

# **MD Simulations and QM/MM Calculations Reveal the Key Mechanistic Elements which are Responsible for the Efficient C-H Amination Reaction Performed by a Bioengineered P450 Enzyme**

Surajit Kalita,<sup>1</sup> Sason Shaik<sup>2\*</sup> and Kshatresh Dutta Dubey<sup>1\*</sup>

<sup>1</sup>Department of Chemistry and Center for Informatics, School of Natural Sciences, Shiv Nadar University, Dadri, Gautam Buddha Nagar, Uttar Pradesh, 201314, India.

<sup>2</sup> Institute of Chemistry, The Hebrew University of Jerusalem, Edmond J Safra Campus, Givat Ram, Jerusalem, 9140401 Israel.

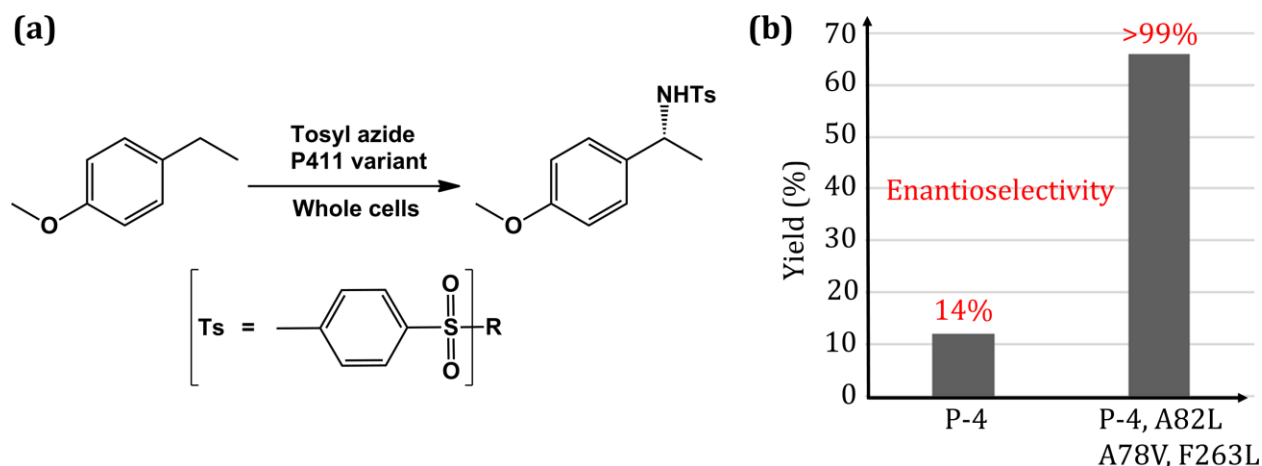
**Abstract:** An enzyme which is capable of catalyzing C-H amination reactions, is considered to be a dream tool for chemists due to its pharmaceutical potential and greener approach. Recently, the Arnold group achieved this feat using an engineered CYP411 enzyme, which further undergoes a random directed evolution which increases its efficiency and selectivity. The present study provides the mechanistic insight and the root cause of the success of these mutations to enhance the reactivity and selectivity of the mutant enzyme. This is achieved by means of comprehensive MD simulations and hybrid QM/MM calculations. The study shows that the efficient C-H amination by the so-engineered CYP411 is a combined outcome of electronic and steric effects. The mutation of axial cysteine ligand to serine relays electron density to the Fe ion in the heme, and thereby enhances the bonding capability of the heme-iron to the nitrogen atom of the tosyl azide. By comparison, the native cysteine-ligated P450, cannot bind the tosyl azide. On top, the A78V and A82L mutations in P411 provide ‘bulk’ to the active site which thereby increases the enantioselectivity via steric effect. At the same time, the QM/MM calculations elucidate the C-H amination by the iron nitrenoid, revealing a mechanism analogous to Compound I in the native-C-H hydroxylation by P450.

## 1. Introduction

Enzymes are efficient nanomachines which have usually evolved for some specific functions. Therefore, tweaking enzymes for functional versatility, and harnessing their catalytic efficiency for commercial applications, has become a holy grail for bioengineers. Due to the versatility in functions and the capability of activating the C-H bond which is a commercially important process, Cytochrome P450 (CYP450) provides an ideal scaffold for bioengineering through directed evolution.<sup>1-11</sup> The native CYP450 uses molecular oxygen and attaches one oxygen to the substrate while the second oxygen is reduced as a water molecule.<sup>12,13</sup> An axial thiolate ligand (Cysteine) that controls the electron density via the *push-pull* effect is the hallmark residue of all CYP450 enzymes.<sup>14,15</sup> Among the members of the P450 family, CYP450<sub>BM3</sub> possesses the widest and most exposed substrate-access channel, and exhibits as such the highest degree of promiscuity among CYP450s. As such, CYP450<sub>BM3</sub> has been widely used as a scaffold for bioengineering of non-native reactions such as Carbene- and Nitrene- transfer reactions.<sup>16,17</sup>

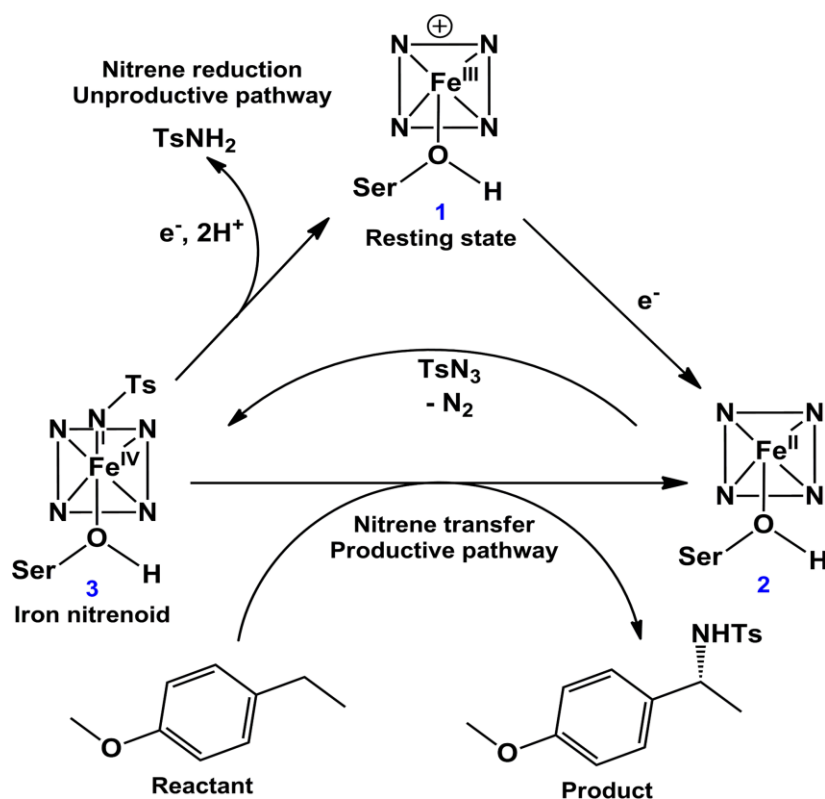
Generally, the naturally occurring CYP450s perform C-H activation through monooxygenation but none of the natural enzymes exhibits in its repertoire C-H bond amination. Since more than 75% of all drugs involve an N-containing heterocyclic ring, this has started a race among biochemists to develop an effective biocatalyst for C-N bond formation using inert C-H bonds.<sup>18,19</sup> Such bioengineering was demonstrated by Gellman in 1985 using a porphyrin mimetic, and then by the Fasan group in 2014 using intra-molecular C-H amination catalyzed by CYP450, albeit with a low yield.<sup>20,8</sup> Ultimately, the Arnold group bioengineered an efficient enzyme, P411, which is a variant of CYP450<sub>BM3</sub>, by mutating the most conserved axial-ligand cysteine to serine.<sup>21</sup> This newly engineered CYP450 variant was sufficiently powerful to accomplish C-H amination reaction, however, the regioselectivity remained uncontrolled. In a subsequent feat of engineering,

the Arnold group used P411 as a scaffold, and reported the first-ever intermolecular C-H amination with significant enantioselectivity.<sup>17</sup> This required the following three key mutations in the P411 scaffold, in Scheme 1.



**Scheme 1:** (a) Intermolecular C-H amination reaction scheme catalyzed by engineered whole-cell P450, (b) Reactivity plot showing the percentage of yield and enantioselectivity for two different mutated variants of P450. Here P4 is an engineered P411.

The C-H amination reaction in Scheme 1a is supposed to be mediated by an active iron-nitrenoid oxidant (complex **3** in Scheme 2), in a catalytic cycle shown in Scheme 2 (note that **3** is a Compound I (Cpd I) analog). As can be seen, the scheme involves three main catalytic steps that begin with a single electron reduction of the resting ferric complex, **1**. The so-formed reduced ferrous complex, **2**, readily reacts with the nitrene source (tosyl azide) and forms a short-lived active oxidant ‘iron nitrenoid’, **3**, that directly facilitates the C-H activation. The third step may bifurcate into either an unproductive nitrene reduction or the productive nitrene transfer, which affects the efficacy of the so engineered enzyme. The root cause of this bifurcation remains an enigma, which is the focus of this work.



**Scheme 2:** A proposed<sup>17</sup> catalytic cycle of P450 variant for intermolecular C-H amination reaction

Thus, this great feat of bioengineering of C-H amination by mutating the axial cysteinate ligand in CYP450, raises several mechanistic puzzles: 1) How does the so assumed iron-nitrenoid active species differ from Cpd I, and how does the swapping of the axial thiolate with serine bring about the unorthodox C-H amination reactions? 2) How do the three-point mutations drastically increase the reactivity and enantioselectivity of the P411 enzyme (Scheme 1b)?

Guided by the above mechanistic questions, we carried out several MD simulations, Density Functional Theory (DFT) calculations, and hybrid QM/MM calculations. We have performed a comprehensive and sequential study starting with the characterization of electronic states of different catalytic steps in Scheme 2, studied the topology of key protein residues by several MD simulations, verified the mechanism of C-H amination by hybrid QM/MM calculation, and

revealed the root cause that triggers the unorthodox C-H amination due to serine mutation. We will see how theoretical calculations coherently explain the elegant choreography of protein matrix engineered by directed evolution and ultimately leads to an efficient and selective C-H amination.

## **2. Computational Methods**

We used molecular docking to generate enzyme-cofactor complexes, MD simulations for the conformational sampling of wild type (WT) and mutant complexes, Density Functional Theory (DFT) calculations for characterization of electronic states, and hybrid QM/MM calculations for exploring the catalytic mechanism. Each of these steps is discussed in detail in the subsequent section.

### **2.1 System Setup**

The starting coordinates for the geometry of the CYP450 variant were taken from the protein data bank of PDB id 5UCW<sup>17</sup> and processed with MODELLER<sup>22</sup> to add missing non-terminal residues. Hydrogen atoms in protein were added using the LEAP module of AMBER20 employing the ff14SB force field. Parametrization for the metal coordinated cluster (iron porphyrin and axial serine) was performed using python based AMBER20 inbuilt Metal Centre Parameter Builder tool.<sup>23</sup> Since the metal coordination of the engineered P411 enzyme is different from its parent CYP450<sub>BM3</sub> enzyme, we characterized the correct ground state geometry of the ferrous complex **2**. The details of the optimized geometry can be found in the SI (c.f. Figure S1). Since the triplet-state is the ground state, charges and other parameters for the subsequent MD simulations were generated for this state. The ligands tosyl azide and 4-Ethylanisole were docked in the active site of protein using AutoDock Vina<sup>24</sup>, and the best pose was considered for MD simulations. The forcefield parameters for ligands were produced using a generalized AMBER force field (GAFF2) in the Antechamber module of AMBER20. The associated partial atomic

charges were also generated by applying the restraint electrostatic potential (RESP) method<sup>25,26</sup> of QM calculated charges at the HF/6-31G(d) level of theory. Subsequently, the systems were solvated in an octahedral box of TIP3P<sup>27</sup> waters extending up to 10 Å from the protein surface. Based on the overall charge of the prepared solvated system, a corresponding number of Na<sup>+</sup> ions were added to neutralize it.

## 2.2 MD Simulations

After proper system setup, the target complexes were subjected to minimization in two steps to remove the poor contacts during system setup. In the first step, only water molecules were minimized while in the second step entire complex was minimized using 5000 steps of steepest descent and subsequently 5000 of conjugate gradient algorithm. Afterward, the systems were gently heated from 10 to 300 K using an NVT ensemble for 50 ps. Following that, we normalized the system under NPT ensemble for 1 ns at a target temperature and pressure of 300 K and 1.0 atm using the Langevin thermostat<sup>28</sup> and Berendsen barostat<sup>29</sup> respectively. Along with that, 2 ps of collision frequency was also applied where the pressure relaxation time was 1 ps. Systems were then equilibrated for the next ~3 ns at the same conditions. The equilibrated systems underwent a further productive MD run of at least 100 ns (depending on the system) using a multi-trajectory approach in which we restarted the simulation after completion of each 50 ns of simulation at a random velocity. The algorithms SHAKE<sup>30</sup> and particle mesh Ewald (PME)<sup>31</sup> were used to constrain the hydrogen bonds and treat the long-range electrostatic forces, respectively. All MD simulations were carried out in the GPU version of the AMBER20 package<sup>32</sup>.

## 2.3 QM/MM Calculations

For the mechanistic study, we used QM/MM calculations employing Chemshell<sup>33,34</sup> that combines Turbomole,<sup>35</sup> for the QM region, and DL\_POLY<sup>36</sup> using the AMBER force field, for the MM part.

All QM/MM calculations were performed on the representative snapshots taken from the MD simulation of complexes **2** and **3** (cf. Scheme 2). In all cases, a truncated heme-porphyrin ring, and the proximal serine (HO-C<sub>2</sub>H<sub>5</sub>) residue were kept in the QM zone along with the reactive ligand of the respective complexes. The representative snapshots were based on the closest available distance of interest of *the most populated MD trajectories*.

The QM optimizations were performed using the UB3LYP/def2-SVP level of theory<sup>37-41</sup> followed by a single point energy calculation using UB3LYP/def2-TZVP as a higher level of theory. The basis set and QM theory were employed here based on similar previous studies in P450 chemistry.<sup>42-44</sup> The energetics were further improved using ZPE (zero-point energy) corrections followed by frequency calculations of optimized reactants (RC), transition state (TS), and product (PC) geometries at the UB3LYP/def2-SVP level of theory. Grimme dispersion (G-D3)<sup>45</sup> was used to add dispersion correction in energetics. The part of protein and water molecules residing up to 8 Å from the QM zone were considered as active atoms and their electrostatic, as well as van der Waal effects, were accounted for by QM calculations. Moreover, an electronic embedding scheme<sup>46</sup> was employed to account for the polarizing effect of the enzyme environment on the QM region. While treating the QM/MM boundary, we used hydrogen link atoms with the charge-shift model.<sup>33,34</sup>

## 2.4 QM only DFT Calculations

The QM-only DFT calculations were performed in *Gaussian 09* software<sup>47</sup> using UB3LYP/B1 level of theory where B1 is defined as LANL2DZ basis set for Fe atom<sup>48-50</sup> and 6-31G(d) for all other atoms.<sup>51,52</sup> The optimized energies were further refined by calculating single-point energy using an all-electron basis set def2-TZVP coupled to UB3LYP. All reported energies are zero-point energy (ZPE) and Grimme dispersion (G-D3) corrected where ZPE values were obtained

from frequency calculation. We optimized the geometry in the gas phase to determine the lowest energy ground state multiplicity followed by a further re-optimization of energetically lowest geometry in chlorobenzene solvent using the SMD solvent model.<sup>53</sup> We chose the chlorobenzene solvent to mimic the non-polar nature of the enzyme environment based on the previous studies.<sup>54,55</sup> The natural- and spin natural -orbital calculations were carried out to identify the presence of singly occupied molecular orbitals and the nature of the electron spin.

### 3. Results and Discussions

We start our study by decoding the enhanced C-H amination activity and regiospecificity due to several site mutations as depicted in Scheme 1b.

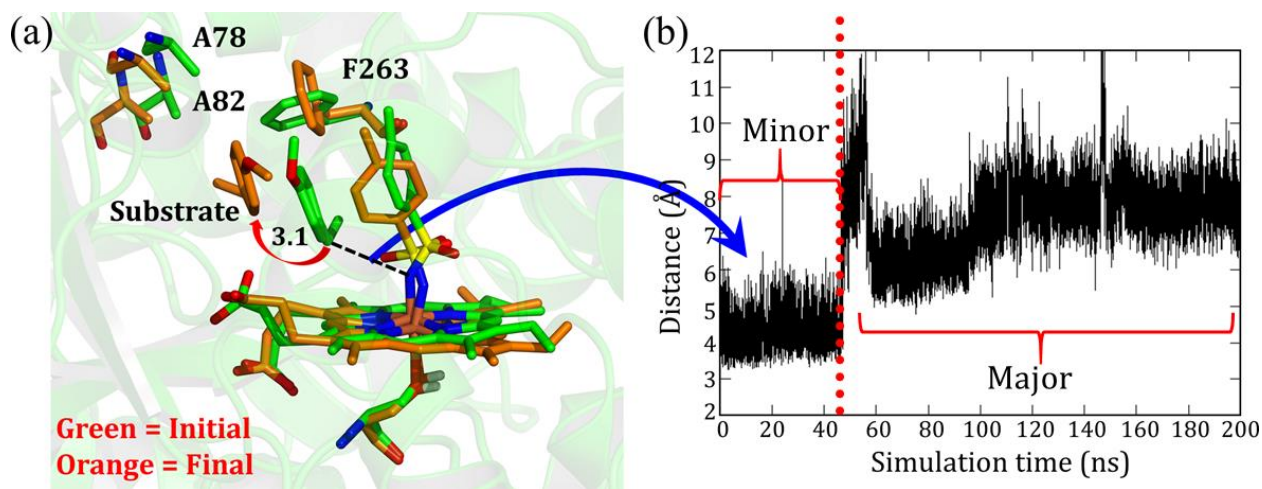
#### 3.1. Decoding enhanced activity due to site-directed mutations in the P411 enzyme

As mentioned, the site-directed mutations (c.f. Scheme 1b) of engineered P411 enzyme enhances the catalytic turnover of C-H amination by several folds and also provides an enantioselective product.<sup>17</sup> However the rationale for the increased activity and selectivity is not apparent and requires elucidation. As such, we intend to show here how theory complements the directed-evolution experiment by providing the underlying mechanistic principles which drive these effects. For simplicity, we named the P-4 variant as **variant 1** while P-4 with additional mutations of A82L, A78V, and F263L as **variant 2**. Note that **variant 1** is less reactive and less enantioselective vis-à-vis **variant 2**.

The simulation of **variant 1** reveals two conformations; (a) initial and less populated (~20%) conformations which we refer to as the minor basin, and is shown by green coloring in Figure 1a; and (b) the highly populated conformations (80%) which is the major basin, shown by orange in Figure 1a. In the minor basin, the substrate is close to the iron nitrenoid (~3.5 Å), and at



the same time, an active site residue F263 is located perpendicular to the substrate. The perpendicular orientation of F263 (green in Figure 1a) applies a restraint on the substrate and limits its flexibility. On the other hand, as shown by the orange color in Figure 1a, in the major basin the substrate moves away from the active oxidant (7-10 Å), and subsequently, the F263 residue flips to a parallel position vis-à-vis the substrate. This reorientation of the F263 frees the substrate from any restraint and, in turn, the substrate acquires flexibility. This might be the root cause for the low activity and less specificity of the substrate in **variant 1**. It is apparent, therefore, that *the MD simulation concisely explains the low activity and specificity for variant 1*. In summary, the Phenylalanine residue (F263) acts as a ringmaster which controls the substrate movement inside the active site by changing its conformation from perpendicular to parallel orientation.

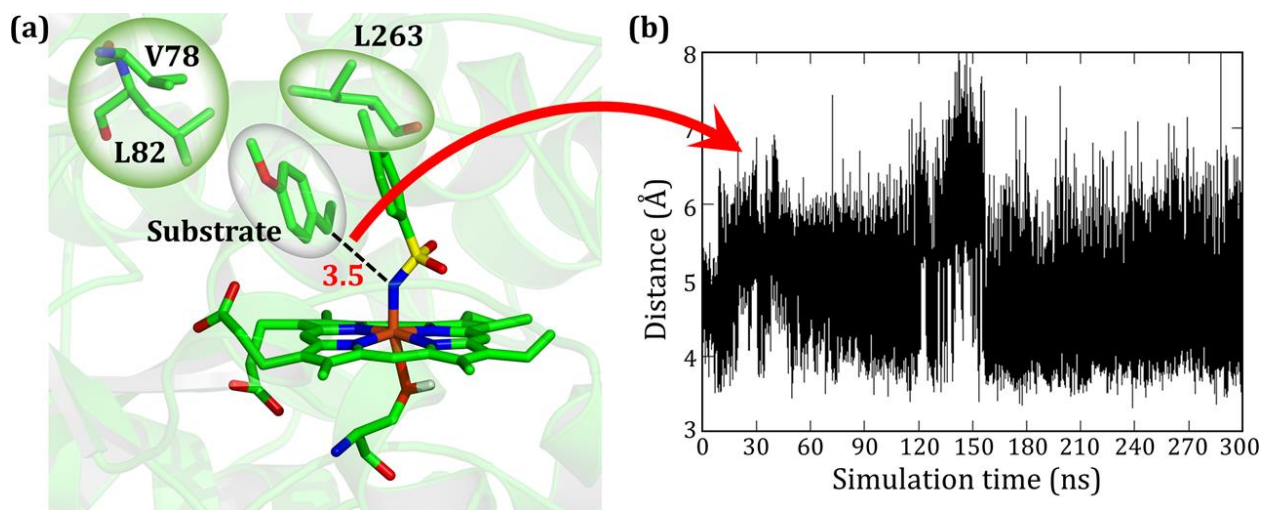


**Figure 1:** (a) Superimposed diagram showing two different conformations of **variant 1** obtained at two different time scales of the simulation, (b) A plot of the distance over time, between the benzylic carbon of the substrate and the nitrogen of nitrenoid.

As stated earlier, the mutations of A82L, A78V, and F263L in **variant 2** significantly enhance the C-H amination activity and enantioselectivity (>99%) relative to **variant 1**. Therefore, we performed MD simulations for this variant to uncover the roots for this change in activity.

Interestingly, during the MD simulations of **variant 2**, the substrate stays close to the oxidant ( $\sim 4\text{\AA}$ ) for more than 90% of the entire 300 ns simulations and remains quite stable (See Figure 2).

As seen in **variant 1**, the substrate was captivated by F263 (Phe 263) via a strong  $\pi$ - $\pi$  interaction, and therefore a mutation of Phe to Leu in **variant 2** removes the  $\pi$ - $\pi$  interaction and allows the substrate to change its orientation. At the same instant, the substrate finds a new  $\pi$ - $\pi$  interaction *with the aromatic ring of tosyl moiety of iron nitrenoid*. Due to the new  $\pi$ - $\pi$  interaction, the substrate snuggles close to the tosyl moiety of the oxidant for the entire simulation. *Therefore, the F263L mutation exerts a binding advantage that contributes to the enhanced activity.*



**Figure 2:** (a) A representative MD snapshot of variant 2 showing the probable interaction between mutated residues and substrate in the reactive position. The different bubbles represent the hydrophobic space occupied by the respective moiety and their interaction. (b) Distance between substrate and nitrogen of nitrenoid for the entire time of the simulation.

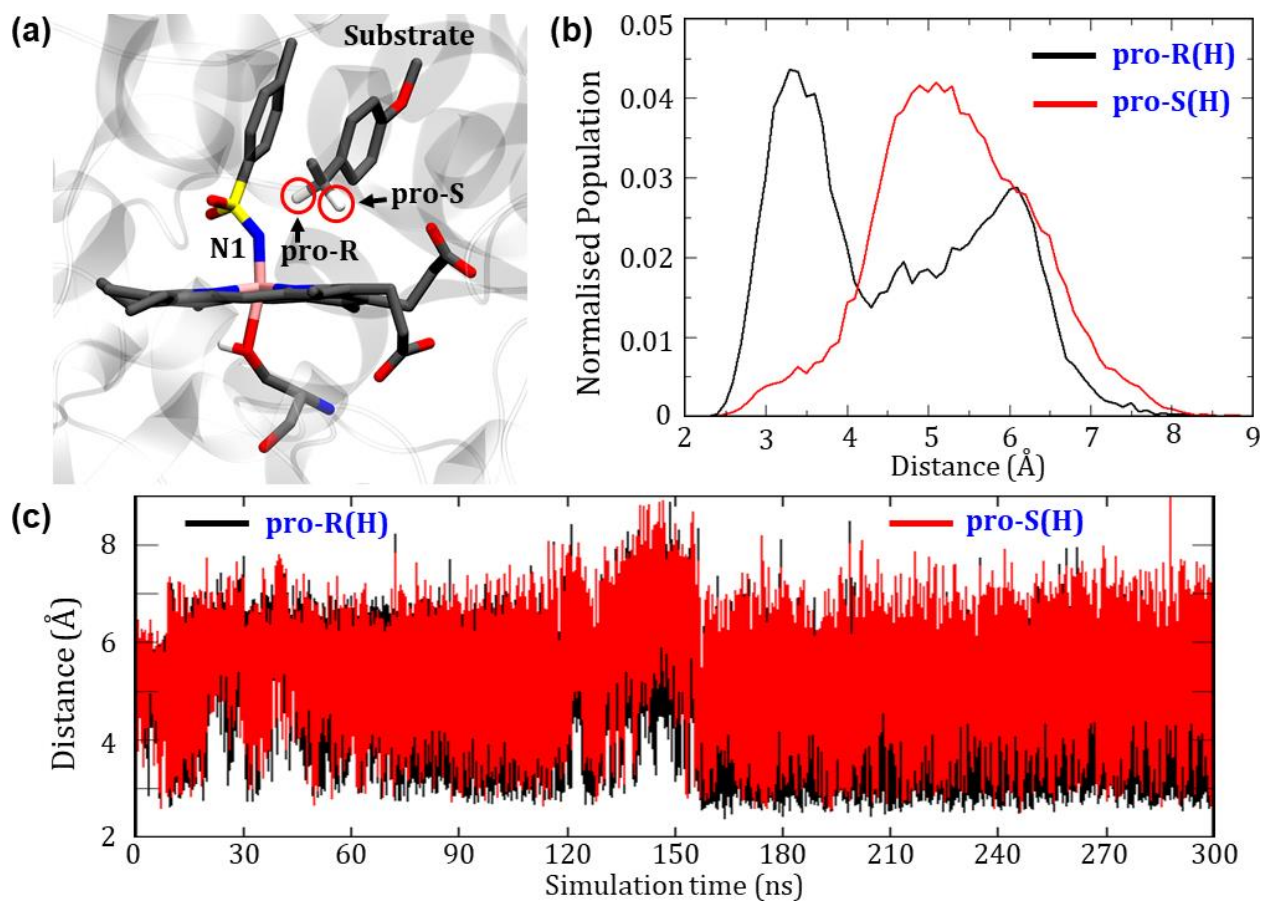
How do the mutations of A78V and A82L augment the enantioselectivity of the reaction? Being non-polar residues, valine (V) and leucine (L) does not change the electrostatic and polar environment of the active site, and at the same time, these mutations increase the rigidity of the active site due to elongated side chains vis-à-vis Alanine (A). This extra “filling” of the active site

is necessary for enantioselectivity. Thus, *the smart bioengineering which enhances the C-H amination is efficiently decoded by the MD simulation.*

### ***MD Simulation explains the product enantioselectivity***

Can the simulation also predict the observed pro-*R* selectivity over pro-*S*? The answer is yes, and this is shown in Figure 3.

Figure 3a depicts a representative snapshot from the MD simulations and highlights the pro-*R* and pro-*S* hydrogens. Figure 3 shows the evolution of distances of these hydrogens from the reactive N1 atom of the oxidant. It is therefore apparent that the pro-*R* hydrogen is significantly closer to N1 compared with the pro-*S* hydrogen. We further calculated the Boltzmann population of the pro-*R* and pro-*S* distances over the entire 300 ns as shown in Figure 3b. Using Figure 3b, it is quite clear that the pro-*R*(H) is populated close to the region of 3 Å for most of the simulation time while pro-*S*(H) stays at 5-6 Å distance from N1. This prediction of enantioselectivity of pro-*R*(H) is in good agreement with the experimental observation of Arnold *et. al.*<sup>17</sup> and hence showing *that our MD simulations are sufficiently accurate to mimic the experimental enantioselectivity.*

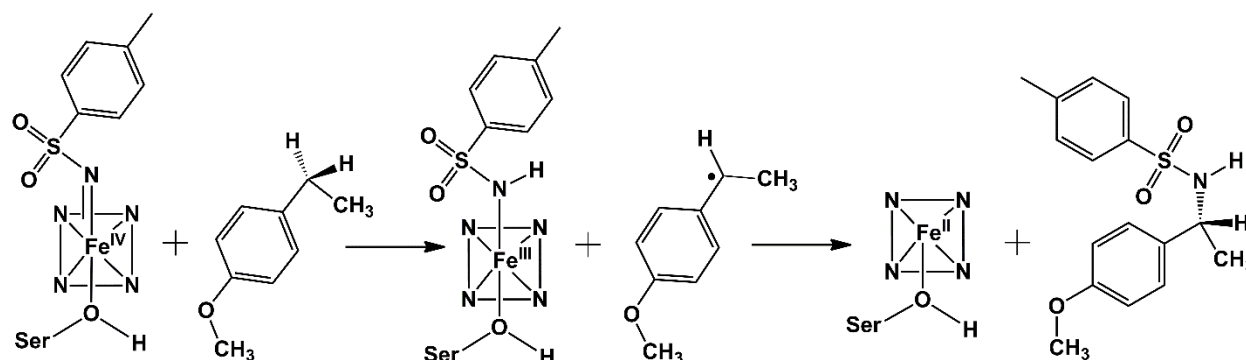


**Figure 3:** (a) A representative MD snapshot displaying the pro-*R* and pro-*S* hydrogens of the substrate. (b) The Boltzmann population of the pro-*R* and pro-*S* distances over the entire 300 ns simulation (c) Distance plots between these hydrogens and N1 of nitrenoid.

### 3.2 QM/MM mechanistic investigation of intermolecular C-H amination reaction

As can be seen, the engineered P411 is entirely different from its parent P450 due to its novel serine-ligated heme-porphyrin structure, therefore, the electronic features which dictate the catalytic mechanism of P411 should be established by means of *quantum mechanical calculations*. We, therefore, performed a comprehensive mechanistic study of C-H amination using hybrid QM/MM calculations.

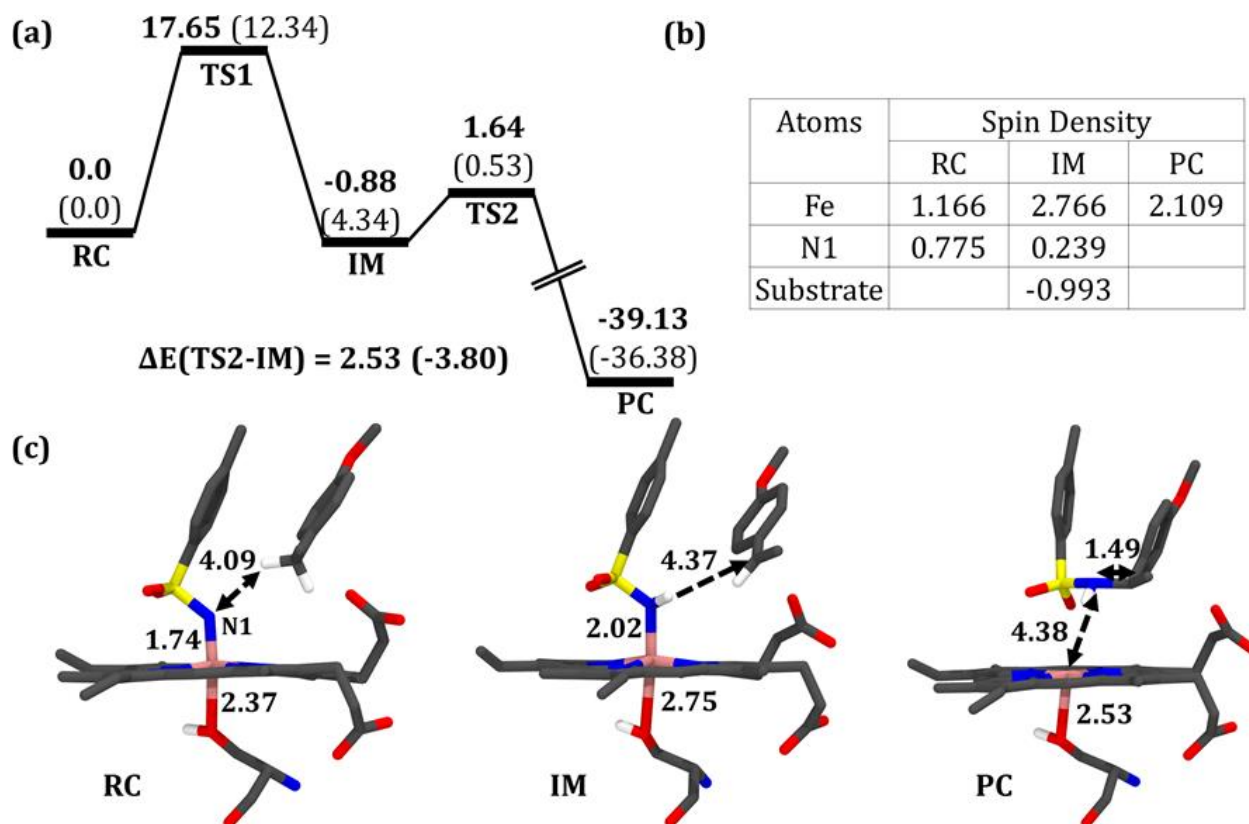
Scheme 3 shows a putative mechanism of this reaction. Initially, the nitrogen atom (N1) abstracts the benzylic C<sub>sp</sub><sup>3</sup>-H atom and forms a reactive intermediate and a radical substrate. Subsequently, these two newly formed species mutually couple to generate the C-H aminated product and a ferrous complex of P411.



**Scheme 3.** The plausible mechanism of C-H amination.

To validate this mechanism, we started our QM/MM calculations by optimizing a representative MD snapshot from the simulation of **variant 2**. The snapshot was chosen based on the closest distance between the benzylic pro-R(H) of substrate and N1 of nitrenoid. An energy scanning was carried out for abstracting the pro-R(H) and leading to the formation of a highly reactive intermediate complex as well as a radical substrate. *Subsequent energy scanning resulted in product formation via rebound mechanism as found in native P450 enzymes.* The energy profile diagram and the key geometries are organized in Figure 4.

In the first step, the reactive intermediate complex (IM) is formed by abstracting the hydrogen at the cost of a moderate energy barrier of 17.7 kcal/mol, which is lowered to 12.3 kcal/mol using the more extensive basis set. This less exothermic step is rate-determining. Subsequently, IM proceeds through the radical rebound mechanism that possesses a tiny energy barrier of 2.5 kcal/mol and forms the C-H aminated product.

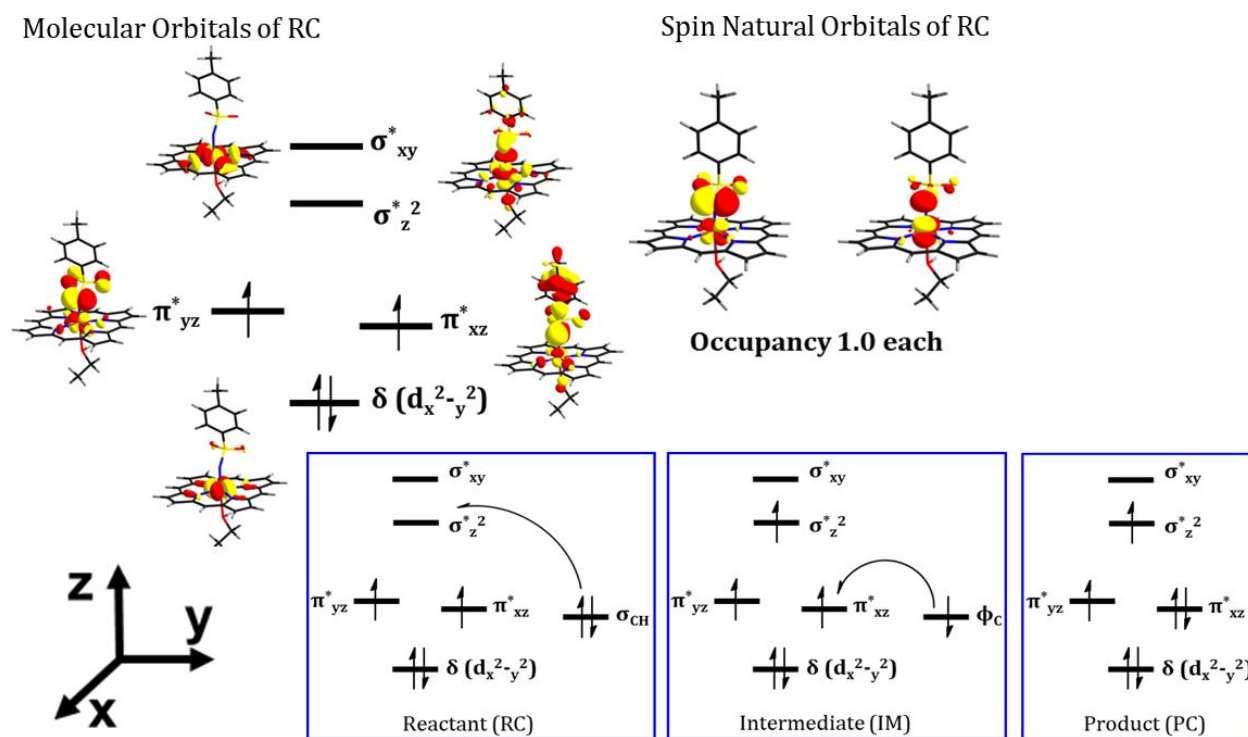


**Figure 4:** (a) A complete reaction profile for the intermolecular C-H amination. Energies (in kcal/mol) are relative to the reaction complex (RC). Values in parenthesis are single-point energies in the better basis set. All energies are corrected for zero-point energy (ZPE) and G-D3 dispersion. (b) Spin densities in RC, the reaction intermediate (IM), and the product cluster (PC). (c) Optimized geometries for RC, IM, and PC (from left to right); respective bond lengths are in Å unit. The optimized geometry of TS1 can be found in the SI.

As can be seen, the *QM/MM* calculations show that the mechanism of the C-H amination reaction by engineered P411 is essentially similar to the C-H oxidation mechanism by the native P450 enzyme. However, whether it is completely identical to the native P450 enzyme including the involvement of porphyrin radical cation and compound II type intermediate, is not clear from the energy profile. Therefore, we calculated the spin density of the species RC, IM, and PC in Figure 4b and detailed electronic structures of RC. The calculations reveal in Figure 5 two unpaired electrons at the antibonding  $\pi$  orbitals of Fe-N bond in RC which is also supported by the Spin



Natural orbital calculations shown in Figure 5. This electronic structure of RC (iron nitrenoid) resembles compound I except for a radical cation at the porphyrin.<sup>56</sup>



**Figure 5.** The electronic-structure details of RC and the ensuing changes in orbital occupation during the amination reaction. The singly occupied orbitals on the righthand-side are the  $\pi^*$  spin natural orbitals (SNOs) of the active oxidant (Iron Nitrenoid).

Using the spin densities as shown in Figure 4b we further depicted the occupation of the key orbitals throughout the reaction pathway shown in Figure 5. In the H-abstraction step an electron, initially in a  $\sigma_{CH}$  orbital of substrate, shifts to the unoccupied high energy  $\sigma_z^{*2}$  orbital of the active oxidant and produces the intermediate IM. In this species, there are three identical-spin electrons (due to orbital delocalization only 2.8 according to population analysis), while one down-spin electron is localized at the benzylic C-atom of the substrate, with a small extent of delocalization to the phenyl ring (hence, population analysis shows -0.993). In the rebound step, the substrate

formally donates its electron to the Fe atom resulting in the formation of the product molecule and the ferrous heme-porphyrin complex.

As such, the active species of P411 is an analog of the hydroxo-iron(IV) Cpd II intermediate in native P450s, having two singly occupied  $\pi^*$  orbitals, which here acts as an H-abstractor. Thus, *QM/MM mechanistic studies provide us the strong energetic and electronic evidence in favor of our proposed pathway and reveal a native P450-like mechanism despite the absence of a Cpd I-like species.*

### 3.3. Formation of active oxidant-Iron Nitrenoid in P411

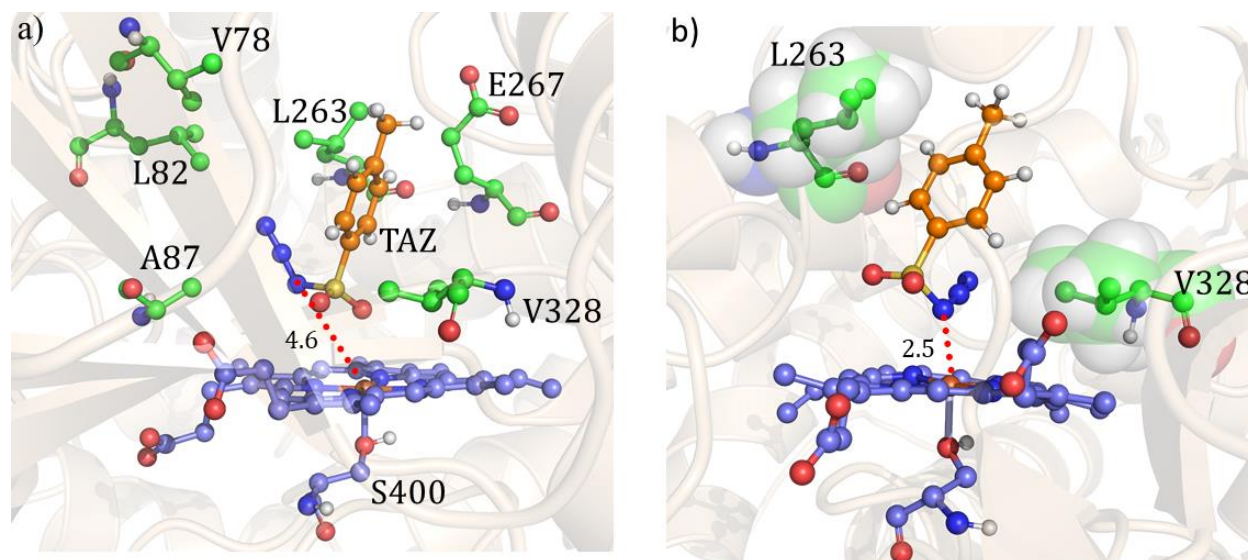
While we achieved an understanding of the C-H amination reaction by the bioengineered P411, this creates an additional mechanistic puzzle: Why is the native cysteine ligand unable to promote the C-H amination? This obviously requires us to understand the role of the mutation of the most conserved cysteine residue to serine. To this end, we performed several MD simulations and QM/MM calculations which are discussed below.

We believe that the key to solving the above mechanistic puzzle might be associated with the ease of formation of the iron-nitrenoid active oxidant. *We therefore proceeded to compare the mechanisms of formations of the serine-ligated vs. cysteine-ligated iron-nitrenoid P411 species.*

Figure 6 shows the two conformations of the distal tosyl azide (TAZ) of P411; before and after MD simulations. As can be seen from Figure 6a, the TAZ is initially far from the heme iron (the respective distance between N1 and Fe is 4.6 Å). However, during the simulation, the distance reduces to 2.53 Å (See Figure 6b) for 30% of the sampled MD trajectory. A closer inspection of the MD trajectory also shows that the proximity of the distal ligand with heme-iron is strongly correlated with the juxtapositions of L263 and V328 (See Figure S2 for graphs showing the

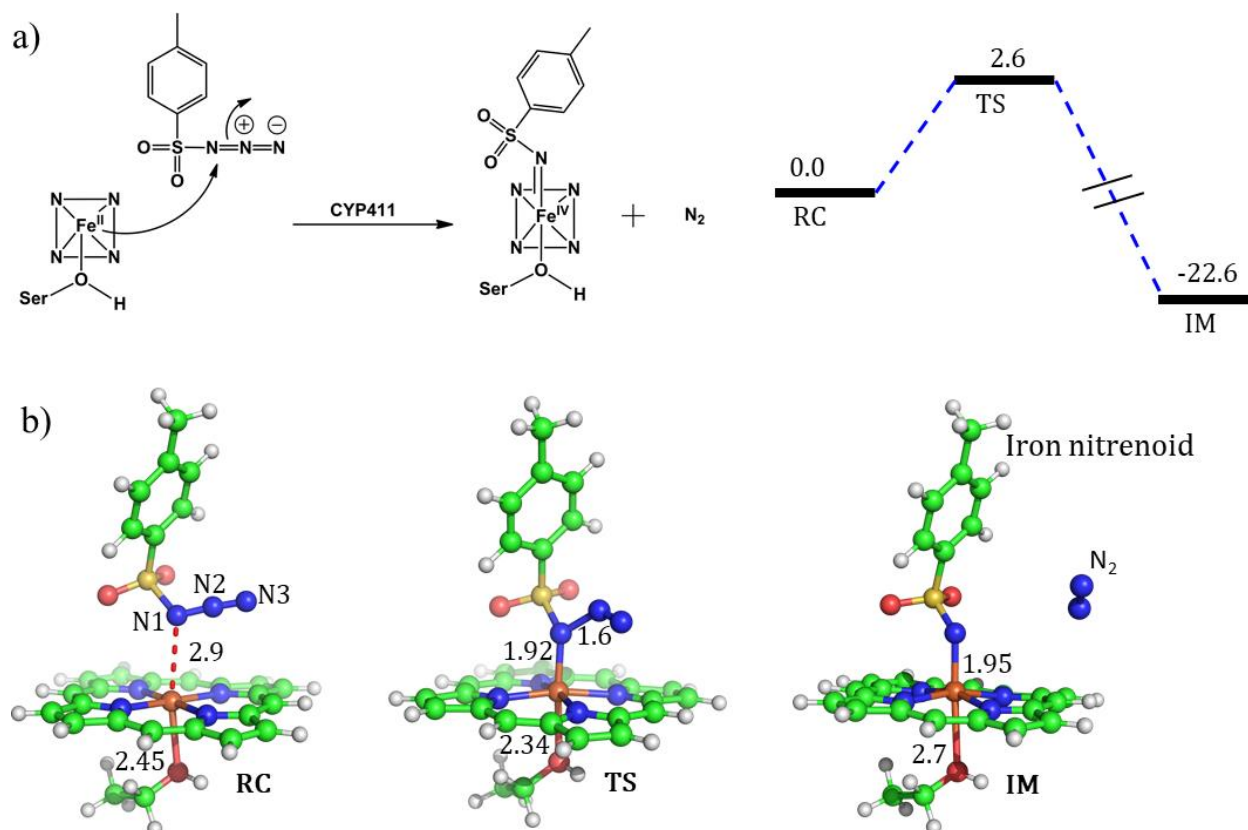


correlation with distance). It is apparent that these residues provide a tight packing to the distal ligand, and therefore, the relative position of these residues directly affects the orientation of the ligand.



**Figure 6.** The precursor enzyme with a serine axial ligand (S400): (a) Geometry of the docked Tosyl azide (TAZ), and the identified active site residues based on reference 5b. (b) A representative MD snapshot showing the most probable interaction of the TAZ ligand with different residues of enzyme.

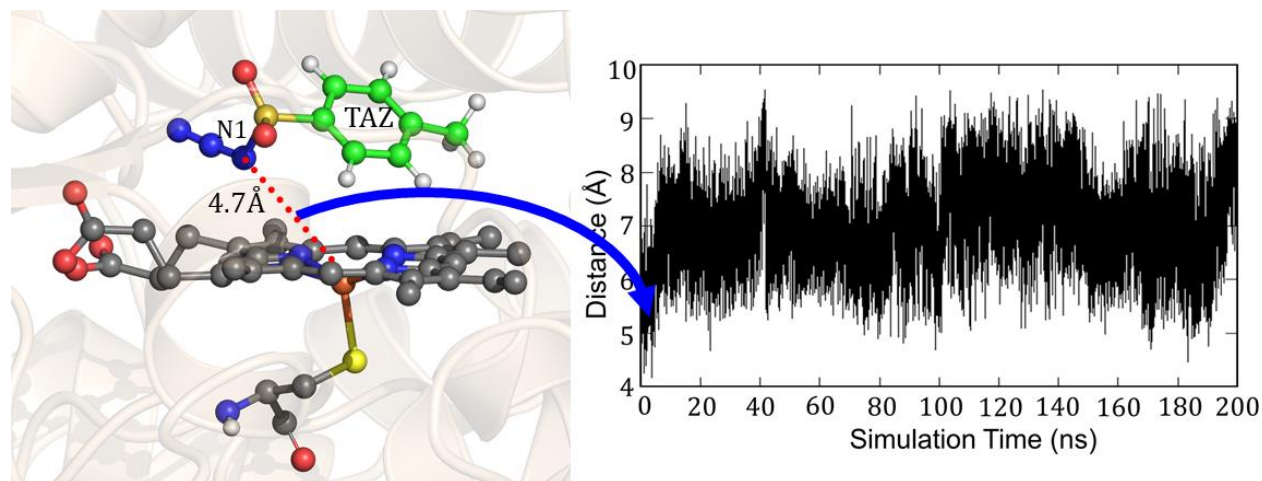
For the mechanism of formation of the active oxidant, iron nitrenoid, we followed with QM/MM calculations for a representative snapshot from MD simulations. We started the calculations with the optimization of the reactant followed by potential energy scanning to trace the reaction coordinate for the formation of the iron nitrenoid. The energy profile for the reaction is shown in Figure 7. As can be seen, the activation barrier for the formation of active oxidant i.e., iron nitrenoid, is just 2.6 kcal/mol. Moreover, this process takes place in a concerted displacement reaction; the Fe—N1 bond is formed and at the same time the N1—N2 bond is broken leaving behind the active oxidant iron nitrenoid and molecular nitrogen.



**Figure 7:** (a) A schematic mechanism for the formation of iron nitrenoid complex, and the corresponding reaction profile calculated by hybrid QM/MM calculations at B3LYP-D3/def2-SVP level of theory. Reported energies are ZPE corrected from the subsequent frequency calculation at the same level of theory. Energies are in kcal/mol and relative to the reactant complex (RC), (b) The optimized geometries of the RC, TS, and IM species during the reaction mechanism; respective bond distances are in Å unit.

As such, our QM/MM calculations show that the rate of formation of the iron nitrenoid active oxidant is by far faster than that of the analogous process which generates Cpd I for the native CYP450<sub>BM3</sub> enzyme where cysteine is the axial ligand.<sup>44</sup> The corresponding barrier for this Cpd I formation process is 15.7 kcal/mol.<sup>44</sup> Hence, our theoretical mechanistic investigation shows that the engineered enzyme produces the iron nitrenoid more efficiently than its functional analog Cpd I in the native P450 enzyme.

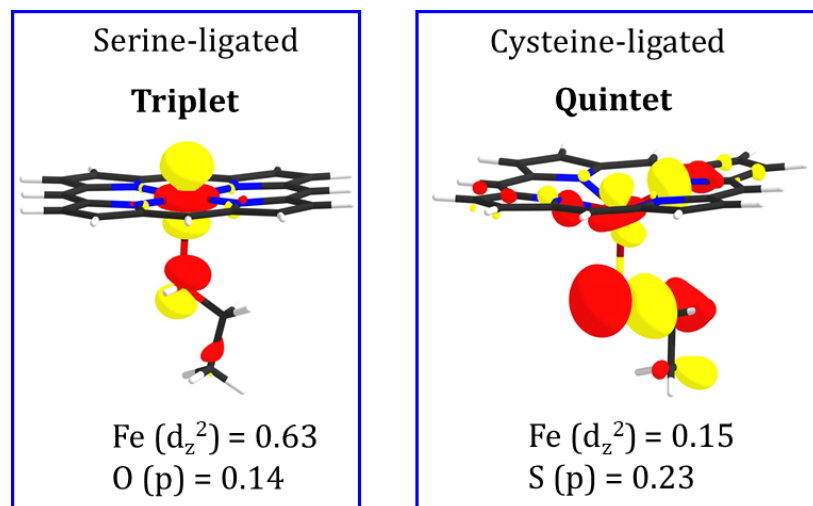
But why does the native enzyme with the cysteine ligand fails to create the iron nitrenoid oxidant? To answer this question, we mutated the proximal serine to cysteine in the engineered P411 and performed 200 ns of MD simulation. Interestingly, now, the tosyl azide ligand never approaches the heme-porphyrin during the entire 200 ns of simulation of cysteine ligated P411 complex. As can be seen in Figure 8, the average distance between Fe and N1 is  $\sim 7$  Å and the lowest possible distance is 4.7 Å. In fact, the QM/MM optimization (see Figure S3) also reveals that the ligand moves away from its original position by a large distance much the same as the MD results.



**Figure 8:** A representative MD snapshot indicating one of the shortest distances between N1 of tosyl azide (TAZ) and the Fe ion of the heme-porphyrin and the variation of this distance during the simulation.

To pinpoint the trigger of this change in the distance of  $\text{Fe}^{\text{II}}\text{---TAZ}$  when serine is replaced by cysteine, we plotted in Figure 9 the molecular orbitals which are responsible for the  $\text{Fe}^{\text{II}}\text{---N1}$   $\sigma$  bonds between the ferrous ion and TAZ. Thus, the serine-ligated complex exhibits a bond-making orbital which is well-located on the  $\text{Fe}^{\text{II}}$  ion (see Figure 9; the weight contribution of Fe to the  $d_z^2$  MO is 0.63). In contrast, the cysteine-ligated ferrous complex has a quintet ground spin state (cf. Figure S3), and its  $\text{Fe}^{\text{II}}\text{---N1}$  bond making orbital has a small weight contribution of  $\text{Fe}^{\text{II}}$  (0.15) in

the respective MO. It is apparent therefore that the corresponding iron ion, in the cysteine ligated heme, will coordinate the TAZ very feebly. On the other hand, the high orbital density for the serine-ligated iron, creates a stronger binding site for TAZ.



**Figure 9:** Molecular orbitals which participate in  $\sigma$  bond formation of  $\text{Fe}^{\text{II}}$  with N1 of TAZ. The orbitals are drawn to the same scale, and the relative sizes on the iron reflect the respective orbital weight. The orbital on the left-hand side is for the serine-ligated heme, while the orbital for the cysteine-ligated heme is depicted on the right-hand side. The spins in the respective ground states are indicated near the orbital drawings. The numbers underneath the MOs are weight contribution contributions of Fe to the  $d_z^2$  molecular orbital.

In a nutshell, *our theoretical investigation beautifully explains the enhanced activity of the C-H amination in cysteine  $\rightarrow$  serine mutation and complements the experimentally observed results.*<sup>17</sup>

#### 4. Conclusions:

The present study provides a rationale and logical explanation of the highly successful engineering that leads the unorthodox C-H amination reaction. Using MD simulations and hybrid QM/MM calculations we have shown that the enhanced C-H amination activity and its enantioselectivity are jointly determined by well-defined electronic and steric effects. The mutation of cysteine $\rightarrow$ serine of the proximal ligand in the engineered P411 enzyme provides a favorable

electronic effect that increases the orbital population on Fe atom vis-à-vis native cysteine ligated P450, and in turn, triggers the C-H amination reactions in P411 enzyme. Similarly, the mutation of A78V and A82L in variant 2 of P411 enzyme provides a, ‘extra filling’ to the active site which increases the enantioselectivity and thus have steric effect. Moreover, MD simulations beautifully explain how a mutation of F263 to L263 can significantly enhance the reactivity by switching its interacting partner from substrate 4-Ethylanisole to distal ligand. Our study supplemented by QM/MM calculations provides a valuable insight that engineered enzyme P411 follows a native P450- like mechanism where iron-nitrenoid acts as active oxidant analogous to Cpd II, but with a more potent ability to abstract H.

As such, the present study shows that the MD simulations and QM/MM calculations complement the bioengineering involved in directed evolution, and elucidate the factors which make this engineering so successful.

## **5. Acknowledgement**

KDD acknowledges the Department of Biotechnology Govt of India for the Ramalingaswami re-entry grant. SS acknowledges Israel Science Foundation for ISF 520/18 grant. SK acknowledges the Council of Scientific and Industrial Research for JRF.

## **Corresponding Authors Details:**

1. Prof. Sason Shaik, Institute of Chemistry, The Hebrew University, Edmond. J. Safra Campus, Givat Ram, Jerusalem 9190400, Israel. [sason@yfaat.ch.huji.ac.il](mailto:sason@yfaat.ch.huji.ac.il) ORCID 0000-0001-7643-9421,
2. Dr. Kshatresh Dutta Dubey, Center for Informatics, Department of Chemistry, School of Natural Science, Shiv Nadar University Delhi-NCR, NH91 Tehsil Dadri, Greater

Noida, Uttar Pradesh 201314, India [kshatresh.dubey@snu.edu.in](mailto:kshatresh.dubey@snu.edu.in) [0000-0001-8865-7602]

3. Surajit Kalita, Department of Chemistry, School of Natural Science, Shiv Nadar University Delhi-NCR, NH91 Tehsil Dadri, Gautam Buddha Nagar, Uttar Pradesh 201314, India. ORCID: 0000-0001-5343-0694

**Author's Contribution:** SK performed all calculations. KDD and SS designed the project and wrote the manuscript.

**Conflict of Interest:** Authors declare no conflict of interest.

## References:

1. F. H. Arnold, *Acc. Chem. Res.*, 1998, **31**, 125-131.
2. P. A. Romero and F. H. Arnold, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**, 866-876.
3. F. H. Arnold, *Angew. Chem. Int. Ed.*, 2019, **58**, 14420-14426.
4. K. Chen and F. H. Arnold, *Nat. Catal.*, 2020, **3**, 203-213.
5. J. B. Wang, G. Li and M. T. Reetz, *Chem. Commun.*, 2017, **53**, 3916-3928.
6. H. Renata, Z. J. Wang and F. H. Arnold, *Angew. Chem. Int. Ed.*, 2015, **54**, 3351-3367.
7. J. A. McIntosh, P. S. Coelho, C. C. Farwell, Z. J. Wang, J. C. Lewis, T. R. Brown and F. H. Arnold, *Angew. Chem. Int. Ed.*, 2013, **52**, 9309-9312.
8. R. Singh, M. Bordeaux and R. Fasan, *ACS Catal.*, 2014, **4**, 546-552.
9. C. K. Prier, T. K. Hyster, C. C. Farwell, A. Huang and F. H. Arnold, *Angew. Chem. Int. Ed.*, 2016, **55**, 4711-4715.
10. S. Kille, F. E. Zilly, J. P. Acevedo and M. T. Reetz, *Nat. Chem.*, 2011, **3**, 738-743.
11. M. T. Reetz, *J. Am. Chem. Soc.*, 2013, **135**, 12480-12496.

12. *Cytochrome P450: Structure, Mechanism and Biochemistry*, O. de Montellano and P. R. (Ed.), 2<sup>nd</sup> Edition, Plenum Press: New York, 1995.
13. B. Meunier, S. P. de Visser and S. Shaik, *Chem. Rev.*, 2004, **104**, 3947–3980.
14. J. H. Dawson, *Science*, 1988, **240**, 433-439.
15. J. T. Groves, *Nat. Chem.*, 2014, **6**, 89-91.
16. P. S. Coelho, E. M. Brustad, A. Kannan and F. H. Arnold, *Science*, 2013, **339**, 307-310.
17. C. K. Prier, R. K. Zhang, A. R. Buller, S. B. Chen and F. H. Arnold, *Nat. Chem.*, 2017, **9**, 629-634.
18. N. Kerru, L. Gummidi, S. Maddila, K. K. Gangu and S. B. Jonnalagadda, *Molecules*, 2020, **25**, 1909.
19. D. Schroder and H. Schwartz., *Proc. Natl. Acad. Sci.*, 2008, **105**, 18114-18119.
20. E. W. Svastits, J. H. Dawson, R. Breslow and S. H. Gellman, *J. Am. Chem. Soc.*, 1985, **107**, 6427–6428.
21. P. S. Coelho, Z. J. Wang, M. E. Ener, S. A. Baril, A. Kannan, F. H. Arnold and E. M. Brustad, *Nat. Chem. Biol.*, 2013, **9**, 485-487.
22. A. Šali and T. L. Blundell, *J. Mol. Biol.*, 1993, **234**, 779-815.
23. P. Li, K. M. Merz, and Jr., *J. Chem. Inf. Model.*, 2016, **56**, 599–604.
24. O. Trott, and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455-461.
25. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
26. W. D. Cornell, P. Cieplak, C. I. Bayly and P. A. Kollman, *J. Am. Chem. Soc.*, 1993, **115**, 9620–9631.

27. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
28. J. A. Izaguirre, D. P. Catarello, J. M. Wozniak and R. D. Skeel, *J. Chem. Phys.*, 2001, **114**, 2090–2098.
29. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
30. J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
31. T. Darden, D. York and L. Pederson, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
32. R. S. Ferrer, A. W. Götz, D. Poole, S. L. Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878–3888.
33. 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.
34. S. Metz, J. Kästner, A. A. Sokol, T. W. Keal and P. Sherwood, *Comput. Mol. Sci.*, 2014, **4**, 101–110.
35. R. Ahlrichs, M. Bär, M. Häser, H. Horn and C. Kölmel, *Chem. Phys. Lett.*, 1989, **162**, 165–169.
36. W. Smith and T. R. Forester, *J. Mol. Graphics*, 1996, **14**, 136–141.
37. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
38. A. D. Becke, *J. Chem. Phys.*, 1992, **96**, 2155–2160.
39. C. Lee, W. Yang, R. and G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
40. F. Weigenda, R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.



41. F. Weigenda, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057-1065.
42. S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar and W. Thiel, *Chem. Rev.*, 2010, **110**, 949–1017.
43. K. D. Dubey, B. Wang and S. Shaik, *J. Am. Chem. Soc.*, 2016, **138**, 837–845.
44. S. Kalita, S. Shaik, H. K. Kisan and K. D. Dubey, *ACS Catal.*, 2020, **10**, 11481–11492.
45. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456.
46. D. Bakowies and W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580–10594.
47. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, *Gaussian 09*, Revision **D.01**, Gaussian, Inc., Wallingford CT, 2013.
48. P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 270-283.
49. W. R. Wadt and P. J. Hay, *J. Chem. Phys.*, 1985, **82**, 284-298.
50. P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 299-310.
51. W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257-2261.

52. M. J. Frisch, J. A. Pople and J. S. Binkley, *J. Chem. Phys.*, 1984, **80**, 3265-3269.
53. A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378-6396.
54. D. Usharani, D. Janardanan and S. Shaik, *J. Am. Chem. Soc.*, 2011, **133**, 176–179.
55. D. A. Sharon, D. Mallick, B. Wang and S. Shaik, *J. Am. Chem. Soc.*, 2016, **138**, 9597–9610.
56. Y. Moreau, H. Chen, E. Derat, H. Hirao, C. Bolm and S. Shaik, *J. Phys. Chem. B*, 2007, **111**, 10288-10299.

## TOC

