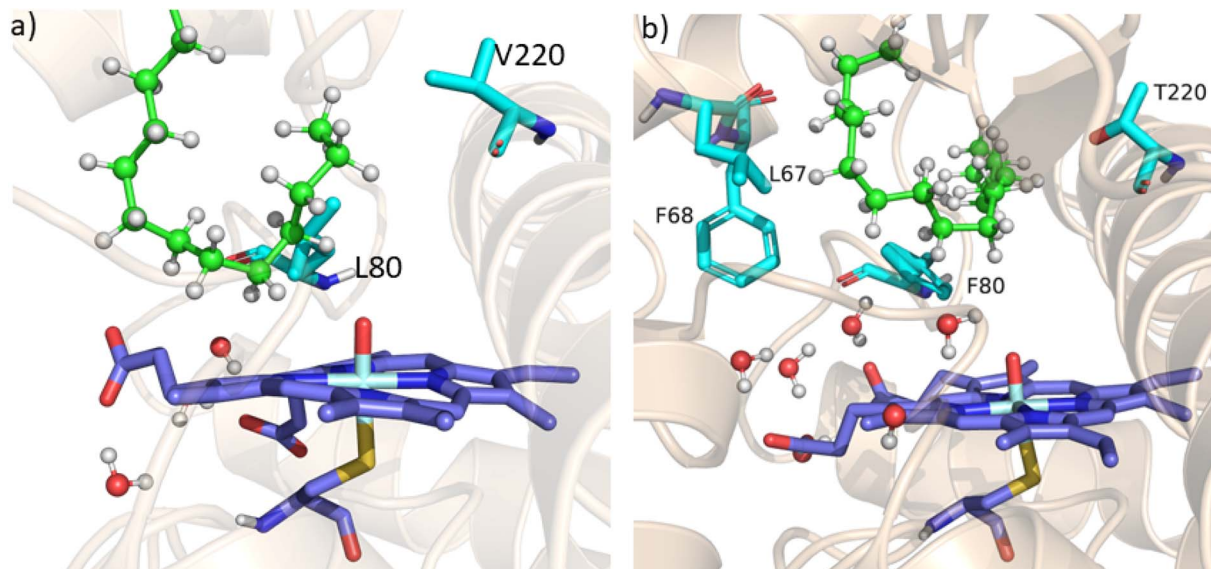


Fig. 2 (a) Figure showing low water content inside the pocket of the LF mutant of CYP175A1. (b) Figure showing high water content inside the active pocket of the LFFT mutant of CYP175A1 (see supporting video VS3 for more details; † images taken from MD simulation data).

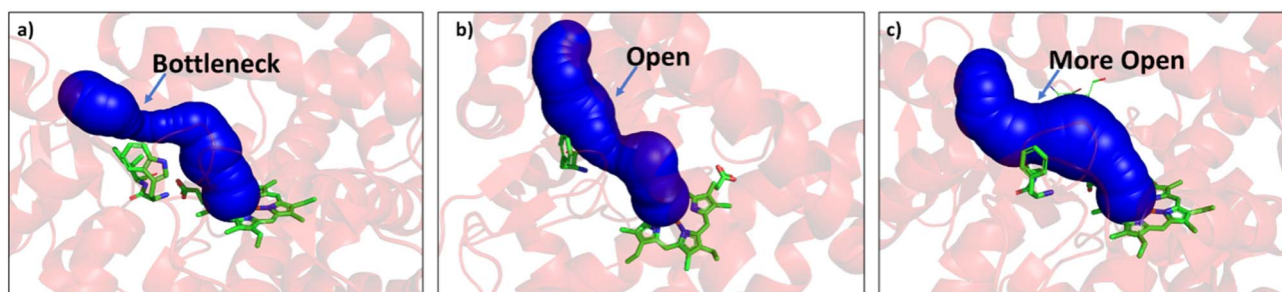


Fig. 3 (a) Figure showing the bottleneck at the substrate entry channel in the wild type CYP175A1 using CAVER. (b) W269F/I270A double mutant of wild type CYP175A1 shows appreciable opening of the bottleneck at the substrate entry channel. (c) Q67L/Y68F/W269F/I270A tetra mutant of wild type CYP175A1 showing the opening of the bottleneck at the substrate entry channel.

verified by MD simulations, and we found the substrate was quite stable inside the active site during the MD simulation as well.

In brief, the simulations identified three strategic sites/hot spots for mutations in the WT enzyme based on: (1) substrate binding mutation, (2) water channel activation, and (3) substrate entry channel. Table 1 summarises these functional sites and the mutants we have made in this study. Mutations were performed using PCR (Table S6†) and validated by plasmid sequencing results (Fig. S8 in the ESI†). All the mutant proteins showed a characteristic CO binding spectrum under reduced conditions (Fig. S9 in the ESI†). To validate the computation prediction, we experimentally verified the catalytic efficiency of these mutants through the peroxide shunt pathway (see Fig. 4, S10–15, and section 21 in the ESI†). As can be seen in Fig. 4, the water channel mutant (FT) alone didn't show any detectable catalytic activity. This signifies that the substrate probably does not bind inside the active pocket due to an unfavorable environment for the hexadecane substrate. The LF mutant showed better activity compared to the FA mutant which signifies that



Fig. 4 Hexadecane hydroxylation activity of the mutants of CYP175A1 through the peroxide shunt pathway using  $H_2O_2$ . \*Relative fold increment in the TON with respect to the least performing FA mutant.



substrate binding mutations (Q67L/Y68F) are more effective and crucial compared to substrate entry channel mutations (W269F/I270A). Due to the opening of the bottleneck, the FA mutant alone showed a small enhancement of activity. As shown in Fig. 4, hexadecane hydroxylation activity gradually started increasing upon sequential mutations signifying the synergistic effect of all the mutations necessary for the

enhanced catalytic activity. This signifies that rational bioengineering is an effective way to expand the substrate scope of the enzyme. Furthermore, we were successfully able to achieve a 19-fold increase in the catalytic activity in the final mutant LFFTFFA compared to the least-performing mutant FA. This signifies that rational bioengineering can not only expand the substrate scope of an enzyme but can also improve the catalytic activity. For

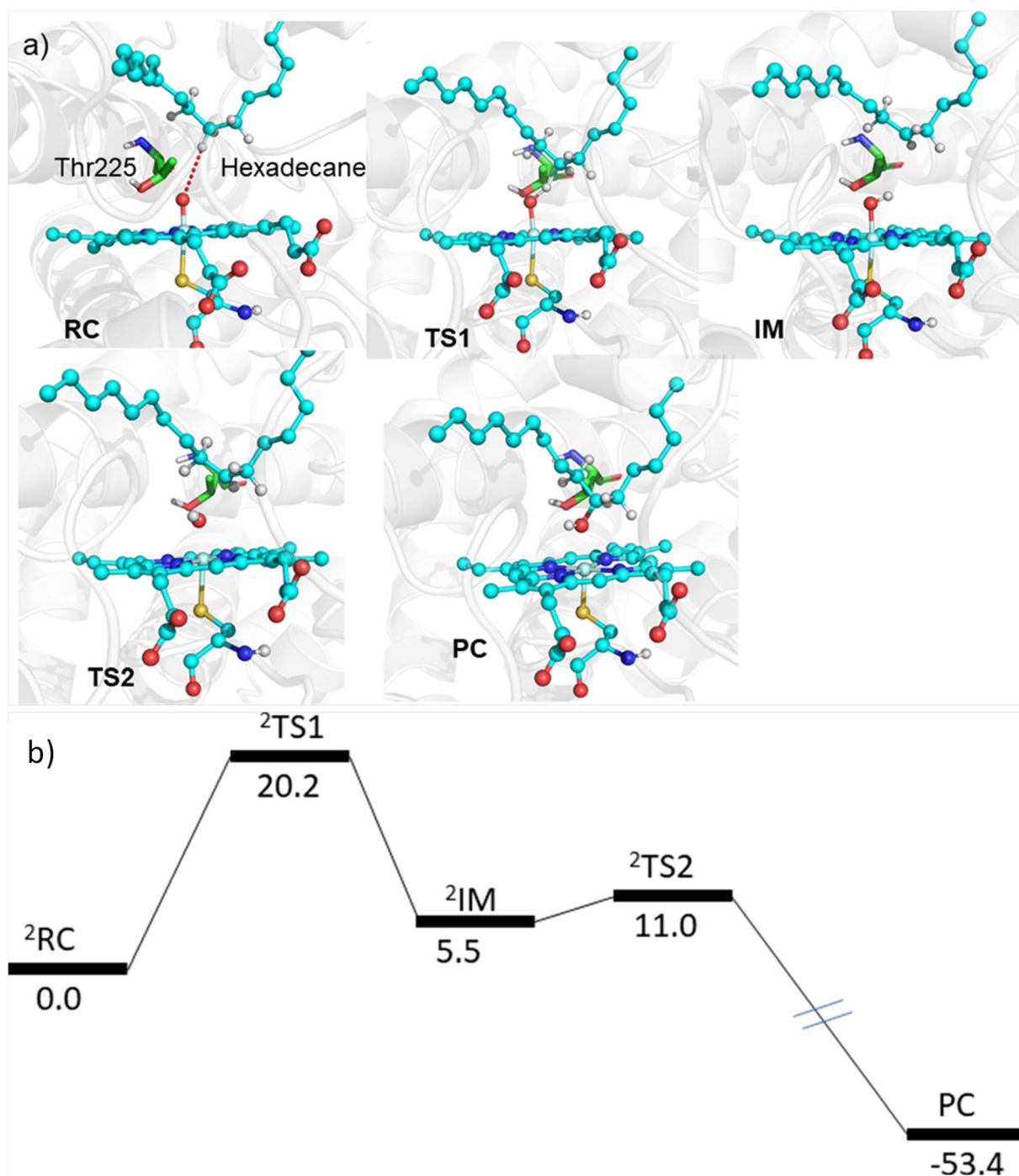


Fig. 5 (a) The geometry of different stationary points during the hydroxylation of hexadecane *via* compound I inside the active pocket of the LFFTFFA mutant, (water molecules near the active site of the protein are not shown for the ease of representation). (b) The energy profile of the reaction obtained from QM/MM potential energy surface scanning using the UB3LYP-D3/def2-TZVP level of the theory. All energy values are in kcal mol<sup>-1</sup> and the energy for PC is not scaled.





a comparison, our developed final mutant performs better than the other hexadecane hydroxylating cytochrome P450 CYP52A4 reported earlier.<sup>52</sup> These findings signify that engineering of all the hot spots in an enzyme is necessary for higher catalytic activity. The cumulative effect of all the mutations has significantly increased the catalytic reaction rate. The effect of mutations was also significant on the product distribution as evident from Table 2. The LF mutant showed 92% formation of the C7 hydroxylated product and the FA mutant showed 83% formation of the C7 hydroxylated product. The FTFA mutant has C6 hydroxylation as the major product (61%), which is probably due to the displacement of the substrate towards the entry channel by L80F/V220T mutation. There were some other unidentified peaks on the GC chromatograph possibly corresponding to the overoxidation product (see Fig. S15 in the ESI†) formed during the reaction.

#### 2.4 Computational analysis by QM/MM studies on the mutants of CYP175A1

The MD simulations combined with the experiments show that the mutants can efficiently catalyze the C–H hydroxylation of hexadecane. However, classical simulations cannot explain the reaction mechanism, so we performed hybrid QM/MM calculations to validate the reaction mechanism of C–H hydroxylation. For doing so, we selected a representative snapshot from the MD simulation of the most efficient variant, *i.e.*, LFFTFA, and performed the QM/MM optimization of the reactant geometry. Compound I (RC) mediated P450 reactions can be

observed in doublet and quartet spin states. Earlier studies have shown that the doublet spin state is the favoured one.<sup>47,53</sup> The calculations were performed using the B3LYP/def2-TZVP level of theory (See the ESI† the QM/MM calculation protocol). The calculated geometry of RC, TS1, IM, TS2 and PC corresponding to different stationary points during the hydroxylation of hexadecane *via* compound I (RC) inside the active pocket of the LFFTFA mutant of CYP175A1, is shown in Fig. 5. As can be seen in Fig. 5 (and Fig. S6, Tables S4 and S5 in the ESI†), the process starts from the hydrogen atom transfer (HAT) reaction from the C7 position of the substrate to form an unstable intermediate IM. The spin density calculations confirm the species IM as Cpd II as mentioned in earlier studies.<sup>47,53</sup> The energy barrier of this reaction is 20.2 kcal mol<sup>-1</sup>, which is higher than the energy barrier for HAT for the reaction of fatty acid catalyzed by other cytochrome P450s<sup>44,54</sup> reported earlier. In the subsequent steps, IM follows the typical rebound mechanism<sup>55</sup> to form the hydroxylated product (PC) *via* a relatively small energy barrier of 5.5 kcal mol<sup>-1</sup> and resets the catalytic cycle at the resting state for the next turnover of catalysis. The low energy barriers observed in the computational studies for the reactions possibly signify that the diffusion of the substrate to the active site is the rate-limiting step for the C–H activation reaction. The reaction proceeds to completion very rapidly as soon as the substrate reaches the active site. The theoretical calculations on the analogous fatty acid substrates were reported previously and they agree well with the present results.<sup>56</sup> Furthermore, we also checked the reaction profile in the quartet spin state, and the reaction profile for the same is shown in Fig. S6 (in the ESI†). As

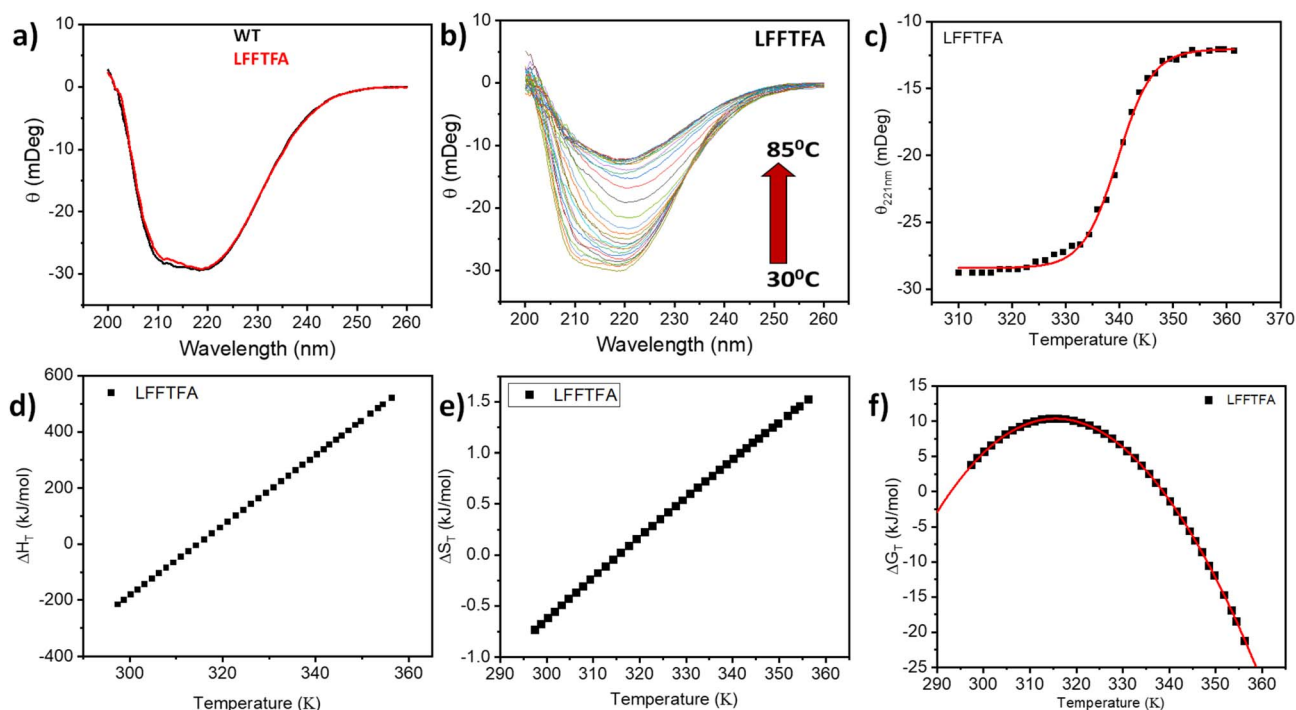


Fig. 6 (a) CD spectrum of wild-type (WT) CYP175A1 (black) and the evolved variant LFFTFA of CYP175A1 (red) in the far UV region. (b) Thermal unfolding of the LFFTFA mutant: CD spectrum at different temperatures from 30 °C to 85 °C. (c) Plot of the CD signal (at 220 nm) at different temperatures of the LFFTFA mutant. (d) Enthalpy change (e) entropy change and (f) free energy change during thermal denaturation of the LFFTFA mutant.





can be seen in Fig. S6 (in the ESI†), the energy barrier for HAT from the RC in the quartet spin state is significantly higher than that of the doublet state ( $23.9 \text{ kcal mol}^{-1}$  in the quartet *vs.*  $20.2 \text{ kcal mol}^{-1}$  in the doublet); therefore, we believe the reaction will proceed *via* the doublet state as reported earlier.<sup>47,53</sup>

## 2.5 Detailed characterization and evaluation of the thermal stability of the mutants of CYP175A1

Reports on enzyme design by directed evolution through random mutagenesis have produced highly catalytically active evolved enzymes, but most often these reports lack detailed characterization of the evolved enzyme. To have a practical biocatalyst for industrial application, the evolved catalyst should be stable. We evaluated the detailed thermal stability of the evolved variants of CYP175A1 using circular dichroism (CD) spectroscopy. The CD spectrum for the secondary structure region (200–260 nm) of the wild type and mutants of CYP175A1 was recorded at 50 °C. The results showed that all the mutants were well folded and had a similar secondary structure to that of the wild-type protein. Fig. 6a shows the CD spectrum of the wild-type and LFFTFFA mutant of CYP175A1, and overlapping spectra signify that there is not much appreciable change in the secondary structure of the protein upon mutations. The thermal

unfolding of all the mutant proteins was also studied by recording the CD spectrum at different temperatures. All the temperature-dependent CD spectra were recorded from 30–85 °C, and Fig. 6b shows the thermal unfolding of the LFFTFFA mutant. As shown in Fig. 6c, for the LFFTFFA mutant, the CD signal at 221 nm was plotted as a function of temperature, and the data were analyzed using modified Gibbs Helmholtz equations.<sup>57</sup> The values of the thermodynamic parameters obtained from the analyses of the data for all the mutants are given in Table S7 in the ESI.† Analysis of the data showed that the midpoint temperature,  $T_m$  for the unfolding of the secondary structure of the mutants of CYP175A1 has decreased compared to that of the wild-type protein but still the final mutant *i.e.* LFFTFFA having a  $T_m$  value of 66 °C was found to be more thermally stable than the mesophilic analogs such as CYP101A1 and CYP102A1. Every mutation in the wild-type protein has led to a decrease in the overall thermal stability of the protein (Fig. S16 and Table S7 in the ESI†). To probe the effect of individual amino acids on the thermal stability of the protein, we have analyzed the thermal unfolding data of all the mutants. The decrease in the thermal stability of the QY mutant was attributed to the loss of hydrogen bonding due to the mutation of Y68 with Phe because the Y68 residue exhibited hydrogen bonding with the heme group and surrounding amino acids as understood from the crystal structure of the wild-type CYP175A1 (Fig. S17 in the ESI†). Interestingly the temperature of maximum stability (the temperature at which the slope of the free energy curve is zero, Fig. 6f) for the LFFTFFA mutant is increased to around 33 °C compared to that of wild-type protein which has a temperature of maximum stability of 29 °C. This suggests that the final mutant is still thermally very stable and can work very well under ambient temperature conditions. The CD spectrum in the visible region (Fig. 7) was analyzed to probe the effect of the mutations on the tertiary structure of the protein. The results showed a distinct shift in the position of the Soret band (413 nm) in the LFFTFFA mutant compared to that of the wild-type protein (408 nm), whereas there was no shift in the delta band position.

The Soret to delta band ratio ( $S/D$ ) in the LFFTFFA mutant was found to be (0.66) lower than that of the wild-type protein signifying the reduction in the polarity and hydrogen bonding near the active site of the protein. These observations have also been reported previously and agree well with the previous literature.<sup>57–59</sup> Table 3 shows the position and the ratio of the Soret and delta bands in various cytochrome P450s. Reduction



Fig. 7 CD spectrum of substrate free wild-type (black) and the LFFTFFA mutant (red) of CYP175A1 in the 300–450 nm region showing the Soret and delta bands. [D] denotes the delta band and [S] denotes the Soret band.

Table 3 Active site spectral features of various cytochrome P450s in the substrate-free state, showing the position of Soret and delta bands and their ratio

| S.No. | Enzyme               | Soret band (nm) | Delta band (nm) | Ratio ( $S/D$ ) | Ref.      |
|-------|----------------------|-----------------|-----------------|-----------------|-----------|
| 1     | CYP101A1             | 409             | 351             | 1.4             | 58        |
| 2     | T101V-CYP101A1       | 406             | 352             | 1.1             | 58        |
| 3     | CYP102A1             | 410             | 350             | 2.2             | 59        |
| 4     | P450 <sub>terp</sub> | 410             | 350             | 2.0             | 59        |
| 5     | CYP175A1             | 408             | 350             | 1.22            | This work |
| 6     | LFFTFFA-CYP175A1     | 413             | 350             | 0.66            | This work |



in the  $S/D$  in LFFTTFA as shown in Table 3 signifies that the active pocket is highly non-polar and perfect for the incorporation of non-polar substrates such as linear chain alkanes.

### 3 Conclusions and future outlook

The present work demonstrates that the substrate scope of an enzyme can be expanded by rationally designing its active pocket, and the catalytic rate can be further increased by the engineering of various important channels such as the substrate entry channel and water channel. We have successfully developed a novel strategy for the rational evolution of enzymes based on computational tools. Computational data provided the rationale behind the absence of catalytic activity for hexadecane hydroxylation by wild-type enzyme. It is widely known that interactions beyond the primary coordination sphere also play a vital role in the enzymatic activity.<sup>60</sup> Various reports have demonstrated the role of the second coordination sphere<sup>61–63</sup> but it remains difficult to rationalise the factors responsible for modulating the activities of natural as well as artificial metalloenzymes. Our findings provide an in depth understanding and rationale to the role of these weak nonbonding interactions in the activity of the enzyme. An unfavorable environment inside the active pocket possibly due to the presence of polar residues (Q67 and Y68) was not allowing the binding of the non-polar substrate. Mutations of these polar residues with non-polar residues have shown the binding of the substrate, as evident from the detectable product formation. We found out that due to the presence of the bottleneck at the substrate entry channel (due to the bulky amino acids), the active site was not easily accessible by the substrate. Tunnel analysis using CAVER has shown that the replacement of bulky amino acids at the substrate entry channel (W269 and I270) with small amino acids can open up the bottleneck. An increase in the catalytic rate after performing W269F/I270A mutations confirms that the bottleneck has opened up. The residues near the substrate entry channel influencing the catalytic activity and selectivity of the enzyme have been previously reported, and the results agree well with our findings.<sup>64</sup> The peroxide shunt pathway is known to be an inefficient method for cytochrome P450 catalysis. We envisioned that the catalytic rate through the peroxide shunt pathway can be increased by engineering the water channel in CYP175A1. To the best of our knowledge, there are only very few reports which could demonstrate that water channel engineering can lead to a higher catalytic rate through the peroxide shunt pathway. The mutation of non-polar residue V220 with the polar residue threonine, coupled with the mutation of L80 with phenylalanine has shown an increase in the catalytic rate by the shunt pathway. The thermal and thermodynamic stability of all the mutants were analyzed using CD spectroscopy. The results showed that the evolved enzymes were well folded and fairly thermally stable and suitable for working under ambient conditions. The secondary structure of the evolved enzymes remains unchanged but there was a significant change in the active site structure and polarity as evident from the CD data in the visible region. The present studies of structure-based rational design of the CYP175A1 enzyme can

also be used for other enzymatic systems to expand the substrate scope and to increase the catalytic rate.

### Data availability

All relevant supporting data are provided in the ESI.†

### Author contributions

MT, KDD and SM conceptualized the problem, MT carried out all experimental work and part of the computational work, KDD carried out computational work, MT, KDD and SM interpreted the results and wrote the manuscript.

### Conflicts of interest

The authors declare no conflict of interest.

### Acknowledgements

The authors thank the Tata Institute of Fundamental Research for funding and infrastructural support. This work was supported by the Department of Atomic Energy, Government of India, under Project Identification No. RTI4003 through the Tata Institute of Fundamental Research, Mumbai (to S. M.). KDD acknowledges the Department of Biotechnology, Ministry of Science and Technology, Government of India, for the Ramalingaswami re-entry grant (BT/Re-entry/RLF/10/2017) and Shiv Nadar Foundation for financial assistance.

### References

- 1 J.-M. Choi, S.-S. Han and H.-S. Kim, *Biotechnol. Adv.*, 2015, **33**, 1443–1454.
- 2 A. R. Alcántara, P. Domínguez de María, J. A. Littlechild, M. Schürmann, R. A. Sheldon and R. Wohlgenuth, *ChemSusChem*, 2022, **15**, e202102709.
- 3 R. E. Cobb, R. Chao and H. Zhao, *AIChE J.*, 2013, **59**, 1432–1440.
- 4 R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam and S. Muley, *Nat. Biotechnol.*, 2007, **25**, 338–344.
- 5 M. T. Reetz and J. D. Carballeira, *Nat. Protoc.*, 2007, **2**, 891–903.
- 6 R. A. Chica, N. Doucet and J. N. Pelletier, *Curr. Opin. Biotechnol.*, 2005, **16**, 378–384.
- 7 Z. Li, Y. Jiang, F. P. Guengerich, L. Ma, S. Li and W. Zhang, *J. Biol. Chem.*, 2020, **295**, 833–849.
- 8 D. Valikhani, J. M. Bolivar and J. N. Pelletier, *ACS Catal.*, 2021, **11**, 9418–9434.
- 9 P. S. Coelho, E. M. Brustad, A. Kannan and F. H. Arnold, *Science*, 2013, **339**, 307–310.
- 10 S. V Athavale, S. Gao, A. Das, S. C. Mallojjala, E. Alfonzo, Y. Long, J. S. Hirschi and F. H. Arnold, *J. Am. Chem. Soc.*, 2022, **144**, 19097–19105.
- 11 S. Karmakar, S. K. Nag, M. Taher, B. T. Kansara and S. Mazumdar, *Protein J.*, 2022, **41**, 659–670.



- 12 O. Shoji, T. Kunimatsu, N. Kawakami and Y. Watanabe, *Angew. Chem., Int. Ed.*, 2013, **52**, 6606–6610.
- 13 R. Saroay, G.-D. Roiban, L. M. Alkhalaf and G. L. Challis, *ChemBioChem*, 2021, **22**, 2262–2265.
- 14 J. Huang, Z. Liu, B. J. Bloomer, D. S. Clark, A. Mukhopadhyay, J. D. Keasling and J. F. Hartwig, *Nat. Chem.*, 2021, **13**, 1186–1191.
- 15 H. M. Key, P. Dydio, Z. Liu, J. Y.-E. Rha, A. Nazarenko, V. Seyedkazemi, D. S. Clark and J. F. Hartwig, *ACS Cent. Sci.*, 2017, **3**, 302–308.
- 16 T. K. Hyster and F. H. Arnold, *Isr. J. Chem.*, 2015, **55**, 14–20.
- 17 H. Renata, Z. J. Wang and F. H. Arnold, *Angew. Chem., Int. Ed.*, 2015, **54**, 3351–3367.
- 18 P. Urban, T. Lautier, D. Pompon and G. Truan, *Int. J. Mol. Sci.*, 2018, **19**, 1617.
- 19 S. Meng, Y. Ji, L. Liu, M. D. Davari and U. Schwaneberg, *ChemSusChem*, 2022, **15**, e202102434.
- 20 N. Shi, Q. Zheng and H. Zhang, *ACS Chem. Neurosci.*, 2022, **13**, 1526–1533.
- 21 N. Krefß, J. M. Halder, L. R. Rapp and B. Hauer, *Curr. Opin. Chem. Biol.*, 2018, **47**, 109–116.
- 22 P. Kokkonen, D. Bednar, G. Pinto, Z. Prokop and J. Damborsky, *Biotechnol. Adv.*, 2019, **37**, 107386.
- 23 P. Hlavica, *J. Inorg. Biochem.*, 2023, **241**, 112150.
- 24 S. Meng, R. An, Z. Li, U. Schwaneberg, Y. Ji, M. D. Davari, F. Wang, M. Wang, M. Qin, K. Nie and L. Liu, *Bioresour. Bioprocess.*, 2021, **8**, 26.
- 25 Z. Li, S. Meng, K. Nie, U. Schwaneberg, M. D. Davari, H. Xu, Y. Ji and L. Liu, *ACS Catal.*, 2022, **12**, 12800–12808.
- 26 E. L. Bell, W. Finnigan, S. P. France, A. P. Green, M. A. Hayes, L. J. Hepworth, S. L. Lovelock, H. Niikura, S. Osuna, E. Romero, K. S. Ryan, N. J. Turner and S. L. Flitsch, *Nat. Rev. Methods Primers*, 2021, **1**, 46.
- 27 Y. Yang, J. Liu and Z. Li, *Angew. Chem., Int. Ed.*, 2014, **53**, 3120–3124.
- 28 N. Kawakami, O. Shoji and Y. Watanabe, *Angew. Chem., Int. Ed.*, 2011, **50**, 5315–5318.
- 29 M. Bordeaux, A. Galarneau, F. Fajula and J. Drone, *Angew. Chem., Int. Ed.*, 2011, **50**, 2075–2079.
- 30 E. G. Funhoff, B. Ulrich, G.-R. Inés, W. Bernard and B. J. B. van, *J. Bacteriol.*, 2006, **188**, 5220–5227.
- 31 A. Glieder, E. T. Farinas and F. H. Arnold, *Nat. Biotechnol.*, 2002, **20**, 1135–1139.
- 32 K. Yonemura, S. Ariyasu, J. K. Stanfield, K. Suzuki, H. Onoda, C. Kasai, H. Sugimoto, Y. Aiba, Y. Watanabe and O. Shoji, *ACS Catal.*, 2020, **10**, 9136–9144.
- 33 F. E. Zilly, J. P. Acevedo, W. Augustyniak, A. Deege, U. W. Häusig and M. T. Reetz, *Angew. Chem., Int. Ed.*, 2011, **50**, 2720–2724.
- 34 P. C. Cirino and F. H. Arnold, *Angew. Chem., Int. Ed.*, 2003, **42**, 3299–3301.
- 35 S. Goyal, S. Banerjee and S. Mazumdar, *Biochemistry*, 2012, **51**, 7880–7890.
- 36 E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor and J. Damborsky, *PLoS Comput. Biol.*, 2012, **8**(10), e1002708.
- 37 A. Seifert, M. Antonovici, B. Hauer and J. Pleiss, *ChemBioChem*, 2011, **12**, 1346–1351.
- 38 A. Seifert, S. Vomund, K. Grohmann, S. Kriening, V. B. Urlacher, S. Laschat and J. Pleiss, *ChemBioChem*, 2009, **10**, 853–861.
- 39 A. Rentmeister, F. H. Arnold and R. Fasan, *Nat. Chem. Biol.*, 2009, **5**, 26–28.
- 40 J. K. Yano, F. Blasco, H. Li, R. D. Schmid, A. Henne and T. L. Poulos, *J. Biol. Chem.*, 2003, **278**, 608–616.
- 41 H. Li and T. L. Poulos, *Nat. Struct. Biol.*, 1997, **4**, 140–146.
- 42 A. Mondal, M. Das and S. Mazumdar, *Inorg. Chim. Acta*, 2019, **487**, 398–404.
- 43 P. C. Cirino and F. H. Arnold, *Adv. Synth. Catal.*, 2002, **344**, 932–937.
- 44 K. D. Dubey, B. Wang and S. Shaik, *J. Am. Chem. Soc.*, 2016, **138**, 837–845.
- 45 T. I. Oprea, G. Hummer and A. E. García, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 2133–2138.
- 46 R. Raag, S. A. Martinis, S. G. Sligar and T. L. Poulos, *Biochemistry*, 1991, **30**, 11420–11429.
- 47 S. Shaik, D. Kumar, S. P. de Visser, A. Altun and W. Thiel, *Chem. Rev.*, 2005, **105**, 2279–2328.
- 48 K. D. Dubey, B. Wang, M. Vajpai and S. Shaik, *Chem. Sci.*, 2017, **8**, 5335–5344.
- 49 K. D. Dubey and S. Shaik, *Acc. Chem. Res.*, 2019, **52**, 389–399.
- 50 P. Zhao, F. Kong, Y. Jiang, X. Qin, X. Tian and Z. Cong, *J. Am. Chem. Soc.*, 2023, **145**, 5506–5511.
- 51 P. J. Winn, S. K. Lüdemann, R. Gauges, V. Lounnas and R. C. Wade, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 5361–5366.
- 52 U. Scheller, T. Zimmer, E. Kärger and W.-H. Schunck, *Arch. Biochem. Biophys.*, 1996, **328**, 245–254.
- 53 S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar and W. Thiel, *Chem. Rev.*, 2010, **110**, 949–1017.
- 54 R. Ramanan, K. D. Dubey, B. Wang, D. Mandal and S. Shaik, *J. Am. Chem. Soc.*, 2016, **138**, 6786–6797.
- 55 S. Shaik, S. Cohen, S. P. de Visser, P. K. Sharma, D. Kumar, S. Kozuch, F. Ogliaro and D. Danovich, *Eur. J. Inorg. Chem.*, 2004, **2004**, 207–226.
- 56 A. S. Faponle, M. G. Quesne and S. P. de Visser, *Chem.–Euro. J.*, 2016, **22**, 5478–5483.
- 57 R. K. Behera and S. Mazumdar, *Int. J. Biol. Macromol.*, 2010, **46**, 412–418.
- 58 S. K. Manna and S. Mazumdar, *Biochemistry*, 2006, **45**, 12715–12722.
- 59 L. A. Andersson and J. A. Peterson, *Biochem. Biophys. Res. Commun.*, 1995, **211**, 389–395.
- 60 C. Van Stappen, Y. Deng, Y. Liu, H. Heidari, J.-X. Wang, Y. Zhou, A. P. Ledray and Y. Lu, *Chem. Rev.*, 2022, **122**, 11974–12045.
- 61 J. C. Fontecilla-Camps and A. Volbeda, *Chem. Rev.*, 2022, **122**, 12110–12131.
- 62 S. I. Mann, T. Heinisch, T. R. Ward and A. S. Borovik, *Chem. Commun.*, 2018, **54**, 4413–4416.
- 63 S. P. de Visser, *Chem.–Euro. J.*, 2020, **26**, 5308–5327.
- 64 A. Eichler, E. Grigman, S. Herter, P. P. Kelly, N. J. Turner, J. Pleiss and S. L. Flitsch, *ChemBioChem*, 2016, **17**, 426–432.

