

# **An atomic picture of ligand migration in toluene 4-monooxygenase**

**Ali Hosseini<sup>1</sup>, Moran Brouk<sup>2</sup>, Maria Fatima Lucas<sup>1</sup>, Fabian Glaser<sup>3</sup>, Ayelet Fishman<sup>\*2</sup>, and Victor Guallar<sup>\*1,4</sup>**

<sup>1</sup>Department of Life Sciences, Barcelona Supercomputing Center, Nexus II Building, Barcelona, Spain

<sup>2</sup>Department of Biotechnology and Food Engineering, Technion-Israel Institute of Technology, Haifa, 32000, Israel

<sup>3</sup>Bioinformatics Knowledge Unit, Lorry I. Lokey Interdisciplinary Center for Life Sciences and Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel

<sup>4</sup>Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys 23, E-08010 Barcelona, Spain

\* To whom correspondence should be addressed. E-mail: [victor.guallar@bsc.es](mailto:victor.guallar@bsc.es), [afishman@tx.technion.ac.il](mailto:afishman@tx.technion.ac.il)

## **Abstract**

Computational modeling combined with mutational and activity assays were used to underline the substrate migration pathways in toluene 4-monooxygenase, a member of the important family of bacterial multicomponent monooxygenases (BMM). In all structurally defined BMM hydroxylases, several hydrophobic cavities in the  $\alpha$ -subunit map a preserved path from the protein surface to the diiron active site. Our results confirm the presence of two pathways by which different aromatic molecules can enter/escape the active site. While the substrate is observed to enter from both channels, the more hydrophilic product is withdrawn mainly from the shorter channel ending at residues D285 and E214. The long channel ends in the vicinity of S395, whose variants have been seen to affect activity and specificity. These mutational effects are clearly reproduced and rationalized by the *in silico* studies. Furthermore, the combined computational and experimental results highlight the importance of residue F269, which is located at the intersection of the two channels.

**Keywords;** Computational modeling, Activity assays, Aromatic molecules, Substrate, *In silico* studies.

## Introduction

Bacterial multicomponent monooxygenases (BMMs)<sup>1</sup> are a family of proteins containing a non-heme carboxylate-bridged diiron center capable of activating molecular oxygen (O<sub>2</sub>) for the oxidation of several hydrocarbon substrates. These proteins present a common architecture with 3 or 4 components including a multi-subunit dimeric hydroxylase component, where the diiron center and catalytic active site are contained.<sup>1,2</sup> Toluene 4-monooxygenase (T4MO) is a soluble four-component BMM that oxidizes toluene with ~95% regioselectivity at the *para* position.<sup>3,4</sup> The reaction occurs in the toluene 4-monooxygenase hydroxylase which contains alpha, beta and gamma subunits<sup>5</sup>. T4MO is of particular interest in industry given the high number of substrates that can be oxidized along with the elevated specificity.<sup>4</sup>

Different studies of these systems, and related diiron active site enzymes, show that mutagenesis in the active site can change the regioselectivity for reactions with toluene and other non-physiological substrates.<sup>6-8</sup> Moreover, using directed evolution, the T4MO variant S395C showed different activity and specificity.<sup>4</sup> The production of hydroxytyrosol, a phenol with high antioxidant and anticarcinogenic activities obtained via double hydroxylation of 2-phenylethanol (PEA), showed a 15-fold improvement in the mutated protein in comparison to the wild type specie. In contrast to other cases, this residue is located 30Å away from the active site, near the interface of subunits  $\alpha$  and  $\gamma$ , suggesting that its catalytic influence should be the result of a change in the active site dynamics or in the ligand delivery. Several crystallographic studies of the BMM superfamily, proposed a common channel through the  $\alpha$  subunit connecting the diiron center to the surface; this pathway involves surface residues D285 and E214 in T4MO<sup>9,10</sup>. Additional reports

described two other hydrophobic cavities, one near the active site pocket and another which is located near the interface of the  $\alpha$  and  $\gamma$  subunits<sup>1,5,10</sup>. Measurements of oxygen migration by the formation rates of a peroxodiiron(III) intermediate, suggested that the two cavities may connect via movement of the protein components thus regulating oxidation rate<sup>10</sup>. This was further supported by the findings describing changes in the volume of the pockets upon binding of the effector protein to the hydroxylase, thus restricting free access from the solvent to the active site<sup>1,5</sup>.

In order to understand the atomic detail of ligand migration in T4MO as well as the involvement of S395 in the catalytic rate enhancement, we have modeled the ligand migration pathways with all-atom computational techniques. Recently, due to its implications in drug design and enzyme engineering there has been an increasing interest in applying molecular dynamics techniques for mapping ligand diffusion, entrance and exit.<sup>11,12</sup> To this aim, many biased approaches, such as steered molecular dynamics<sup>13</sup> or metadynamics<sup>14</sup>, have been applied. Using special purpose machines or graphical processors units, a non-biased search accessing microsecond time scale simulations is also possible.<sup>15,16</sup> These computational approaches, however, still represent a significant computational cost (out of the reach from a typical lab) when dealing with complex systems such as T4MO. Our Monte Carlo approach PELE (Protein Energy Landscape Exploration), allows for an unbiased search of the protein-ligand dynamics, which includes mapping the exit and entrance pathways at an accessible computational cost.<sup>17-19</sup>

Along the simulations, two different substrates were evaluated: the reactant PEA and the first hydroxylation product, *p*-tyrosol (*p*-tyr), see Figure 1B. Our results confirm the presence of two pathways by which aromatic molecules can enter/escape the active site.

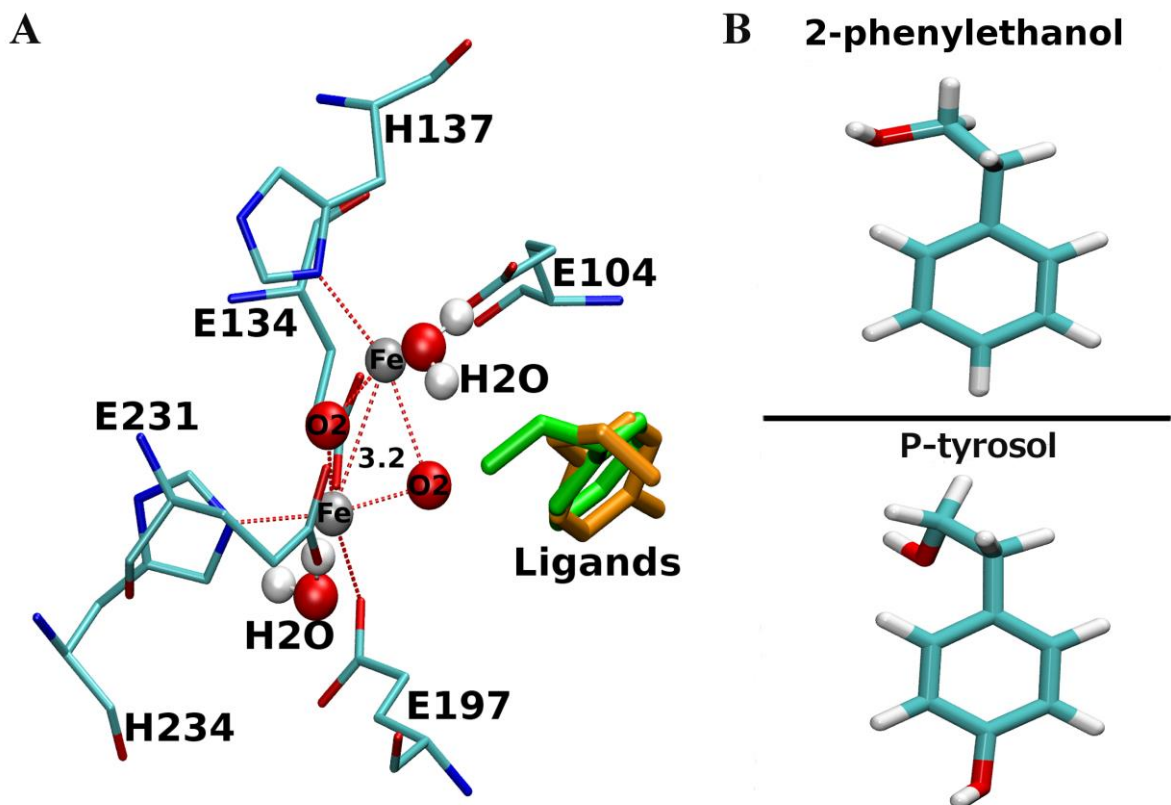
While the substrate enters from both channels, the more hydrophilic products are withdrawn mainly from the shorter channel ending at residues D285 and E214. Furthermore, computational studies were complemented with mutational and activity assays. The combined results highlight the importance of residue F269, which is located at the intersection of the two channels.

## Methods and Materials

### Computational work

The starting coordinates were taken from the Protein Data Bank (PDB) with entry 3DHG<sup>1</sup>. The protein preparation wizard algorithm<sup>20</sup> implemented in the Schrödinger software package was employed to prepare the initial structure. The algorithm predicts protonation states of histidines, aspartic acid and glutamic, and optimizes the hydrogen bond network. In this work, all charged residues (at pH 7) kept their initial state. Histidines 4, 36, 37, 137, 163, 174, 247, 329, 332 and 484 were epsilon protonated and 328, 421 and 447 were double protonated; all other histidine residues kept their default delta protonation state. Two water molecules, generally found in the active site of T4MO, one in the first coordination sphere of an iron atom and the other in the active site pocket, were kept in the system.

All distances in the iron coordination sphere were constrained to their initial crystal value. Likewise, the Fe–Fe distance was maintained at 3.2 Å. As seen in Figure 1A, reactant and product molecules were placed in a similar position in the active site. This assumption is based on the high degree of complementarity of the T4MO active site that shows that toluene and other aromatic ring substrates bind in a similar fashion.<sup>21,22</sup>



**Figure 1.** Panel A: T4MO active site view. Panel B: the two substrates (reactant and product) used in this work.

Ligand diffusion was studied with our in house program PELE.<sup>18,23</sup> This technique is based on a Monte Carlo algorithm, where new trial configurations are produced with sequential ligand and protein perturbation, side chain prediction and minimization steps. Trial configurations are then filtered with a Metropolis acceptance test, where the energy is described with an all-atom OPLS force field<sup>24</sup> with a surface generalized Born solvent model.<sup>25</sup> PELE has recently shown to provide more accurate induced fit results than state of the art commercial software,<sup>23</sup> and to reproduce the conformational sampling obtained in microsecond molecular dynamics trajectories with a two order reduction in computational cost.<sup>19</sup>

Two different sets of simulations, describing the ligand exit and entrance were performed for the wild type and the S395C, F269V and F269W mutants.

**Ligand exit**, a total of 10 independent runs were produced for each simulation, where the ligand initially placed in the active site explores randomly possible exit paths. Each run is based on 12 trajectories (running on different cores) that search collectively for an exit route. The collective search uses the distance from the ligand center of mass to the iron metal (iron 501 in the PDB file) to couple the different trajectory search. In particular, if a processor is 3 Å behind the leading processor (the one with the furthest distance to the iron), it will abandon its current coordinates and receive the leading ones. Thus, there is no bias in the direction that the ligand should take. All 10 runs were interrupted after 24h of CPU or when (if) the ligand reached a solvent accessible surface area larger than 0.4, i.e. the ligand enters the solvent.

**Ligand entrance**, studies were also done for the migration of the substrate from the surface of the protein to the active site for both channels identified in the exit exploration. In this case the ligands were placed at the solvent and in the vicinity of the exit point as described in the exits paths. Again, we performed 10 independent runs with 12 trajectories in each run, where the trajectories search collectively to move towards the active site (with a 3 Å distance to the iron criteria).

Fpocket,<sup>26,27</sup> was employed to measure the cavities size and their hydrophobic nature. This method uses a Monte Carlo approach to find a path where a sphere of variable radius can pass. Moreover, Fpocket also allows pocket detection on a large set of structures forming a trajectory.

## **Experimental work**

### Chemicals



2-Phenylethanol (PEA), *o*-, *m*-, *p*-tyrosol, were purchased from Sigma-Aldrich Chemical Co. (Sigma-Aldrich, Rehovot, Israel). Hydroxytyrosol was obtained from Cayman Chemical Co. (MI, USA). All standards were prepared as stock solutions in ethanol. All materials used were of the highest purity available and were used without further purification.

#### Bacterial strains and growth conditions

*Escherichia coli* TG1 (*supE hsdΔ5 thi Δ(lac-proAB) F' [traD36 proAB<sup>+</sup> lacI<sup>q</sup> lacZΔM15]*) with the plasmid constructs was routinely cultivated at 37°C in Luria-Bertani (LB) medium<sup>28</sup> supplemented with kanamycin at 100 μg/ml to maintain the plasmids. To stably and constitutively express the toluene monooxygenase genes from the same promoter, the expression vector pBS(Kan)T4MO (henceforth T4MO), was constructed as described earlier<sup>4</sup>. All experiments were conducted as described previously.<sup>4,29</sup> Shortly, overnight cells were diluted to an optical density (OD) at 600 nm of 0.1 and grown to an OD of 1.3. The exponentially-grown cells were centrifuged (8,000 g for 10 min at 25°C) and re-suspended in potassium phosphate buffer (100 mM, pH 7.0). Cells were subsequently used in biotransformation protocols as described later on.

#### Construction of T4MO mutants

Variant S395C was obtained through random mutagenesis as described by Brouk and Fishman<sup>4</sup>. Variants F269V and F269W were constructed using site-directed mutagenesis at the T4MO *tmoA* gene via overlap extension PCR<sup>29</sup>. Briefly, three oligonucleotide primer pairs were designed (Supplementary Table 1) to generate the desired mutations. For generating the F269V and F269W mutations, the first mutated PCR fragment was amplified using primers T4MO-befEcoRI-front and T4MO\_269V\_rear, T4MO-befEcoRI-

front and T4MO\_269W\_rear, respectively. The second mutated PCR fragment was amplified using T4MO\_269V\_front and T4MO-ABRear, and T4MO\_269W\_front and T4MO-ABRear, respectively. The PCR program consisted of an initial denaturation at 94°C for 2 min, followed by 25 cycles of 94°C for 45 s, 55°C for 45 s, and 72°C for 2.2 min, with a final extension at 72°C for 8 min. The two fragments were combined during the final reassembly PCR in a 1:1 molar ratio using the outer primers T4MObefEcoRI Front and T4MOABRear. The assembling PCR was programmed similarly to the above PCR program, with extension at 72°C for 3.15 min instead of 2.2 min. The assembled PCR fragment was ligated into WT T4MO, after the double digestion of both vector and insert with EcoRI and AatII, replacing the corresponding fragment in the original plasmid. The resulting plasmid library was electroporated into *E. coli* TG1 cells. Plasmid DNA was isolated using a Mini Kit (Qiagen, CA, USA), and verification of the mutations was done by sequencing with primer T4MO seq 1.

#### Whole-cell enzymatic biotransformations

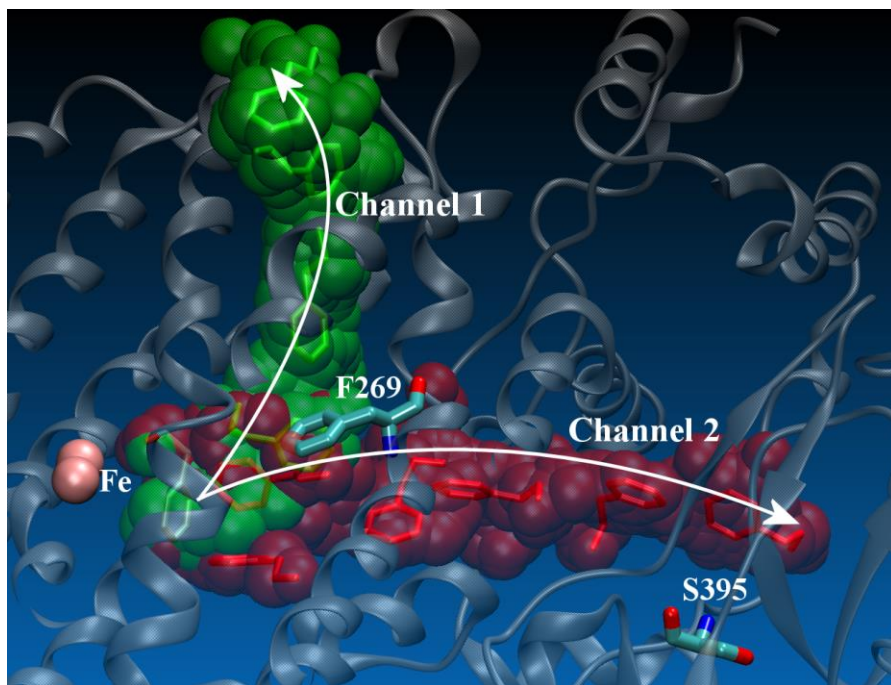
Whole-cell activity assays were performed as described previously.<sup>4,29</sup> The biotransformation was carried out in screw-capped 16-ml glass vials containing 2 ml cells and 0.25 mM substrate (added from a 100 mM stock solution in ethanol). All the vials were shaken at 600 rpm (Vibramax 100, Heidolph, Nuremberg, Germany) at 30°C. The reaction was stopped periodically (a vial was sacrificed) by filtration of the cells and analysis by HPLC. The negative control used in these experiments was TG1/pBS(Kan) (a plasmid without the monooxygenase). The initial transformation rates were determined by sampling at 3-20 min intervals during the first 2-5 hours. The specific activity (nmol/min/mg protein) was calculated as the ratio of the initial transformation rate and the total protein content,

0.24 [mg protein/ml/OD<sub>600nm</sub>].<sup>10,30,31</sup> Activity data reported in this paper are based on at least three independent results. Analytical methods were described previously.<sup>4,29</sup>

## Results

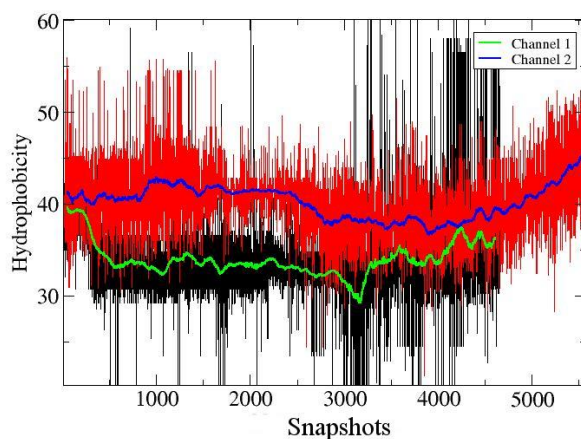
### Ligands migration in the wild type

We started our computational study by searching possible migration pathways connecting the active site and the protein surface. For this purpose, we placed both the reactant (PEA) and the first hydroxylation product (p-tyr) at the active site cavity and modeled their exit using PELE. As seen in Figure 2, we found two different channels crossing T4MO's  $\alpha$ -subunit and adopting a nearly 90° disposition. One channel (hereafter known as channel 1) is a traverse of about 23 Å to the protein's surface providing the shortest path between the active site and solvent. This channel is formed mainly by: H96, F269, Q204, D211, E214, and D285. The second route, channel 2, is significantly longer, 35 Å, and presents a larger hydrophobic nature: F269, P390, V335, W167, W338, and S395. As seen in Figure 2, both channels share the active site pathway section and bifurcate around F269.



**Figure 2.** The two observed migration paths obtained with the PELE simulations. Channel 1 is shown in green while channel 2 is shown in red.

From the 10 simulations performed for the reactants, we observe a similar number of migrations in either pathway (40% by channel 1 and 60% by channel 2). In the case of the products, however, we see that 90% of the runs leave by channel 1 while only 10% follow channel 2 (see Table 1). As mentioned above, visual inspection of both channels indicate the presence of more polar residues in channel 1. This might be the main reason why the products prefer this exit pathway; the product presents an extra hydroxyl group in the para position. To quantify better the polar nature of each pathway we used Fpocket,<sup>26,27</sup> which allows us to compute the hydrophobicity of both channels. As seen in Figure 3, channel 1 is significantly less hydrophobic than channel 2.



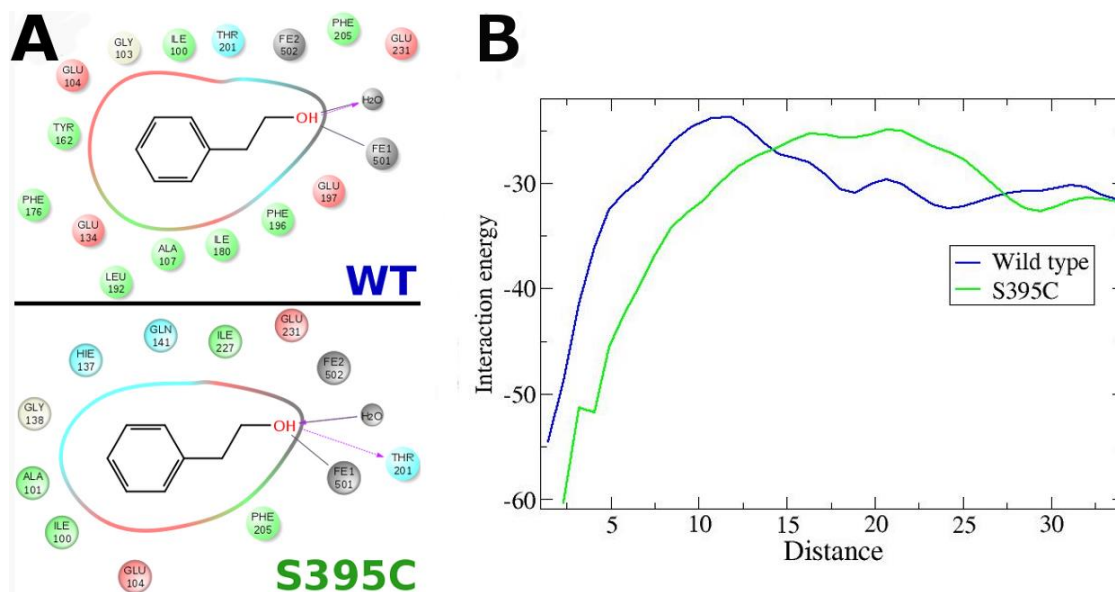
**Figure 3.** Comparison of channel 1 and channel 2 hydrophobicities for the different snapshots produced by PELE along the exit pathways.

### Role of S395

Figure 2 indicates the location of S395, in close vicinity to the solvent exit point in channel 2. As mentioned previously, variant S395C, obtained by directed evolution, shows a 15-fold improvement in PEA oxidation in comparison to the wild type specie.<sup>4</sup> Performing

saturation mutagenesis at position 395 revealed three additional variants with improved activity on PEA. The product distribution obtained from PEA oxidation displayed a decrease in the regioselectivity by the variants and ability to form *o*-tyrosol which wild type cannot (results not shown). Thus, residue S395, which is distant from the active site, has strong influence on both activity and selectivity of T4MO.

In order to understand the possible influence of the S395C mutation on ligand delivery, we computed the average protein-substrate interaction energy along channel 2 for the wild type and the mutant. Figure 4 shows both interaction energy plots, presenting clear binding profiles for both systems, with lower energies at the active site distance ( $\sim 5\text{\AA}$ ). The main differences between the wild type and the mutant, however, are observed precisely at the active site and at the 15-25  $\text{\AA}$  distance segment from the active site, where residues W167 and W338 are located. These two residues play an important role as gatekeepers in ligand egression through channel 2. Replacement of the serine by a cysteine reduces the interaction of the ligand with the protein (less favorable interaction) at this 15-25 $\text{\AA}$  segment, facilitating in this way its passage through the channel. At the active site, however, we observe the opposite effect, an increase in interaction energy for the mutated species. Interestingly, the minimum at the active site is shifted in the mutant, pointing to an increase in volume. Panel A in Figure 4 compares the main contacts of the protein with the substrate at the active site, where we observe a slight increase in the number of interactions in the mutant.

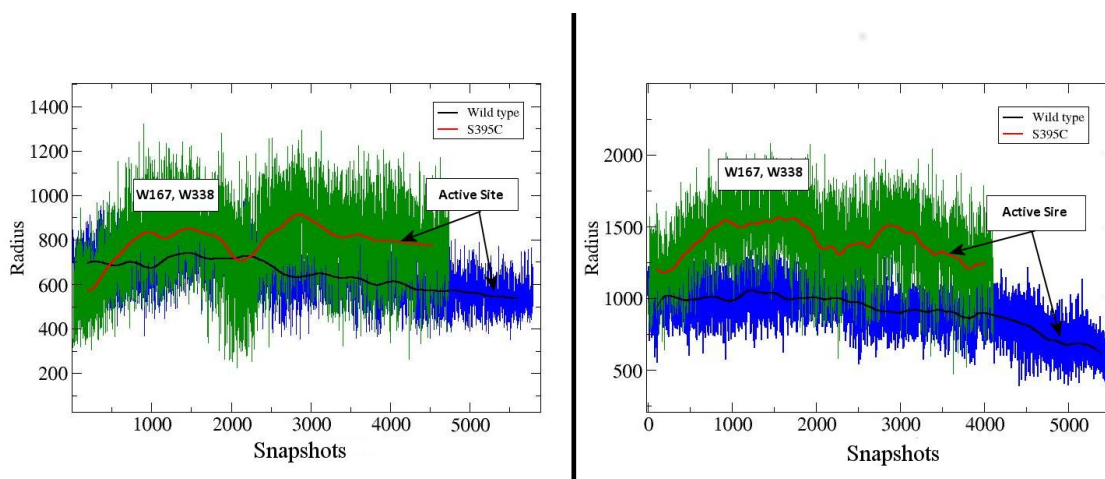


**Figure 4.** Panel A: different contacts observed for the reactant in the active site of WT and variant S395C. Panel B: interaction energies versus distance for the substrate entrance along channel 2. The distances, in Å, correspond to the distance of the ligand to its active site position. Energies are in Kcal/mol.

PELE uses a random walk approach, producing hundreds of new conformations that are accepted or rejected based on their energy change (following a Metropolis algorithm). Under the same conditions, the number of steps required to accomplish a task indicates qualitatively its “level of difficulty”. Thus, for example, by comparing the number of steps required by different ligands, or by introducing a mutation in the protein, we can obtain an approximation to the system dynamics. In the case of the wild type the substrate reached the active site in 700 steps while for S395C only 500 steps are needed (average numbers for the 10 runs). This translates in a significantly faster migration for S395C when compared to the wild type systems, in agreement with the lower protein-substrate interaction energy observed in Figure 4.

To complete the analyses we have inspected the differences in the cavity radius along the entrance in channel 2. Figure 5 shows the radius for two different runs, where we combined all 12 trajectories for each run. As mentioned above, the number of steps to reach the active

site is clearly lower in the mutant. Moreover, and in agreement with the interaction energy and the quicker substrate migration, the pathway for the mutant presents a larger cavity radius. The main variation occurs close to positions W167/W338 and at the entrance to the active site, the regions where we observed the larger interaction energy differences. Thus, the Fpocket analysis confirms an increase in volume for the mutant at the active site.



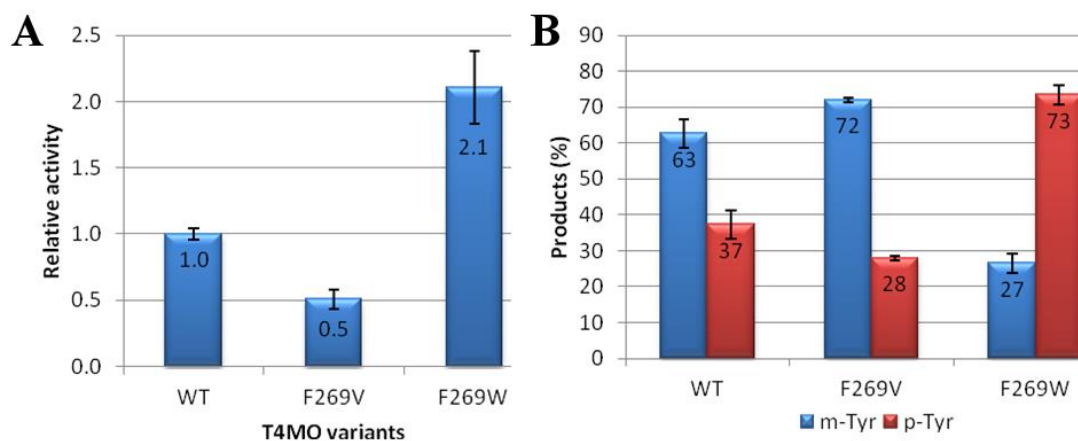
**Figure 5.** Cavity radius in two different runs for all snapshots along the channel 2 entrance as viewed for wild type and variant S395C. The presence of the ligand (along the snapshots) in the active site and next to W167/W338 is underlined with boxes

## Role of F269

Due to the strategic position of F269 at the bifurcation point observed in the modeled pathways, we performed the following *in vitro* and *in silico* mutations: F269V and F269W. We substituted position F269 to valine and tryptophan with the expectation that the smaller residue would allow better and faster discharge of the product, whereas the larger residue would slow down catalysis. In fact, the experimental results show quite the opposite (Figure 6A). The F269V mutant was very slow and quite inactive. The product distribution was similar to wild type (Figure 6B). However, the F269W variant was more active than WT (2.1-fold) and the regioselectivity changed. The enzyme favors *p*-tyrosol rather than *m*-



tyrosol. This latter fact is unique, since all the mutants we have generated until now (probably tens of different variants)<sup>4,29,32</sup> were always pro-*meta* except for position 395 which also influenced the regioselectivity. The change in selectivity is the result of different formation rates of the two products (Table 1).



**Figure 6.** Relative activity on PEA (A) of WT and T4MO F269 variants and the product distribution after 24 h (B). The activity was determined via HPLC analysis over a 5-h time period and regioselectivity was determined via GCMS analysis after 24 h, with initial PEA concentration of 0.25mM.

**Table 1.** The initial PEA oxidation rate and product formation rates of WT T4MO and the F269 variants

Rate (nmol/min/mg protein)	T4MO variant		
	WT	F269W	F269V
PEA oxidation	0.062 ± 0.002	0.131 ± 0.017	0.032 ± 0.004
<i>m</i> -tyr formation	0.141 ± 0.008	0.060 ± 0.0007	0.010 ± 0.001
<i>p</i> -tyr formation	0.120 ± 0.002	0.223 ± 0.003	ND

ND - The *p*-tyr formation rate by the F269V could not be calculated due to its low activity.

To further understand the influence of residue 269, we performed *in silico* analysis of PEA and *p*-tyr movement after placing them in the active site. As observed for the wild type protein, for both F269 mutants we find again channel 1 and channel 2 as the only possible migration pathways. Moreover, as observed previously, the products have a strong preference for channel 1. Interestingly, however, the preferred channel for PEA's migration changes with the mutation at residue 269. In particular, our calculations show that upon mutation of phenylalanine 269 to tryptophan, the main exit pathway shifts from an almost even ratio to exclusively channel 1. A summary of all results is depicted in Table 2.

**Table 2.** Distribution of ligand exit by the two channels for the reactant and product species. Values are in %.

	PEA		<i>p</i> -tyr	
	Channel 1	Channel 2	Channel 1	Channel 2
Wild type	40	60	90	10
F269V	50	50	70	30
F269W	100	0	90	10

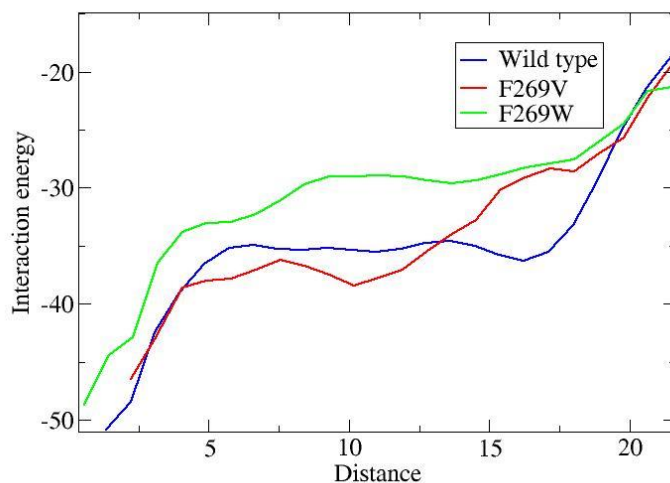
Besides shifting the ratio of migration pathways, we find that the mutation at position 269 strongly affects the dynamics of both ligands in channel 1. Table 3 lists the number of simulation steps needed to displace the ligand from the active site to the protein's surface for all species and ligands along channel 1. Clearly the product requires significantly less steps, which correlates with its larger polarity, capable of creating new contacts along the polar cavity. The most interesting result, however, is how mutation at F269 influences the ligand's diffusion. In the case of the wild type the reactant molecule finds the solvent in an

average of 175 steps while for F269V 210 steps are needed. Migration was particularly fast (~100 steps) for the F269W when compared to the other two systems.

**Table 3.** Average steps required for the migration of the ligands through channel 1.

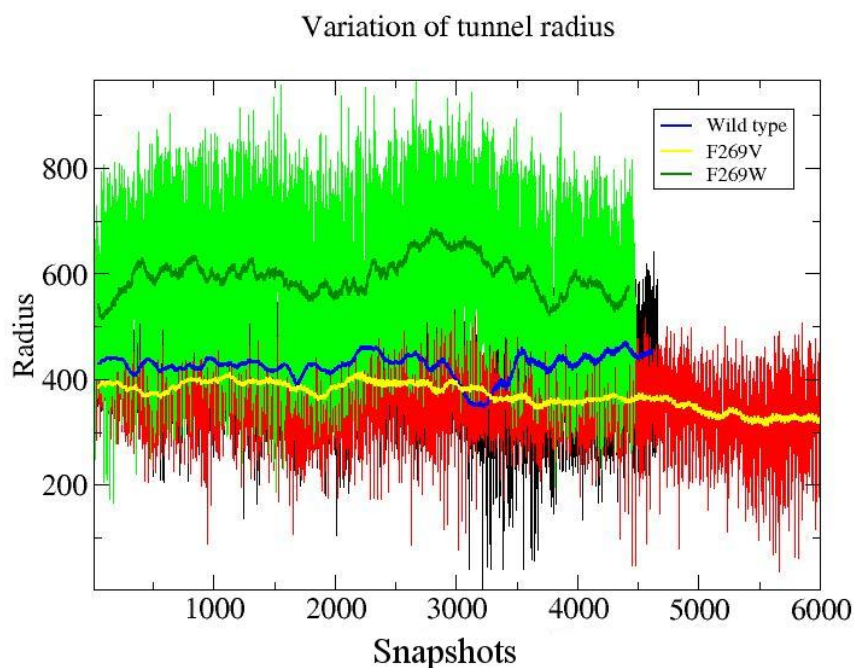
	PEA	p-tyr
Wild type	175	130
F269V	210	150
F269W	100	75

The difficulty in ligand migration shown by PELE agrees nicely with the experimental relative activities. To gain more insights about the substrate diffusion towards the active site, we modeled PEA's entrance along channel 1 for the wild type and the two mutants. Figure 7 shows the interaction energy along the entrance process, where we observe considerable differences among the different systems. Placing a tryptophan residue results in substantial lower interaction energy, facilitating the migration of the reactants from the solvent into the catalytic active site. For the valine mutant, however, we see a slight increase in the interaction energy, which explains the increase in simulation time required to complete the migration.



**Figure 7.** Interaction energies versus the distance for the substrate entrance through channel 1 for wild type and F269 variants. The distances, in Å, correspond to the distance of the ligand to its active site position. Energies are in Kcal/mol.

Finally, Figure 8 shows the changes in the radius for the entrance along channel 1 for one run (other runs give analogous results). While the volume of the tunnel is largely increased for the tryptophan mutant, changing the phenylalanine to a valine slightly decreases the radius along the migration cavity. Thus, contrary to our initial intuition, inserting a tryptophan at the “bifurcation” position 269 resulted in a substantial increase of the volume along channel 1, easier migration of the substrate to the active site and overall increase of the activity.



**Figure 8.** Cavity radius for the wild type and the F269 variants along the entrance simulations in channel 1.

## Discussion

Our dynamic simulation experiments confirm the presence of two important passages for ligand entry/exit to the active site of T4MO. One begins in the active site pocket and extends to the protein surface close to E214 and E285 (channel 1), which was described previously by Sazinsky et al<sup>9</sup> for the analogous ToMO. The second (channel 2) is a longer passage connecting the active site to the protein surface by passing W167 and exiting close to S395, and is in good agreement with recent experimental data.<sup>1,22</sup> While Song et al. suggested that this latter channel is the major pathway by which dioxygen enters the active site<sup>20</sup>, our experimental and computational work suggest that it is also a means for substrate entrance/exit. The differences in hydrophobicity of these two channels, explain the preference of the more hydrophilic product to exit by channel 1, the less hydrophobic among the two. PELE simulations also revealed the involvement of W167 and W338 in the

restriction of aromatic molecules within the enzyme. The presence of cysteine rather than serine at the entrance of channel 2 in the variant, caused an increase in the radius of the channel precisely around W167 and W337 (Figure 5) and this was corroborated with less favorable interactions with the substrate (Figure 4). Faster rates of reaction with PEA may be possible, as well as different orientation and positioning of the substrate in the active site, leading to the differences observed in regiospecificity. In particular, an increase in the volume at the active site of the mutant could enable several options for orientation of the substrate thus changing the regiospecificity. These two residues are conserved in ToMO and indeed a site specific mutation in this enzyme, W167E, enabled a three-fold increase in oxygen transfer rate.<sup>22</sup> The crystal structure of the wild type and the variant revealed that the indole ring of W167 obstructs the channel, while the glutamic acid residue formed hydrogen bonds with adjacent residues causing an opening of the channel.<sup>22</sup>

While in the previous case Monte Carlo simulations were used to explain the experimental results observed for S395C, the opposite scenario occurred with residue F269. This position, conserved in both T4MO and ToMO, has not been reported in the literature. However, visual inspection of the T4MO's active site, revealed that residue F269 is positioned in the intersection of the two channels. This finding motivated further experimental studies and the mutants F269V and F269W were produced. Interestingly, modification to the smaller valine and the larger tryptophan revealed apparent contradictory results. While we expected a correlation between the size of this residue and the difficulties in migration through the channel, the opposite was seen. To rationalize these findings additional PELE simulations were performed for the mutated proteins. The results revealed local and propagated conformational changes in the protein's structure which explain the

experimental results. The tryptophan at position 269, increased the radius of channel 1 rather than reduced it, leading to better access to and from the active site by this pathway, and possible different orientation within the pocket which may explain the changes in regioselectivity of variant F269W. Determination of the crystal structure of this variant could substantiate this assumption.

## **Conclusion**

It is concluded, that mutations in distant locations in the protein may alter the activity and selectivity of non-heme monooxygenases. The combination of experimental and computational approaches facilitates understanding of structure-function correlations, and assists in designing improved biocatalysts.

## **Supporting Information Paragraph**

Variants F269V and F269W were constructed using site-directed mutagenesis at the T4MO *tmoA* gene via overlap extension PCR<sup>25</sup>. Briefly, three oligonucleotide primer pairs were designed (Table 1) to generate the desired mutations.

## **Acknowledgement**

This work was supported by grants from the European Research Council—2009-Adg25027-PELE European project, the Spanish Ministry “Juan de laCierva” for M.F.L.

## **References**

- (1) Bailey, L. J.; McCoy, J. G.; Phillips, G. N.; Fox, B. G. *Proceedings of the National Academy of Sciences* **2008**, *105*, 19194.
- (2) Lindqvist, Y.; Huang, W.; Schneider, G.; Shanklin, J. *The EMBO journal* **1996**, *15*, 4081.
- (3) Lountos, G. T.; Mitchell, K. H.; Studts, J. M.; Fox, B. G.; Orville, A. M. *Biochemistry* **2005**, *44*, 7131.
- (4) Brouk, M.; Fishman, A. *Food Chemistry* **2009**, *116*, 114.
- (5) McCormick, M. S.; Sazinsky, M. H.; Condon, K. L.; Lippard, S. J. *Journal of the American Chemical Society* **2006**, *128*, 15108.
- (6) Notomista, E.; Scognamiglio, R.; Troncone, L.; Donadio, G.; Pezzella, A.; Di Donato, A.; Izzo, V. *Applied and Environmental Microbiology* **2011**, *77*, 5428.

- (7) Pikus, J. D.; Studts, J. M.; McClay, K.; Steffan, R. J.; Fox, B. G. *Biochemistry* **1997**, *36*, 9283.
- (8) Vardar, G.; Wood, T. K. *Applied and Environmental Microbiology* **2004**, *70*, 3253.
- (9) Sazinsky, M. H.; Bard, J.; Di Donato, A.; Lippard, S. J. *Journal of Biological Chemistry* **2004**, *279*, 30600.
- (10) Fishman, A.; Tao, Y.; Rui, L.; Wood, T. K. *Journal of Biological Chemistry* **2005**, *280*, 506.
- (11) Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T. M.; Onufriev, A. *V. Proceedings of the National Academy of Sciences* **2008**, *105*, 9204.
- (12) Lin, T.-L.; Song, G. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79*, 2475.
- (13) Grubmuller, H.; Heymann, B.; Tavan, P. *Science* **1996**, *271*, 997.
- (14) Gervasio, F. L.; Laio, A.; Parrinello, M. *Journal of the American Chemical Society* **2005**, *127*, 2600.
- (15) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. *Journal of the American Chemical Society* **2011**, *133*, 9181.
- (16) Buch, I.; Giorgino, T.; De Fabritiis, G. *Proceedings of the National Academy of Sciences* **2011**.
- (17) Takahashi, R.; Gil, V. A.; Guallar, V. *Journal of Chemical Theory and Computation* **2013**, *10*, 282.
- (18) Borrelli, K.; Vitalis, A.; Alcantara, R.; Guallar, V. *Journal of Chemical Theory and Computation* **2005**, *1*, 1304.
- (19) Cossins, B. P.; Hosseini, A.; Guallar, V. *Journal of Chemical Theory and Computation* **2012**, *8*, 959.
- (20) Song, W. J.; Gucinski, G.; Sazinsky, M. H.; Lippard, S. J. *Proceedings of the National Academy of Sciences* **2011**, *108*, 14795.
- (21) Moe, L. A.; Hu, Z.; Deng, D.; Austin, R. N.; Groves, J. T.; Fox, B. G. *Biochemistry* **2004**, *43*, 15688.
- (22) Bailey, L. J.; Acheson, J. F.; McCoy, J. G.; Elsen, N. L.; Phillips, G. N.; Fox, B. G. *Biochemistry* **2012**, *51*, 1101.
- (23) Borrelli, K. W.; Cossins, B.; Guallar, V. *Journal of Computational Chemistry* **2010**, *31*, 1224.
- (24) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225.
- (25) Yu, Z. Y.; Jacobson, M. P.; Josovitz, J.; Rapp, C. S.; Friesner, R. A. *Journal of Physical Chemistry B* **2004**, *108*, 6643.
- (26) Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. *Bioinformatics* **2011**.
- (27) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. *BMC Bioinformatics* **2009**, *10*, 168.
- (28) Sambbrook J, R. D.; 3 ed.; Cold Spring Harbor Laboratory Press, N. Y., Ed. 2011.
- (29) Brouk, M.; Derry, N.-L.; Shainsky, J.; Ben-Barak Zelas, Z. ; Boyko, Y.; Dabush, K.; Fishman, A. *Journal of Molecular Catalysis B-Enzymatic* **2010**, *66*, 72.
- (30) Tao, Y.; Fishman, A.; Bentley, W. E.; Wood, T. K. *Journal of Bacteriology* **2004**, *186*, 4705.
- (31) Vardar, G.; Wood, T. K. *Appl. Environ. Microbiol.* **2004**, *70*, 3253.
- (32) Brouk, M.; Nov, Y.; Fishman, A. *Applied and Environmental Microbiology* **2010**, *76*, 6397.



## Table of Contents

