

# Structural Insights into the Mechanism for Recognizing Substrate of the Cytochrome P450 Enzyme TxtE

Feng Yu<sup>1\*</sup>, Minjun Li<sup>1</sup>, Chunyan Xu<sup>1</sup>, Zhijun Wang<sup>1</sup>, Huan Zhou<sup>1</sup>, Min Yang<sup>1</sup>, Yaxing Chen<sup>1</sup>, Lin Tang<sup>2</sup>, Jianhua He<sup>1\*</sup>

**1** Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai, China, **2** Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

## Abstract

Thaxtomins, a family of phytotoxins produced by *Streptomyces spp.*, can cause dramatic plant cell hypertrophy and seedling stunting. Thaxtomin A is the dominant form from *Streptomyces scabies* and has demonstrated herbicidal action. TxtE, a cytochrome P450 enzyme from *Streptomyces scabies* 87.22, catalyzes direct nitration of the indolyl moiety of L-tryptophan to L-4-nitrotryptophan using nitric oxide, dioxygen and NADPH. The crystal structure of TxtE was determined at 2.1 Å resolution and described in this work. A clearly defined substrate access channel is observed and can be classified as channel 2a, which is common in bacteria cytochrome P450 enzymes. A continuous hydrogen bond chain from the active site to the external solvent is observed. Compared with other cytochrome P450 enzymes, TxtE shows a unique proton transfer pathway which crosses the helix I distortion. Polar contacts of Arg59, Tyr89, Asn293, Thr296, and Glu394 with L-tryptophan are seen using molecular docking analysis, which are potentially important for substrate recognition and binding. After mutating Arg59, Asn293, Thr296 or Glu394 to leucine, the substrate binding ability of TxtE was lost or decreased significantly. Based on the docking and mutation results, a possible mechanism for substrate recognition and binding is proposed.

**Citation:** Yu F, Li M, Xu C, Wang Z, Zhou H, et al. (2013) Structural Insights into the Mechanism for Recognizing Substrate of the Cytochrome P450 Enzyme TxtE. PLoS ONE 8(11): e81526. doi:10.1371/journal.pone.0081526

**Editor:** Andrea Cavalli, University of Bologna & Italian Institute of Technology, Italy

**Received:** June 8, 2013; **Accepted:** October 14, 2013; **Published:** November 25, 2013

**Copyright:** © 2013 Yu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the National Natural Science Foundation (Grant No. 31100529 and 31100528, <http://www.nsf.gov.cn>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

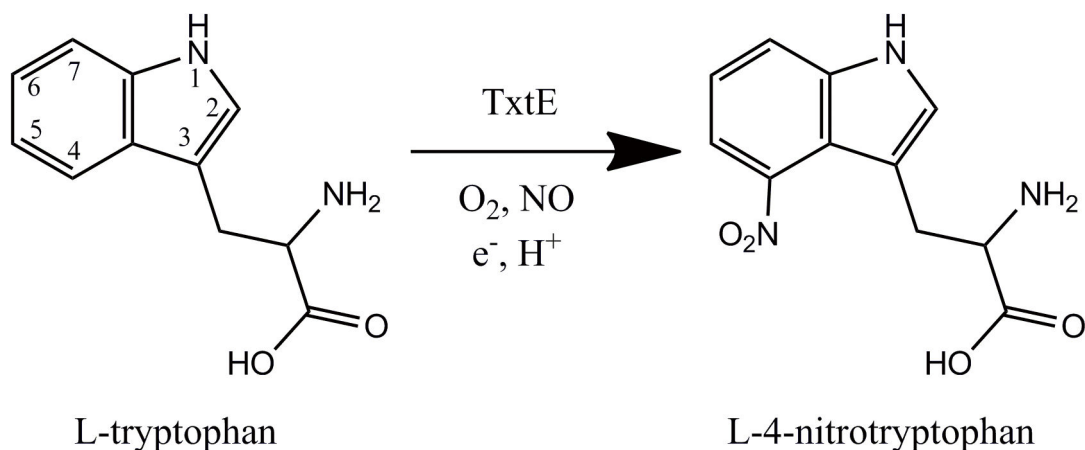
\* E-mail: yufeng@sinap.ac.cn (FY); hejh@sinap.ac.cn (JH)

## Introduction

Cytochromes P450 superfamily (CYP) is a large and diverse group of enzymes containing a heme cofactor which participate in a variety of physiological roles including hormone and secondary metabolite biosynthesis, as well as xenobiotic metabolism[1,2]. Most of CYPs catalyze the insertion of an oxygen atom from dioxygen into the carbon-hydrogen bonds of organic compounds, while the other oxygen atom reacts with protons to form water. This reaction is referred to as the monooxygenase reaction. However, many other types of catalytic activity of CYPs have been demonstrated, including C-C bond cleavage, oxidative dealkylation, C-C bond formation, various ring alterations, and the reduction of nitric oxide to nitrous oxide[1,3,4].

Thaxtomins, a family of phytotoxins produced by *Streptomyces spp.*, can cause dramatic plant cell hypertrophy and seedling stunting[5]. Thaxtomin A is the dominant form from *Streptomyces scabies* which also has demonstrated herbicidal action. Until recently, six proteins were identified in the Thaxtomin A biosynthesis pathway: TxtA and TxtB are non-

ribosomal peptide synthases[6,7]; TxtC[8] and TxtE[9] are CYPs; TxtD is a nitric oxide synthase that produces nitric oxide from L-arginine[10]; TxtR is a cellobiose-responsive pathway specific activator[11]. TxtE catalyzes direct nitration of the indolyl moiety of L-tryptophan (L-Trp) to L-4-nitrotryptophan using nitric oxide, dioxygen and NADPH (Figure 1), which is the key step of Thaxtomin A biosynthesis. Although NO-related nitration is not an uncommon chemical process in organisms, particularly for tyrosine nitration[12], TxtE is the first reported enzyme that catalyzes a direct nitration reaction specifically in a biosynthetic pathway and thus it can potentially be developed for industrial applications. TxtE catalyzes regiospecific nitration of L-tryptophan. D-tryptophan is not the substrate of TxtE. Tryptophan nitration can take place at positions 1, 2, 4, 5, 6 and 7 of the indolyl moiety of tryptophan[13]. However, TxtE only catalyzes a nitration reaction at the 4-position of indolyl moiety. In the reaction products, one oxygen atom of the nitro group and the oxygen atom of the water molecule come from dioxygen. Combined with the catalytic mechanism of CYPs, it was proposed that TxtE may activate dioxygen and nitric oxide to form a peroxyxynitrite[9].



**Figure 1. TxtE catalyzes direct nitration of the indolyl moiety of L-tryptophan to L-4-nitrotryptophan.**

doi: 10.1371/journal.pone.0081526.g001

In this paper, we report the crystal structure of TxtE in substrate-free form. TxtE shows an open conformation. A clearly defined substrate access channel and a potential proton transfer pathway were observed. According to docking and mutagenesis experiments, we propose a possible substrate recognition and binding mechanism.

## Materials and Methods

### Protein expression and purification

The full-length TxtE DNA sequence from *S. scabiei* 87.22 (406 amino acids) was subcloned into plasmid pET-28a, which produced an N-terminal His-tagged protein with a thrombin digestion site. The R59L (TxtE-R59L), Y89F (TxtE-Y89F), N293L (TxtE-N293L), T296L (TxtE-T296L) and E394L (TxtE-E394L) mutants were produced by applying mutagenic PCR to the pET28a-TxtE vector according to QuikChange protocol.

Wild-type TxtE protein was expressed in *Escherichia coli* Rosetta(DE3) plysS cells in LB broth medium for 12 hours at 20°C with 1 mM IPTG induction. The resulting cell pellet was re-suspended in buffer A (20 mM Tris-HCl pH8.0, 300 mM NaCl, 5% Glycerol). Cells were lysed by passing through a microfluidizer (18,000 psi) twice, and then the lysate was centrifuged at 30,000 g for 30 minutes. The supernatant was loaded into a 5 mL Ni-IDA column (GE Healthcare) which had been equilibrated with buffer A, and then eluted by using buffer B (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM imidazole, 5% glycerol). After dilution with buffer C (20 mM Tris-HCl pH 8.0, 5% glycerol), the sample was loaded into HiTrap Q FF and eluted by using a linear gradient. Finally, TxtE was loaded into a Superdex 75 16/60 column which was equilibrated with buffer D (25 mM Tris-HCl pH 8.0, 150 mM NaCl and 5% glycerol). The resulting protein peak was concentrated to 15 mg/mL by using a 10 kDa MWCO Amicon Ultra. Aliquots were snap frozen in liquid nitrogen and stored at -80°C until they were used for crystallization. TxtE-R59L, TxtE-Y89F, TxtE-N293L, TxtE-T296L and TxtE-E394L mutant proteins were expressed and purified using the same method as mentioned above.

### Crystallization and structure determination

Before crystallization, fresh DTT and imidazole pH 6.5 were added to protein solution. The final concentrations were 2 mM and 200 mM, respectively. Crystals of TxtE were grown by the hanging-drop vapor diffusion method at 18°C. Protein solution was mixed with equal volumes of reservoir solution containing 20% MPD and 20% PEG6000. Crystals were cryo-protected with cryoprotectant before freezing in liquid nitrogen.

X-ray diffraction data were collected at beamline BL17U1 of the Shanghai Synchrotron Radiation Facility. Data were processed using the HKL2000 software suite[14]. The TxtE structure was solved by molecular replacement using Phaser[15] with P450BIOI[16] (PDB: 3EJB) as the search model. After molecular replacement, maximum likelihood-based refinement of the atomic positions and temperature factors were performed with PHENIX[17], and the atomic model was fit with the program Coot[18]. The stereochemical quality of the final model was assessed with MolProbity[19]. Crystallographic statistics for the final model are shown in Table 1. The substrate access channel was analyzed using Caver[20]. Figures were prepared with PyMOL[21].

### Substrate docking calculations

L-tryptophan was docked into the structure of TxtE in the zwitterionic form using AutoDock 4.2[22]. Imidazole and water molecules were removed before docking. All side chains were set to rigid body and grid spacing was set to 1 Å. Other parameters were with their default values. The three lowest-energy docking solutions from 2,500,000 search results were chosen since they all showed relevant enzyme-substrate interactions.

### L-tryptophan binding assays of wild-type TxtE, TxtE-R59L, TxtE-T296L and TxtE-E394L

Purified wild-type TxtE, TxtE-R59L, TxtE-T296L and TxtE-E394L were exchanged to buffer E (25 mM Tris-HCl pH 8.0) by ultrafiltration. Finally, protein was diluted to about 15 μM using buffer E. Spectra were recorded between 325 nm and 500 nm

**Table 1.** Data collection and refinement statistics.

Parameter	Values
<b>Data collection</b>	
Space group	P1
a (Å)	46.3
b (Å)	52.6
c (Å)	88.2
α, β, γ (°)	77.8, 82.2, 68.4
Resolution (Å)	50.00 – 2.10 (2.14 – 2.10) <sup>a</sup>
R <sub>merge</sub> (%)	4.6 (37.7) <sup>a</sup>
<math>\langle I \rangle / \langle \sigma \rangle</math>	28.7 (5.5) <sup>a</sup>
Completeness (%)	96.7 (96.5) <sup>a</sup>
Redundancy	3.9 (3.9) <sup>a</sup>
<b>Refinement</b>	
No. reflections	40,644
R <sub>work</sub> /R <sub>free</sub> (%)	19.64/23.43 <sup>b, c</sup>
No. atoms	
Protein	6141
Ligand	101
Water	339
RMSD	
Bond lengths (Å)	0.005
Bond angles (°)	1.255
Ramachandran plot (%)	
Favored	97.70
Allowed	2.05
Outliers	0.26

<sup>a</sup>. Values in parentheses are for the highest resolution shell.

<sup>b</sup>.  $R = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} F_{obs}$

<sup>c</sup>. R<sub>free</sub>, calculated the same as R<sub>work</sub>, but from a test set containing 5% of data excluded from the refinement calculation.

doi: 10.1371/journal.pone.0081526.t001

with or without L-tryptophan using a UV-vis spectrophotometer (Chirascan, Applied PhotoPhysics, UK). The final concentration of L-tryptophan was 1 mM.

### Determination of dissociation constants of L-Tryptophan from wild-type TxtE, TxtE-Y89F and TxtE-N293L

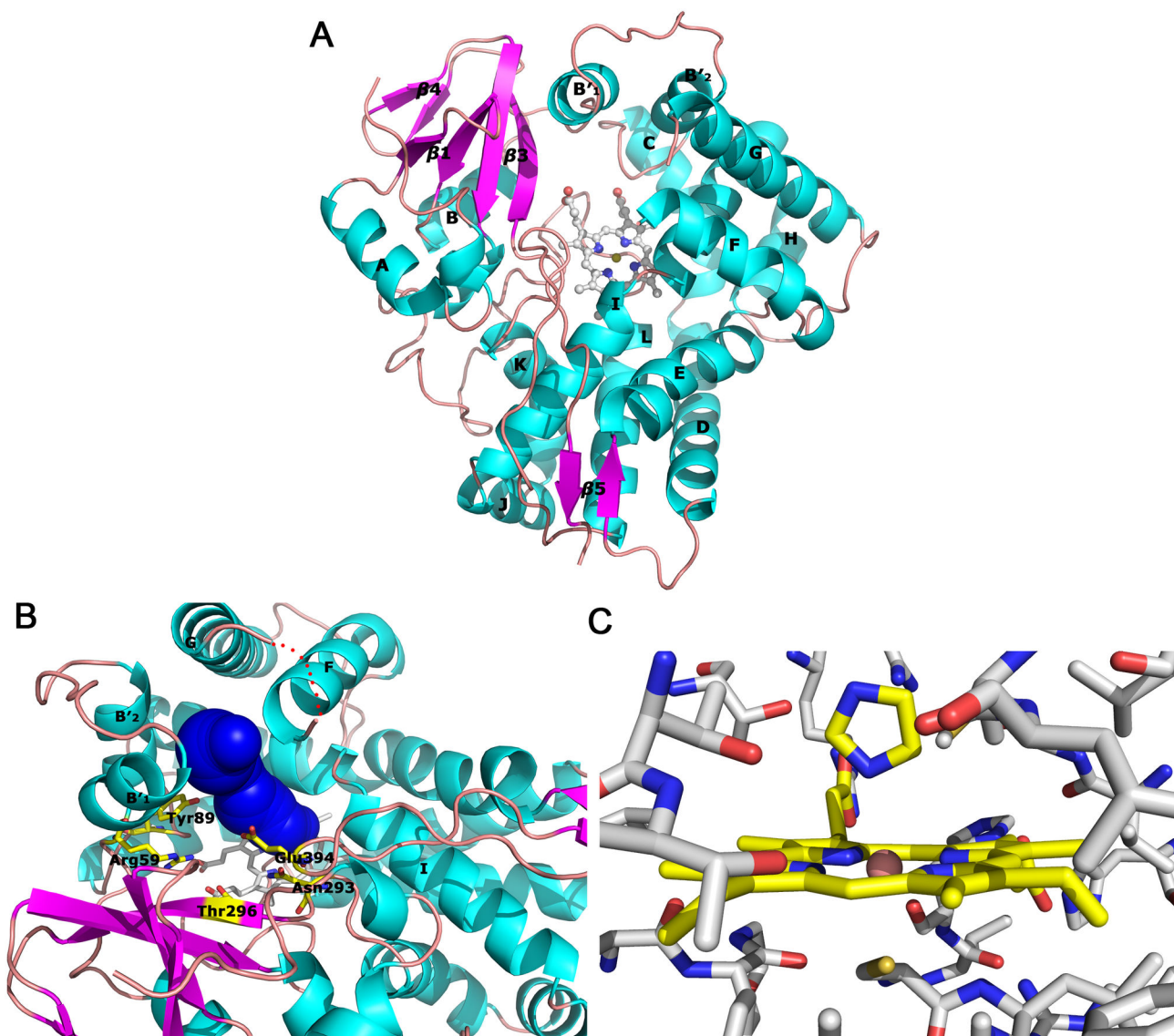
Purified wild-type TxtE, TxtE-Y89F, and TxtE-N293L were exchanged to buffer E (25 mM Tris-HCl pH 8.0) by ultrafiltration. Finally, the protein was diluted to about 8 μM using buffer E. L-tryptophan was added to the sample cuvette until the protein was saturated. The added substrate solution did not exceed 2% of the total volume. The protein solution without L-tryptophan was served as control. This procedure was repeated three times for each sample. The average values of the difference in absorbance of each spectrum at 390 nm and 420 nm were plotted against the concentration of the substrate. The data were fitted to a one site binding model using Qtiplot.

## Results

### Overall structure and the substrate access channel

The purified wild-type TxtE used in this study is catalytically active. L-4-nitrotryptophan could be separated and identified in the reaction products using LC-MS (Figure S1, Supporting Information). The TxtE crystals belong to space group P1, with unit-cell parameters  $a = 46.3$  Å,  $b = 52.6$  Å,  $c = 88.2$  Å,  $\alpha = 77.8^\circ$ ,  $\beta = 82.2^\circ$  and  $\gamma = 68.4^\circ$ . Detailed crystallographic data statistics for TxtE are shown in Table 1. Although gel filtration results show that TxtE is a monomer in the solution, two molecules are found in the asymmetric unit. TxtE presents a typical trigonal-prism like cytochrome P450 fold with a length of ~50 Å and a thickness of ~30 Å. The overall structure of TxtE is shown in Figure 2A. The N-terminal His-tag, residues 1 of chain A, residues 1 - 4 of chain B, residues 175 - 183 and the last residue of both chains cannot be observed as there is no clear electron density for these residues. This structure also contains three imidazole molecules, two of which are coordinated with the iron atoms of heme. The third imidazole molecule binds with Glu187 of both chains simultaneously.

Compared with the arrangement of the secondary structure elements in the P450 fold, four of the five β-strands (β1, β3, β4



**Figure 2. Structure of TxtE.** A Overall structure of TxtE. The helices and strands are noted. B The substrate access channel is present as blue tunnel and locates between the B<sub>1</sub>' helix, the F helix and the G helix, which can be classified as channel 2a according the common nomenclature. Arg59, Tyr89, Asn293, Thr296 and Glu394 are also present. F-G loop is draw as red dots. C Imidazole coordinates to the iron atom of heme.

doi: 10.1371/journal.pone.0081526.g002

and  $\beta 5$ ) in TxtE (Figure 2A) can be traced following the order first defined in the structure of CYP101A1[23]. Because of the absence of backbone hydrogen bonds,  $\beta 2$  strand is replaced by a loop in TxtE. Most of  $\alpha$ -helices (A - L) sit at positions commonly found in the P450 fold. However, two B' helices (B<sub>1</sub>' and B<sub>2</sub>') are separated by an inserted loop (residues 69 - 77). The helix B<sub>1</sub>' of TxtE is well overlapped with helix B' of CYP101A1. The helix B<sub>2</sub>' locates at the similar position of the B-C loop of CYP101A1, and is parallel to the helix B<sub>1</sub>' of TxtE. Among the reported CYP structures, it is unusual to find two B' helices separated by a short loop. Two B' helices together with B<sub>2</sub>'-C loop, F helix and G helix are located in the distal side of

the heme, and are involved in substrate access and binding. In both molecules, the F-G loop (residues 175 - 183) is disordered. These residues may form a lid on the top surface of the substrate access channel (Figure 2B). This channel is defined by Arg59, Val63, Lys67, Phe68 from the B<sub>1</sub>' helix; Phe79 and Trp82 from the B<sub>2</sub>' helix; Met88, Tyr89 from the B<sub>2</sub>'-C loop; Met173, Thr174 from the F helix; Val184, Thr185, Ile189 from the G helix; Leu241, Ala245 and Thr250 from the I helix; and Asn293, Thr296, Trp297 and Glu394 abound  $\beta 5$  strand. Residues Arg59, Met88, Leu241, Ala245, Thr250, Asn293, Thr296 and Glu394 sit near the bottom of the channel, closely

to the heme, and therefore are considered to be a part of the substrate binding pocket.

### The structure of active site and L-tryptophan binding

The orientation of heme in TxtE is similar to that in other CYPs (Figure 2C). The thiolato-sulfur of Cys357 is the proximal ligand (Fe–S = 2.65 Å). Imidazole is the sixth heme ligand (Fe–N = 2.74 Å).

Because we could not find crystallization conditions without imidazole, attempts to co-crystallize TxtE with L-tryptophan were unsuccessful. The program AutoDock was therefore used to explore potential L-tryptophan binding poses. Three lowest-energy docking solutions from 2,500,000 searching results were chosen (Figure 3). The calculated binding energies were between –4.63 and –4.95 kcal/mol, and the relative binding poses were similar, especially for the first two poses. Interestingly, we found that, in all three poses, the substrate L-tryptophan carboxylate group interacted with the guanyl group of Arg59 and the hydroxy group of Thr296, and its amino group interacted with the carboxylate group of Glu394 via salt bridges or hydrogen bonds (Figure 3). Besides these three residues, Tyr89 could form polar contacts with the carboxylate group of L-tryptophan, and Asn293 could contact with the imino or carboxylate group of L-tryptophan.

To validate this docking model of L-tryptophan binding, we designed five single-residue mutation experiments. In which, Arg59, Asn293, Thr296 and Glu394 were mutated to leucine, and Tyr89 was mutated to phenylalanine. In principle, substrate binding of CYPs results in a shift in the UV-vis spectrum of the heme chromophore within the protein[24]. In the type I binding spectrum, the peak is blue-shifted. After substrate binding, the maximum absorption wavelength changes from 420 nm to 390 nm. In this case, the substrate does not coordinate to the iron atom of the heme. In the type II binding spectrum, the peak is red-shifted. Upon substrate binding, the maximum absorption wavelength changes from 420 nm to 420 - 435 nm, which indicates substrate binds directly to the iron atom of the heme. The wild-type TxtE shows a type I binding spectrum after L-tryptophan added. As shown in Figure 4A, after adding L-tryptophan to TxtE solution, the absorption at 390 nm increased, while the absorption at 420 nm decreased. For TxtE-R59L, TxtE-T296L and TxtE-E394L mutants, no shift was observed after L-tryptophan added (Figure 4B, 4C and 4D), which suggests that Arg59, Thr296 and Glu394 are essential for binding the substrate L-tryptophan. TxtE-Y89F and TxtE-N293L still could bind L-tryptophan, but the binding ability decreased significantly, especially for TxtE-N293L. The dissociation constants ( $K_d$ ) of TxtE-Y89F and TxtE-N293L were  $120.63 \pm 4.30 \mu\text{M}$  and  $1.70 \pm 0.28 \text{ mM}$ , respectively, while the  $K_d$  of the recombinant wild-type TxtE used in this study was  $43.59 \pm 2.26 \mu\text{M}$  (Figure 5). Therefore Asn293 is also likely a key residue for binding L-tryptophan.

### The proton and electron transfer pathway

Residues Ala245 - Thr250 in TxtE, which are located at helix I, strongly overlap with Gly248 - Thr252 in CYP101A1. In CYP101A1, the hydrogen bond between the Thr252 hydroxyl group and the Gly248 carbonyl group causes the kink of helix I,

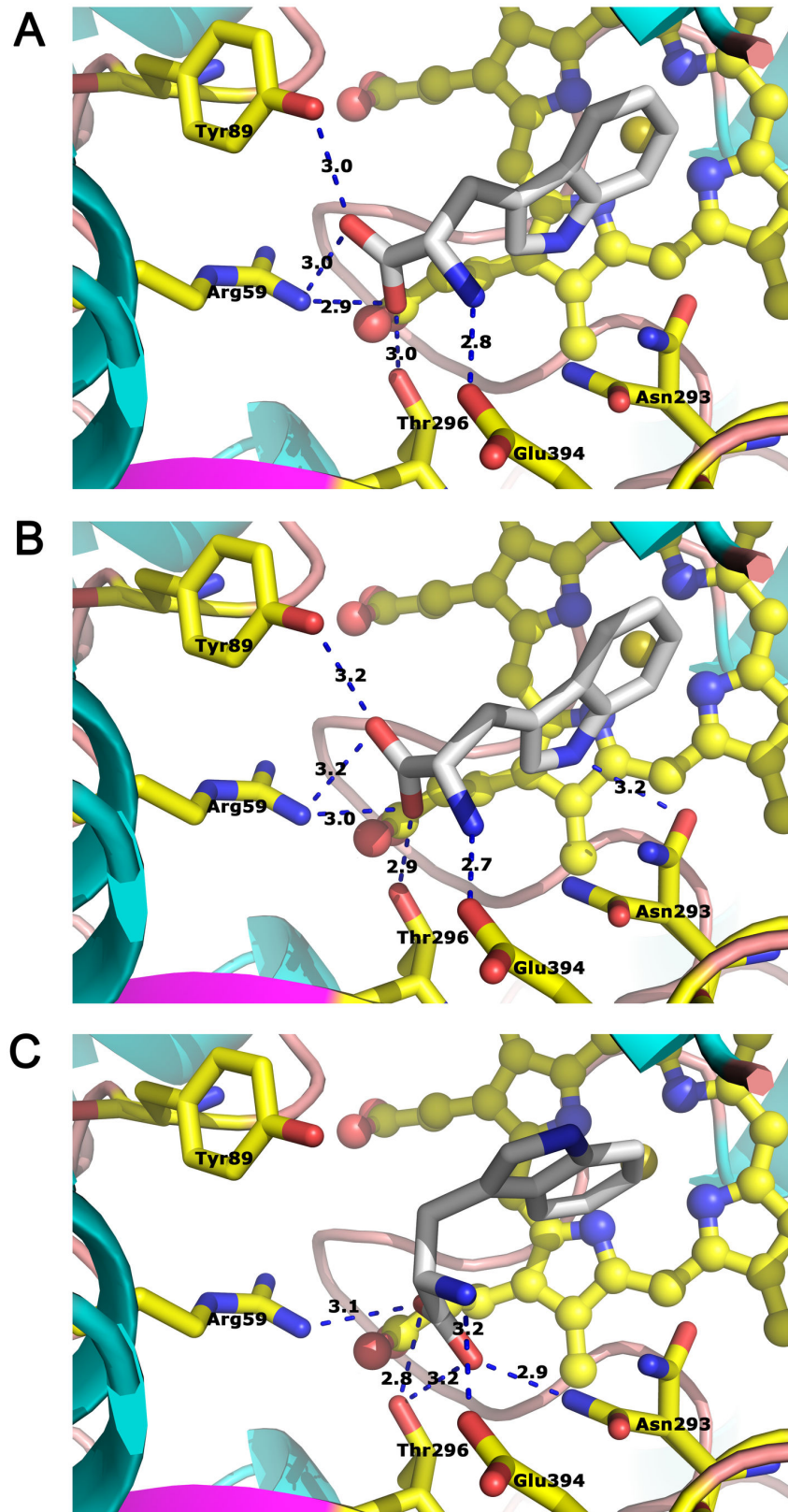
which is common in CYPs, and is believed to play important roles in oxygen activation[25]. In contrast to CYP101A1, in TxtE there is one more residue between Ala245 and Thr250. Therefore, the kink between Ala245 and Thr250 is widened which opens up the entrance between the active site and the external solvent (Figure 6). In addition, two water molecules (HOH38 and HOH309) are found to participate in hydrogen bond chain formation between the Thr250 hydroxyl group and the Ala245 carbonyl group.

Proton transfer is required for the CYPs functions. In the case of TxtE, protons are likely to protonize an Fe(III)-OH intermediate, which is formed at the end of the reaction cycle, to regenerate the resting state of the enzyme. However, in many known CYP structures, the complete proton transfer pathway is not clear with the exception of CYP158A2[26] and P450BM3[27]. Interestingly, in the structure of TxtE a continuous hydrogen bond chain from the active site to the external solvent is observed due to the I helix distortion. This hydrogen bond chain is from HOH309 to HOH219. The carbonyl group of Ala248 forms hydrogen bonds with HOH38 and HOH73 simultaneously. HOH73 links the external solvent (HOH123 and HOH219). HOH309 is near the active site, and links to HOH38 by the carbonyl group of Ala245 and the hydroxyl group of Thr250. However, the hydrogen bond network, as well as the arrangement and number of water molecules in such putative proton transfer pathways, are known to change in the different reaction stages.

The heme proximal face, where the iron atom is closest to the enzyme surface, is the likely binding site for the electron transfer complex. *In vitro*, TxtE can accept electrons from spinach ferredoxin (Fd) and ferredoxin reductase (Fr). Usually, in P450 enzyme reactions, there are two electrons transferred from NAD(P)H through the Fd and FrFr pathway, which can be divided into two transfer processes. The first one happens between the CYP-substrate complex and Fd, and the second one happens between CYP-dioxygen-substrate complex and Fd. According to the theoretical and experimental results, in CYP101A1, Arg112 plays important roles in the first electron transfer[28]. Compared with the heme proximal face residues of TxtE and CYP101A1, we found that Arg112 is located at the identical location in TxtE. This implies that TxtE may have the same first electron transfer pathway as that in CYP101A1.

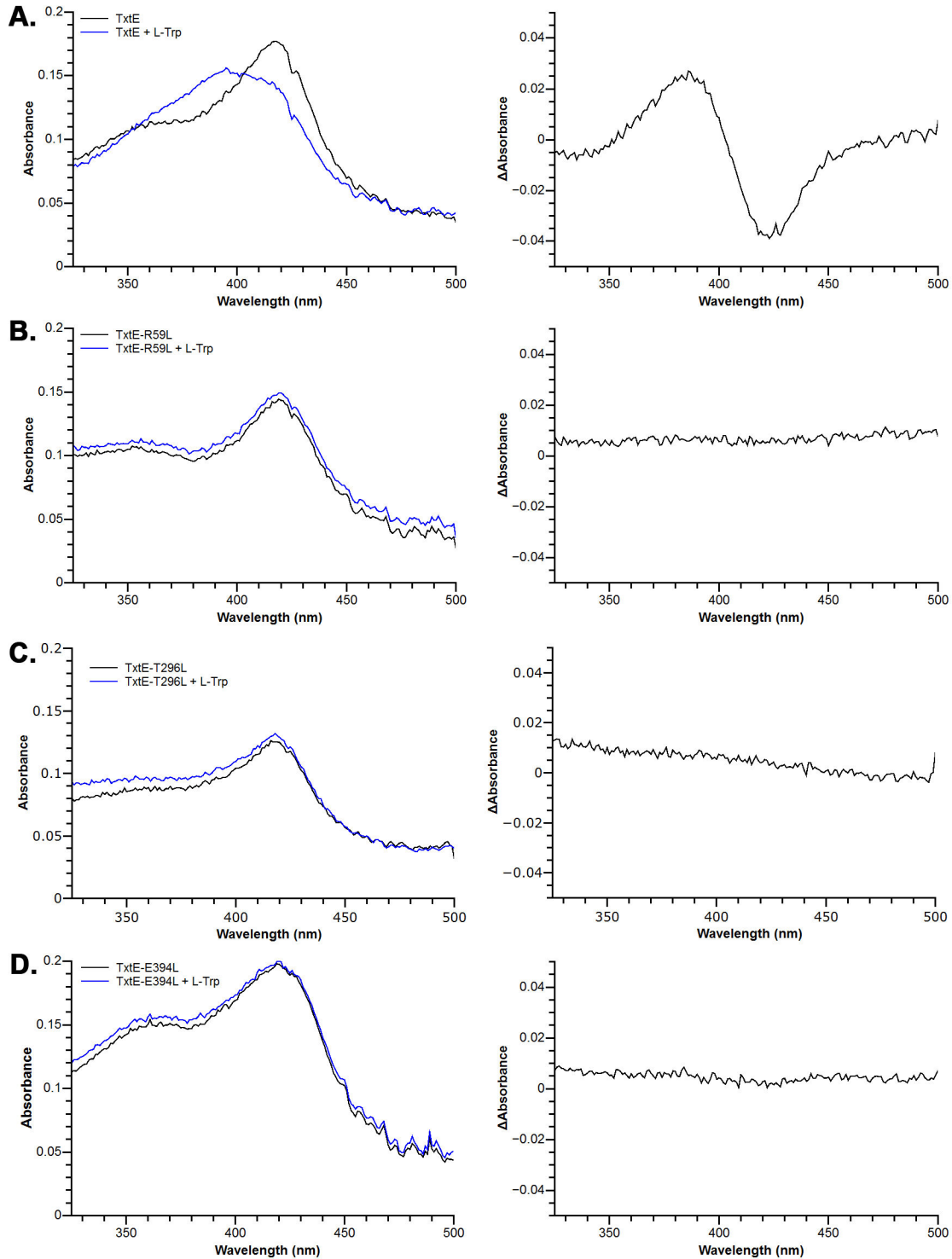
### Discussion

Although the overall folds of CYPs are strongly conserved, the precise position of various structure elements differs substantially, especially in the regions controlling substrate specificity. In the structure reported here, TxtE has the characteristic P450 fold. A clearly defined substrate access channel was observed between helix B' and helix G which can be classified as channel 2a according to the common nomenclature[29]. Channel 2a is common in bacteria CYPs. TxtE has two B' helices, which is an uncommon conformation. Besides TxtE, there are five known CYP structures which have two B' helices, but only AurH[30], CYP121[31] and CalO<sub>2</sub>[32] form two B' helices bundle similar to TxtE. This B' helices bundle creates a hydrophobic "wall" that shields the substrate



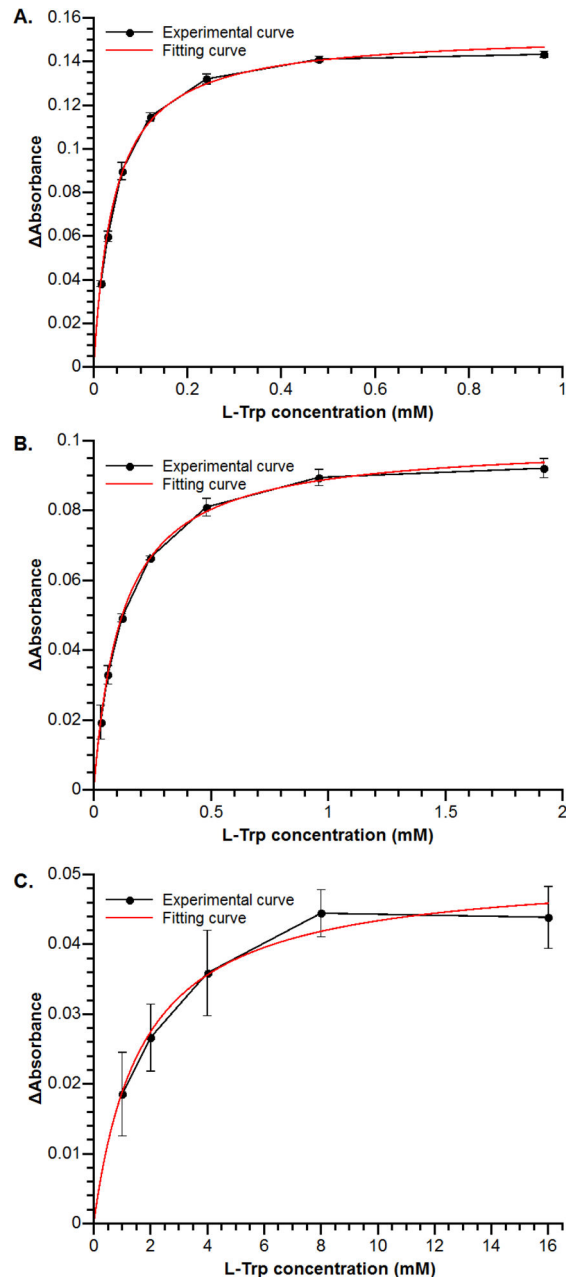
**Figure 3. The molecular docking results.** A, B and C are the three lowest binding poses, respectively. The polar contacts are presented as blue dashes and distances are also labeled.

doi: 10.1371/journal.pone.0081526.g003



**Figure 4. L-tryptophan binding assays of wild-type TxtE, TxtE-T296L and TxtE-E394L.** A The UV-vis spectra of TxtE with or without L-tryptophan (Left). TxtE shows an obvious type I different spectrum (Right). B The UV-vis spectra of TxtE-R59L with or without L-tryptophan (Left). No shift can be observed (Right). C The UV-vis spectra of TxtE-T296L with or without L-tryptophan (Left). No shift can be observed (Right). D The UV-vis spectra of TxtE-E394L with or without L-tryptophan (Left). No shift can be observed (Right).

doi: 10.1371/journal.pone.0081526.g004



**Figure 5. Dissociation constants of L-Tryptophan from wild-type TxtE, TxtE-Y89F and TxtE-N293L.** The average values of the difference in absorbance of each spectrum at 390 nm and 420 nm were plotted against the concentration of the substrate. A Wild-type TxtE (8  $\mu$ M) was titrated with L-tryptophan from 15  $\mu$ M to 960  $\mu$ M. B TxtE-Y89F (8  $\mu$ M) was titrated with L-tryptophan from 30  $\mu$ M to 1920  $\mu$ M. C TxtE-N293L (8  $\mu$ M) was titrated with L-tryptophan from 1 mM to 16 mM.

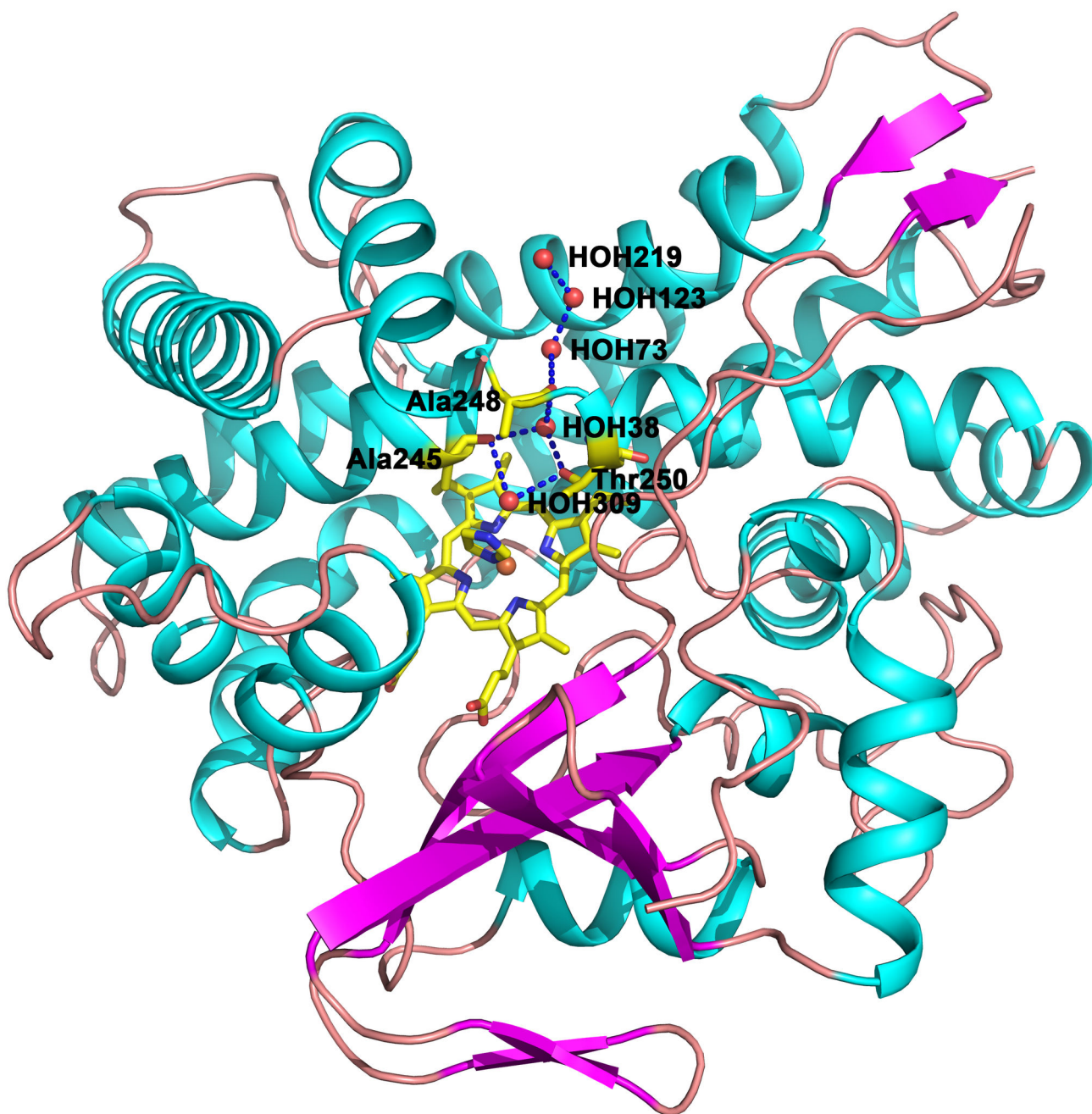
doi: 10.1371/journal.pone.0081526.g005

binding pocket from solvent, and seals the substrate binding pocket. Other examples with two B' helices are CYP119[33] and P450epoK[34]. In CYP119 B'<sub>2</sub> helix only has one turn, and in P450epoK these two helices are nearly vertical.

In the reaction of CYPs, protons should reach the active site through a hydrogen bond chain involving water molecules. Unfortunately, few CYP structures show a clear proton transfer

pathway from the active site to the external solvent except CYP158A2[26] and P450BM3[27]. The continuous hydrogen bond chains from the active site to the external solvent were observed in the CYP158A2-dioxygen-substrate complex and the P450BM3-reductase domain complex, in which the hydrogen bond chain is nearly vertical to the heme plane. Compared with these two structures, TxtE shows a unique





**Figure 6. The potential protons transfer pathway.** The kink between Ala245 and Thr250 is widened and opens up an entrance between the active site and the external solvent. The hydrogen bonds are presented as blue dashes.

doi: 10.1371/journal.pone.0081526.g006

proton transfer pathway which crosses the helix I distortion and is nearly parallel to the heme plane. The mechanism of oxygen activation of CYP101A1 was elaborated [25]. TxtE shows a possible mechanism of oxygen activation, which is consistent with that of CYP101A1. In CYP101A1, Thr252<sub>CYP101A1</sub> and a water molecule are required to form hydrogen bonds with dioxygen simultaneously for the cleavage of dioxygen. In TxtE, Thr250 and HOH309 may have the same roles in oxygen

activation although the precise positions of HOH309 and Thr250 may be changed after dioxygen binding due to the rearrangement of water molecules. In CYP101A1, protons are also transferred through widened helix I, but the water chain only extend to Glu366<sub>CYP101A1</sub>. Compared with CYP101A1, the kink of the helix in TxtE is wider than that in CYP101A1. Therefore it is easier to form a complete water chain through this kink in the reaction process.

The molecular docking suggests that carboxyl group and amino group of L-tryptophan form salt bridges or hydrogen bonds to Arg59, Tyr89, Asn293, Thr296 and Glu394. The bond lengths (from proton donor to acceptor) are shown in Figure 3. Substrate binding assays show Arg59, Asn293, Thr296 and Glu394 are key residues for binding L-tryptophan. In our simulation results, the distance between the carbon atom at 4-position of indolyl moiety and the iron atom of heme is between 5.50 and 6.22 Å, which is a reasonable distance in CYPs. To date, it has not been possible to obtain crystals of the substrate-bound form of TxtE, so we cannot determine whether this conformation of the side chain is the actual conformation in protein-substrate complex. Even so, because tryptophan has the largest side chain in all of amino acids, no side chain atom can be at the corresponding position of 4-position of L-tryptophan, even with the conformation of the side chains changing. In this proposed model, the formation of the TxtE-substrate complex mainly depends on the amino and carboxyl groups of substrates, but not the side chain. This provides the possibility to modify TxtE or substrates for producing nitration products of amino acids analogs. Nitration products of amino acids and analogs are important drug synthesis intermediates. Our findings can help to develop enzymatic synthesis pathway of these intermediates, and avoid the risk of chemical synthesis. However, our docking model is based on the substrate-free conformation. CYPs usually undergo a conformational change upon substrate binding. To clarify the substitute binding mechanism of TxtE completely, it still need to

determine the crystal structure of TxtE and L-tryptophan complex.

## Supporting Information

**Figure S1. The enzyme activity assay.** The enzyme reaction products were analyzed by using LC-MS. The *m/z* of L-tryptophan is 205, and *m/z* of L-4-nitrotryptophan is 250. Ions at *m/z* = 250 were detected in the products which were catalyzed by using un-boiled TxtE, but could not be detected in the products which were catalyzed by using boiled TxtE. (PDF)

**Materials and Methods S1.** (PDF)

## Acknowledgements

The structure coordinate and structure factors have deposited to Protein Data Bank, and accession code is 4L36. We thank Dr. Thomas Earnest for his help in revising the paper.

## Author Contributions

Conceived and designed the experiments: FY JH. Performed the experiments: FY. Analyzed the data: FY ML. Contributed reagents/materials/analysis tools: ZW HZ MY YC. Wrote the manuscript: FY ML CX LT JH.

## References

- de Montellano PRO (2004) Cytochrome P450: structure, mechanism, and biochemistry. Springer.
- Pikuleva IA, Waterman MR (2013) Cytochromes P450: Roles in Diseases. *J Biol Chem* 288: 17091-17098. doi:10.1074/jbc.R112.431916. PubMed: 23632021.
- Shoun H, Fushinobu S, Jiang L, Kim SW, Wakagi T (2012) Fungal denitrification and nitric oxide reductase cytochrome P450nor. *Philos Trans R Soc Lond B Biol Sci* 367: 1186-1194. doi:10.1098/rstb.2011.0335. PubMed: 22451104.
- Guengerich FP, Munro AW (2013) Unusual Cytochrome P450 Enzymes and Reactions. *J Biol Chem* 288: 17065-17073. doi:10.1074/jbc.R113.462275. PubMed: 23632016.
- Fry BA, Loria R (2002) Thaxtomin A: evidence for a plant cell wall target. *Physiological and Molecular Plant Pathology* 60: 1-8. doi: 10.1006/pmpp.2001.0371.
- Johnson EG, Krasnoff SB, Bignell DRD, Chung WC, Tao T et al. (2009) 4-Nitrotryptophan is a substrate for the non-ribosomal peptide synthetase TxtB in the thaxtomin A biosynthetic pathway. *Mol Microbiol* 73: 409-418. doi:10.1111/j.1365-2958.2009.06780.x. PubMed: 19570136.
- Healy FG, Wach M, Krasnoff SB, Gibson DM, Loria R (2000) The txtAB genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Mol Microbiol* 38: 794-804. doi:10.1046/j.1365-2958.2000.02170.x. PubMed: 11115114.
- Healy FG, Krasnoff SB, Wach M, Gibson DM, Loria R (2002) Involvement of a cytochrome P450 monooxygenase in thaxtomin A biosynthesis by *Streptomyces acidiscabies*. *J Bacteriol* 184: 2019-2029. doi:10.1128/JB.184.7.2019-2029.2002. PubMed: 11889110.
- Barry SM, Kers JA, Johnson EG, Song LJ, Aston PR et al. (2012) Cytochrome P450-catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nat Chem Biol* 8: 814-816. doi:10.1038/nchembio.1048. PubMed: 22941045.
- Kers JA, Wach MJ, Krasnoff SB, Widom J, Cameron KD et al. (2004) Nitration of a peptide phytotoxin by bacterial nitric oxide synthase. *Nature* 429: 79-82. doi:10.1038/nature02504. PubMed: 15129284.
- Joshi MV, Bignell DRD, Johnson EG, Sparks JP, Gibson DM et al. (2007) The AraC/XylS regulator TxtR modulates thaxtomin biosynthesis and virulence in *Streptomyces scabies*. *Mol Microbiol* 66: 633-642. doi: 10.1111/j.1365-2958.2007.05942.x. PubMed: 17919290.
- Radi R (2004) Nitric oxide, oxidants, and protein tyrosine nitration. *Proc Natl Acad Sci U S A* 101: 4003-4008. doi:10.1073/pnas.0307446101. PubMed: 15202765.
- Sala A, Nicolis S, Roncone R, Casella L, Monzani E (2004) Peroxidase catalyzed nitration of tryptophan derivatives - Mechanism, products and comparison with chemical nitrating agents. *European Journal of Biochemistry* 271: 2841-2852. doi:10.1111/j.1432-1033.2004.04219.x. PubMed: 15206949.
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* 276: 307-326. doi:10.1016/S0076-6879(97)76066-X.
- Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40: 658-674. doi:10.1107/S0021889807021206. PubMed: 19461840.
- Cryle MJ, Schlichting I (2008) Structural insights from a P450 Carrier Protein complex reveal how specificity is achieved in the P450(Biol) ACP complex. *Proc Natl Acad Sci U S A* 105: 15696-15701. doi: 10.1073/pnas.0805983105. PubMed: 18838690.
- Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66: 213-221. doi: 10.1107/S0907444909052925. PubMed: 20124702.
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486-501. doi:10.1107/S0907444910007493. PubMed: 20383002.
- Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66: 12-21. doi: 10.1107/S0907444909042073. PubMed: 20057044.
- Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J et al. (2012) CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput Biol* 8: e1002708. PubMed: 23093919.

21. Schrodinger LLC (2012) The PyMOL Molecular Graphics System, version 1.5.0.3.
22. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK et al. (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem* 30: 2785-2791. doi:10.1002/jcc.21256. PubMed: 19399780.
23. Poulos TL, Finzel BC, Howard AJ (1987) High-Resolution Crystal-Structure of Cytochrome-P450cam. *J Mol Biol* 195: 687-700. doi: 10.1016/0022-2836(87)90190-2. PubMed: 3656428.
24. Barry SM, Challis GL (2012) Tailoring Reactions Catalyzed by Heme-Dependent Enzymes: Spectroscopic Characterization of the L-Tryptophan-Nitrating Cytochrome P450 TxtE. *Methods Enzymol* 516: 171-194. doi:10.1016/B978-0-12-394291-3.00001-0. PubMed: 23034229.
25. Schlichting I, Berendzen J, Chu K, Stock AM, Maves SA et al. (2000) The catalytic pathway of cytochrome P450cam at atomic resolution. *Science* 287: 1615-1622. doi:10.1126/science.287.5458.1615. PubMed: 10698731.
26. Zhao B, Guengerich FP, Voehler M, Waterman MR (2005) Role of active site water molecules and substrate hydroxyl groups in oxygen activation by cytochrome P450 158A2 - A new mechanism of proton transfer. *J Biol Chem* 280: 42188-42197. doi:10.1074/jbc.M509220200. PubMed: 16239228.
27. Sevrioukova IF, Li HY, Zhang H, Peterson JA, Poulos TL (1999) Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc Natl Acad Sci U S A* 96: 1863-1868. doi:10.1073/pnas.96.5.1863. PubMed: 10051560.
28. Kuznetsov VY, Poulos TL, Sevrioukova IF (2006) Putidaredoxin-to-cytochrome P450cam electron transfer: Differences between the two reductive steps required for catalysis. *Biochemistry* 45: 11934-11944. doi:10.1021/bi0611154. PubMed: 17002293.
29. Cojocaru V, Winn PJ, Wade RC (2007) The ins and outs of cytochrome P450s. *Biochim Biophys Acta* 1770: 390-401. doi:10.1016/j.bbagen.2006.07.005. PubMed: 16920266.
30. Zocher G, Richter MEA, Mueller U, Hertweck C (2011) Structural Fine-Tuning of a Multifunctional Cytochrome P450 Monooxygenase. *J Am Chem Soc* 133: 2292-2302. doi:10.1021/ja110146z. PubMed: 21280577.
31. Hudson SA, McLean KJ, Surade S, Yang YQ, Leys D et al. (2012) Application of Fragment Screening and Merging to the Discovery of Inhibitors of the Mycobacterium tuberculosis Cytochrome P450 CYP121. *Angew Chem Int Ed Engl* 51: 9311-9316. doi:10.1002/anie.201202544. PubMed: 22890978.
32. McCoy JG, Johnson HD, Singh S, Bingman CA, Lei IK et al. (2009) Structural characterization of CaO<sub>2</sub>: A putative orsellinic acid P450 oxidase in the calicheamicin biosynthetic pathway. *Proteins* 74: 50-60. doi:10.1002/prot.22131. PubMed: 18561189.
33. Yano JK, Koo LS, Schuller DJ, Li HY, de Montellano PRO et al. (2000) Crystal structure of a thermophilic cytochrome P450 from the archaeon *Sulfolobus solfataricus*. *J Biol Chem* 275: 31086-31092. doi:10.1074/jbc.M004281200. PubMed: 10859321.
34. Nagano S, Li HY, Shimizu H, Nishida C, Ogura H et al. (2003) Crystal structures of epothilone D-bound, epothilone B-bound, and substrate-free forms of cytochrome P450epoK. *J Biol Chem* 278: 44886-44893. doi:10.1074/jbc.M308115200. PubMed: 12933799.