Homology Modeling and In Vitro Analysis for Characterization of *Streptomyces peucetius* CYP157C4

Hemraj Rimal¹, Sang-Cheol Yu¹, Jong Hwa Jang²*, and Tae-Jin Oh¹*

¹Department of Pharmaceutical Engineering, SunMoon University, Asan 336-708, Republic of Korea
²Department of Dental Hygiene, Hanseo University, Seosan 336-706, Republic of Korea

In this study, we tried to characterize *Streptomyces peucetius* CYP157C4 with homology modeling using three cytochrome P450 (CYP) structures (CYP157C1, CYP164A2, and CYP107L1), having discovered that CYP157C4 lacks the ExxR motif that was considered invariant in all CYPs. We used Discovery Studio 3.5 to build our model after first assessing the stereochemical quality and side-chain environment, and a 7-ethoxycoumarin substrate was docked into the final model. The model-substrate complex allowed us to identify functionally important residues and validate the active-site architecture. We found a distance of 4.56 Å between the 7-ethoxycoumarin and the active site of the heme, and cloning and an *in vitro* assay of the CYP157C4 showed the dealkylation of the substrate. Since the details regarding this group of CYP structures are still unknown, the findings of this study may provide elucidation to assist with future efforts to find a legitimate substrate.

**Keywords:** Cytochrome P450, 7-ethoxycoumarin, homology modeling, *in vitro* assay, *Streptomyces peucetius*

**Introduction**

The extensive cytochrome P450 (CYP) enzyme system catalyzes the monoxygenation of a wide variety of structurally diverse substrates [10, 18]. The enormous biodiversity of the CYP superfamily has been found in bacteria, fungi, insects, worms, plants, fish, and mammals, as shown in the following URLs: http://drnelson.utmem.edu/CytochromeP450.html; http://p450.antibes.inra.fr; and http://Arabidopsis-P450.bioteic.uiuc.edu. CYPs also participate in oxidation, peroxidation, and reduction reactions, and metabolize exogenous compounds, including therapeutic drugs and xenobiotics [3, 19].

Current CYP research seeks to find the reason behind the specificity of the enzyme in regio- and stereoselective reactions to obtain improved chemical or drug yields that are difficult to produce with conventional methods. For the rational engineering of the CYP specificity, the methodology of homology modeling should be used to obtain a basic structural understanding [4]. Since X-ray crystallography and nuclear magnetic resonance (NMR) methodologies are very expensive and time-consuming techniques for finding protein structures, bioinformatics, as demonstrated in this study, could play a major role in *in silico* analyses of biological queries.

The invariant residues in most CYPs are glutamic acid and arginine, which are located within the conserved ExxR motif found in the K-helix [21]. Both glutamic acid and arginine, and either a histidine, arginine, or asparagine from the meander region, generate a salt–bridge interaction that helps to form the final tertiary structure of the CYP, suggesting that a CYP polypeptide connects the ExxR to a heme-binding complex [7]. The ExxR motif is not necessary in all CYP tertiary structures, and just a single cysteine residue, which coordinates as the fifth thiolate ligand to the P450 heme iron, is constant in all CYP structures [22]. Completely improper and inactive protein isoforms were observed after site-directed mutagenesis of cysteine, glutamic acid, or arginine individually [8, 26].

Coumarins form the largest group of secondary metabolites...
and are found mainly in plants, with lesser quantities found in bacteria [15]. Previous research studies have been carried out regarding the dealkylating activities of CYP enzymes based on O-dealkylation of 7-ethoxycoumarin with high efficiency [27, 28, 31, 32] (Fig. 1). In this study, we found a unique CYP subfamily (CYP157C4) of *Streptomyces peucetius* without the ExxR motif that was considered compulsory in all CYPs. As very few CYP157C4-like proteins existed prior to this study, we characterized CYP157C4 by using homology modeling and an in vitro enzyme assay; we used 7-ethoxycoumarin for the assay as it is a generally used substrate for investigating the functions of most CYPs [28]. Our findings could help facilitate the identification of the actual substrate and the crystallization of this CYP type (CYP157C4).

**Materials and Methods**

**Construction of the Homology Model**

Homology modeling was used to build the model of CYP157C4. The protein is 476 amino acids long and the gene bank accession number is AJ605537. Homology modeling and molecular docking were carried out in the Discovery Studio 3.5 (DS 3.5; Accelrys, San Diego, USA) [2, 6]. The sequences were retrieved from the NCBI (National Center for Biotechnology Information) website, and the searching of template was made against the protein data bank (PDB) using the ExPASy server of PSI-BLAST (Comparison matrix, BLOSUM 62; E-threshold, 10) [30]. Three protein templates were used for the construction of our model. The proteins were CYP154C1 from *S. coelicolor* A3 (2) [12], CYP164A2 from *Mycobacterium smegmatis* str. MC2 155 [1], and CYP107L1 from *S. venezuelae* [25], and their PDB IDs are 1GW1, 3R9C, and 2BVJ, respectively. The reasons behind choosing these three proteins are that they have a well-known crystal structure, and multi-template models may produce better structure understanding than a single-template model. The MODELER computer program was used to generate the protein model [23]. Heme coordinates were obtained from 2BVJ and positioned in the template. The model structure was refined by minimizing the energy used for the molecular docking undertaken with DS 3.5 [5]. In terms of the refined model, the Accelrys CHARMM forcefield [13] was used for the simulation, and the ProSA 2003 z-score [29] and Ramachandran plot were used for the validation [13]. SUPERPOSE was used for a statistical analysis of the template-based predicted model [14].

**Protein and Ligand Interaction**

For the docking of the 7-ethoxycoumarin, the DS 3.5 package was used with the refined model. We used a sketch toolbar application to draw the substrate structure, ensuring that we produced a stereochemically correct configuration, and used the Deriding-like forcefield in DS 3.5 to optimize the structure. The optimized ligand molecule was docked into the refined protein model using the DS 3.5 function “LigandFit” [33] and residues in active sites were identified. For a better understanding of the protein-ligand interactions, 60 poses of the ligand were generated and the DS scoring function was used for score calculation.

**Bacterial Strains, Plasmids, and Growth Conditions**

Luria-Bertani (LB) agar plates or LB broth at 37°C were used to culture the *E. coli* strains [17]. *E. coli* XL1-Blue MRF’ (Stratagene, La Jolla, CA, USA) was used for DNA manipulation, and *E. coli* BL21 (DE3) was used as the overexpression host. Ampicillin measuring 100 μg/ml was the concentration selection marker that was used to select the recombinant strains of the *E. coli* transformants, whereas 0.4 mM isopropyl-β-thiogalactopyranoside (IPTG) and 40 μl of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside were used for the blue-white screening of the transformants. PGEM-T Easy (Promega, USA) was used as the cloning vector, and pET32a(+) was used as the overexpression vector (Novagen, USA). All of the chemicals were purchased from Sigma Co., and standard protocols were followed during all DNA manipulations, including ligations and enzyme digestions [11, 24].

**Cloning and Overexpression of CYP157C4**

A polymerase chain reaction (PCR) was performed in a thermocycler (Takara, Japan). The amplification conditions were 94°C for 7 min, followed by 30 cycles of each of the following conditions: 94°C for 1 min, 60°C for 1 min, 72°C for 1 min, and 72°C for 10 min. The total PCR mixture consisted of 20 μl containing a 5 μl PCR Mix (Genotech, Korea), 1 μl forward primer, 1 μl reverse primer, 1 μl template DNA, 2 μl dimethyl sulphoxide, and 10 μl distilled water. We used 5′-CCG AAT TCG ATC ACC ACC CCC GAG CCC TCC GTG-3′ (EcOR) and 5′-AAG CTT TCA GCC ACC CCG GTG GAA CAG GGA CCT-3′ (HindIII) as the forward primer and reverse primer, respectively. The computer-based programs BLAST, FASTA, ClustalW, and GeneDoc were used for the analysis and comparison of the nucleotide and amino acid sequences. After the PCR product was purified, we ligated it.
with a T-vector for DNA amplification prior to its transformation into E. coli XL1 blue. After digestion with the selected restriction enzymes, ligation was repeated with pET32a(+) and the product was transformed into E. coli BL21(DE3). The E. coli transformants were spread on an LB plate containing an appropriate amount of ampicillin, and the plates were then maintained at 37°C overnight.

After 1 ml of seed culture was taken, it was inoculated into a 250 ml flask containing 50 ml of LB broth, with ampicillin serving as an antibiotic. The culture was grown at 37°C until the cell density reached 0.6 at OD600. Then, 250 mM FeCl₃, and 1 mM 5-α-aminoluvulinic acid were added, followed by a temperature decrease to 20°C. After 30 min, 1 mM of IPTG was added for the expression at 200 rpm. After 30 h, the cell pellets were harvested and centrifuged at 3,500 rpm for 15 min, washed twice using Tris-HCl buffer with 10% glycerol (pH7.4), and finally mixed with 1 ml of respective buffer. We used an ultrasonicator to lyse the cells and 12,000 rpm centrifugation, for 25 min at 4°C, to separate the soluble protein. We used nickel resin to further purify the soluble protein and conducted an analysis with 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

**CO-Reduction Assay**

The CO-reduction assay of CYP157C4 was conducted by following a pre-existent protocol [16]. In accordance with this method, the protein was diluted in 50 mM potassium phosphate buffer containing 10% glycerol. A small amount of sodium dithionite was added, followed by a division of the dilution into two cuvettes; one cuvette was used as a reference and the other was saturated with 50–60 bubbles of CO at a rate of 1 bubble per second. The dilution was then scanned between 400 and 490 nm at room temperature using a Shimadzu 1601PC spectrophotometer. The CYP content was measured according to the difference in the absorbance values at 450 and 490 nm, and an extinction coefficient of 91 mm⁻¹cm⁻¹ was used.

**In Vitro Assay of 7-Ethoxycoumarin Conversion into Hydroxylated Product by CYP157C4**

The in vitro assay of the CYP157C4 was performed using 7-ethoxycoumarin as a substrate. The reaction mixture consisted of 1 µM CYP157C4, 5 µM FDX (spinach ferredoxin), 1 µM FDR (spinach ferredoxin reductase), 4 mM glucose-6-phosphate, 5 U glucose-6-phosphate dehydrogenase, 10 mM MgCl₂, 350 µM 7-ethoxycoumarin, and 50 mM sodium phosphate buffer. The reaction was initiated by adding 250 µM NADPH, followed by incubation at 30°C for 2 h. The reaction mixing process was then stopped to perform an extraction with double volumes of ethyl acetate. The extracted sample was dried under nitrogen gas before methanol was used in further mixing. High performance liquid chromatography (HPLC) and mass spectrometry (MS) were used to analyze the final product.

**Results**

**Template Alignment and CYP157C4 Sequence Analysis**

A multiple sequence alignment was used to check the sequence conservation and signature motifs of the CYP57C4, and a PSI-BLAST search (NCBI server) was used for finding the homologous protein sequences. The enzyme activity, organism, chain length, percent sequence identity, and E-value for the homology of CYP157C4 compared with other CYPs are presented in Table 1. The Align3D tool was used to complete the sequence alignment among all of the PDB templates shown in Table 1, and the Align2D tool was used to complete the alignment between the CYP157C4 sequence and the pre-aligned templates (Fig. 2A). The sequence identities between the target and the template structures 1GWI, 3R9C, and 2BVJ were 31%, 28%, and 29%, respectively, and were reasonable for model building, CYP157C4 conserved the various signature motifs. The structural motifs of CYP157C4 (from the N-terminal region of the protein) are pentapeptides I-helix (AxxET), heptapeptides K-helix (ALxQxxW) (Fig. 2B), and L-helix (xxxxRxaxxxxxxL), and the characteristic signature motif for the CYP subfamily is FxxGxFxxCxxxxR. The signature motif showed a conserved cysteine residue that ligated to the Fe heme. Permanently bound heme with protein helps in binding the compounds that can bind divergent iron (Fe). The bound molecules can modulate the function of CYPs. These structural and characteristic reserve motifs verified the authenticity of the CYP157C4 model construction based on sequence alignments.

**Table 1.** Template alignment and sequence analysis of CYP157C4.

<table>
<thead>
<tr>
<th>Protein</th>
<th>PDB ID</th>
<th>Enzyme activity</th>
<th>Organism</th>
<th>Sequence identity (%)</th>
<th>Chain length</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP154C1</td>
<td>1GW1</td>
<td>Oxidoreductase</td>
<td><em>Streptomyces coelicolor</em> A3(2)</td>
<td>31</td>
<td>411</td>
<td>4.71972e-38</td>
</tr>
<tr>
<td>CYP164A2</td>
<td>3R9C</td>
<td>Oxidoreductase</td>
<td><em>Mycobacterium smegmatis</em> str. MC2 155</td>
<td>28</td>
<td>418</td>
<td>2.83952e-26</td>
</tr>
<tr>
<td>CYP107L1</td>
<td>2BVJ</td>
<td>Oxidoreductase</td>
<td><em>Streptomyces venezuelae</em></td>
<td>29</td>
<td>436</td>
<td>1.07277e-25</td>
</tr>
</tbody>
</table>

The sequence conservation and signature motifs of CYP157C4 were examined using multiple sequence alignment (MSA). The templates for the MSA were obtained by scanning the protein sequence of CYP157C4 against 3D structures that were deposited in the protein data bank with a BLAST search (NCBI server).
Homology Models and Validation

The structural templates with the most homologous sequences regarding CYP157C4 were used for comparative homology modeling. The templates obtained through two-dimensional (2D) alignment were a compatible match with CYP157C4 in terms of both alignment and conserved residues. The conserved residues shown in the dark-blue color (identity, 39.3%; similarity, 55.2%) and the 3D CYP157C4 model were generated using the MODELER module in DS 3.5 (Fig. 3). We obtained the backbone root-mean-square deviation (RMSD) for the model after superimposing CYP157C4 onto the most-homologous template CYP157C1 (1GW1, chain A, red), and the template crystal structure was finally 1.92 Å (Fig. 4A). The Ramachandran plot (\(\phi/\psi\)) distribution of the backbone conformation angles for each of the residues of the refined structure revealed that 84.9%, 9.2%, and 5.9% were in the favored region, allowed region, and outlier region, respectively (Fig. 4B). We therefore considered our optimized model as satisfactory and reliable for the remainder of the study.

Structural Features of the CYP157C4 Model and Ligand Docking

The CYP157C4 model consists of helices A–L, five beta sheets, and the connecting loops that link the two of them to each other. A comparison with other templates showed
that our CYP model contains almost every fold and the following structurally conserved regions (SCR) [7]: helices E, I, K, and L. CYP157C4 also comprises structurally variable regions (SVR), which are also known as “plastic regions.” The locations of the heme are between the K-helix, L-helix, and the N-terminal loop. The highly conserved and functionally important residues that are found mostly in CYPs are existent in CYP157C4; the consensus sequence F(G/S)xGx(H/R)xCxGxx(I/L/F)A(FSxGxHxCxGxxIG) in CYP157C4 contains the Cys responsible for heme binding.

By using the LigandFit method, we docked ligands into an active site with a shape filter and generated Monte Carlo algorithms for ligand conformation. The docked poses were optically minimized using CHARMm and evaluated with a set of scoring functions. The 7-ethoxycoumarin was bound almost entirely via hydrophobic interactions with amino acid side-chains. The residues whose side-chain faces the substrate-binding pocket are CYS115, HIS123, LEU254, VAL257, THR262, LEU305, THR309, and SER408 (Fig. 5). It is worth noting that almost all of the residues involved in the binding of 7-ethoxycoumarin were located in the substrate recognition sites (SRSs) [9]. We generated 60 poses and selected the best-suited docking modes for 7-ethoxycoumarin. Our familiarity with the dealkylation sites of the 7-ethoxycoumarin allowed us to select the lowest energy conformation by considering the distance between the heme and a 7-ethoxycoumarin dealkylation site with a high consensus score. We also generated poses wherein allylic C7 faced the heme with a distance of 4.564 Å (Fig. 5).

Cloning and Overexpression of CYP157C4 and CO-Reduction Assay

Heterologous overexpression of CYP157C4 in E. coli BL21 (DE3) presented a considerable yield of target protein. Using the SDS-PAGE tool, we found the actual size of the CYP157C4 to be about 51 kDa (Fig. 6A). A cytosolic fraction that was obtained from the heterologously

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig4}
\caption{Validation of the model. (A) Superimposition of the CYP157C4 model (blue) over CYP154C1 (1GW1, red). (B) Ramachandran plot (\(\phi/\psi\)) distribution. The most favored regions are represented inside the blue line, additional allowed regions are inside the pink line, and outlier regions are white.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig5}
\caption{Structure of the amino acid surrounding the ligands. Allylic C7 (yellow) of the ligand is facing the heme within the active site. The heme is shown as a black circle. The green line represents the distance between the heme and the carbon.}
\end{figure}
expressed CYP157C4 exhibited a CO-reduced difference maximum at 450 nm. The shift of about 30 nm observed in CYP157C4 represents a significant perturbation of the electron-density distribution of the heme (Fig. 6B).

**Biotransformation of 7-Ethoxycoumarin**

In a bacterial CYP system, the auxiliary protein of ferredoxin reductase and ferredoxin are needed to couple the electron flow from NAD(P)H to the terminal CYP components; as with P450CAM, bacterial class I CYP receives electrons from ferredoxin and NADH ferredoxin reductase [10, 20]. The probable activity of CYP157C4 was detected by HPLC/LC-MS. After the injection of the reaction product, the HPLC showed a possible dealkylated peak at a retention time of 9.9 min (Fig. 7A). The LC-MS analysis of the peak fairly indicated a monohydroxylated product with a mass of 163.0397 in positive mode (Fig. 7B).

**Discussion**

The absence of an ExxR motif in other CYPs led to the misfolding and inactive isoforms [8, 26]. Although this CYP series lack the ExxR motif, they showed the proper folding and active form [22]. In this study, the CYP157C4 model was built using DS 3.5 and was validated with the appropriate programs for stereochemical and amino acid positions. The active site was studied by docking with a 7-ethoxycoumarin substrate, and the discovery of Q332VLW335 instead of the ExxR motif in CYP157C4 disclosed a unique subfamily type. Although the exact biological function of CYP157C4 is still unknown, a comparison with the genetic arrangements of other CYPs from the same *S. peucetius* organism provides a minor clue. The endogenous substrate...
of CYP157C4 is still a matter of study; we used 7-ethoxycoumarin as a model substrate and observed an O-dealkylation of the 7-ethoxycoumarin in an in vitro assay. Our results demonstrated that the absence of the ExxR motif may not be a critical factor for CYP activity.

The characterization of this type of CYP enzyme revealed a remarkable diversity of protein-arrangement types, and this can be further used for electron transport and catalysis. Owing to a wide variety of chemical-reaction mechanisms, the substrate preferences of different CYPs are unpredictable. A successful development of the crystal structure of the CYP enzyme may assist with finding the actual substrate of this enzyme. Furthermore, the data obtained from our modeling and docking study of CYP157C4 may be useful in the rational design of regioselective proteins, which in turn assists with the conversion of natural products into medicinally important drugs.

Acknowledgments

This work was carried out with the support of “Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ010457052015)” Rural Development Administration, Republic of Korea.

References