The Performance of Different Water Models on the Structure and Function of Cytochrome P450 Enzymes

Shalini Yadav†, Vandana Kardam†, Ankita Tripathi, Shruti T G and Kshatresh Dutta Dubey*

Department of Chemistry, School of Natural Sciences, Shiv Nadar University Delhi-NCR, Gautam Buddha Nagar, U.P. 201314, India.

† Authors have equal contribution.

*Corresponding Author: kshatresh.dubey@snu.edu.in, kshatresh@gmail.com

Abstract: Modelling approaches and modern simulations to investigate the biomolecular structure and function rely on various methods, one among which is the choice of water model. Water molecules play a crucial role in all sorts of chemistry. Cytochrome P450 (CYP450), in particular, water molecules are crucial for the formation of active oxidants which perform the oxidation and metabolism of several substrates. Computational chemistry tools such as MD simulations and QM/MM calculations, nowadays, have become complementary tools to study the structure and functions of CYP450 enzymes, and therefore, accurate modeling of water molecules is crucial. In the present study, we have highlighted the behavior of the three most widely used water models—TIP3P, SPC/E, and OPC for three different CYP450 enzymes—$CYP450_{BM3}$, $CYP450_{OleT}$, and $CYP450_{BS\beta}$ during MD simulations and QM/MM calculations. We studied the various properties such as RMSD, RMSF, H-bond, water occupancy in the first solvation shell, and Hydrogen Atom Transfer (HAT) using QM/MM calculations and compared them for all the three water models. Our study shows that, the stability of the enzymes structure is well maintained in all the three water models. However, OPC water model performs well for the polar active sites, i.e., in $CYP450_{OleT}$, $CYP450_{BS\beta}$ while the TIP3P water model is superior for the hydrophobic site such as $CYP450_{BM3}$.

Keywords: Water Models; MD simulations; Cytochrome P450; QM/MM calculations; Reactivity

1. Introduction:

Active site water molecules are considered an important component in biological systems.[1–4] They can affect the binding affinity of a protein-ligand complex by bridging ligand and protein through hydrogen bonds, and can also catalyze several biologically relevant reactions via shuttling proton channels.[5–10] In cytochrome P450 (CYP450), in particular, the active site water molecules play a crucial role in the catalytic functions and mechanism.[11–15] In fact, the water molecule starts playing a role from the first step of CYP450 itself where it vacates the sixth coordination place for the substrate binding.[16] In the proceeding steps of the catalytic cycle it also plays a role as a proton donor during the formation of the Cpd 0 and Cpd I.[17,18] Furthermore, the presence of a water molecule can significantly lower the transition state barrier in CYP450s.[13] With an ever-increasing use of computational chemistry tools to study the mechanism and functions of CYP450, accurate modeling of the water molecules in the CYP450s has become an important challenge for a computational biochemist. In the last four decades, several water models have been proposed where the three interaction site models such as TIP3P (Transferrable Intermolecular Potential 3P), SPC (Simple Point Charge), and SPC/E (Extended Simple Point Charge) –have been used extensively.[19–21] Although, all these water models represent similar interaction site, they show significant differences in the bulk properties.[22] Very recently, Onufriev's group developed a new water model, OPC which differs significantly in the charges and the bond length.[23] This new water model resulting into 3-charge, 4-point rigid water model, reproduces a comprehensive set of bulk properties significantly more accurate than commonly used water models (see ref).[23]
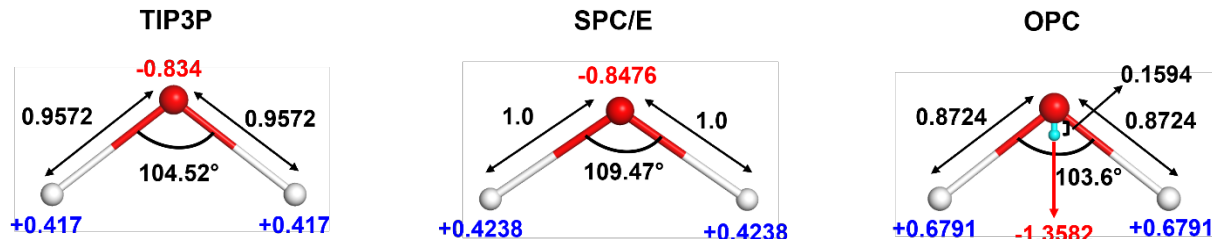
Figure 1. Structure and parameters (bond length, angle, charges) for different water models.

In the present study, using SPC/E, TIP3P, and OPC water models, we have highlighted the effect of different water models on the active site orientations and reactivity of CYP450 enzymes. For so doing, we have chosen CYP450$_{BM3}$,[24] CYP450$_{BSβ}$,[25], and CYP450$_{OleT}$ [26] as model systems since they catalyze similar fatty acids but possess entirely different reactivity patterns. For instance, in the CYP450$_{BM3}$ the fatty acid enters the active site via its hydrophobic tail part which lies above the heme plane. The presence of an alkyl substrate tail and F87 residue in the heme cavity makes the reaction center of CYP450$_{BM3}$ hydrophobic (cf. Figure 2a).
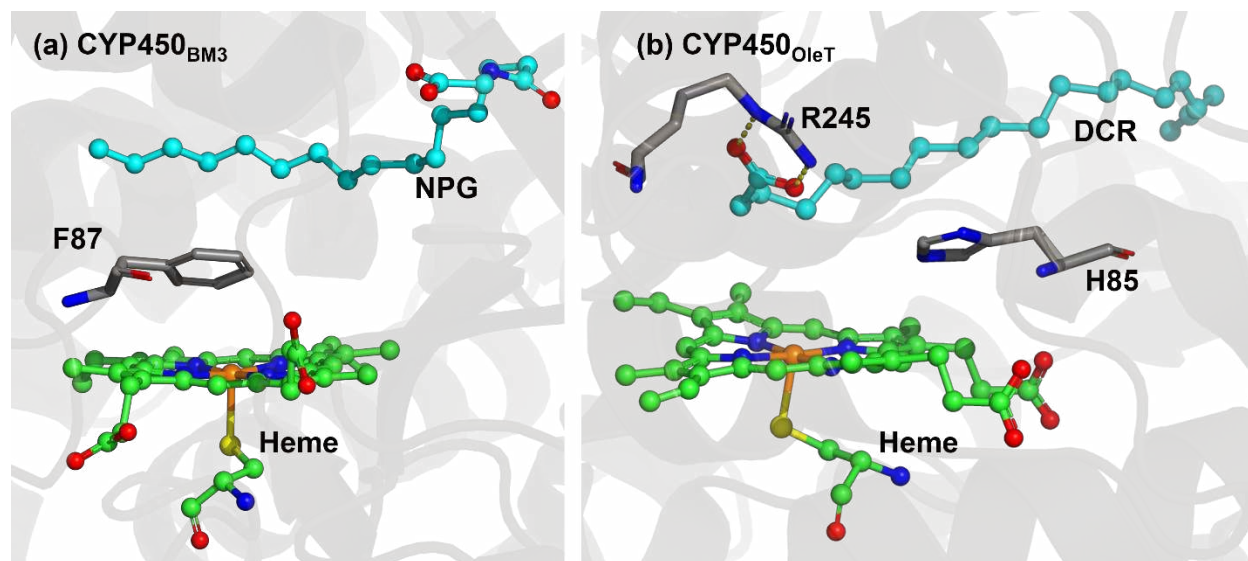

Figure 2. Different active site environments in two different families of CYP450. **(a)** The presence of an alkyl substrate tail and F87 above the heme creates a hydrophobic environment around the heme in bacterial CYP450$_{BM3}$. **(b)** In CYP450 peroxygenases, the presence of carboxylate end of the substrate and two polar residues like Arg245 and His85 make the heme environment hydrophilic.

In contrast, the fatty acid enters the heme cavity in CYP450$_{BSβ}$ and CYP450$_{OleT}$ via its carboxylate end and interacts with the guanidinium group of R242/245 (CYP450$_{BSβ}$/CYP450$_{OleT}$) close to the active site,[27] (cf. Figure 2b), which produces a more polar environment vis-à-vis CYP450$_{BM3}$. Therefore, due to the different polarities of the active site in these enzymes, they could be an ideal system to benchmark the performance of water for similar substrates but of different reactivity.

2. Methodology

2.1. System setup

The study focuses on three CYP450 enzymes; CYP450$_{BM3}$, CYP450$_{OleT,}$ and CYP450$_{BSβ}$. The initial coordinates for these enzymes were taken from the crystal structures with PDB IDs: 1JPZ (CYP450$_{BM3}$),[24] 4L4O (CYP450$_{OleT}$)[26], and 1IZO (CYP450$_{BSβ}$).[25] These structures were co-crystalized with the following substrates: N-Palmitoglycine (NPG) in CYP450$_{BM3}$, Icosanoic acid (DCR) in CYP450$_{OleT,}$ and Palmitoloeic acid (PAM) in CYP450$_{BSβ}$. Missing hydrogen atoms and some heavy atoms were added by the leap module of Amber 20. The force fields for the heme moiety were taken from the already published parameters for the Cpd I state.[28] The Amber ff19SB force field was employed for all the enzymatic systems.[29] The associated partial atomic charges and missing parameters for the substrates were generated by applying the restraint electrostatic potential (RESP) method of QM calculated charges at the HF/6-31G(d) level of theory.[30,31] The corresponding parameters for substrates were generated using Generalized Amber Forcefields (GAFF2) in the Antechamber module of Amber 20. A few Na+ ions were added to the protein surfaces to neutralize the total charge of the system. Finally, the resulting systems were solvated in an octahedral box of three different water models: TIP3P, OPC, and SPC/E, each extended up to a minimum cut-off of 10 Å from the protein boundary. The catalytic cycle of CYP450 has many

states. However, here we focus on the Cpd I state which is the ultimate oxidant species that leads to catalyze the substrate.

2.2. MD Simulation.

After proper parametrization of the system, to remove bad contacts, minimization was performed in two stages using a combination of steepest descent (5000 steps) and conjugate gradient (5000 steps) methods. In the first stage water position and conformations are relaxed keeping the protein fixed. Thereafter the whole complex was minimized. Subsequently, the system is gently annealed up to 300K under the NVT ensemble for 50ps. After that, 1ns of density equilibration was performed under an NPT ensemble at a target temperature of 300K and pressure 1 atm by using Langevin thermostat[32] and Berendsen barostat[33] with a collision frequency of 2ps and pressure relaxation time of 1ps. This 1 ns density equilibration is a weakly restrained MD simulation in which the system is slowly released to achieve uniform density after heating under periodic boundary conditions. Then after we remove all the restraints applied before and the system gets equilibrated for 3 ns to get a well-settled pressure and temperature for chemical and conformational analyses. Thereafter, a productive MD simulation was performed using the Monte Carlo barostat[34] for 100 ns for each complex system. We used three different replicas starting from different initial velocities each for 300 ns. Therefore, we performed a total of 300 ns simulations including all three replicas for each system. During all the MD simulations, covalent bonds containing hydrogens were constrained using the SHAKE[35] algorithm, and Particle Mesh Ewald (PME)[36] method was used to treat long-range electrostatic interactions with the cut-off set as 10 Å. All MD simulation was performed with the GPU version of the AMBER 20 package.[37] The MD trajectory analysis was done with the CPPTRAJ[38] module of AMBER20. The visualization

of the MD trajectories was performed by VMD[39]. *All MD simulations were re-run for 3 different replicas starting from different initial velocities (see SI).*

2.3. QM/MM Methodology

Hydrogen atom transfer (HAT) reactions were calculated using QM/MM calculations for the representative snapshots from MD simulations. In CYP450$_{BM3}$, the QM-region involves substrate NPG and heme, whereas, the QM region of peroxygenases CYP450$_{OleT}$ and CYP450$_{BS\beta}$ include truncated heme-porphyrin, truncated substrates, and R245/R242, H85/Q85 and two water molecules. The active region in QM/MM calculations in all the systems involves the protein residues and water molecules present within the cutoff of 8Å from the active oxidant heme. The atoms in the selected "active region" (mainly from the MM part) interact with the QM zone through electrostatic and van der Waals interactions and the corresponding polarization effects were considered in the subsequent QM/MM calculations. All QM/MM calculations were performed with ChemShell,[40,41] by combining the Turbomole[42,43] for the QM part, and DL_POLY[44] for the MM part. The MM part was described using ff19SB forcefield. To account for the polarizing effect of the protein environment on the QM region, an electronic embedding scheme was used. Hydrogen link atoms with the charge shift model were employed for treating QM/MM boundary. During QM/MM geometry optimizations, the QM region was treated using the hybrid UB3LYP functional with two basis sets, B1 and B2, where B1 stands for def2-SVP and B2 stands for def2-TZVP. The B1 basis set was used for geometry optimization, potential energy surface scanning, and frequency calculations. The energies were further corrected with the Grimme dispersion correction. All of the QM/MM transition states were located by relaxed potential energy surface (PES) scans followed by full TS optimizations using the P-RFO[45] optimizer implemented in the HDLC code. The energies were further corrected with the large all-electron basis set, designated

as def2-TZVP. The zero-point energies (ZPE) were calculated for all the species and the corresponding final energies were reported as UB3LYP/B2-D3+ZPE. Since s=1/2 (doublet) and s=3/2 (quartet) spin state of Cpd I generally exhibit similar reactivities in CYP450 enzymes. For the sake of comparison to the already reported mechanistic studies, we have performed the calculation in the same spin state i.e., s=1/2 (doublet) used in earlier studies.[13]

3. Results and Discussion

We benchmarked the effect of different water models based on: a) Molecular Dynamics Simulations to study the structural effects and b) QM/MM calculations to study the effect on reactivity of Compound I (Cpd I) for three CYP450 members. *The structural effects were probed using RMSD, RMSF, SASA, dipole moment, etc., while reactivity patterns were studied by the hybrid QM/MM calculations for the structure generated by three solvent models*. The performance of different water models were ranked on the qualitative comparison depending on above parameters. In CYP450 chemistry, it is well established that the 'water occupancies' close to the heme site is crucial for the catalytic activity of the P450 enzymes. In fact, it is well known that destroying the water channel through the acid-alcohol pairs (for example Glu267-Thr268 in P450$_{BM3}$) can inhibit the enzyme function. Therefore, we have used the population of water molecules close to the heme site as our standard parameter to measure the relative performance of each water model. In this way- a simulated system with more ordered and populated water molecules has been ranked better and vice versa.

3.1. Effect on Structural Parameters:

RMSD and RMSF: RMSD is commonly used by computational chemists to validate the convergence and the stability of the simulated complex while RMSF is widely used to study the

thermal fluctuations of atoms/residues during the simulations. To study whether the different water models have any effect on the convergence and the stability of the simulated complex, we, therefore, analyzed the RMSD and RMSF of all three water models for CYP450$_{BM3}$, CYP450$_{BS\beta}$, and CYP450$_{OleT}$. A comparison of these RMSDs and RMSFs is shown in Figure 3. As can be seen from Figure 3a, the three water models for all three CYP450s have similar convergence and they show very small overall deviations. However, all three enzymes show marginally fewer RMSD deviations with OPC/E water model relative to TIP3P and OPC models. The RMSF of all three water models for three CYP450 enzymes is shown in Figure 3b. Similar to RMSDs, we can see that all three models *qualitatively* show similar thermal fluctuations, however, the residue-wise thermal fluctuations are slightly higher in the TIP3P model relative to OPC and SPC/E.
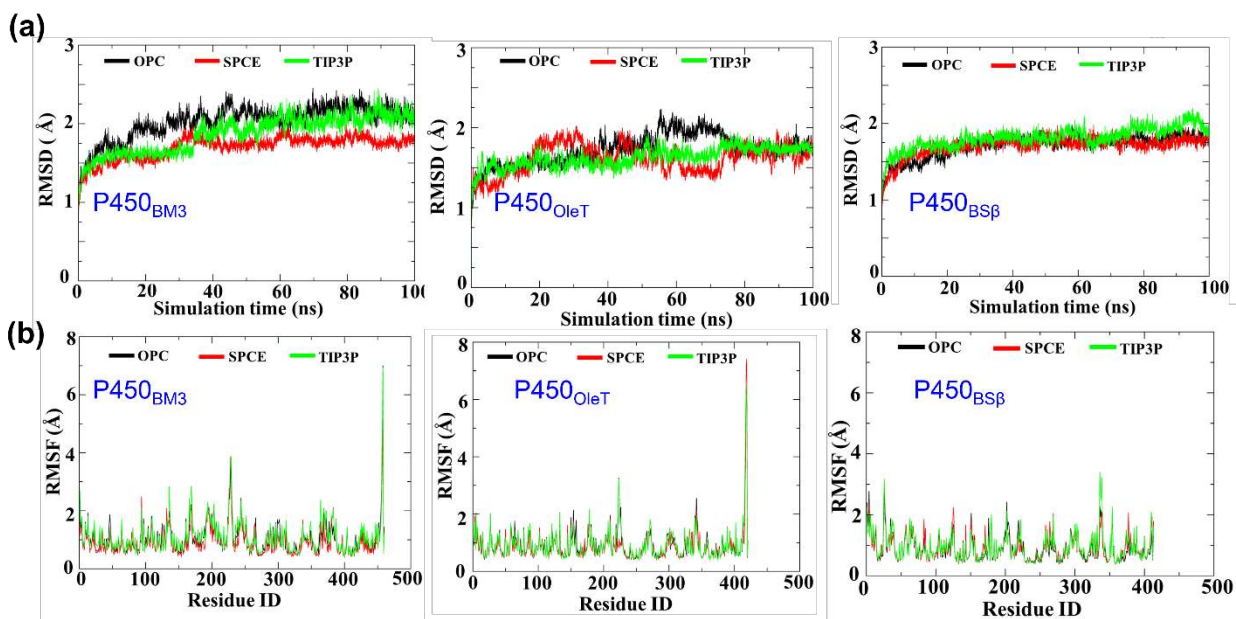

Figure 3. **(a)** RMSD(Å) and **(b)** RMSF (Å) of all the three systems (CYP450$_{BM3}$, CYP450$_{OleT,}$ and CYP450$_{BS\beta}$), calculated over 100ns of simulation time.

In summary, we can conclude that all three models have similar RMSD and RMSF values and *water models have little or no effect on the stability of the CYP450 enzymes.*

9

Solvent Accessible Surface Area (SASA): Since the hydration of the active site of the heme-binding plays a crucial role in the function of CYP450s, we calculated the solvent-accessible surface area (SASA) around 5Å of the heme molecule. A comparison of the SASA for the three models is shown in Figure 4. As can be seen, the SASA for the OPC model performs slightly better compared to TIP3P and SPC/E models. In addition, the values of SASA of all three models are large in CYP450$_{BM3}$ relative to CYP450$_{BS\beta}$, and CYP450$_{OleT}$.
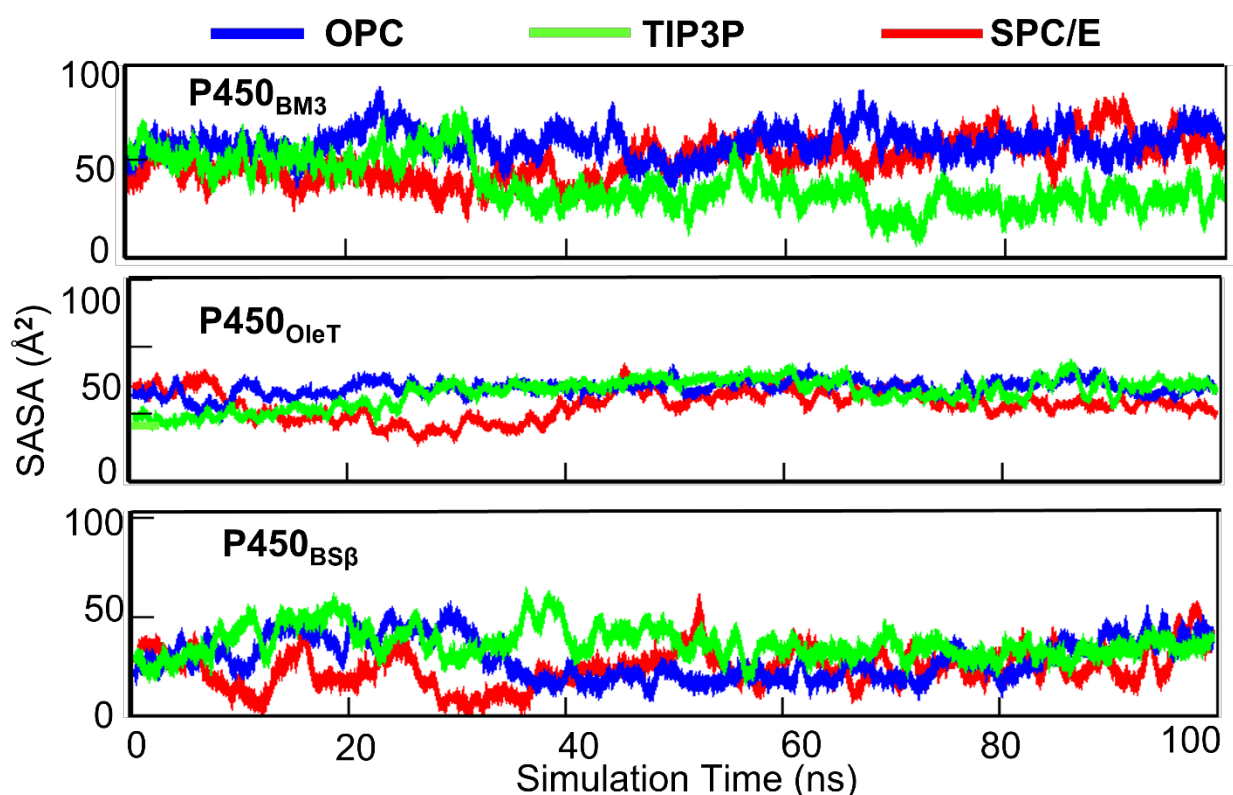

Figure 4. The solvent-accessible surface area around 5 Å of heme residue in CYP450$_{BM3}$, CYP450$_{OleT}$, and CYP450$_{BS\beta}$.

Populations of Water and H-bonds: In CYP450s, it is well known that the heme propionate forms a well-organized chain connected through the H-bonds, therefore, we performed water population analysis (SI) and water-heme H-bond analysis around 5 Å of Heme for different water models for

all three complexes. Interestingly, among all the water models, the TIP3P water model has the maximum water count for CYP450$_{BM3}$ while the TIP3P water count observed to be less for the CYP450$_{BSβ}$, and CYP450$_{OleT}$. At the same time, the OPC water model exhibits greater number of water molecules in case of CYP450$_{OleT}$ and CYP450$_{BSβ}$. We note that the hydrophobicity of the active site is higher in CYP450$_{BM3}$ due to the Phe87 and substrate hydrophobic tail while, in contrast, the active site in CYP450$_{OleT}$ and CYP450$_{BSβ}$ are more polar due to charged end of the substrate and an Arg245/Arg242 residues. Therefore, it could be anticipated that due to the more accurate distribution of point charges in OPC which mimic the optimal electrostatics of water models and the hydrophilic nature of active sites (in CYP450$_{OleT}$ and CYP450$_{BSβ}$) supports the electrostatic interactions like H-bonds. Moreover, for CYP450$_{BM3}$ where the active site is hydrophobic in nature, higher water populations (SI) and the H-bonds were observed for TIP3P water model. Therefore, we may conclude that *TIP3P and SPC/E could be more suitable models for the hydrophobic active site while the OPC water model is more suitable for the polar active sites*.
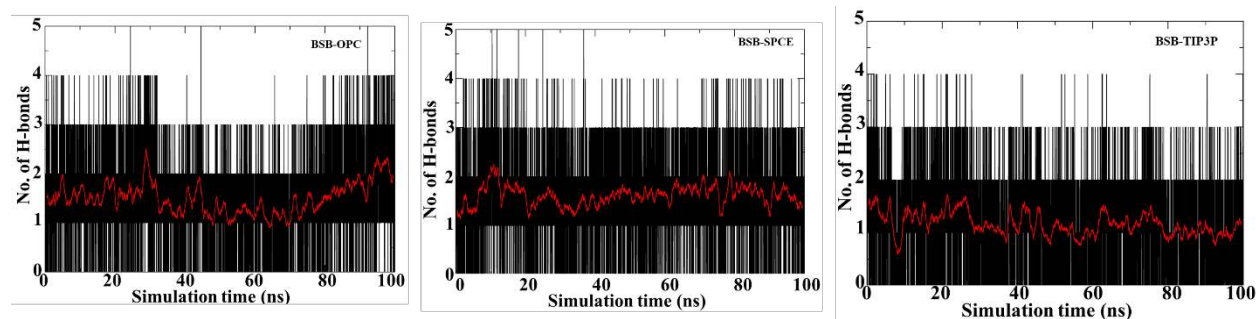

11
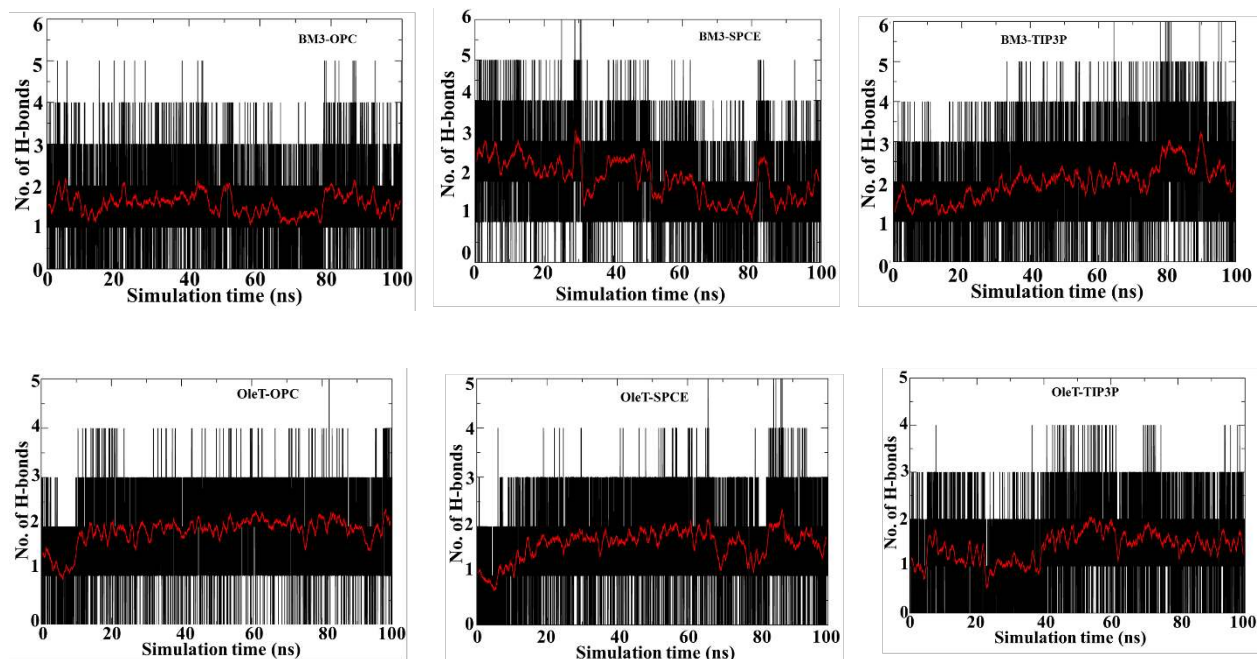
Figure 5. Hydrogen bonds calculated between active oxidant heme and the solvent water molecules present near 5Å. The graph shown at the top represents the number of H-bonds in CYP450$_{BM3}$ simulated with the three different water models. Below one signifies the same for CYP450$_{OleT}$ and CYP450$_{BS\beta}$.

Dipole moment: In a recent study, we have shown that the local electric field of the CYP450 enzyme could dictate the specific functions, and the dipole moment induced by the enzymatic local electric field could significantly catalyze the reaction by lowering the activation energy.[46] Therefore, it is intriguing to study how the different water models affect the electrostatics of the active site residues. For so doing, we selected the 11 snapshots each at the interval of 10 ns (0ns, 10ns, 20ns, 30ns, - - - - -100ns) and calculated the dipole moment due to the water molecules present near the 5 Å of the active site heme fragment. Table 1 represents the average dipole moment for each model for all three enzymes. As can be seen, the dipole moment due to the water molecules is higher in the case of TIP3P for CYP450$_{BM3}$ only, while OPC shows higher dipole moment for both peroxygenases, i.e., CYP450$_{OleT}$ and CYP450$_{BS\beta}$. The rationale for these different dipole moments could again be correlated with the hydrophobicity and the polarity of the active sites in all three enzymes.

Table 1. Average Dipole moment calculation in CYP450$_{OleT}$, CYP450$_{BS\beta}$ and CYP450$_{BM3}$ for three Water models.

Enzyme / Water Model	Avg. value (calculated over diff. frames)	CYP450$_{OleT}$	CYP450$_{BS\beta}$	CYP450$_{BM3}$
TIP3P	D.M. (μ)	9.89	14.19	7.8
	Water count	9.45	11.54	14.4
OPC	D.M. (μ)	10.29	19.64	6.8
	Water count	10.72	12.45	12.09
SPC/E	D.M. (μ)	9.32	15.14	7.4
	Water count	10.18	11.54	11.36

3.2. Effect on the Conformational Properties:

In the above sections, we have seen the effect of different water models on some structural parameters. In the proceeding section, we will see if the water models also affect the conformation of the substrate and active site residues.

Conformation of CYP450$_{BM3}$: CYP450$_{BM3}$ catalyzes the hydroxylation of long-chain fatty acid and the conformational changes responsible for its functions have already been explained with MD simulation using the TIP3P water model.[47] Therefore, we compared the key conformational features e.g., the orientation of Phe87 and substrate dynamics using different water models. It is well established that the Phe87 flips from parallel to perpendicular plane relative to the heme plane, therefore we compared the propensity of the rotation of Phe87 in different water models. Interestingly, the propensity to switch to a perpendicular plane was significantly higher in the OPC

model relative to SPC/E and TIP3P. Similarly, the substrate can easily slide close to the heme site in the MD simulation with the OPC model while in other water models it requires more MD sampling to access the heme site.
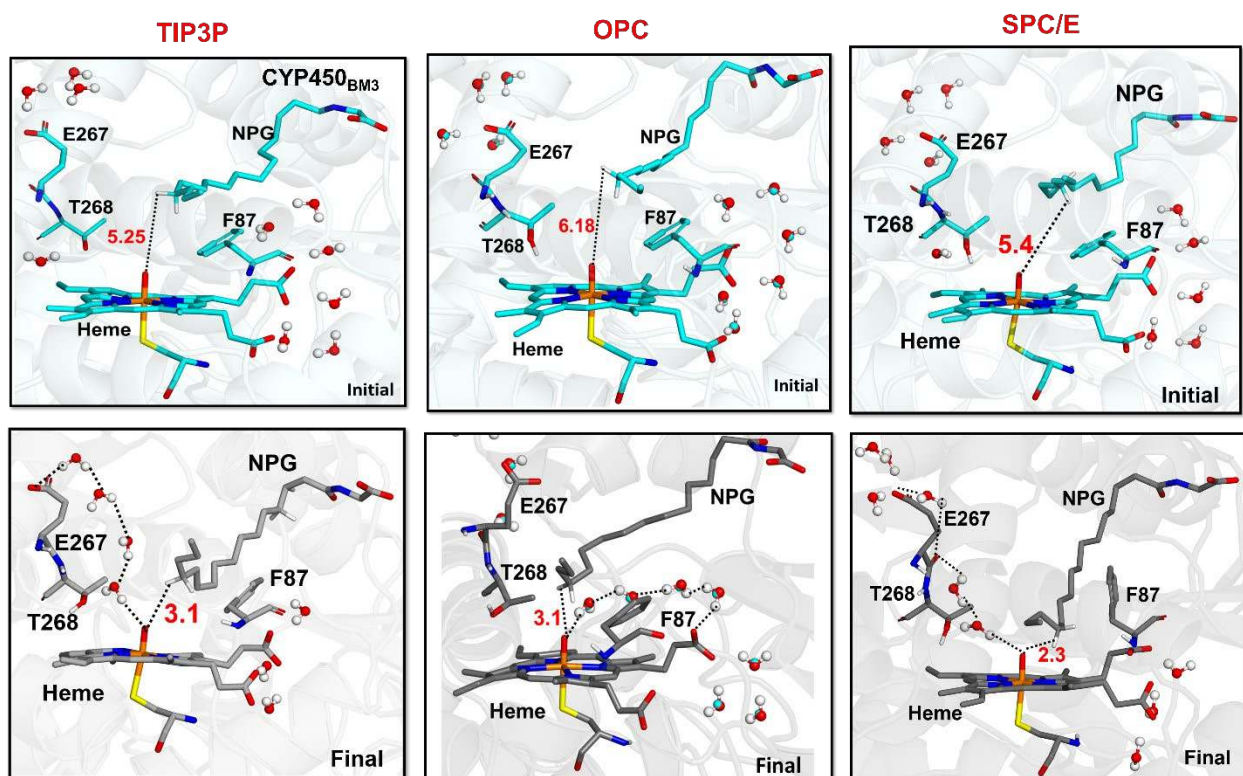

Figure 6. Representative MD snapshots in case of CYP450$_{BM3}$ system for three different water models.

Conformation of CYP450 peroxygenases i.e., CYP450$_{OleT}$ and CYP450$_{BS\beta}$: As discussed earlier, the active site of the peroxygenases is relatively polar (with respect to CYP450$_{BM3}$) due to carboxylate tail and an Arginine close to heme site. During the MD Simulations of both peroxygenases, the water occupancy was almost similar for all three water models (TIP3P, OPC, and SPC/E) which are gated by propionate side chain as shown in Figure 7.
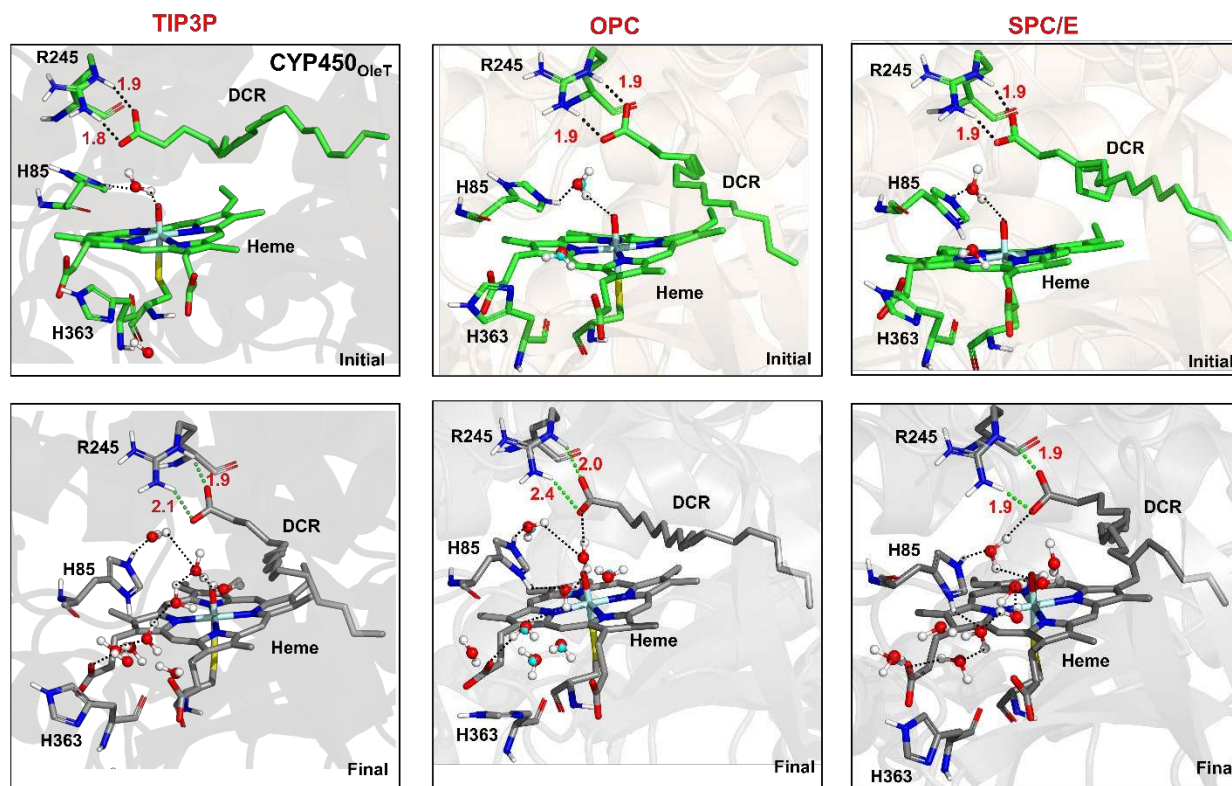
Figure 7. Origin of water molecules approaching to heme. Note that flipping of H363 opens the water aqueduct and accumulation of waters in the active site affects the substrate orientation in CYP450$_{OleT}$.

Though the aqueducts are rather similar to all the water modeled systems, we have found a significant difference in the behavior of the fatty acid substrate. During the MD with TIP3P and SPC/E modeled aqueous medium, the substrate strongly binds to the Arg245 residue through its carboxylate end and shows significant rigidity during the simulations in both peroxygenases. In contrast, during the MD simulations with the OPC modeled aqueous medium, the substrate slightly changes conformation over the course of time in both peroxygenases. Initially, the His363/H361 (CYP450$_{OleT}$/CYP450$_{BSβ}$) restricts the water accessibility, and the substrate interacts strongly through its carboxylate end with the guanidinium group of Arg245/Arg242. As the simulation progresses, the H363/H361 flips its orientation and allows an easy approach of water to the heme.

Due to the change in the water influx, the interaction of the substrate with Arg245 weakens and it reclines towards the heme. Therefore, comparing the three water models in three different CYP450 enzymes we may conclude that the *role of the water model becomes imperative in the polar environment and the OPC model may have a significant impact on the conformation in polar active sites.*
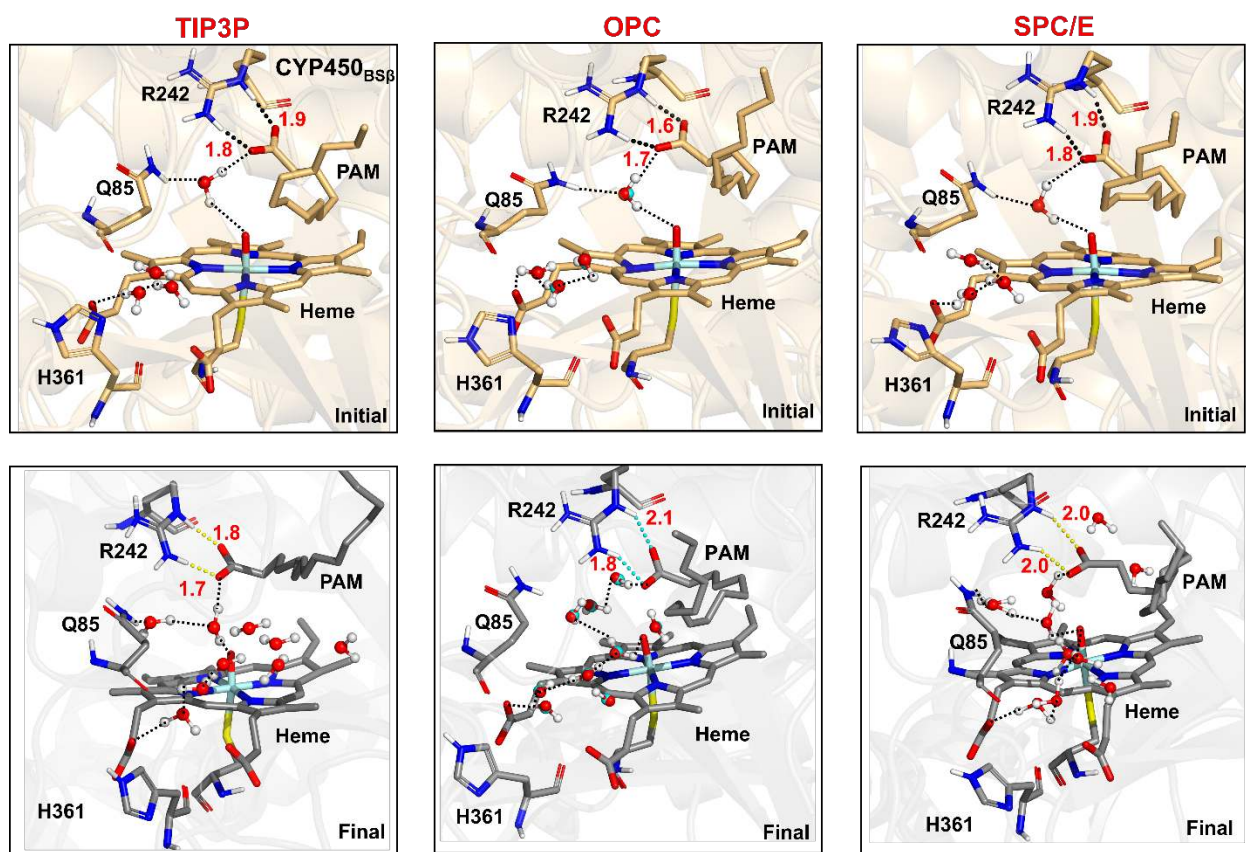

Figure 8. Representative snapshots for the CYP450$_{BS\beta}$ simulated in TIP3P, OPC, and SPC/E water model.

3.3 Effect on Reactivity

In the previous section, we have seen that the water model may affect the substrate orientation via altered water influx which, in turn, may affect the activation energy of the reaction.

We, therefore, performed QM/MM calculations for representative snapshots from the MD trajectories. Since the hydrogen atom transfer (HAT) from the substrate to Cpd I is the key step in the substrate oxidation, we performed the Potential Energy Surface (PES) scan for the HAT step for each enzyme for different water models.

HAT in CYP450 peroxygenases, CYP450_{OleT} and CYP450_{BSβ}: We started the QM/MM calculations with the geometry optimization of the reactant complex (RC), Cpd I followed by PES scanning of the optimized geometry. The reaction coordinate of the HAT reaction along with the key stationary points is shown in Figure ~~8~~ 9.

a) CYP450_{OleT} in OPC water model

c)
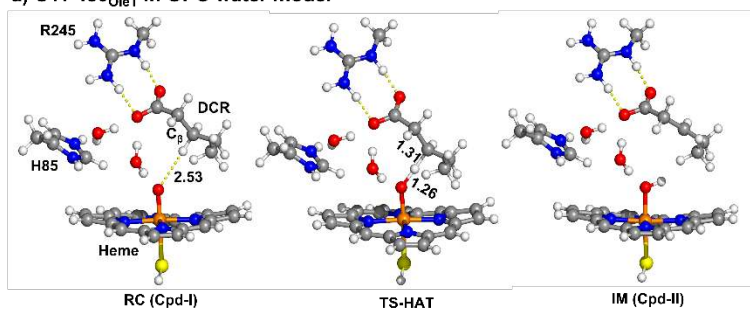
Spin							
	Fe-O-S	Porphyrin	Sub	Arg	His	W1	W2
RC	1.78	- 0.87	0	0	0	0	0
IM	2.14	- 0.21	- 0.97	0	0	0	0
Charge							
RC	0.16	0.22	- 0.62	0.75	0.70	0.08	-0.03
IM	- 0.03	- 0.05	- 0.61	0.74	0.71	0.09	-0.04

b) CYP450_{OleT} in SPC/E water model

d)
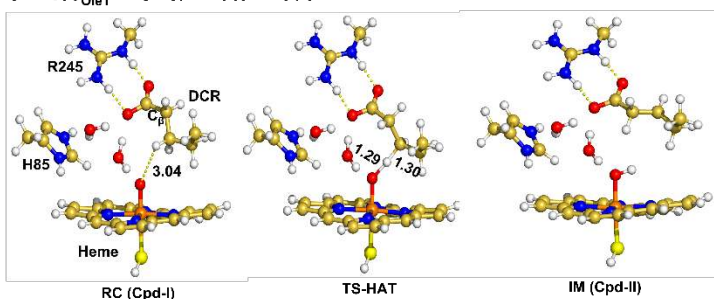
Spin							
	Fe-O-S	Porphyrin	Sub	Arg	His	W1	W2
RC	1.87	- 0.96	0	0	0	0	0
IM	2.14	- 0.21	- 0.97	0	0	0	0
Charge							
RC	- 0.19	0.29	- 0.63	0.73	0.73	0.08	- 0.03
IM	0.09	0.22	0.62	0.73	0.73	0.09	- 0.04

e)

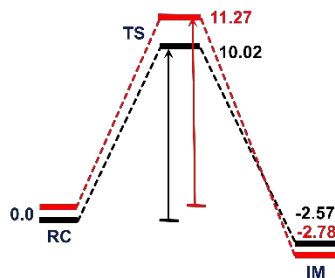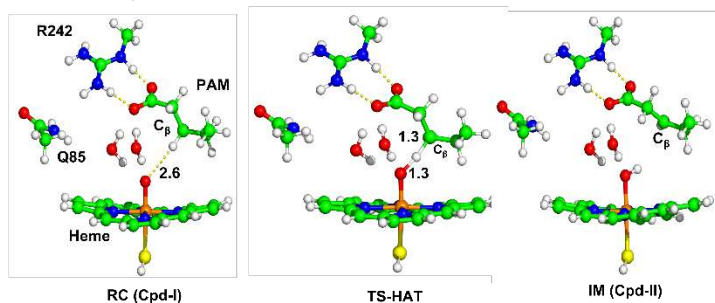Figure 9. **(a,b)** Optimized geometries of RC, TS, and IM during the HAT in CYP450$_{OleT}$ in case of OPC and SPC/E water model respectively. **(c,d)** Charge and spin density values for CYP450$_{OleT}$ in OPC and SPC/E water model respectively. **(e)** Energy profile obtained during the PES scanning by QM/MM at the B3LYP-D3/def2-TZVP level of theory. Energies reported have the unit, kcal/mol. Note that the OPC water model used here exhibits the direction of OPC and structural properties of TIP3P.

The energy profile for this step shows a barrier of 10.02 and 11.27 kcal/mol for OPC and SPC/E water models, respectively. In both models, the intermediate IM is exothermic which is in agreement with previously published data. Similarly, for CYP450$_{BS\beta}$ energy profile for the HAT reaction from the C_β of the fatty acid substrate, exhibit a barrier of 22.01 kcal/mol and 22.49 for OPC and SPC/E water models, respectively. Interestingly, in both the calculation (OPC and SPC/E) a finding of a less stable intermediate reflects the endothermic behavior of HAT in CYP450$_{BS\beta}$ which is the root cause of the diversification for hydroxylation vs decarboxylation reactions in CYP450 peroxygenases (see ref).[46]
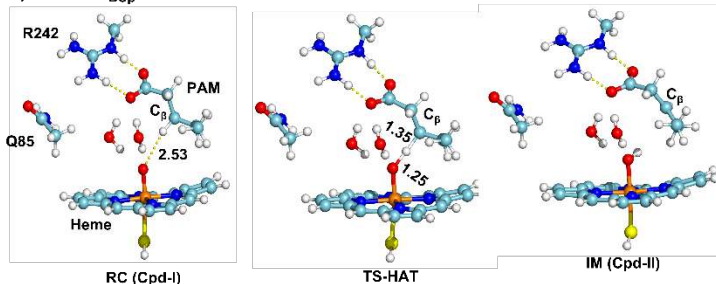
a) CYP450$_{BS\beta}$ in OPC water model

RC (Cpd-I)   TS-HAT   IM (Cpd-II)

b) CYP450$_{BS\beta}$ in SPC/E water model

RC (Cpd-I)   TS-HAT   IM (Cpd-II)

c)

Spin	Fe-O-S	Porphyrin	Sub	Arg	Gln	W1	W2
RC	1.87	- 0.95	0	0	0	0	0
IM	1.01	- 0.8	0.72	0	0	0	0
Charge							
RC	- 0.18	0.30	- 0.61	0.70	0	0	0
IM	- 0.1	0.03	- 0.24	0.80	0	0	0

d)

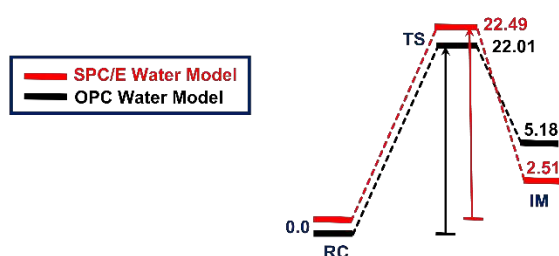| Spin	Fe-O-S	Porphyrin	Sub	Arg	Gln	W1	W2
RC	1.81	- 0.90	0	0	0	0	0
IM	1.01	- 0.74	0.67	0	0	0	0
Charge							
RC	- 0.13	- 0.07	- 0.63	0.70	0	0	0
IM	- 0.25	- 0.26	- 0.22	0.80	0	0	0

e)

Figure 10. **(a,b)** Optimized geometries of RC, TS, and IM during the HAT in P450$_{BS\beta}$ in the case of OPC and SPC/E water model. **(c,d)** Charge and spin density values for CYP450$_{BS\beta}$ in OPC and SPC/E water model respectively. **(e)** Energy profile obtained during the PES scanning by QM/MM at the B3LYP-D3/def2-TZVP level of theory. Energies reported have the unit kcal/mol. Note that the OPC water model used here exhibits the direction of OPC and structural properties of TIP3P.

The electronic structure of the reaction geometries i.e., Cpd I (RC) and Cpd II (IM) CYP450$_{OleT}$ and P450$_{BS\beta}$ were confirmed by the spin densities and Mulliken charge analysis as shown in tabular form in Figure 9 and 10. Optimized geometries during HAT in the SPC/E water model (for CYP450$_{OleT}$ and for CYP450$_{BS\beta}$) can be seen in SI.

HAT P450BM3: We have seen from the MD results that the pro-RH hydrogen of ω-3 carbon terminal is the closest to oxo of heme fragment to a greater extent for the OPC water model followed by SPC/E and TIP3P water model. Therefore, to see whether such conformation of the substrate due to different water models has some impact on the reactivity, we performed QM/MM calculations for the representative snapshots taken from the MD trajectories and performed the PES scanning to get the reaction profile of the HAT reaction. The energy profile diagram for the Hydrogen Atom Transfer (HAT) process for CYP450$_{BM3}$ simulated in OPC and SPC/E water models have been shown in figure 11. The HAT process in CYP450$_{BM3}$ exhibits a barrier of 12.2, 12.6 and 15.4 (units Kcal/mol) in OPC, SPC/E and TIP3P water (see reference 47) models.

a) CYP450$_{BM3}$ in OPC water model

RC (Cpd-I)   TS-HAT   IM (Cpd-II)

c)

Spin			
	Fe-O-S	Porphyrin	Sub
RC	1.94	- 0.99	0
IM	1.85	- 0.13	- 0.73
Charge			
RC	- 0.19	0.26	0.07
IM	- 0.19	- 0.20	0.36

b) CYP450$_{BM3}$ in SPC/E water model

RC (Cpd-I)   TS-HAT   IM (Cpd-II)

d)
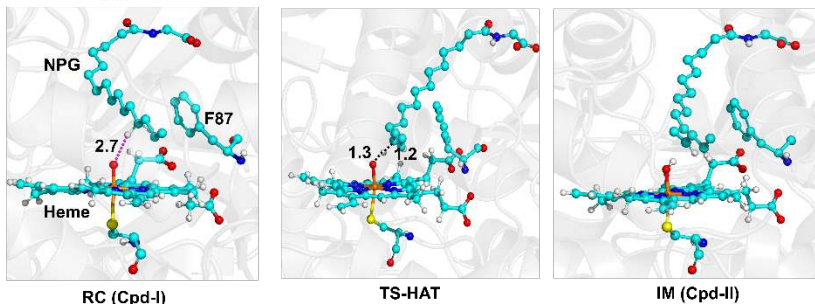
Spin			
	Fe-O-S	Porphyrin	Sub
RC	1.94	- 1.00	0
IM	1.05	- 0.50	0.37
Charge			
RC	- 0.22	0.02	0.20
IM	- 0.32	- 0.40	0.73

e)

SPC/E Water Model
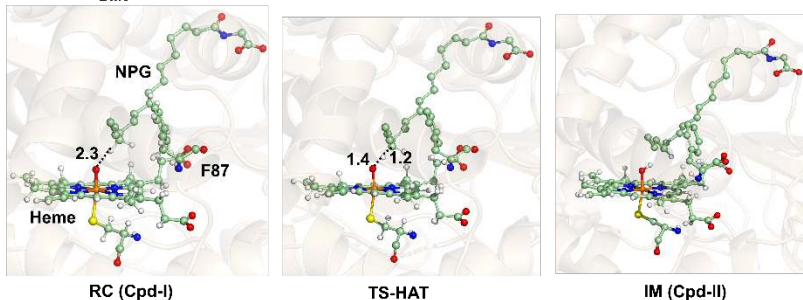OPC Water Model

TS 12.63
    12.22

RC 0.0

IM 1.48
    -1.63

20

Figure 11. **(a,b)** Optimized geometries of RC, TS, and IM during HAT process in CYP450$_{BM3}$ in the case of OPC and SPC/E water model respectively. **(c,d)** Charge and spin density values for CYP450$_{BM3}$ in OPC and SPC/E water model respectively. **(e)** Energy profile obtained during the PES scanning by QM/MM at the B3LYP/def2-TZVP level of theory. Energies reported have the unit kcal/mol.
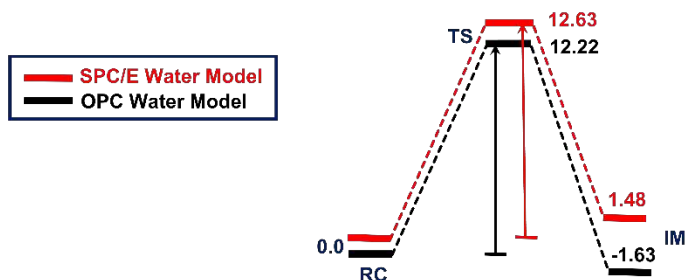
Table 2. Hydrogen atom transfer barrier for different water models in CYP450$_{OleT}$, CYP450$_{BSβ}$ and CYP450$_{BM3}$.

Water Model	TIP3P	OPC (modeled as TIP3P)	SPC/E
CYP450$_{OleT}$	9.6 kcal/mol	10.02 kcal/mol	11.27 kcal/mol
CYP450$_{BSβ}$	15.5 kcal/mol	22.01 kcal/mol	22.49 kcal/mol
CYP450$_{BM3}$	15.4 kcal/mol	12.2 kcal/mol	12.6 kcal/mol

In a nutshell, based on the HAT barrier shown in the above table in the TIP3P (see ref)[46,47] OPC, and SPC/E water model in CYP450$_{OleT}$ and CYP450$_{BSβ}$, one could conclude that in peroxygenase, TIP3P water model acts more efficiently, while in case of monooxygenase, OPC water model acts more efficiently. On the other hand, the SPC/E water model is the least efficient for peroxygenase and TIP3P is least efficient in case of monooxygenase among all three water models. However, it is important to note that we were unable to fully implement the OPC water model in Chemshell due to the unavailable parameter of the fourth site of OPC water model. Therefore, we believe if the complete implementation of the OPC water model in QM/MM will be done then the OPC model could get better results in the field of activation energy barrier calculation due to the implementation of a better charge model.

3.4. *Effect of Water Model on the Enzymatic Local electric Field*

It is well established that the enzymatic local electric field (LEF) can be a descriptor of the enzymatic activity,[48–51] therefore, we probed the effect of different water models on the LEF of the enzyme. The LEF of the enzyme was calculated on the oxo moiety of heme "O" and in the direction Fe=O axis using TITAN[52] Programs of QM/MM optimized RC, TS, and IM geometries.
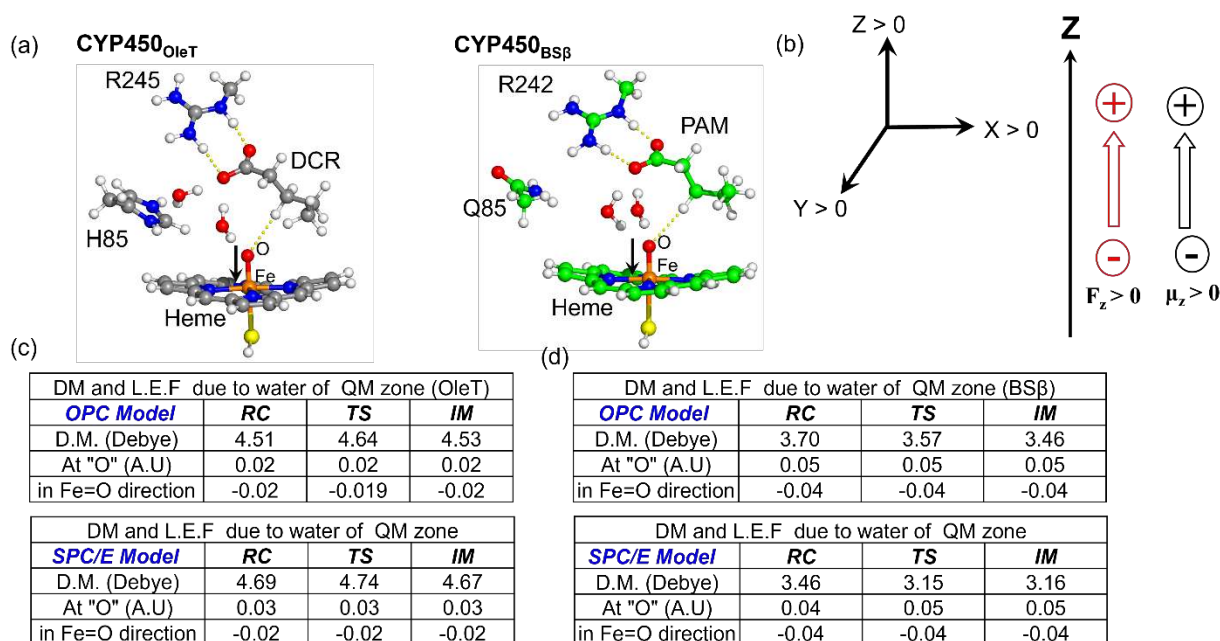

Figure 12. **(a)** QM zone of CYP450$_{OleT}$ and CYP450$_{BS\beta}$. **(b)** Gaussian convention for positive field and dipole moment vectors along Z-axis. **(c,d)** Dipole moment and local electric field calculation for the RC, TS, and IM in CYP450$_{OleT}$ and CYP450$_{BS\beta}$ respectively, in the case of the OPC and SPC/E water model. Note that L.E.F has been calculated along the Fe=O axis at the point "O".

DM and L.E.F due to water of QM zone (OleT)			
OPC Model	*RC*	*TS*	*IM*
D.M. (Debye)	4.51	4.64	4.53
At "O" (A.U)	0.02	0.02	0.02
in Fe=O direction	-0.02	-0.019	-0.02

DM and L.E.F due to water of QM zone			
SPC/E Model	*RC*	*TS*	*IM*
D.M. (Debye)	4.69	4.74	4.67
At "O" (A.U)	0.03	0.03	0.03
in Fe=O direction	-0.02	-0.02	-0.02

DM and L.E.F due to water of QM zone (BSβ)			
OPC Model	*RC*	*TS*	*IM*
D.M. (Debye)	3.70	3.57	3.46
At "O" (A.U)	0.05	0.05	0.05
in Fe=O direction	-0.04	-0.04	-0.04

DM and L.E.F due to water of QM zone			
SPC/E Model	*RC*	*TS*	*IM*
D.M. (Debye)	3.46	3.15	3.16
At "O" (A.U)	0.04	0.05	0.05
in Fe=O direction	-0.04	-0.04	-0.04

As can be seen from Figure 12, the LEF remains constant despite using the different water models i.e., OPC and SPC/E in RC, TS, and IM, however, there is a notable difference in the dipole moments of the reactive species. We found an increase in the dipole moment of 0.13 debye from RC to TS using the OPC water model for CYP450$_{OleT}$, while in the case of the SPC/E water model this increment in dipole moment is 0.05 debye only.

Interestingly in the case of CYP450$_{BS\beta}$, this dipole moment trend is completely opposite. In both the water models (OPC and SPC/E) dipole moment due to the waters of the QM zone, get

decreased from RC to TS. This decrement of dipole moment is 0.13 debye and 0.31 debye in OPC

and SPC/E water models, respectively. Considering the stabilization energy $\Delta E_{LEF} = 4.8$ (LEF) (μ)

cos θ (where LEF is the local electric field, μ is the dipole moment and θ is the angle between LEF

vector and μ vector), the change in dipole moment can significantly affect the total stabilization of

the state. In a positively charged polar environment of CYP450$_{OleT}$ due to the OPC water model,

an increase in the dipole moment stabilizes the TS and we get lower activation energy. On the

other hand, a decrease in the dipole moment of TS in CYP450$_{BS\beta}$ destabilizes the TS vis-à-vis

CYP450$_{OleT}$, and therefore we see relatively high activation energy (cf. Table 2).

In summary: The effectiveness of various water models on three different enzyme systems has

been evaluated using computational methods while taking into account various characteristics. We

have calculated several properties with the classical MD simulation tool and all the outcomes has

been summarized in the table given below.

Table 3. Evaluation of different water models for CYP450$_{OleT}$, CYP450$_{BS\beta}$ and CYP450$_{BM3}$. Note
that the water models have been enlisted based on the best performance for various properties.
Here N.D represents No difference.

Properties Enzymes	RMSD	RMSF	SASA	H-bonds	Dipole moment	Water occupancy
CYP450$_{OleT}$	N.D	N.D	OPC	OPC	OPC	OPC
CYP450$_{BS\beta}$	N.D	N.D	OPC	OPC	OPC	OPC
CYP450$_{BM3}$	N.D	N.D	OPC	TIP3P	TIP3P	TIP3P

As can be seen that the use of different water models creates no difference (N.D) on the RMSD

and RMSF parameters of the enzymes whereas significant differences was observed for other

properties mentioned in the Table 3. In summary, for the peroxygenases $CYP450_{OleT}$ and $CYP450_{BS\beta}$ OPC water models performs well whereas for the $CYP450_{BM3}$ better results were observed for the TIP3P water model. Effect of the different water models on the reactivity is describes in section 3.3 and Table 2.

4. Conclusions

In the present study, we have comprehensively studied the impact of three widely used water models on the structure and reactivity of three CYP450 enzymes using MD simulations and hybrid QM/MM calculations. MD simulations show that the three water models, TIP3P, SPC/E, and OPC do not have much impact on the RMSD and RMSF values, however, the population of water and dipole moment depends upon the polarity of the active sites of the CYP450 enzymes. For instance, for hydrophobic active sites such as $CYP450_{BM3}$, the TIP3P, and SPC/E water models have better water populations while for the polar active sites such as $CYP450_{OleT}$ and $CYP450_{BS\beta}$ the OPC model performs well. Due to altered water count, the local dynamics in the active site such as Phe87 and substrate orientations in $CYP450_{BM3}$ were found to be slightly different in OPC and it requires less conformational sampling with respect to TIP3P and SPC/E water models. The QM/MM calculations show that the TIP3P and OPC water models perform better than the SPC/E water model, although an accurate implementation of the fourth site of the OPC water model was not possible in the Chemshell program. In nutshell, TIP3P and SPC/E water models may perform better than the OPC model in the hydrophobic site while the OPC water model is recommended for the polar environments.

5. Data and Software Availability:

All molecular structures and its coordinates are available as the supporting information. We didn't use any new program/software.

6. Acknowledgements:

7. Dedications

The women in India always tackle challenges on multiple fronts –caste, religion, gender, and equality. Therefore, we dedicate this work to all Indian women working in Chemistry.

8. References:

(1)   Loftfield, R. B.; Eigner, E. A.; Pastuszyn, A.; Lövgren, T. N.; Jakubowski, H. Conformational Changes during Enzyme Catalysis: Role of Water in the Transition State. *Proc. Natl. Acad. Sci. U. S. A.* **1980**, *77*, 3374–3378.

(2)   Stillman, T. J.; Baker, P. J.; Britton, K. L.; Rice, D. W. Conformational Flexibility in Glutamate Dehydrogenase: Role of Water in Substrate Recognition and Catalysis. *J. Mol. Biol.* **1993**, *234*, 1131–1139.

(3)   Kumar, D.; Altun, A.; Shaik, S.; Thiel, W. Water as Biocatalyst in Cytochrome P450. *Faraday Discuss.* **2011**, *148*, 373–383.

(4)   Meyer, E. Internal Water Molecules and H-bonding in Biological Macromolecules: A

Review of Structural Features with Functional Implications. *Protein Sci.* **1992**, *1*, 1543–1562.

(5)   Kalita, S.; Shaik, S.; Kisan, H. K.; Dubey, K. D. A Paradigm Shift in the Catalytic Cycle of P450: The Preparatory Choreography during O2 Binding and Origins of the Necessity for Two Protonation Pathways. *ACS Catal.* **2020**, *10*, 11481–11492.

(6)   Fishelovitch, D.; Shaik, S.; Wolfson, H. J.; Nussinov, R. How Does the Reductase Help to Regulate the Catalytic Cycle of Cytochrome P450 3A4 Using the Conserved Water Channel? *J. Phys. Chem. B* **2010**, *114*, 5964–5970.

(7)   Efimov, I.; Badyal, S. K.; Metcalfe, C. L.; MacDonald, I.; Gumiero, A.; Raven, E. L.; Moody, P. C. E. Proton Delivery to Ferryl Heme in a Heme Peroxidase: Enzymatic Use of the Grotthuss Mechanism. *J. Am. Chem. Soc.* **2011**, *133*, 15376–15383.

(8)   Vidakovic, M.; Sligar, S. G.; Li, H.; Poulos, T. L. Understanding the Role of the Essential Asp251 in Cytochrome P450cam Using Site-Directed Mutagenesis, Crystallography, and Kinetic Solvent Isotope Effect. *Biochemistry* **1998**, *37*, 9211–9219.

(9)   Cukierman, S. Et Tu, Grotthuss! And Other Unfinished Stories. *Biochim. Biophys. Acta - Bioenerg.* **2006**, *1757*, 876–885.

(10)   Agmon, N. The Grotthuss Mechanism. *Chem. Phys. Lett.* **1995**, *50*, 456–462.

(11)   Meunier, B.; De Visser, S. P.; Shaik, S. Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes. *Chem. Rev.* **2004**, *104*, 3947–3980.

(12)   Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. Theoretical Perspective on the Structure and Mechanism of Cytochrome P450 Enzymes. *Chem. Rev.* **2005**, *105*, 2279–

2328.

(13)   Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. P450 Enzymes: Their

Structure, Reactivity, and Selectivity□ Modeled by QM/MM Calculations. *Chem. Rev.*

2010, *110*, 949–1017.

(14)   Dubey, K. D.; Shaik, S. Cytochrome P450 - The Wonderful Nanomachine Revealed through

Dynamic Simulations of the Catalytic Cycle. *Acc. Chem. Res.* **2019**, *52*, 389–399.

(15)   Shaik, S.; Dubey, K. D. The Catalytic Cycle of Cytochrome P450: A Fascinating

Choreography. *Trends Chem.* **2021**, *3*, 1027–1044.

(16)   Oprea, T. I.; Hummer, G.; García, A. E. Identification of a Functional Water Channel in

Cytochrome P450 Enzymes. *Proc. Natl. Acad. Sci.* **1997**, *94*, 2133–2138.

(17)   Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet,

R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. The Catalytic Pathway of Cytochrome

P450cam at Atomic Resolution. *Science* **2000**, *287*, 1615–1622.

(18)   Ramanan, R.; Dubey, K. D.; Wang, B.; Mandal, D.; Shaik, S. Emergence of Function in

P450-Proteins: A Combined Quantum Mechanical/Molecular Mechanical and Molecular

Dynamics Study of the Reactive Species in the H2O2-Dependent Cytochrome P450SPα and

Its Regio-and Enantioselective Hydroxylation of Fatty Acids. *J. Am. Chem. Soc.* **2016**, *138*,

6786–6797.

(19)   Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models

at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.

(20)   Zielkiewicz, J. Structural Properties of Water: Comparison of the SPC, SPCE, TIP4P, and

TIP5P Models of Water. *J. Chem. Phys.* **2005**, *123*, 104501.

(21)   Fuhrmans, M.; Sanders, B. P.; Marrink, S.-J.; de Vries, A. H. Effects of Bundling on the Properties of the SPC Water Model. *Theor. Chem. Acc.* **2010**, *125*, 335–344.

(22)   Kinnaman, L. J.; Roller, R. M.; Miller, C. S. Comparing Classical Water Models Using Molecular Dynamics to Find Bulk Properties. *J. Chem. Educ.* **2018**, *95*, 888–894.

(23)   Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.

(24)   Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A. Pivotal Role of Water in the Mechanism of P450BM-3. *Biochemistry* **2001**, *40*, 13456–13465.

(25)   Lee, D.-S.; Yamada, A.; Sugimoto, H.; Matsunaga, I.; Ogura, H.; Ichihara, K.; Adachi, S.; Park, S.-Y.; Shiro, Y. Substrate Recognition and Molecular Mechanism of Fatty Acid Hydroxylation by Cytochrome P450 from Bacillus Subtilis: Crystallographic, Spectroscopic, and Mutational Studies. *J. Biol. Chem.* **2003**, *278*, 9761–9767.

(26)   Belcher, J.; McLean, K. J.; Matthews, S.; Woodward, L. S.; Fisher, K.; Rigby, S. E. J.; Nelson, D. R.; Potts, D.; Baynham, M. T.; Parker, D. A. Structure and Biochemical Properties of the Alkene Producing Cytochrome P450 OleTJE (CYP152L1) from the Jeotgalicoccus Sp. 8456 Bacterium. *J. Biol. Chem.* **2014**, *289*, 6535–6550.

(27)   Munro, A. W.; McLean, K. J.; Grant, J. L.; Makris, T. M. Structure and Function of the Cytochrome P450 Peroxygenase Enzymes. *Biochem. Soc. Trans.* **2018**, *46*, 183–196.

(28)   Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham III, T. E. Quantum Mechanically Derived AMBER-compatible Heme Parameters for Various States of the Cytochrome P450

Catalytic Cycle. *J. Comput. Chem.* **2012**, *33*, 119–133.

(29) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2019**, *16*, 528–552.

(30) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(31) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **2002**, *115*, 9620–9631.

(32) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. Langevin Stabilization of Molecular Dynamics. *J. Chem. Phys.* **2001**, *114*, 2090–2098.

(33) Berendsen, H. J. C.; Postma, J. P. M. van; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(34) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular Dynamics Simulations of Water and Biomolecules with a Monte Carlo Constant Pressure Algorithm. *Chem. Phys. Lett.* **2004**, *384*, 288–294.

(35) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J.*

Comput. Phys. **1977**, *23*, 327–341.

(36) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N· Log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(37) Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.

(38) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(39) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(40) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. QUASI: A General Purpose Implementation of the QM/MM Approach and Its Application to Problems in Catalysis. *J. Mol. Struct. THEOCHEM* **2003**, *632*, 1–28.

(41) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell-a Modular Software Package for QM/MM Simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 101–110. https://doi.org/10.1002/wcms.1163.

(42) Balasubramani, S. G.; Chen, G. P.; Coriani, S.; Diedenhofen, M.; Frank, M. S.; Franzke, Y.

J.; Furche, F.; Grotjahn, R.; Harding, M. E.; Hättig, C.; Hellweg, A.; Helmich-Paris, B.; Holzer, C.; Huniar, U.; Kaupp, M.; Marefat Khah, A.; Karbalaei Khani, S.; Müller, T.; Mack, F.; Nguyen, B. D.; Parker, S. M.; Perlt, E.; Rappoport, D.; Reiter, K.; Roy, S.; Rückert, M.; Schmitz, G.; Sierka, M.; Tapavicza, E.; Tew, D. P.; Van Wüllen, C.; Voora, V. K.; Weigend, F.; Wodyński, A.; Yu, J. M. TURBOMOLE: Modular Program Suite for Ab Initio Quantum-Chemical and Condensed-Matter Simulations. *J. Chem. Phys.* **2020**, *152*, 184107.

(43)   Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(44)   Smith, W.; Forester, T. R. DL_POLY_2.0: A General-Purpose Parallel Molecular Dynamics Simulation Package. *J. Mol. Graph.* **1996**, *14*, 136–141.

(45)   Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A* **2009**, *113*, 11856–11865.

(46)   Yadav, S.; Shaik, S.; Siddiqui, S. A.; Kalita, S.; Dubey, K. D. Local Electric Fields Dictate Function: The Different Product Selectivities Observed for Fatty Acid Oxidation by Two Deceptively Very Similar P450-Peroxygenases OleT and BSβ. *J. Chem. Inf. Model.* **2022**, *62*, 1025–1035.

(47)   Dubey, K. D.; Wang, B.; Shaik, S. Molecular Dynamics and QM/MM Calculations Predict the Substrate-Induced Gating of Cytochrome P450 BM3 and the Regio- and Stereoselectivity of Fatty Acid Hydroxylation. *J. Am. Chem. Soc.* **2016**, *138*, 837–845.

(48) Siddiqui, S. A.; Dubey, K. D. Can the Local Electric Field Be a Descriptor of Catalytic Activity? A Case Study on Chorismate Mutase. *Phys. Chem. Chem. Phys.* **2022**, *24*, 1974–1981.

(49) Visser, S. P. De; Mukherjee, G.; Sastri, C. V. Local Charge Distributions, Electric Dipole Moments, and Local Electric Fields In Fl Uence Reactivity Patterns and Guide Regioselectivities in α - Ketoglutarate-Dependent Non-Heme Iron Dioxygenases. *Acc. Chem. Res.* **2022**, *55*, 65-74.

(50) Fried, S. D; Boxer, S.G. Electric Fields and Enzyme Catalysis, E. HHS Public Access. **2018**, 387–415. https://doi.org/10.1146/annurev-biochem-061516-044432.

(51) Shaik, S.; Visser, S. P. De; Kumar, D. External Electric Field Will Control the Selectivity of Enzymatic-Like Bond Activations. *J. Am. Chem. Soc.* **2004**, *126*, 11746–11749.

(52) Stuyver, T.; Huang, J.; Mallick, D.; Danovich, D.; Shaik, S. TITAN: A Code for Modeling and Generating Electric Fields—Features and Applications to Enzymatic Reactivity. *J. Comput. Chem.* **2020**, *41*, 74–82.

For Table of Contents Only.