

Structure-based modeling reveals molecular basis for CYP153A6's novel activity toward toluene derivatives

Received: 31 December 2025

Accepted: 2 February 2026

Published online: 06 February 2026

Cite this article as: Wei Y., Donzella S., Foiadelli S. *et al.* Structure-based modeling reveals molecular basis for CYP153A6's novel activity toward toluene derivatives. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-38986-7>

Yao Wei, Silvia Donzella, Sara Foiadelli, Francesco Molinari, Uliano Guerrini & Ivano Eberini

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

Structure-based modeling reveals molecular basis for CYP153A6's novel activity toward toluene derivatives

Yao Wei¹, Silvia Donzella², Sara Foadelli², Francesco Molinari², Uliano Guerrini¹, Ivano Eberini^{1*}

¹Dipartimento di Scienze Farmacologiche e Biomolecolari "Rodolfo Paoletti", Università degli Studi di Milano, Via Giuseppe Balzaretto 9, 20133, Milano, Italy

² Dipartimento di Scienze per gli Alimenti, la Nutrizione e l'Ambiente (DeFENS), Università degli Studi di Milano, Via Luigi Mangiagalli 25, 20133, Milano, Italy

*Corresponding author email: ivano.eberini@unimi.it

Abstract

CYP153A6 from *Mycobacterium* sp. strain HXN-1500 is a monooxygenase that catalyzes the selective hydroxylation of terminal methyl groups in various alkanes. This study explored CYP153A6 activity towards a set of toluene derivatives, along with the underlying molecular recognition. Initial *in vivo* evaluation of CYP153A6 activity, conducted using both whole cells and cell-free extracts, showed efficient conversion of toluene derivatives with apolar or slightly polar substituents, while no detectable activity was observed for derivatives bearing more polar groups. A homology model of CYP153A6 3D structure was built and validated, revealing key structural features and molecular tunnels. An ensemble docking strategy was used to identify the most effective docking setup. Molecular dynamics simulations and binding free energy calculations further confirmed the hydrophobic nature of the active site. QM/MM calculations supported the different reactivity observed between p-chlorotoluene and p-nitrotoluene. Toluene derivatives bearing a hydroxyl or nitro group on the aromatic ring exhibit reduced binding affinity, adopting unfavorable orientations and non-productive distances between the methyl group and the enzyme's heme iron center. These computational findings agree with experimental data. Overall, this study provides valuable insights into CYP153A6 molecular recognition mechanism and lay a strong foundation for future protein engineering to extend CYP153A6 enzyme substrate scope and/or enhance the product yield.

Key Words: CYP153A6; Toluene Derivatives; Molecular Modeling; Molecular Recognition

1. Introduction

Selective hydroxylation of methyl groups is a highly valuable transformation in synthetic chemistry with particular attention to green chemistry. This process enables the functionalization of the inert C-H bonds, which is critical for enhancing chemical solubility, and generating novel molecular frameworks [1][2]. Environmentally, hydroxylation of hydrophobic hydrocarbons facilitates detoxification and biodegradation of industrial pollutants, particularly toluene derivatives frequently encountered in fuels, solvents, and chemical manufacturing [3].

The selective hydroxylation of benzylic positions in toluene derivatives represents a critical transformation in synthetic and pharmaceutical chemistry, offering a sustainable alternative to traditional chemical oxidation methods [4]. Enzymes, particularly cytochrome P450 (CYP450) monooxygenases and flavin-dependent oxygenases, have emerged as powerful biocatalysts for these reactions due to their regio- and stereo-selectivity under mild conditions. A notable example is that CYP102A1 converts ethylbenzene to (S)-1-phenylethanol with high enantioselectivity (95% ee) [5]. Similarly, 4-hydroxyphenylpyruvate dioxygenase (HPPD) catalyzes the hydroxylation of 4-hydroxyphenylpyruvate to homogentisate via benzylic oxidation [6]. Non-heme iron oxygenases can also hydroxylate benzylic carbons in aromatic compounds such as ethylbenzene, yielding (S)-1-phenylethanol [7]. Flavin-dependent monooxygenases, including styrene monooxygenase, oxidize styrene derivatives at the benzylic position to form epoxides or alcohols [8]. Together, these approaches underscore a broader shift toward environmentally benign and functionally selective hydroxylation methods – particularly relevant for the modification of pharmaceuticals and natural products.

CYP153A monooxygenases belong to a distinct subfamily of cytochrome P450 enzymes (CYPs) and are primarily involved in the oxidation of aliphatic and aromatic hydrocarbons, with a strong

preference for terminal hydroxylation of n-alkanes and similar substrates [9][8][10]. Among them, CYP153A6 from *Mycobacterium* sp. HXN-1500 is a soluble monooxygenase that exhibits selective activity toward terminal methyl groups in a variety of linear alkanes and alicyclic substrates [11][12]. Notably, it selectively hydroxylates various monoterpene derivatives with high regioselectivity at the allylic methyl group attached to the aliphatic ring. This catalytic feature enables the efficient conversion of limonene into perillyl alcohol, a biotransformation of significant industrial relevance [13]. In an earlier study, CYP153A6 was shown to exhibit activity toward p-cymene, although the resulting product was not characterized; this observation suggests a potentially broader substrate scope, including the ability to target benzylic methyl groups. The selective oxidation of benzylic positions remains a significant challenge in both chemical and enzymatic catalysis. Benzylic hydroxylation offers a valuable route to benzyl alcohols, which are important intermediates in the pharmaceutical, fragrance, and fine chemical industries [14][15][16]. Given its inherent regioselectivity and terminal methyl preference, CYP153A6 emerges as a potential promising candidate for the selective hydroxylation of methyl substituents on the aromatic rings.

In this study, we investigate the catalytic potential of CYP153A6 for benzylic hydroxylation in series of para-, meta-, and ortho-substituted derivatives. Preliminary experimental work was carried out using whole-cell biocatalysis and crude cell-free extracts of recombinant *E. coli* expressing CYP153A6. To guide the future protein engineering efforts, we adopted a set of computational modeling strategies. This included homology modeling, molecular dynamics (MD) simulations, substrate tunnel calculations, ensemble docking, protein-ligand interaction fingerprints analysis, MM-GBSA energy analysis, and QM/MM calculations. The aim of this study is to elucidate the CYP153A6 molecular recognition mechanism and provide insights into the toluene derivatives binding modes and the driven potential. This will lay the groundwork for expanding the catalytic ability of CYP153A6 toward high-value aromatic compounds through selective toluene derivative hydroxylation.

2. Results and Discussion

2.1 Activity on toluene derivatives

Preliminary experiments were performed using whole cells (WC) of recombinant CYP153A6-*E. coli* and crude cell free extract (CFE). We first evaluated the hydroxylation of p-cymene (1a) using a nominal

substrate concentration of 6 mM; reactions were carried out at a formal substrate concentration of 6 mM, exceeding the aqueous solubility of the compound (**Figure 1**).

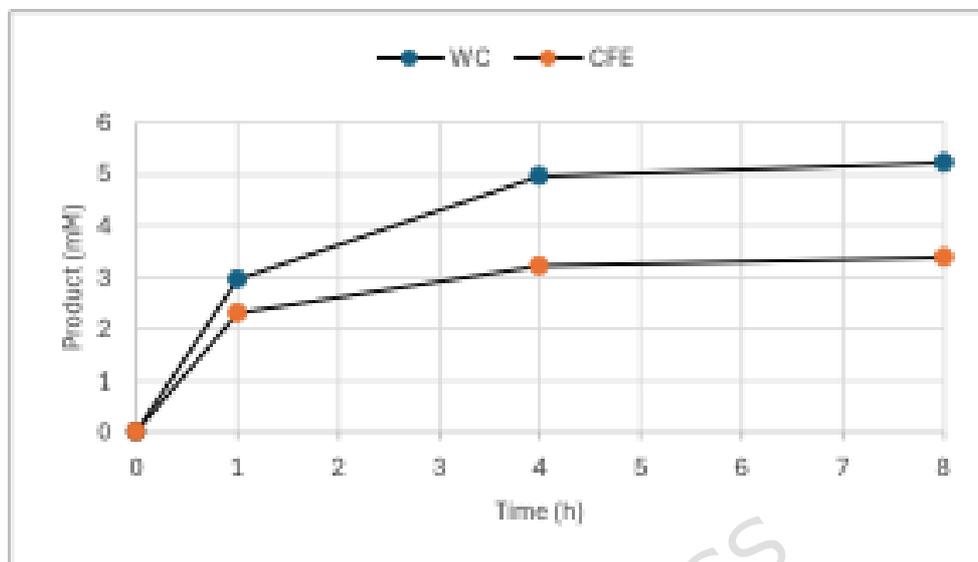
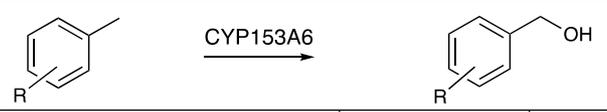


Figure 1. Biotransformation of *p*-cymene using WC and CFE (concentration of CYP153A6 1.8 μ M) of *E. coli* at pH 7.0, 6 mM substrate, 28°C. Reaction with CFE was carried out in the presence of 6 mM NADH.

p-Cymene underwent highly selective hydroxylation to yield *p*-isopropylbenzyl alcohol as the only detectable product, with conversions of 85% using whole cells and 57% using the cell-free extract (CFE) after 8 h. Kinetic parameters were measured using CFE containing substrate concentrations in the range of 0.5-10.0 mM, showing that CFEs were used in further experiments where the activity of CYP153A6 was investigated using different *p*-substituted toluene derivatives (1b-1h, **Table 1**). Because these substrates are highly hydrophobic, the kinetic constants are apparent values calculated from nominal substrate concentrations, potentially leading to an underestimation of the intrinsic enzymatic affinity. Hydroxylation occurred exclusively at the benzylic position, leading to the formation of the corresponding benzyl alcohols. The most reactive substrate was *p*-methylanisole (1d), which showed the highest specific activity (1.75 U mg⁻¹) and, being the most water-soluble among the active substrates, exhibited the lowest apparent K_m . Although the activity on toluene (1b) was sluggish, hydroxylation of *p*-xylene (1c) proceeded at a good rate; notably, only one of the two benzylic methyl groups of 1c was hydroxylated, even after prolonged incubation, highlighting the enzyme's remarkable selectivity. Good conversion was also observed for *p*-chlorotoluene

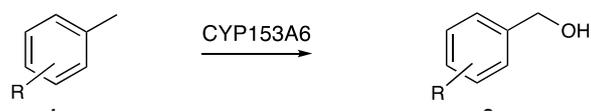
(1f), while *p*-cresol, *p*-methyl-benzylalcohol, and *p*-nitrotoluene showed no detectable activity. Because all these substrates are highly hydrophobic, the kinetic constants represent apparent values calculated from nominal substrate concentrations, and the elevated apparent K_M values are consistent with the limited aqueous solubility of most of the substrates. These experimental results suggest that CYP153A6 activity was strongly influenced by substituent polarity, favoring apolar or slightly polar toluene derivatives over more polar analogues.

Table 1. Specific activity and kinetic parameters of CYP153A6 (1.8 μ M) towards *p*-substituted toluene derivatives. Specific activity was measured using CFE at pH 7.0, 6 mM substrate, 28 °C.^a n.d. = not determined.

					
Substrate	R ₁	Specific activity (U mg ⁻¹)	k _{cat} (s ⁻¹)	K _M (mM)	k _{cat} /K _M (s ⁻¹ mM ⁻¹)
1a	CH(CH ₃) ₂	1.15 ± 0.05	2.54 ± 0.06	12.0 ± 0.8	0.21 ± 0.07
1b	H	< 0.15	n.d. ^a	n.d.	n.d.
1c	CH ₃	1.05 ± 0.05	1.44 ± 0.08	9.4 ± 0.9	0.15 ± 0.04
1d	OCH ₃	1.75 ± 0.08	2.14 ± 0.03	3.4 ± 0.2	0.63 ± 0.02
1e	OH	-	-	-	-
1f	Cl	0.99 ± 0.04	1.75 ± 0.04	9.3 ± 0.9	0.18 ± 0.03
1g	CH ₂ O	-	-	-	-
1h	NO ₂	-	-	-	-

CYP153A6 was also employed for the oxidation of a series of meta- and ortho-substituted toluene derivatives (1i-1o, **Table 2**); the substrates were chosen to have meaningful indication about the influence of the substituent position on the overall activity.

Table 2. Specific activity and kinetic parameters of CYP153A6 (1.8 μ M) towards *p*-substituted toluene derivatives. Specific activity was measured using CFE at pH 7.0, 6 mM substrate, 28 °C.

					
Substrate	R ₁	Specific activity (%) ^a	k _{cat} (s ⁻¹)	K _M (mM)	k _{cat} /K _M (s ⁻¹ mM ⁻¹)
1i	<i>o</i> -OH	< 5			
1j	<i>m</i> -OH	< 5			
1k	<i>o</i> -CH ₃	< 5			
1l	<i>m</i> -CH ₃	0.98	1.44	9.4	0.15
1m	<i>o</i> -CH ₂ OH	5			
1n	<i>m</i> -CH ₂ OH	< 5			

Substrates contain a hydroxyl group, either as a phenol (1i and 1j) or a primary alcohol (1m and 1n), exhibited minimal or undetectable activity, with a trend like that observed for the *p*-substituted toluene analogues. The activity observed across the series of xylene derivatives suggests that steric hindrance plays a critical role: *o*-xylene, due to its steric bulk, is unreactive, whereas *p*- and *m*-xylene exhibit comparable levels of reactivity.

2.2 Homology modeling of CYP153A6 3D structure

CYP153A6 structure has not yet been experimentally solved. Therefore, its 3D structure (**Figure 2** (a)) was modeled basing on the experimentally solved structure of CYP153A_{P,SP} (PDB ID 7A07) by X-ray diffraction [17]. This homologous template enzyme shares similar functionality with CYP153A6, and sequence alignment revealed a global identity of 78.6 % and a similarity of 86.5 %.

Upon superimposing the CYP153A6 and CYP153A_{P,SP} structures, the root mean square deviation (RMSD) of the C α atoms were calculated to be smaller than 2 Å, indicating a good structural overlap. The Ramachandran plot (**Figure 2** (b)) further supports the high quality of the CYP153A6 homology model, with most residues occupying favorable conformational regions and only a few outliers. Notably, the plot closely resembles that of the experimentally solved template structure, in which a similar number of outliers is also observed. These results suggest that the model is of high structural quality [18].

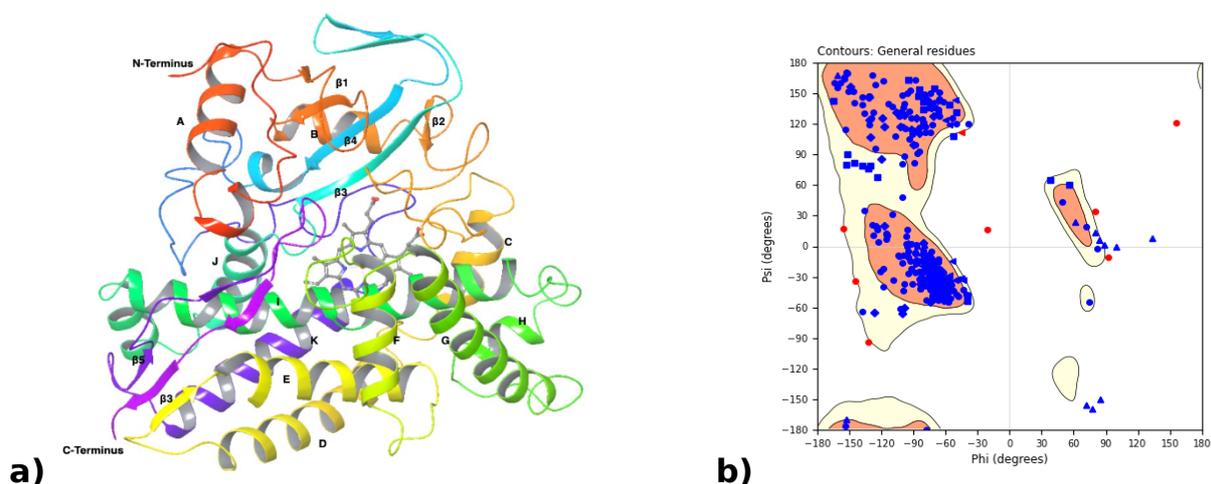


Figure 2. a) CYP153A6 homology model: Cofactor heme is shown in gray sticks, 11 α -helices are labeled with A-K, 5 β -sheets are labeled with β 1-5. **b) Ramachandran plot of the CYP153A6 homology model.** (Figures were produced with the Schrödinger Maestro Suite (release 2024-2) [19], and protein secondary structures were labeled using Diagram program: <https://app.diagrams.net/>)

To assess the structural stability of the CYP153A6 homology model, classical molecular dynamics (MD) simulations were performed. The RMSD profile of the protein (**Figure 3**), calculated on the C α atoms of the amino acids, stabilized around $\sim 3\text{\AA}$, indicating that the backbone structure reached equilibrium during the simulation. This suggests that the homology model maintains a reasonable overall fold, providing confidence in the generated 3D structure for subsequent analyses. No ligands were included in this simulation, and the primary goal was to verify the structural reliability of the protein model.

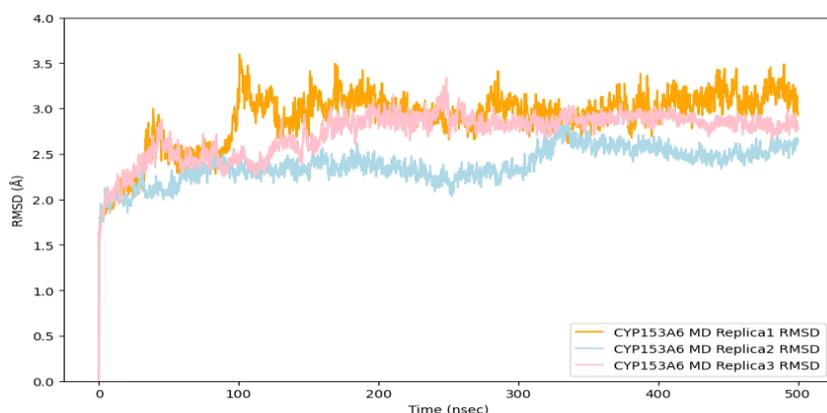


Figure 3. C α RMSD values of CYP153A6 from three replicate MD simulations.

2.3 Molecular Tunnels Contributes to Enzyme Activity

The active site of CYP450s is deeply buried within the protein core [20]. These enzymes possess molecular tunnels that connect the active site to the surrounding solvent environment, allowing substrate access and product release. These tunnels play a crucial role in determining substrate specificity [21]. Typically, tunnels involved in the uptake of hydrophobic, membrane-associated substrates are oriented toward the membrane, while those associated with the release of hydrophilic molecules – such as oxidized products or water – extend into the solvent [22].

Tunnel analysis of CYP153A6 revealed three distinct molecular tunnels (**Figure 4**): two are likely substrate access tunnels (designated as tunnels 2b and 2c), and one is predicted to work as a water tunnel (designated as tunnel W). Substrate tunnel 2b has a throughput score of 0.48 and a bottleneck radius of 1.15 Å. It involves residues GLY86, GLY87, ILE88, THR89, ALA96, ALA97, SER98, LEU99, MET101, ILE103, and ALA104, with ALA104 identified as the bottleneck residue. Substrate tunnel 2c and water tunnel W share the same set of lining residues: THR89, MET91, ASN94, ALA97, THR186, THR187, ALA188, SER405, PHE407, and VAL408, with MET91 acting as bottleneck residue in both tunnels. The throughput score values for tunnels 2c and W are 0.34 and 0.31, respectively, and both a bottleneck radius of 1.00 Å. To *in-silico* validate tunnel W as a water pathway, water molecules were generated using the three-dimensional reference interaction site model (3D-RISM). This method provides time-averaged distribution of hydrogen and oxygen densities in water, as well as free-energy maps that help assess solvent stability and solvation contributions to binding free-energy [23]. All generated water molecules were located within the predicted tunnel (**Figure 5**), supporting its functional consequence.

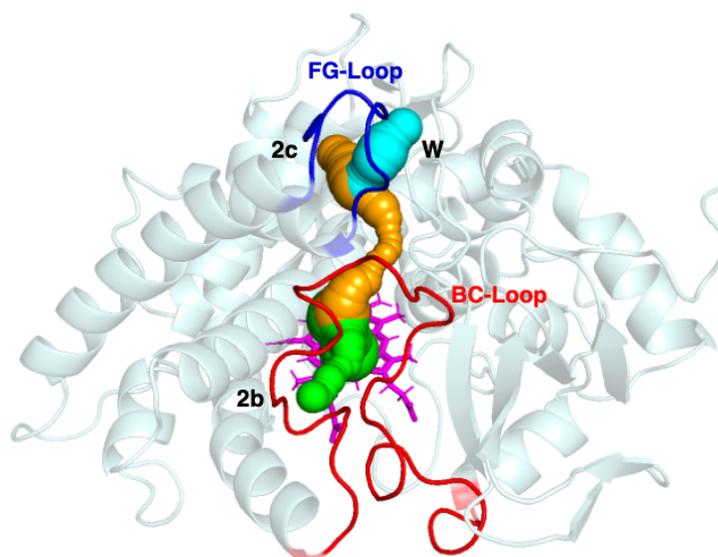


Figure 4. Detected molecular tunnels in CYP153A6. The paths of tunnels 2b (green), 2c (orange), W (cyan) are depicted. Enzyme cofactor heme is shown in magentas sticks, FG loop is shown in blue, BC loop is shown in red (Protein structure was displayed by PyMOL [24], molecular tunnels were generated using CAVER 3.03 plugin software [25], protein secondary structures were labeled using Diagram program: <https://app.diagrams.net/>)

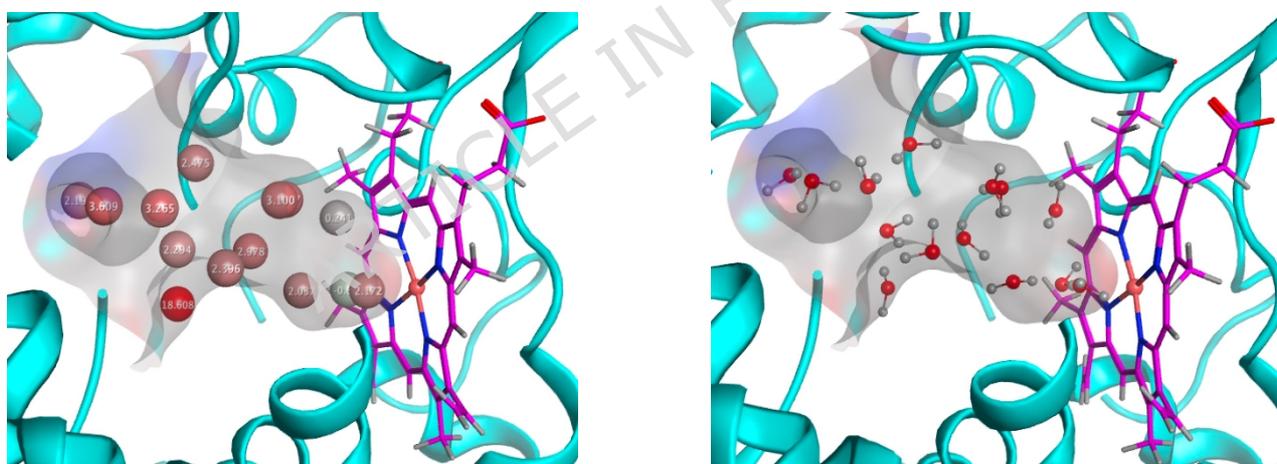


Figure 5. *In silico* solvent analysis of the CYP153A6 water tunnel molecular surface using 3D-RISM. Left: hydration free energy ranking of the predicted water molecules. Right: distribution of 3D-RISM-predicted water molecules within the identified water tunnel. The cofactor heme is shown in magentas sticks, water molecules are shown in the gray (H) - red (O) ball-stick representation, the water tunnel is shown as gray molecular surface. (Protein structures and the *in-silico* solvents were generated using MOE 2022.02 [26])

All three tunnels are structurally framed by the enzyme's FG and BC loops. Consistently with their role in regulating tunnel access, classical MD simulations revealed that these loops are highly flexible, as indicated by the root mean square fluctuation (RMSF) plot (**Figure 6**).

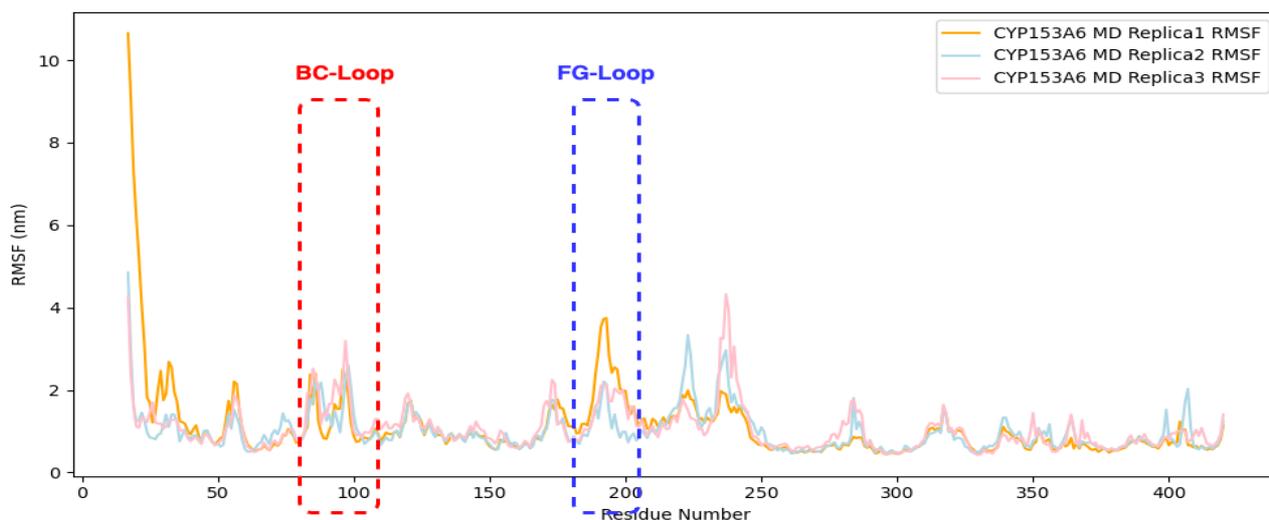


Figure 6. RMSF plot of C α atoms from the CYP153A6 protein during MD simulations.

2.4 CYP153A6 ensemble docking study.

The binding site of CYP153A6 comprises the heme-binding region and an adjacent substrate-binding pocket, a structural hallmark of CYP450 enzymes. This pocket is deeply embedded within the enzyme's core and features an iron atom coordinated within a porphyrin ring [27]. In its resting ferric (Fe^{3+}) oxidation state, the cofactor heme iron is critical for the enzyme's catalytic function. Several loops surrounding the active site exhibit conformational flexibility, allowing them to adapt to various ligand structures during the binding process [28]. In addition, multiple molecular access/egress tunnels - partially involving the substrate recognition sites (SRSs) - have been proposed for this enzyme [29]. Therefore, we employed an ensemble docking strategy [30] to further investigate the molecular recognition mechanism of CYP153A6. This docking workflow can be found in Figure S1.

The generation of protein structural ensemble was based on classical MD simulations. Ensemble docking was carried out under four distinct cofactor heme iron charge states and presence or absence of explicit water molecules in the binding site. Water molecules were predicted using 3D-RISM. The four docking scenarios were: 1) ferric iron (Fe^{3+}) without explicit water molecules in the binding site; 2) ferric iron (Fe^{3+}) with explicit water molecules in the binding site; 3) ferrous iron (Fe^{2+}) without explicit water molecules in the binding site; 4) ferrous iron (Fe^{2+}) with explicit water molecules in the binding site.

Since the CYP153A6 model was generated *via* homology modeling, an *in silico* 3D-RISM solvent analysis was performed to predict solvent distribution. Based on the analysis results, water molecules were then added to the enzyme binding pocket for the following ensemble docking study. The analysis revealed significant hydration free energy contributions around surface regions within 10 Å of the binding pocket. In contrast, the active site - located between the heme cofactor and the ligand - exhibited markedly lower hydration free energy density, suggesting that the CYP153A6 active site possesses typical hydrophobic properties (**Figure 7**).

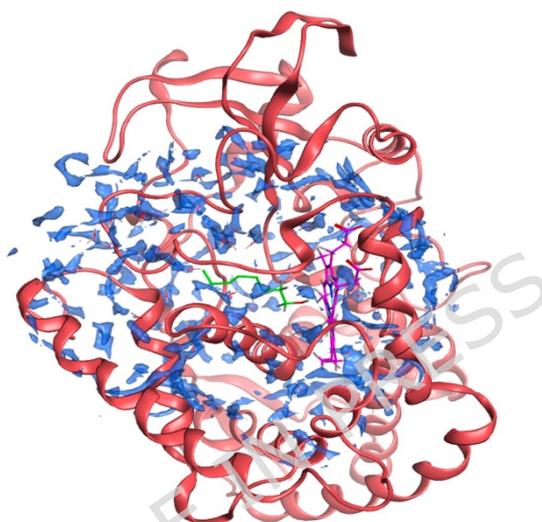
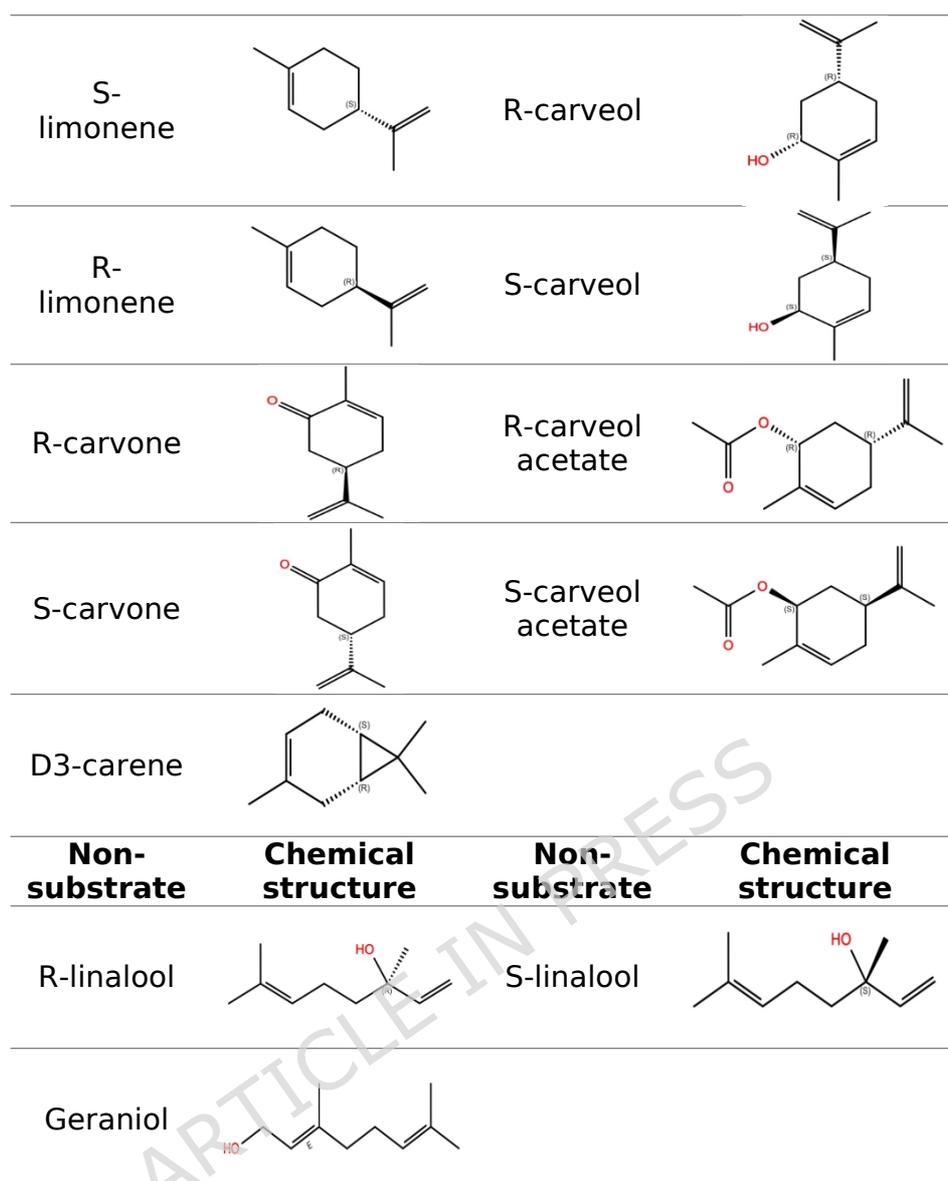


Figure 7. In silico solvent analysis of CYP153A6 surface (within a 10 Å cut off from the binding pocket). Hydration free energy density is shown as a solid blue surface. The enzyme cofactor heme is displayed as magentas sticks, and the ligand is shown as green sticks. (Protein structures and the *in-silico* solvent surface were generated using MOE 2022.02 [26])

A total of 12 already investigated compounds (**Table 3**)[13], comprising both experimentally verified substrates and non-substrates of CYP153A6, were used for evaluating the molecular docking workflow and conditions. Each compound was docked in all selected receptor snapshots under all four conditions. The docking results revealed diverse binding poses. In favorable cases, the reactive methyl group of the ligand was oriented toward the cofactor heme iron at a catalytically relevant distance (approx. 6 Å) [31]. In contrast, other poses positioned the methyl group oriented to the cofactor heme iron out of the 6 Å, or in the opposite direction, suggesting non-reactive binding poses.

Table 3. Chemicals for evaluating CYP153A6 docking conditions

Substrate	Chemical structure	Substrate	Chemical structure
-----------	--------------------	-----------	--------------------



The docking performance was assessed by calculating the proportion of reactive poses – those with the reactive group oriented toward the heme iron – relative to the total number of poses. To account for conformational variability of the enzyme, molecular docking was performed on three independent MD simulation replicates, and the results are summarized as box plots in **Figure 8**. This provided insight for us to select the most suitable docking condition for further CYP153A6 molecular recognition study. Docking poses were ranked using the GlideScore, which is an empirical scoring function that approximates complex-ligand binding free energy. The binding free energy values of the reactive docking poses are around -6.00 kcal/mol and are more negative than the non-reactive docking poses. The docking score distribution box plots can be found in (Figure S2).

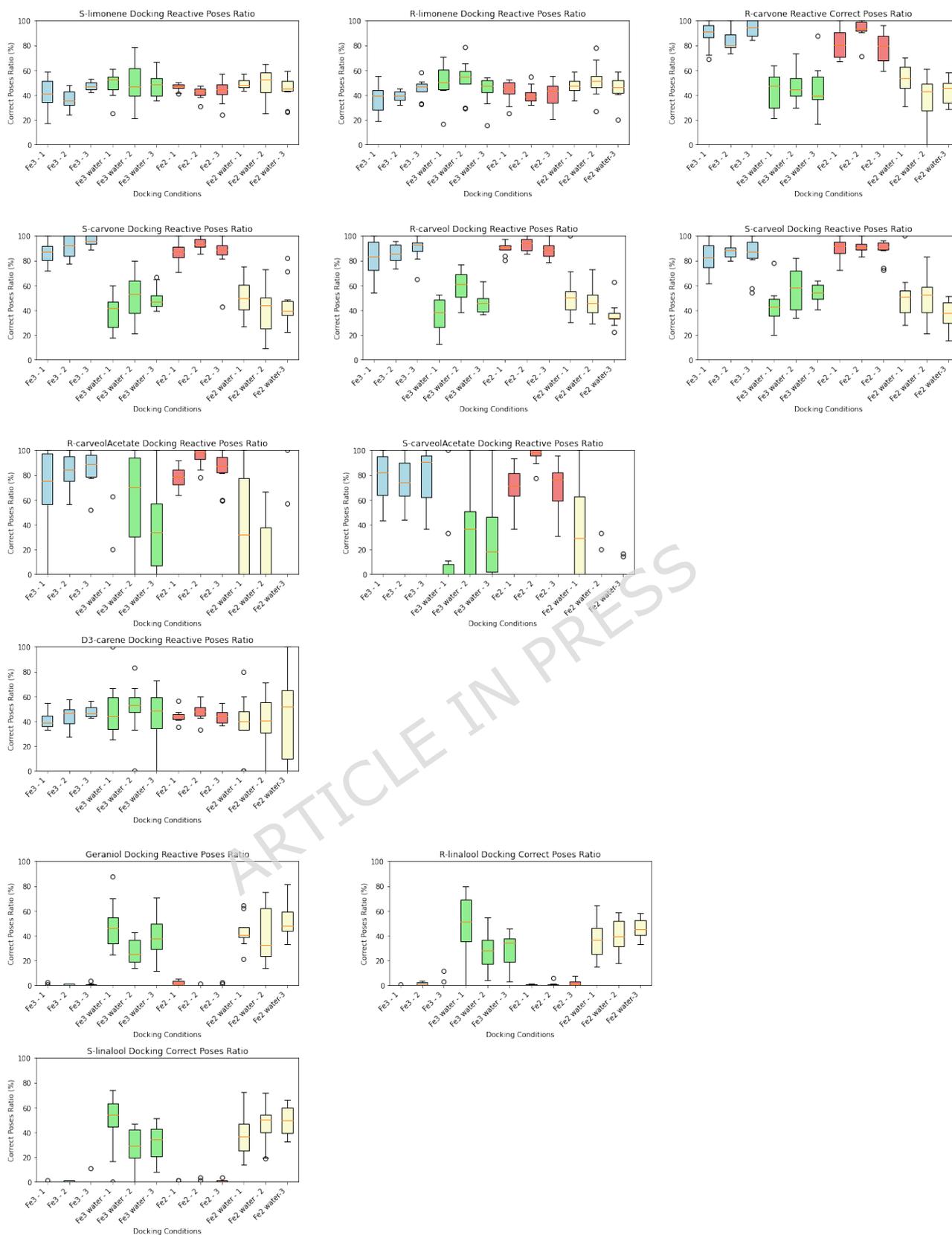


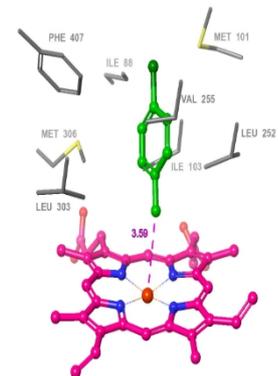
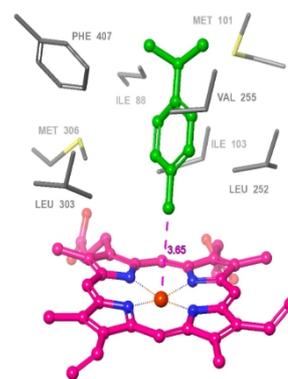
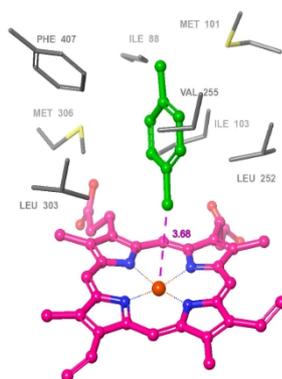
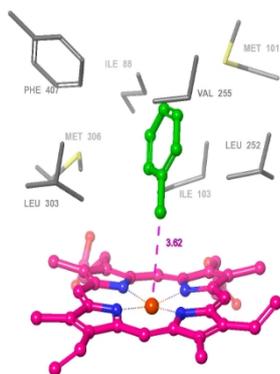
Figure 8. Box plots showing the distribution of reactive docking poses for published CYP153A6 substrates and non-substrates. Docking conditions: 1) Fe³⁺ without water (blue boxes); 2) Fe³⁺ with water (green boxes); 3) Fe²⁺ without water (red boxes); 4) Fe²⁺ with water (yellow boxes)

Among the four tested docking conditions studied, the setup using the cofactor heme iron in the ferric state (Fe^{3+}) without explicit water molecules in the binding site exhibited the best overall performance. This condition most effectively discriminated between CYP153A6 substrates and non-substrates in our test set, as evidenced by narrower box plots, smaller error bars, and fewer outliers for both the ratio of reactive docking poses and the corresponding docking scores. Consequently, this docking setup was selected for further molecular recognition study of CYP153A6.

2.5 Toluene derivatives: molecular characteristics of the CYP153A6 binding site and docking pose analysis.

To further investigate the molecular recognition profile of CYP153A6, 14 toluene derivatives (Table S1) were studied. Our validated molecular docking protocol - ferric heme iron (Fe^{3+}) without explicit water molecules in the binding site - was utilized to predict the binding modes of these toluene derivatives with CYP153A6.

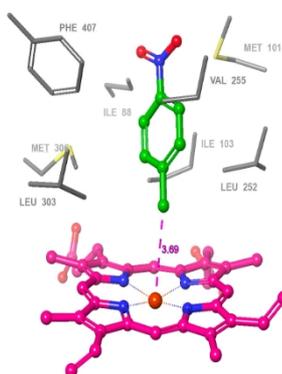
In substrate-bound models of CYP153A6 with toluene derivatives, some residues were found within 4 Å of the ligand (**Figure 9**). These include: ILE88, THR89, MET101, ILE103, THR 186, VAL251, LEU252, VAL255, GLY256, THR260, LEU303, MET306, and PHE407. Among these, hydrophobic residues predominate, such as ILE (2x), MET (2x), VAL (2x), LEU (2x), and PHE (1x), while polar residues such as THR (2x) are less frequent. This composition suggests that hydrophobic interactions are a major driving force in ligand binding within the CYP153A6 active site.



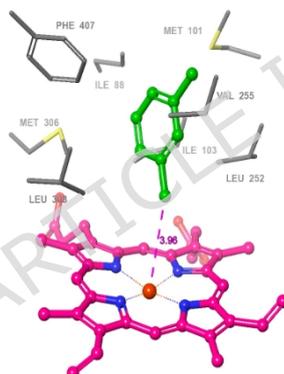
toluene
***p*-chlorotoluene**

***p*-xylene**

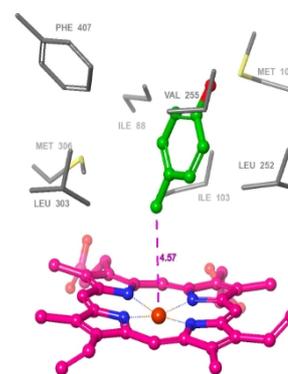
***p*-cymene**



***p*-nitrotoluene**
***p*-cresol**



***m*-xylene**



***m*-cresol**

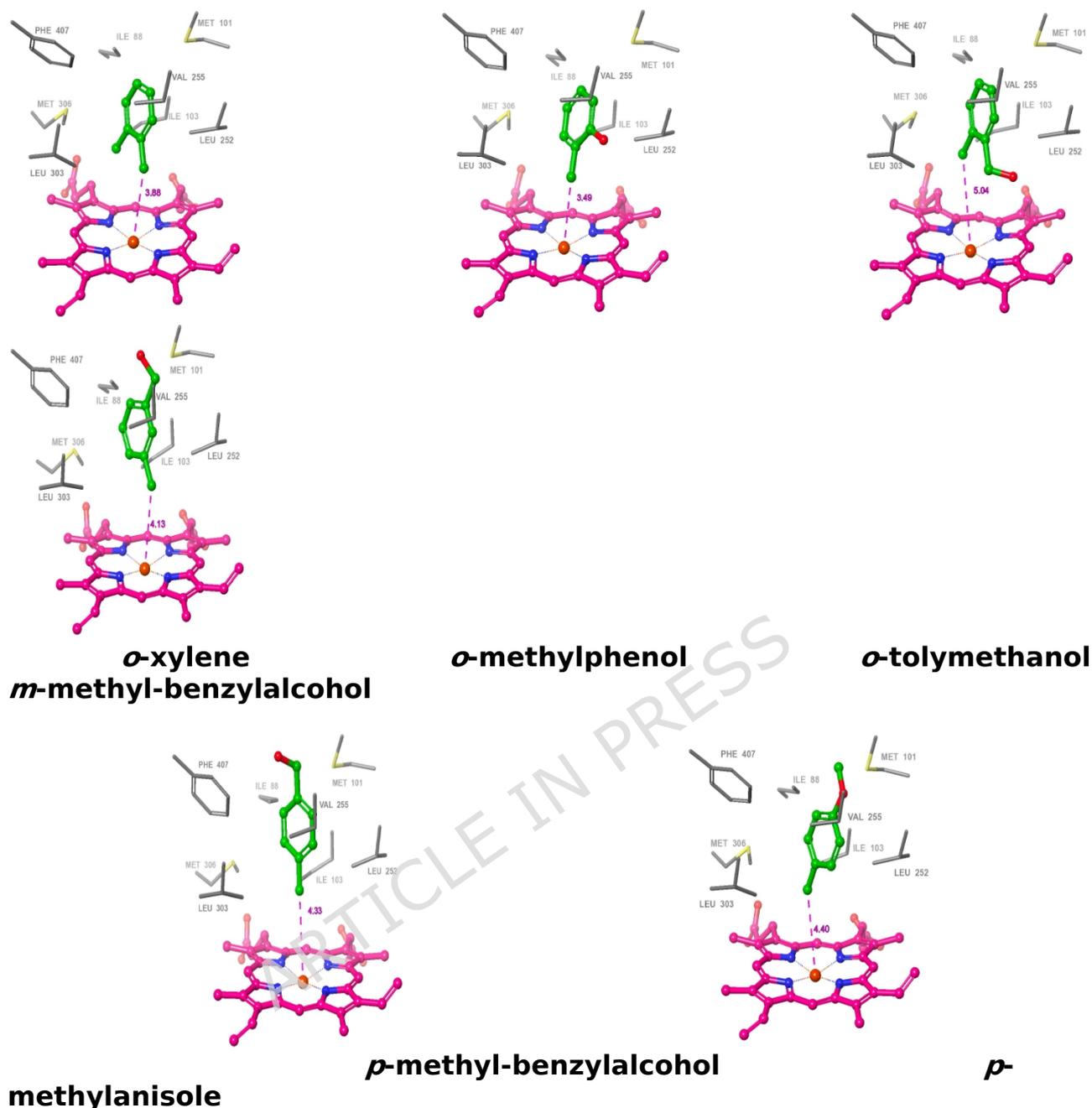


Figure 9. CYP153A6 toluene derivative substrates binding mode. (All the docking poses were produced with the Schrödinger Maestro Suite (release 2024-2) [19])

The binding modes of these toluene derivative substrates, obtained from the top-scoring docking conformations, are illustrated in **Figure 10**. To further characterize the interactions between ligands and active site residues, we employed the Schrödinger Protein-Ligand Interaction Fingerprint tool, which automatically generates visual summaries of significant molecular contacts. The resulting fingerprints revealed a variety of backbone and sidechain interactions, highlighting key residues involved in toluene derivative ligand recognition. Specifically, backbone interactions were primarily mediated by ALA97, ILE101, ILE103, LEU252, VAL255, GLY256,

LEU303, MET306, and PHE407. While sidechain and hydrophobic interactions involved ILE88, ALA97, MET101, ILE103, THR186, VAL251, LEU252, VAL255, LEU303, MET306, and PHE407. Notably, LEU252 and VAL255 engaged in both backbone and sidechain interactions with almost all the studied ligands, underscoring their dual role in substrate stabilization. The residues involved in the CYP153A6 protein-ligand interaction are shown in Figure 10, and the detailed analysis of the CYP153A6 protein-ligand interaction (within 3 Å) fingerprints are shown in Table S2. Table S2 shows the toluene derivative structures, and the ligand binding information, as well as the docking scores.

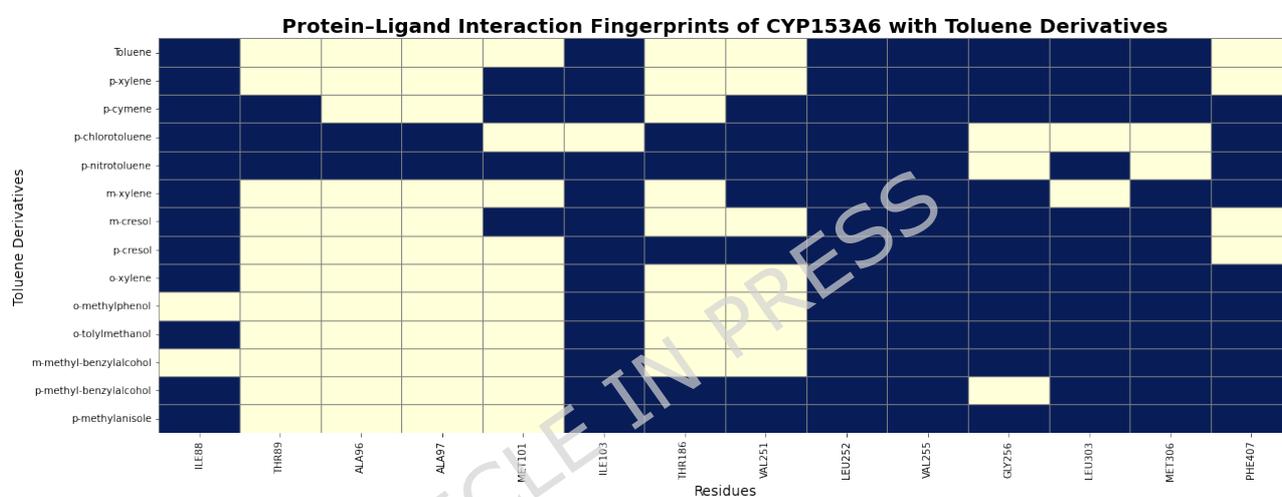


Figure 10. Analysis of protein-ligand interaction fingerprints of CYP153A6 with toluene derivatives.

2.6 Molecular dynamics simulations to verify the molecular recognition of CYP153A6 for toluene derivatives.

To further evaluate the binding behavior of toluene derivatives to CYP153A6, we have conducted three 200 ns replica classical MD simulations for the top scoring ligand binding poses derived from the molecular docking. Each simulation generated 1001 frames, and ligand binding free energy was calculated every 10th frame using MM-GBSA methods. The ligand binding efficacy was primarily assessed based on three criteria: (1) the distance of the ligand's reactive group to the enzyme cofactor heme iron (within a 6.00 Å cut-off), (2) the number of frames meeting this proximity threshold, (3) the reweighted MM-GBSA (Re- ΔG_{GBSA}) calculated using only the frames within the reactive range, (4) ligand RMSD. The detailed results are illustrated in **Table 4**. Additional factors such as the interactions from lipophilic, electrostatic, and ligand packing, were also considered to assess the ligand binding stability and quality, these data can be

found in the Table S3. The protein-ligand interaction of CYP153A6 with the toluene derivative during MD simulations is illustrated in Figure S3.

ARTICLE IN PRESS

Table 4. Ligand RMSDs, Ligand Reactive Methyl Group to the Enzyme Cofactor Heme Iron Distance and Binding Free Energies during MD Simulations for the Complexes of CYP153A6 and Toluene Derivatives (values are presented as mean \pm standard deviation).

	Simulation 1	Simulation 2	Simulation 3	Simulation 1	Simulation 2	Simulation 3
	Comparatively stronger binders					
	<i>p</i>-cymene			<i>p</i>-xylene		
RMSD (Å)	2.62 \pm 0.39	2.07 \pm 0.48	2.29 \pm 0.56	2.88 \pm 1.57	2.67 \pm 0.99	3.15 \pm 1.28
Distance (Å)	3.85 \pm 0.46	4.58 \pm 0.62	4.11 \pm 0.62	4.30 \pm 1.35	4.82 \pm 1.40	4.11 \pm 0.69
ΔG_{GBSA} (kcal/mol)	-14.14 \pm 4.46	-27.97 \pm 6.54	-14.83 \pm 5.18	-13.03 \pm 4.80	-15.40 \pm 6.14	-12.78 \pm 5.18
Frames	1000	993	994	908	884	983
Re- ΔG_{GBSA} (kcal/mol)	-14.14 \pm 4.46	-27.89 \pm 6.53	-14.83 \pm 5.18	-12.35 \pm 3.77	-14.50 \pm 5.41	-12.72 \pm 5.18
	Moderate binders					
	toluene			<i>m</i>-xylene		
RMSD (Å)	3.16 \pm 1.09	4.11 \pm 1.24	3.20 \pm 1.07	3.48 \pm 1.50	2.98 \pm 1.18	4.04 \pm 2.20
Distance (Å)	5.56 \pm 2.20	7.43 \pm 2.46	5.33 \pm 1.92	5.56 \pm 1.72	4.27 \pm 0.62	5.88 \pm 2.39
ΔG_{GBSA} (kcal/mol)	-7.47 \pm 3.59	-12.79 \pm 6.67	-9.21 \pm 4.05	-17.45 \pm 5.08	-12.47 \pm 4.13	-17.66 \pm 7.28
Frames	582	360	645	699	989	582
Re- ΔG_{GBSA} (kcal/mol)	-6.76 \pm 2.51	-7.86 \pm 5.36	-8.63 \pm 4.10	-15.87 \pm 4.67	-12.51 \pm 4.19	-8.41 \pm 4.42
	<i>p</i>-chlorotoluene			<i>p</i>-methylanisole		
RMSD (Å)	5.04 \pm 1.43	4.20 \pm 1.20	2.84 \pm 0.96	4.25 \pm 1.80	3.30 \pm 0.76	4.12 \pm 1.69
Distance (Å)	9.31 \pm 3.08	8.56 \pm 3.20	4.81 \pm 1.81	7.26 \pm 3.28	4.12 \pm 0.59	6.65 \pm 2.87
ΔG_{GBSA} (kcal/mol)	-22.32 \pm 6.00	-17.13 \pm 8.41	-14.22 \pm 4.28	-13.80 \pm 5.24	-15.36 \pm 4.05	-21.85 \pm 6.12
Frames	231	284	831	475	991	561
Re- ΔG_{GBSA} (kcal/mol)	-19.62 \pm 5.16	-19.21 \pm 6.74	-14.98 \pm 3.42	-11.82 \pm 5.64	-15.36 \pm 4.05	-18.69 \pm 5.41
	<i>p</i>-nitrotoluene			<i>p</i>-methyl-benzylalcohol		
RMSD (Å)	5.01 \pm 2.26	2.24 \pm 0.41	2.88 \pm 1.19	3.57 \pm 1.76	5.70 \pm 1.06	5.76 \pm 0.76
Distance (Å)	8.97 \pm 4.26	3.93 \pm 0.47	7.15 \pm 1.98	6.25 \pm 2.78	10.25 \pm 2.03	9.62 \pm 0.99
ΔG_{GBSA} (kcal/mol)	-22.64 \pm 7.58	-14.73 \pm 3.26	-23.17 \pm 7.69	-19.05 \pm 9.37	-28.05 \pm 6.28	-36.95 \pm 3.74
Frames	391	997	271	590	49	20
Re- ΔG_{GBSA} (kcal/mol)	-15.87 \pm 5.39	-14.73 \pm 3.29	-20.63 \pm 5.36	-12.77 \pm 5.45	-20.84 \pm 7.56	-24.98 \pm 9.54
	Weak binders					
	<i>o</i>-methylphenol			<i>o</i>-tolymethanol		
RMSD (Å)	2.98 \pm 0.56	4.34 \pm 1.86	3.98 \pm 1.43	1.52 \pm 0.32	1.74 \pm 0.65	1.89 \pm 0.39
Distance (Å)	4.44 \pm 0.92	7.86 \pm 3.28	7.04 \pm 1.98	4.82 \pm 0.26	5.00 \pm 0.70	5.09 \pm 0.50
ΔG_{GBSA} (kcal/mol)	-8.77 \pm 4.48	-19.82 \pm 5.86	-14.23 \pm 5.30	-3.47 \pm 6.90	-8.30 \pm 8.02	-7.50 \pm 6.24
Frames	945	451	441	999	935	929
Re- ΔG_{GBSA} (kcal/mol)	-8.61 \pm 4.50	-16.96 \pm 3.00	-11.71 \pm 5.11	-3.47 \pm 6.90	-7.78 \pm 7.97	-7.23 \pm 6.18
	<i>o</i>-xylene			<i>m</i>-methyl-benzylalcohol		
RMSD (Å)	4.14 \pm 1.34	4.86 \pm 2.40	5.57 \pm 1.28	6.91 \pm 1.92	4.01 \pm 1.47	4.34 \pm 1.11
Distance (Å)	6.62 \pm 2.18	7.30 \pm 2.95	7.99 \pm 1.92	11.02 \pm 3.13	7.30 \pm 2.00	7.46 \pm 1.65
ΔG_{GBSA} (kcal/mol)	-16.70 \pm 5.71	-18.48 \pm 6.05	-21.95 \pm 6.61	-28.15 \pm 9.15	-18.81 \pm 10.54	-5.57 \pm 8.05
Frames	473	422	160	149	316	183
Re- ΔG_{GBSA} (kcal/mol)	-13.79 \pm 4.45	-14.23 \pm 5.00	-12.16 \pm 6.40	-16.13 \pm 7.10	-20.78 \pm 5.09	-13.92 \pm 7.07

	<i>p</i> -cresol			<i>m</i> -cresol		
RMSD (Å)	4.13 ± 1.81	4.92 ± 0.88	3.96 ± 1.22	2.10 ± 0.54	5.00 ± 0.66	4.51 ± 1.30
Distance (Å)	7.30 ± 3.17	6.06 ± 2.18	4.73 ± 2.21	9.98 ± 0.71	9.29 ± 4.26	8.11 ± 2.39
ΔG_{GBSA} (kcal/mol)	-20.55 ± 11.92	-13.85 ± 6.47	-10.58 ± 6.62	-26.24 ± 3.59	-19.10 ± 9.69	-19.76 ± 8.84
Frames	398	587	893	0	371	207
Re- ΔG_{GBSA} (kcal/mol)	-11.07 ± 8.17	-11.76 ± 7.34	-8.909 ± 4.39	0	-10.02 ± 5.06	-11.65 ± 6.94

The average binding free energy (ΔG_{GBSA}) values across simulations revealed significant variation among the ligands, indicating distinct binding profiles. Among the tested toluene derivatives, *p*-cymene and *p*-xylene showed consistently strong and stable binding. *p*-cymene maintained short reactive methyl group distances to heme iron (3.85 – 4.58 Å) with nearly all frames in this reactive range, and strong Re- ΔG_{GBSA} values (up to -27.89 kcal/mol), indicating a relatively favorable binding mode compared with other ligands. Similarly, *p*-xylene showed general short reactive distances (4.11 – 4.82 Å), with over 884 frames. The Re- ΔG_{GBSA} values ranging from -12.35 to -14.50 kcal/mol.

Moderate binders included toluene, *m*-xylene, *p*-methylanisole, *p*-chlorotoluene, and *p*-nitrotoluene. Two out of three toluene simulations showed the reactive distance around 5.33 – 5.56 Å, with more than 582 frames within a 6.00 Å reactive distance. However, in another simulation, the reactive distance was 7.43 ± 2.46 Å, and all Re- ΔG_{GBSA} values ranged from -13.22 to -2.50 kcal/mol, reflecting relatively weaker binding. *m*-xylene exhibited consistent reactive contact with the heme iron and moderate binding energy, but with slightly elevated RMSD. *p*-methylanisole showed a wide variation in reactive distances, ranging from (4.12 – 7.26 Å), with one simulation with 991 frames within the 6.00 Å reactive distance, while others having 475 and 561 qualifying frames, respectively. In the simulation with good proximity, the binding energy was strong (Re- $\Delta G_{\text{GBSA}} = -15.36 \pm 4.05$ kcal/mol). *p*-chlorotoluene displayed mixed results: in one simulation, the reactive distance was favorable with strong energetic binding (-14.98 ± 3.42 kcal/mol), but in others, the average reactive distances exceeded 6.00 Å, significantly reducing their hydroxylation potential. Despite *p*-nitrotoluene exhibited similar binding energies and reactive distances to *p*-chlorotoluene, however it is not catalyzed by CYP153A6 in the wet lab experiments. This discrepancy suggests that while binding is similar, the nitro group exerts an unfavorable electronic effect, likely reducing the reactivity of the benzylic C-H bond by withdrawing electron density. Nitro substituents are well-known to decrease the susceptibility of

para-position to oxidative reaction due to their strong electron-withdrawing properties [32].

In contrast, *m*-cresol, *o*-xylene, *m*-methylbenzyl alcohol, *o*-methylphenol, and *o*-tolymethanol showed limited binding within the reactive distance. Despite occasionally favorable ΔG_{GBSA} values, they failed to consistently maintain the 6.00 Å reactive distance, with frames often below 391 or reactive distances greater than 8.00 Å, which is too far for effective hydroxylation. *m*-cresol showed zero frame within range in one simulation and very high average distances (> 9.00 Å) in others. *o*-xylene and *p*-methylbenzyl alcohol exhibited poor proximity (reactive distances were consistently greater than 6 Å across all MD simulations) and showed inconsistent energy profiles (the ΔG_{GBSA} values varied substantially among simulations). Similarly, *o*-methylphenol exhibited variable binding behavior, with one simulation maintaining favorable reactive distance (4.44 ± 0.92 Å and 945 frames within cutoff), while the others exceeded the reactive range (> 6.50 Å) with fewer than 500 qualifying frames. Its binding energies varied moderately (Re- ΔG_{GBSA} from -8.61 to -16.96 kcal/mol), suggesting sporadic but insufficiently stable interactions for effective hydroxylation. Although *o*-tolymethanol exhibited relatively closed reactive distances (4.82 – 5.09 Å) and low RMSD values (< 2.00 Å) across simulations, its Re- ΔG_{GBSA} varied from -3.47 to -7.78 kcal/mol, indicating an overall unfavorable protein-ligand interaction.

p-cresol had reaction distance below the cutoff only in the replicate characterized by the least favorable Re- ΔG_{GBSA} , while other replicates showed larger distances and still not much favorable Re- ΔG_{GBSA} . *p*-methylbenzyl alcohol was a largely ineffective binder, particularly unstable with minimal frames with the reactive distance (20 and 49 frames) in two of the three replicas. Although it presented strong ΔG_{GBSA} values in some frames, the lack of reactive proximity undermines its likelihood of effective hydroxylation.

Overall, for effective protein-ligand interaction with CYP153A6 enzyme, the reactive group of ligands must remain close and stable, positioning to the heme iron over time. Given their strong binding free energies, *p*-cymene and *p*-xylene behave as promising candidates for future protein design strategies based on these criteria, while ligands such as *m*-cresol, *o*-xylene, *m*-methylbenzyl alcohol, *o*-methylphenol, *p*-cresol and *p*-methylbenzyl alcohol are less likely to be effective CYP153A6 substrates, due to poor spatial engagement and/or unstable binding.

2.7 QM/MM calculations to explore the effect of *p*-nitro substitution on CYP153A6 catalytical activity.

To further investigate how properties of the *p*-nitro substituent affect CYP153A6 activity, we performed QM/MM calculations. Substrate oxidation in CYP450 enzymes occurs when the enzyme is in the Compound I (Cpd I) intermediate state. The reaction of the iron(IV)-oxo species with the molecule requires an optimal geometry to achieve efficient orbital overlap and effective biotransformation. The reactivity of Cpd I primarily determines the selectivity of CYP450s oxidative reactions [33].

In this study, we employed QM/MM calculations to evaluate the effect of a *p*-nitro group on CYP153A6 reactivity by comparing *p*-nitrotoluene and *p*-chlorotoluene. As shown in Table 4, MD simulations suggest that both chemicals display similar binding behavior, with *p*-nitrotoluene even showing slightly better results. However, experimentally, CYP153A6 exhibits high catalytic reactivity toward *p*-chlorotoluene but almost no reactivity toward *p*-nitrotoluene.

Referring to previous studies [33], the quartet and doublet spin states of Cpd I exhibit comparable energetics during hydroxylation and epoxidation steps. Therefore, we only conducted detailed QM/MM calculations on quartet Cpd I here. For both chemicals, the Cpd I complexes were optimized at the QM/MM level. After optimization of *p*-nitrotoluene-Cpd I complex, the Fe-O bond length is 1.63 Å, the shortest O-H_{benz} distance is 2.95 Å, the corresponding Fe-O-C angle is 140,9° and O-H-C angle is 94.5°. In the *p*-chlorotoluene-Cpd I complex has an Fe-O bond length is 1.62 Å, the shortest O-H_{benz} distance is 2,97 Å, the corresponding Fe-O-C angle is 137,1°, and O-H-C angle is 112,5°. According to Don and colleagues [34] in another QM/MM calculation for CYP2D6 hydroxylation reaction, the O-H-C angle for reactants with the Cpd I spans from approx. 106° to 144°, and this suggests *p*-chlorotoluene is closer to a higher-quality near-reactive conformation. (**Figure 11**)

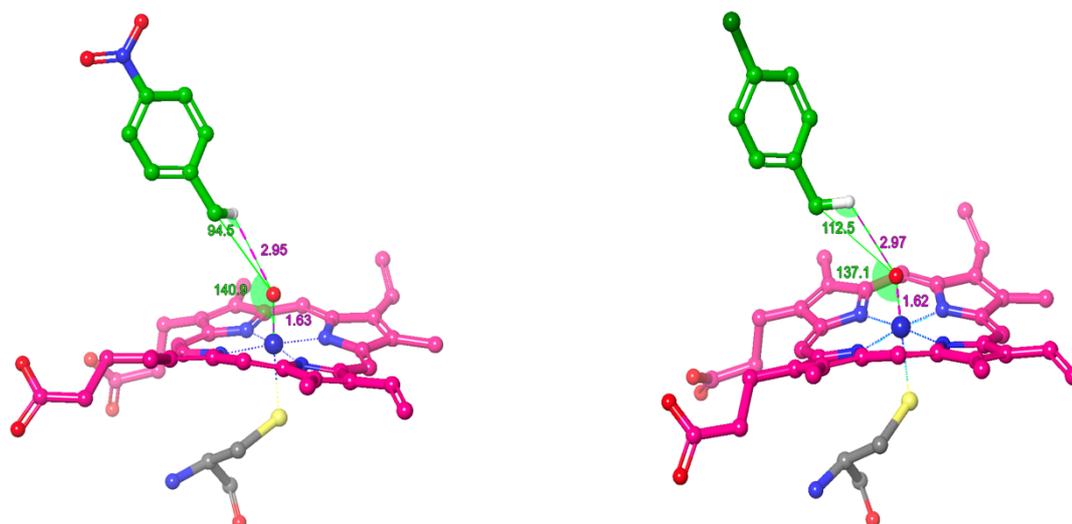


Figure 11. View on the Topology of CYP153A6 Compound I with *p*-nitrotoluene (left) and *p*-chlorotoluene (right). Distances are in Å and angles in degrees. (Figures were produced with the Schrödinger Maestro Suite (release 2025-1) [35])

Mulliken population analysis was performed during the QM/MM calculations to evaluate the hypothesis that CYP153A6 exhibits poor reactivity toward *p*-nitrotoluene. The atomic spin density of the heme iron in Cpd I is comparable in the Cpd I - *p*-nitrotoluene (1.10451) and Cpd I - *p*-chlorotoluene (1.12181) complexes. Similarly, the spin density of the Cpd I oxo oxygen shows no significant difference between the two systems (nitro, 0.83696 vs. chloro, 0.81862). In contrast, pronounced differences are observed for the substrates. The total spin density of *p*-nitrotoluene is very small (0.00226), whereas *p*-chlorotoluene exhibits a substantially higher value (0.35628). This trend is consistent across individual moieties: the aromatic ring (nitro, 0.00063 vs. chloro, 0.26337), methyl group (nitro, 0.00161 vs. chloro, 0.01611), and substituent group (nitro, 0.00002; chloro, 0.07680). The near-zero spin density of *p*-nitrotoluene indicates a minimal delocalization of the spin density of the porphyrin ring to the ligand itself, indicating a much smaller propensity of the site to oxidize the ligand. Differently, *p*-chlorotoluene exhibits higher spin density, indicating the incoming oxidation of the ligand by the CYP153A6 enzyme.

Atomic numbering for both ligands, cofactor heme iron, and Cpd I oxo oxygen is shown in Figure S4, and the corresponding spin density values are summarized in Table S4 and Table S5. QM/MM calculation convergence plots for geometry optimizations are provided in Figure S5.

3. Materials and Methods

3.1 Expression constructs

The synthetic gene sequence encoding for CYP153A6 and its two redox partners, ferredoxin and ferredoxin reductase, has been deposited in NCBI database with accession number OM622424. The pET100-CYP53A6 plasmid was expressed in *Escherichia coli* BL21 (DE3) using heat shock transformation and the recombinant cells were plated on LB (Luria-Bertani: 10 g L⁻¹ tryptone, 5 g L⁻¹ yeast extract, 10 g L⁻¹ NaCl) agar with 100 µg mL⁻¹ ampicillin.

3.2 Enzyme preparation

CYP153A6 was produced as recombinant protein as described before. A single colony was selected and inoculated in 20 mL of LB medium supplemented with 100 µg mL⁻¹ ampicillin and incubated at 37 °C at 120 rpm in 100 ml Erlenmeyer flasks for 16 h. The seed cultures were inoculated at an initial cell density of 0.1 OD₆₀₀ in 1 L baffled flasks containing 200 mL of SB medium (Super Broth: 32 g L⁻¹ tryptone, 20 g L⁻¹ yeast extract and 5 g L⁻¹ NaCl) with 100 µg mL⁻¹ ampicillin. The cultures were incubated at 37°C, 150 rpm until an OD₆₀₀ of 0.6-0.8 was reached, then brought to 4°C for 15 minutes and induced with isopropyl-β-d-thiogalactopyranoside (IPTG) 0.5 mM for 4 hours at 28 °C and 150 rpm. Cell pellets were recovered by centrifugation at 5000 rpm for 20 min at 4 °C, washed once and resuspended in 100 mM potassium phosphate buffer pH 7. Cell free extracts (CFE) were prepared by sonication for 6 cycles 1 minute each at 4°C. The insoluble fraction was removed by centrifugation at 15000 rpm for 1 hour at 4°C. The concentration of CYP153A6 was measured by CO-binding affinity using an extinction coefficient of 91.9 mM⁻¹ cm⁻¹ at 450 nm (Choi et al. 2012). Cell growth was measured both as optical density (OD_{600nm}, Eppendorf BioSpectrometer) and as cell dry weight (mg mL⁻¹) of washed cells coming from a known volume of culture.

3.3 Biotransformations, analyses and kinetics

All chemicals were from Sigma-Aldrich and were used without further purification. Biotransformations with WC were carried out in shaken flasks (100 mL) with 50 mg_{dry cells} mL⁻¹ in 10 mL phosphate buffer (100 mM) containing glucose (50 mM) at pH 7.0. Biotransformations

with CFE (CYP153A6 concentration in the cell extract was $\sim 1.8 \mu\text{M}$) were carried out in 10 ml flasks with tight sealing caps, in a final volume of 1 ml of 100 mM potassium phosphate buffer pH 7, using 10 mg of crude extract and 6 mM NADH. The substrates were prepared as 600 mM stock solutions in ethanol and added at a final concentration of 6 mM. After incubation at 28°C and 160 rpm, the reactions were stopped by extraction with an equal volume of EtOAc two times; the organic solvent was dried over anhydrous Na_2SO_4 and directly injected for GC-FID analysis. GC analyses were performed using a Shimadzu GC-2030 gas chromatographer equipped with a flame ionization detector (250 °C). Chromatographic conditions were as follows: column: Polyethylene Glycol (PEG) Mega-Wax® (30 m \times 0.25 mm); injection volume: 1 μL (split (1/20, 250 °C); injection solvent: EtOAc; carrier: H_2 (0.6 mL/min). Analyses were performed with the following program: (i) gradient from 50 °C to 230 °C (20 °C/min), (ii) isocratic at 230 °C for 3 min. Data were processed with Shimadzu's LabSolutions software. Product concentrations were determined using conversion factors derived from standard curves of authentic samples. The initial rates used for kinetic analysis were measured at eight substrate concentrations ranging from 0.1 to 10 mM. Reaction rates were plotted against substrate concentration, and the kinetic parameters V_{max} and K_m were obtained by non-linear regression of the Michaelis-Menten equation. Errors represent the 95% confidence intervals derived from the variance of the regression fit.

3.4 Homology modeling of CYP153A6 protein structure

The 3D structure of CYP153A6 was constructed using the Homology Modeling tool within the Multiple Sequence Viewer/Editor in the Schrödinger Maestro Suite (release 2024-2) [19]. The template was selected based on the highest sequence identity to CYP153A6, the greatest percentage of BLAST positives, and an e-value close to zero. To guide accurate modeling of the holo form, the ligand bound in the template's active site was retained. After model building, the structure was refined using the Protein Preparation Wizard. Especially, the bond orders were assigned according to the Cambridge Crystallographic Database (CCD) rules for both the protein and the ligand. For the heme cofactor, the four bonds to the central Fe atom were set to zero-order bonds. In addition, a formal charge of -1 was assigned to the two nitrogen atoms in the porphyrin ring that are bonded through only two single bonds. Model quality was evaluated using a Ramachandran plot. Then hydrogen atoms were added according to the primary sequence and refined based on

the predicted ionization states at the experimental pH (7.4). The final homology model was used in the subsequent molecular docking and MD simulations to study the enzyme-substrate recognition.

3.5 Molecular tunnel analysis

CYP153A6 molecular tunnels were identified using CAVER 3.02 software [25], analyzed with CAVER Analyst 2.0 [36], and visualized using PyMOL [24]. The heme iron atom was set as the starting point for tunnel calculations. Detection parameters included a probe radius of 0.9 Å, shell radius of 3.0 Å, shell depth of 4.0 Å, and a clustering threshold of 3.5 Å. Tunnel analysis was performed on frames from a MD simulation on the CAVER software, which considered all protein residues, the heme cofactor, and the bound ligand. Simulations were run with YASARA parameters [37] (density: 0.997 g/cm³, pH: 7.4, temperature: 298.15 K, duration: 10,000 ps, frame interval: 10 ps), generating 1000 frames. The top five tunnel clusters were extracted and analyzed for bottleneck radii and the throughput score.

3.6 *In silico* solvent analysis

To validate the predicted water tunnel in CYP153A6 and generate water molecules for subsequent ensemble docking studies, an *in silico* solvent analysis was performed on the CYP153A6 homology model. The model was first processed using the QuickPrep module in Molecular Operating Environment (MOE 2022.02) [26] for protein structure preparation, energy minimization, and protonation. Subsequently, the 'Solvent Analysis' tool in MOE was used to predict water molecule positions within the enzyme structure. Both protein preparation and solvent analysis were conducted using the AMBER10:EHT force field. A cutoff distance of 10 Å from the retained ligand in the homology model was applied to verify the predicted water tunnel. While a separate 10 Å cutoff from the binding pocket was used to investigate how water molecules influence the enzyme's binding site and to generate water molecules for subsequent ensemble docking studies.

3.7 Ensemble docking study for selecting the best docking method

An ensemble docking strategy (Figure S1) was utilized to investigate CYP153A6 enzyme ligand recognition under four conditions:

- (1) Cofactor heme iron as Fe³⁺ (ferric), without water molecules in the binding site;

- (2) Cofactor heme iron as Fe^{3+} (ferric), with water molecules in the binding site;
- (3) Cofactor heme iron as Fe^{2+} (ferrous), without water molecules in the binding site;
- (4) Cofactor heme iron as Fe^{2+} (ferrous), with water molecules in the binding site.

Protein structural ensembles were generated from 500 ns classical MD simulations, performed in three replicas using Desmond simulation engine in Maestro with the OPLS4 force field. Systems were built with the System Setup panel using an orthorhombic box (SPC solvent model, 15 Å buffer box size), and neutralized with Na^+ ions. Trajectory analyses included RMSD and RMSF analyses. Clustering of the MD trajectory (every 10th frame) was performed with Schrödinger's trajectory_cluster.py (2024-2). The top ten cluster medoids were selected as representative structures for ensemble docking. Protein receptor grid generation was performed using the Grid Generation panel. The Van der Waals radius scaling factor was set to 1.0, with a partial charge cutoff of 0.25. The grid box was centered on the centroid of workspace ligand (the ligand of the holo-CYP153A6 homology model) and defined with a size of 20 Å.

Ligands were prepared with LigPrep panel using OPLS4 force field, generating all relevant ionization states at $\text{pH } 7.4 \pm 2.0$ and retaining specified chirality. Docking was performed using Glide in SP (standard precision) mode with flexible ligand sampling. Epik state penalties were included in the scoring, and the post-docking minimization was applied for all the generated docking poses. All the docking poses were performed post-docking minimization. Finally, the docking poses were ranked by docking score.

3.8 Molecular docking and interaction fingerprints analysis of toluene derivatives

The best docking condition (cofactor heme as Fe^{3+} without water molecules in the binding site) identified in section 3.7 was used to study the substrate recognition of toluene derivatives by CYP153A6. Docking was performed on the best target receptor conformation; protein preparation was the same as described in section 3.4. The docking setup followed the same procedures outlined in section 3.7. Protein-ligand interaction fingerprints were generated using Maestro's Interaction Fingerprints panel. The "Docked poses" option was selected, and all interaction types were included in the analysis. Residues interacting with the ligand were account to create a

heatmap for visualization, the heatmap image was made by Python with the Seaborn and Matplotlib libraries.

3.9 Molecular dynamics studies on toluene derivatives

MD simulations of CYP153A6 complexes with various toluene derivatives were conducted for 200 ns in three replicas, following the same simulation protocol described in section 3.1. Ligand RMSDs were analyzed using Schrödinger's `trj_rmsd.py` script (version 2024-2), the distances between the ligand reactive methyl group to the enzyme cofactor heme iron during the simulations were monitored using Schrödinger `trajectory_asl_monitor.py` script (version 2024-2), ligand energy profiles were obtained using the Schrödinger `thermal_mmgsa.py` script (version 2024-2) on every 10th simulation frame.

The mathematical formula of MM-GBSA calculating a ligand binding free energy (ΔG_{bind}) to a protein are as [38]:

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S \approx \Delta E_{\text{MM}} + \Delta G_{\text{sol}} - T\Delta S$$

Where, ΔE_{MM} , ΔG_{sol} and $-T\Delta S$ are the changes in gas-phase molecular mechanics energy, solvation free energy, and conformational entropy upon binding, respectively. The ΔE_{MM} terms consists of:

- $\Delta E_{\text{internal}}$: bond, angle, and dihedral contributions
- $\Delta E_{\text{electrostatic}}$: electrostatic interactions
- ΔE_{vdW} : van der Waals interactions

The solvation free energy, ΔG_{sol} , is the sum of:

- The electrostatic solvation energy (polar contribution), ΔG_{GB}
- The non-electrostatic solvation component (nonpolar contribution), ΔG_{SA}

3.10 QM/MM calculations setup

QM/MM calculations were performed using the Schrödinger Suite (2025-1) [35]. The QM region was defined as the ligand, the heme, the oxygen atom and the side chain of Cys367 residue that binds to the heme (with a hydrogen cap added between the C α and C β atoms). The QM region was treated with Jaguar at the DFT level using the B3LYP-D3 functional, while the MM region was modeled with the OPLS4 force field, including the electrostatic field generated by MM atomic charges. For both geometry optimizations and single-point energy calculations, the LACVP+ basis set was

applied to the heme iron atom, and the 6-31G** basis set was used for all other atoms.

4. Conclusions

This study combined experimental biocatalysis and multiscale computational modeling to elucidate the molecular recognition and catalytic behavior of CYP153A6 toward toluene derivatives.

Whole-cell and cell-free assays confirmed that CYP153A6 selectively hydroxylates the benzylic methyl group of several substrates, with p-methylanisole displaying the highest turnover frequency (52.3 min^{-1}). Substrates bearing hydrophobic para-substituents, such as p-cymene, p-xylene, and p-chlorotoluene, were effectively converted, whereas derivatives with strongly polar or electron-withdrawing groups, such as p-nitrotoluene and m-cresol, showed negligible activity.

Computational analyses provided a mechanistic explanation for this selectivity. MD simulations and binding free-energy calculations revealed a predominantly hydrophobic active site, stabilized by residues including ALA97, ILE101, ILE103, LEU252, VAL255, LEU303, MET306, and PHE407. Hydrophobic substituents align productively near the heme center, while polar groups disrupt favorable orientations. QM/MM calculations supported the different reactivity observed between p-chlorotoluene and p-nitrotoluene.

Together, these results highlight that hydrophobic complementarity and electronic compatibility with Compound I are the primary determinants of CYP153A6 substrate specificity. This mechanistic insight establishes a rational basis for protein engineering to broaden substrate scope and enhance selective benzylic hydroxylation.

Funding Declaration

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101073546 (MSCA Doctoral Network Metal-containing Radical Enzymes - MetRaZymes), and grants from MUR - "Progetto Eccellenza 2023 - 2027".

Data Availability

Materials are archived in the University of Milan Dataverse (institutional FAIR repository) if applicable at the following link: https://doi.org/10.13130/RD_UNIMI/9GCFAM

Author Contributions

Y.W. designed and performed all molecular modeling research and wrote the bioinformatics section of the manuscript. S.D. and S.F. designed and conducted all enzyme expression and biotransformation experiments and wrote the experimental section. U.G. was responsible for managing the molecular modeling software. F.M. and I.E. supervised the study and were responsible for funding acquisition. All authors participated in the review and revision of the manuscript.

Reference:

- [1] Simonian NG, Brutiu BR, Kaiser D, Maulide N. Cationic, Iodine (III)-Mediated and Directed Diastereoselective Oxidation of Inert C-H Bonds in Cyclic Hydrocarbons. *Angewandte Chemie International Edition*.:e202421872.
- [2] Docherty JH, Lister TM, McArthur G, Findlay MT, Domingo-Legarda P, Kenyon J, Choudhary S, Larrosa I. Transition-metal-catalyzed C-H bond activation for the formation of C-C bonds in complex molecules. *Chemical Reviews*. 2023 May 10;123(12):7692-760.
- [3] Alaidaroos BA. Advancing eco-sustainable bioremediation for hydrocarbon contaminants: challenges and solutions. *Processes*. 2023 Oct 22;11(10):3036.
- [4] Tanwar, L., Börgel, J., & Ritter, T. (2019). Synthesis of benzylic alcohols by C-H oxidation. *Journal of the American Chemical Society*, 141(45), 17983-17988.
- [5] Munch, J., Pullmann, P., Zhang, W., & Weissenborn, M. J. (2021). Enzymatic hydroxylations of sp³-carbons. *Acs Catalysis*, 11(15), 9168-9203.
- [6] HYDROXYLATED, P. O. A. (2011). 4-hydroxyphenylpyruvate dioxygenase catalysis. *The Journal of Biological Chemistry*, 286(29), 26061-26070.
- [7] Sheet, D., & Paine, T. K. (2016). Aerobic alcohol oxidation and oxygen atom transfer reactions catalyzed by a nonheme iron (II)- α -keto acid complex. *Chemical Science*, 7(8), 5322-5331 .
- [8] Martínez-Montero, L., Tischler, D., Süß, P., Schallmeyer, A., Franssen, M. C., Hollmann, F., & Paul, C. E. (2021). Asymmetric azidohydroxylation of styrene derivatives mediated by a

biomimetic styrene monooxygenase enzymatic cascade. *Catalysis Science & Technology*, 11(15), 5077-5085.

- [9] Peters MW, Meinhold P, Glieder A, Arnold FH. Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. *Journal of the American Chemical Society*. 2003 Nov 5;125(44):13442-50.
- [10] Fiorentini F, Hatzl AM, Schmidt S, Savino S, Glieder A, Mattevi A. The extreme structural plasticity in the CYP153 subfamily of P450s directs development of designer hydroxylases. *Biochemistry*. 2018 Nov 6;57(48):6701-14.
- [11] Scheps D, Malca SH, Hoffmann H, Nestl BM, Hauer B. Regioselective ω -hydroxylation of medium-chain n-alkanes and primary alcohols by CYP153 enzymes from *Mycobacterium marinum* and *Polaromonas* sp. strain JS666. *Organic & biomolecular chemistry*. 2011;9(19):6727-33.
- [12] Funhoff, E. G., Bauer, U., García-Rubio, I., Witholt, B., & van Beilen, J. B. (2006). CYP153A6, a soluble P450 oxygenase catalyzing terminal-alkane hydroxylation. *Journal of bacteriology*, 188(14), 5220-5227.
- [13] Cannazza P, Rabuffetti M, Donzella S, De Vitis V, Contente ML, de Oliveira MD, de Mattos MC, Barbosa FG, de Souza Oliveira RP, Pinto A, Molinari F. Whole cells of recombinant CYP153A6-*E. coli* as biocatalyst for regioselective hydroxylation of monoterpenes. *AMB Express*. 2022 Apr 27;12(1):48.
- [14] Ciriminna R, Pandarus V, Béland F, Xu YJ, Pagliaro M. Heterogeneously catalyzed alcohol oxidation for the fine chemical industry. *Organic Process Research & Development*. 2015 Nov 20;19(11):1554-8.
- [15] Neufeld K, Marienhagen J, Schwaneberg U, Pietruszka J. Benzylic hydroxylation of aromatic compounds by P450 BM3. *Green Chemistry*. 2013;15(9):2408-21
- [16] Neufeld K, Marienhagen J, Schwaneberg U, Pietruszka J. Benzylic hydroxylation of aromatic compounds by P450 BM3. *Green Chemistry*. 2013;15(9):2408-21.
- [17] Rapp LR, Marques SM, Zukic E, Rowlinson B, Sharma M, Grogan G, Damborsky J, Hauer B. Substrate anchoring and flexibility reduction in CYP153A M. aq leads to highly improved efficiency toward octanoic acid. *ACS Catalysis*. 2021 Feb 24;11(5):3182-9.
- [18] Park SW, Lee BH, Song SH, Kim MK. Revisiting the Ramachandran plot based on statistical analysis of static and dynamic characteristics of protein structures. *Journal of Structural Biology*. 2023 Mar 1;215(1):107939.
- [19] Schrödinger release 2024. New York: Maestro LLC; 2024

- [20] Urban P, Truan G, Pompon D. Access channels to the buried active site control substrate specificity in CYP1A P450 enzymes. *Biochimica et Biophysica Acta (BBA)-General Subjects*. 2015 Apr 1;1850(4):696-707.
- [21] Kingsley LJ, Lill MA. Substrate tunnels in enzymes: structure-function relationships and computational methodology. *Proteins: Structure, Function, and Bioinformatics*. 2015 Apr;83(4):599-611.
- [22] Šrejber M, Navrátilová V, Paloncýová M, Bazgier V, Berka K, Anzenbacher P, Otyepka M. Membrane-attached mammalian cytochromes P450: An overview of the membrane's effects on structure, drug binding, and interactions with redox partners. *Journal of inorganic biochemistry*. 2018 Jun 1;183:117-36.
- [23] Beglov D, Roux B. An integral equation to describe the solvation of polar molecules in liquid water. *The journal of physical chemistry B*. 1997 Sep 25;101(39):7821-6.
- [24] DeLano WL. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr*. 2002 Mar;40(1):82-92.
- [25] Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, Gora A, Sustr V, Klvana M, Medek P, Biedermannova L. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures.
- [26] Molecular Operating Environment (MOE); Chemical Computing Group Inc. 1010 Sherbooke St. West, Suite# 910: Montreal, QC, Canada, (2022).
- [27] Denisov IG, Makris TM, Sligar SG, Schlichting I. Structure and chemistry of cytochrome P450. *Chemical reviews*. 2005 Apr 21;105(6):2253-78.
- [28] Olsen L, Oostenbrink C, Jørgensen FS. Prediction of cytochrome P450 mediated metabolism. *Advanced drug delivery reviews*. 2015 Jun 23;86:61-71.
- [29] Dueholm B, Krieger C, Drew D, Olry A, Kamo T, Taboureau O, Weitzel C, Bourgaud F, Hehn A, Simonsen HT. Evolution of substrate recognition sites (SRSs) in cytochromes P450 from Apiaceae exemplified by the CYP71A subfamily. *BMC evolutionary biology*. 2015 Dec;15:1-4.
- [30] Amaro RE, Baudry J, Chodera J, Demir Ö, McCammon JA, Miao Y, Smith JC. Ensemble docking in drug discovery. *Biophysical journal*. 2018 May 22;114(10):2271-8.
- [31] Hritz J, de Ruiter A, Oostenbrink C. Impact of plasticity and flexibility on docking results for cytochrome P450 2D6: a combined approach of molecular dynamics and ligand docking. *Journal of medicinal chemistry*. 2008 Dec 11;51(23):7469-77.

- [32] Dey, B., Mehta, S., Mondal, A., Cirera, J., Colacio, E., & Chandrasekhar, V. (2022). Push and Pull Effect of Methoxy and Nitro Groups Modifies the Spin-State Switching Temperature in Fe (III) Complexes. *ACS omega*, 7(43), 39268-39279.
- [33] Lonsdale R, Harvey JN, Mulholland AJ. Compound I reactivity defines alkene oxidation selectivity in cytochrome P450cam. *The Journal of Physical Chemistry B*. 2010 Jan 21;114(2):1156-62.
- [34] Don CG, Smieško M. Deciphering reaction determinants of altered-activity CYP2D6 variants by well-tempered metadynamics simulation and QM/MM calculations. *Journal of Chemical Information and Modeling*. 2020 Dec 3;60(12):6642-53.
- [35] Schrödinger release 2025. New York: Maestro LLC; 2025
- [36] Jurcik A, Bednar D, Byska J, Marques SM, Furmanova K, Daniel L, Kokkonen P, Brezovsky J, Strnad O, Stourac J, Pavelka A. CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics*. 2018 Oct 15;34(20):3586-8.
- [37] Land H, Humble MS. YASARA: a tool to obtain structural guidance in biocatalytic investigations. *Protein engineering: methods and protocols*. 2018:43-67.
- [38] Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of chemical information and modeling*. 2011 Jan 24;51(1):69-82.