J. Microbiol. Biotechnol. (2017), 27(8), 1472–1482
https://doi.org/10.4014/jmb.1706.06013

jmb

Crystal Structure and Functional Characterization of a Cytochrome P450 (*Ba*CYP106A2) from *Bacillus* sp. PAMC 23377

Ki-Hwa Kim[1†], Chang Woo Lee[2,3†], Bikash Dangi[1], Sun-Ha Park[2], Hyun Park[2,3], Tae-Jin Oh[1,4*], and Jun Hyuck Lee[2,3*]

[1]*Department of Life Science and Biochemical Engineering, Sunmoon University, Asan 31460, Republic of Korea*
[2]*Unit of Polar Genomics, Korea Polar Research Institute, Incheon 21990, Republic of Korea*
[3]*Department of Polar Sciences, University of Science and Technology, Incheon 21990, Republic of Korea*
[4]*Department of BT-Convergent Pharmaceutical Engineering, Sunmoon University, Asan 31460, Republic of Korea*

Received: June 7, 2017
Revised: June 17, 2017
Accepted: June 19, 2017

First published online
June 21, 2017

*Corresponding authors
T.-J.O.
Phone: +82-41-530-2677;
Fax: +82-41-530-2279;
E-mail: tjoh3782@sunmoon.ac.kr
J.H.L.
Phone: +82-32-760-5555;
Fax: +82-32-760-5509;
E-mail: junhyucklee@kopri.re.kr

†These authors contributed equally to this work.

pISSN 1017-7825, eISSN 1738-8872

Bacterial cytochrome P450 (CYP) steroid hydroxylases are effectively useful in the pharmaceutical industry for introducing hydroxyl groups to a wide range of steroids. We found a putative CYP steroid hydroxylase (*Ba*CYP106A2) from the bacterium *Bacillus* sp. PAMC 23377 isolated from Kara Sea of the Arctic Ocean, showing 94% sequence similarity with *Bm*CYP106A2 (*Bacillus megaterium* ATCC 13368). In this study, soluble *Ba*CYP106A2 was overexpressed to evaluate its substrate-binding activity. The substrate affinity (K_d value) to 4-androstenedione was 387 ± 37 μM. Moreover, the crystal structure of *Ba*CYP106A2 was determined at 2.7 Å resolution. Structural analysis suggested that the α8–α9 loop region of *Ba*CYP106A2 is intrinsically mobile and might be important for initial ligand binding. The hydroxyl activity of *Ba*CYP106A2 was identified using in vitro enzyme assays. Its activity was confirmed with two kinds of steroid substrates, 4-androstenedione and nandrolone, using chromatography and mass spectrometry methods. The main products were mono-hydroxylated compounds with high conversion yields. This is the second study on the structure of CYP106A steroid hydroxylases, and should contribute new insight into the interactions of bacterial CYP106A with steroid substrates, providing baseline data for studying the CYP106A steroid hydroxylase from the structural and enzymatic perspectives.

Keywords: *Bacillus* sp., crystal structure, cytochrome P450, steroid hydroxylase, X-ray crystallography

Introduction

Cytochrome P450s (CYPs) are well-known heme-containing proteins found in all forms of life. CYPs have been widely studied because of their ability to hydroxylate endogenous and xenobiotic chemicals such as steroids, lipids, and several secondary metabolite-related natural products [1–4]. To catalyze these oxidative processes, the electrons commonly derived from nicotinamide adenine dinucleotide (NAD(P)H) are transferred to bacterial CYPs through electron transfer partners acting in paired form, such as ferredoxin-ferredoxin reductase (FDX-FDR) [5, 6].

The steroids used as substrates for the hydroxylation reaction with CYPs play key roles in the biosynthesis of cholesterol, control of blood pressure, and drug metabolism in humans. They act as hormones, thereby regulating the secretion of sex hormones, and also show anti-inflammatory and anticancer activities [7–9]. The steroid hydroxylases are rare proteins among bacterial CYPs, with only a few such CYPs identified in various strains, including CYP106A1, CYP106A2, and CYP109E1 from *Bacillus megaterium* [9–11], CYP154C3 from *Streptomyces griseus* [12], CYP154C5 from *Nocardia farcinica* [13], and CYP260A1 and CYP260B1 from *Sorangium cellulosum* [14, 15]. In particular, CYP106A2 (*Bm*CYP106A2) from *B. megaterium* ATCC 13368, a member of the bacterial CYP106 family, was reported in the late 1970s [10], and has been shown to exhibit hydroxylation activity not only on various types of steroids, including

3-oxo-$\Delta^{2,4}$-steroids and 3-hydroxy-Δ^{5}-steroids, but also on terpenoids, including diterpenes and triterpenes [16]. CYP106A1 from *B. megaterium* DSM 319 was also reported as a steroid hydroxylase with an activity profile different from that of CYP106A2 [9, 16]. To date, these two CYP106 forms have been reported from only *B. megaterium*, suggesting that members of the CYP106 enzyme family are likely present in other *Bacillus* strains with high activity against several steroids. Recently, some researchers have focused on engineering methods of CYP106 for making more hydroxylated steroids with pharmaceutical significance for regioselectivity [17, 18].

Identification of the protein structure can help to determine the active site and its interaction with various substrates. The crystal structures of free-form and complex-form CYP109E1 and CYP260B1 with steroids have been determined, and the recent identification of the crystal structure of *Bm*CYP106A2 represents the first structural determination of an enzyme in the CYP106 protein family, which now allows for structural interpretation [11, 15, 19]. In the present study, the strain *Bacillus* sp. PAMC 23377 was isolated from Kara Sea of the Arctic Ocean, and we found a new CYP106A2 (*Ba*CYP106A2) with high sequence identity (94%) to *Bm*CYP106A2 at the amino acid level. We further identified the crystal structure of *Ba*CYP106A2 and compared it with the previously determined *Bm*CYP106A2 structure [19]. Biochemical activity analysis of *Ba*CYP106A2 was also examined to characterize its hydroxylation activity towards several steroids. These results can help us to identify the enzymatic reaction mechanism and substrate specificity of *Ba*CYP106A2 that can be applied to biotechnological engineering methods requiring steroid hydroxylation.

Materials and Methods

Materials and Solvents

The steroid substrates 4-androstenedione and nandrolone were purchased from Tokyo Chemical Industry Co., Ltd. (Japan), and isopropyl-1-thio-β-D-galactopyranoside (IPTG), 1,4-dithiothreitol, and kanamycin were obtained from Duchefa Biochemie (Korea). 5α-Aminolevulinic acid (ALA), NADPH, spinach FDX, and spinach FDR were obtained from Sigma-Aldrich (USA), and the restriction enzymes, T4 DNA ligase, DNA polymerase, and dNTPs were purchased from Takara Clontech (Korea); all other chemicals and solvents were high-grade products obtained from commercially available sources.

Cloning, Overexpression, and Purification of *Ba*CYP106A2

The *Ba*CYP106A2 gene is 1,236 bp in length, encoding a 411-amino-acid protein. The gene was amplified by polymerase chain reaction (PCR) using a pair of primers: forward primer, 5′-GTCGACTCATGAAAGAAGTTATAGCAA-3′ (*Sal*I sites underlined); and reverse primer, 5′-AAGCTTTTATAGTGTACGGCATGCC-3′ (*Hin*dIII sites underlined). Genomic DNA from *Bacillus* sp. PAMC 23377 was PCR-amplified through 25 cycles after initial denaturation at 94°C for 5 min. One cycle consisted of 95°C for 30 sec, 60°C for 30 sec, 72°C for 1.5 min, and 72°C for 5 min. After the PCR products were digested with *Sal*I and *Hin*dIII, they were ligated into the pET28a(+) vector and integrated into *Escherichia coli* C41(DE3) host cells for overexpression. Overexpression was carried out by the addition of 1 mM ALA and 0.5 mM $FeCl_3$ in Luria-Bertani medium after the culture reached an optical density of 0.6–0.8 at 600 nm. After a 30-min incubation at 20°C, 0.5 mM IPTG was added for induction. The cultures were incubated for 70 h at 20°C with shaking at 200 rpm. After the cultures were harvested, the samples were washed twice with 50 mM potassium phosphate buffer (pH 7.4) and the cells were dissolved with ultrasonic treatment. To obtain the soluble protein, the samples were centrifuged at 13,000 ×*g* for 20 min at 4°C and mixed with metal affinity Ni^{2+}-NTA resin (Qiagen, Germany). The elution buffer was 50 mM potassium phosphate buffer (pH 7.4) containing 10, 100, or 500 mM imidazole per buffer. The protein was purified by trapping on an Amicon Ultra centrifugal filter (Ultracel-3K; Millipore, USA) equilibrated with 50 mM potassium phosphate buffer (pH 7.4). The enzyme concentration was calculated by the difference in the UV spectra at 450 and 490 nm. The sample was prepared in a saturated state to carbon monoxide reduction, using the molar extinction coefficient of 91 $mM^{-1}cm^{-1}$ [20].

Substrate-Binding Activity

A *Ba*CYP106A2 concentration of 2 μM was used in the entire substrate-binding assay in 50 mM potassium phosphate buffer (pH 7.4). The substrates were prepared in dimethyl sulfoxide and titrated until reaching saturation (0–1,700 μM). The absorbance of all samples was measured using a Biochrome Libra S35PC UV/visible spectrophotometer in the range of 350–490 nm. The absorbance difference (A_{390} - A_{420}) was fit to a quadratic equation and calculated using GraphPad 7.01 software [19, 21], where A_{obs} is the absorbance difference at any ligand concentration, A_{max} is the absorbance difference at ligand saturation, S is the substrate concentration, E_t is the enzyme concentration, and K_d is the dissociation constant for the enzyme–ligand complex.

Crystallization and Data Collection

Initial crystallization screening in a 96-well sitting-drop plate (Emerald Bio, USA) was carried out using a mosquito crystallization robot (TTP LabTech, UK) with commercial crystallization screening kits (MCSG I-IV; Microlytic USA; Wizard Classic I-IV; Emerald Bio). Crystallization droplets were mixed with a protein solution (0.2 μl) and reservoir solution (0.2 μl), equilibrated against 80 μl reservoir solution at 293 K. Optimal crystals were obtained from a condition of 0.1 M sodium cacodylate:HCl (pH 6.5), 1.26 M ammonium sulfate (MCSG2 #A6). A single crystal was harvested

and soaked into Paratone-N oil (Hampton Research, USA) for cryoprotection from a nitrogen-gas stream. X-ray diffraction data containing 120 images were collected on a BL-5C beam line of the Pohang Accelerator Laboratory (PAL, Pohang, Korea). The dataset at a resolution of 2.7 Å was indexed, integrated, and scaled using the HKL-2000 program [22]. The statistics of data collection are given in Table 1.

Structure Determination and Refinement

The structure was solved by molecular replacement using CYP106A2 from *B. megaterium* ATCC 13368 (PDB code 4YT3) as a search model, with the MOLREP program from the CCP4 suite [23, 24]. Iterative rounds of structure rebuilding and refinement were carried out using the Coot program and REFMAC5. The program phenix.refine from PHENIX was used for further refinement [25, 26]. The final model had R_{work} and R_{free} values of 19.1% and 25.8%, respectively, and the model was validated using the MolProbity program [27]. The detailed refinement statistics are presented in Table 1. All structural figures were prepared using PyMOL [28]. Atomic coordinates and structure factor data of *Ba*CYP106A2 have been deposited in the Protein Data Bank [29] with the accession code 5XNT.

In Vitro Assay and Instrumental Analysis

For the in vitro assay, a final volume of 250 µl was constituted with 10 µM of *Ba*CYP106A2, 25 µg of FDX, 0.1 U of FDR, 100 µg/ml of catalase, an NADPH-regeneration system (10 mM glucose-6-phosphate, 1 U glucose-6-phosphate dehydrogenase, and 5 mM $MgCl_2$), and 100 µM of steroid substrates in 50 mM potassium phosphate buffer (pH 7.4). The reaction was initiated when 1 mM of NADPH was added, and proceeded for 2 h at 180 rpm and 30°C. When the reaction was terminated, the sample was extracted twice with 500 µl of ethyl acetate and centrifuged at 12,000 ×*g* for 10 min. The separated supernatants were dried and then dissolved in the HPLC solvent, acetonitrile and water (60:40), for HPLC and LC-MS. The products were analyzed by the Dionex Ultimate 3000 UHPLC$^+$ system (Thermo Fisher Scientific, Germany) using a Mightysil reverse-phase C_{18} GP column (4.6 × 250 mm, 5 µm). The configuration of the HPLC system was an LPG-3400SD pump, ACC-3000 autosampler column compartment, and DAD-3000 diode array detector. The mobile phase consisted of solution A (in HPLC-grade water) and solution B (in HPLC-grade acetonitrile). The flow rate was maintained at 1.0 ml/min, and the oven temperature was kept at 30°C. To analyze the products, the gradient system was operated in the following conditions: percentages of solution B were increased from 15% to 50% (0–10 min), and then increased to 70% (10–20 min), kept at 70% (20–25 min), and decreased to 15% (25–30 min). UV detection was performed at 245 nm to confirm the substrates and their products. The same reaction samples were used to measure molecular weight with LC-electrospray quadrupole-time-of-flight MS (ACQUITY UPLC) connected to a SYNAPT G2-S system (Waters, USA) in positive-ion mode. The mobile phase solution was the same as that used for HPLC.

Results and Discussion

Protein Overexpression, Purification, and Spectral Characterization of *Ba*CYP106A2

The 411-amino-acid sequence of *Ba*CYP106A2 (G+C content of 42%) showed high identity (~94%) with *Bm*CYP106A2 [30] (data not shown). In our previous study, we found that heterologous overexpression of a gene in *E. coli* C41(DE3) allowed for efficient gene insertion and induction of a large amount of a specific protein production [31, 32]. Thus, in the present study, the overexpressed protein was purified through Ni^{2+} resin, and 15% sodium dodecyl sulfate-polyacrylamide gel electrophoresis revealed a molecular mass of about 48 kDa (Fig. 1A). In addition, the purified protein concentration was calculated as 450.55 mg/ml. The iron contained in *Ba*CYP106A2 could be reduced by sodium dithionite, and the addition of carbon monoxide resulted in a shift in the ultraviolet (UV) spectra, with a peak shift from 450 nm to 420 nm (Fig. 1B).

Substrate-Binding Activity of *Ba*CYP106A2

The concentration of *Ba*CYP106A2 was calculated by determining the difference in absorbance at 450 nm and 490 nm; a reference sample was reduced with dithionite, while the target sample was reduced with carbon monoxide and dithionite (Fig. 2A). Substrate binding to the active site

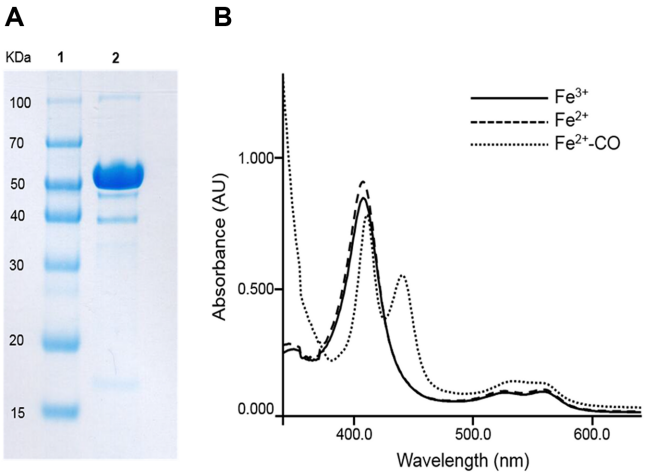

Fig. 1. Characterization of the cytochrome P450 (*Ba*CYP106A2) purified from *Bacillus* sp. PAMC 23377.
(**A**) Purified *Ba*CYP106A2 protein in 15% sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Lane 1, marker; Lane 2, *Ba*CYP106A2. (**B**) Ultraviolet spectra of three *Ba*CYP106A2 forms: solid line, oxidized form of *Ba*CYP106A2; long dotted line, reduced form by the addition of sodium dithionite; and short dotted line, addition of carbon monoxide that shifted the spectrum with a peak at 420 nm.

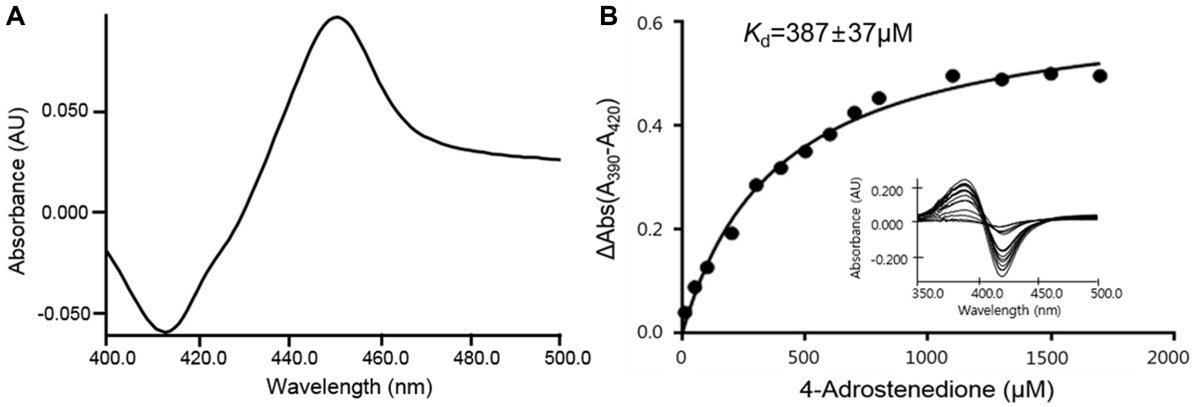

Fig. 2. Substrate binding activity of the cytochrome P450 (*Ba*CYP106A2) purified from *Bacillus* sp. PAMC 23377.
(**A**) Carbon monoxide-reduction spectrum of *Ba*CYP106A2. (**B**) Type I shift spectra with increasing concentrations of the substrate 4-adrostenedione, and the corresponding K_d values. The Type I spectra showed a peak at 390 nm and a trough at 420 nm. This analysis was based on a substrate concentration ranging from 0 to 1,700 μM containing 10 μM *Ba*CYP106A2 in 50 mM potassium buffer (pH 7.4). The binding activity was calculated as the difference in binding affinity between 390 nm and 420 nm.

of CYP induces the high spin state called a type I shift, with peak absorbance at 390 nm and the minimal absorbance detected at 420 nm [33]. This characteristic of CYP is essential in substrate screening for putative enzymes. However, this general property does not necessarily indicate that all substrates exhibit a high spin state. For example, the steroid substrate nandrolone was reported to not display this type I shift despite showing activity against CYP106A2 [16]. Therefore, in the present study, we conducted the binding assay with 4-androstenedione. The equilibrium dissociation constant (K_d value) can be obtained by expressing the affinity with the substrate as a numerical value; the K_d value of 4-androstenedione was calculated to be 387 ± 37 μM (Fig. 2B). Notably, this K_d value was ~5-fold higher than those of other CYP106A proteins (CYP106A1 from *B. megaterium* DSM319 (K_d = 77 ± 2 μM) and CYP106A2 from *B. megaterium* strain ATCC 13368 (K_d = 81 ± 10 μM)) [34, 35].

Overall Structure of *Ba*CYP106A2

The crystal structure of *Ba*CYP106A2 was determined at a resolution of 2.7 Å using the molecular replacement method. The crystal structure of *Bm*CYP106A2 (PDB code 4YT3) was used as a search model [19]. *Ba*CYP106A2 showed high sequence similarity (94% sequence identity) and structural similarity (DALI z-score of 59.6) with *Bm*CYP106A2. The *Ba*CYP106A2 crystal belonged to space group $P6_122$ and contained one protomer in the asymmetric unit. The final structural model included 392 amino acids, 31 water molecules, and one heme cofactor with bound iron. The N-terminal region (residues 1–6) and the loop region (residues 68–80) between α3 and α4 were disordered because the electron density was too weak to be detected for building a structural model in those regions. The crystal structure of *Ba*CYP106A2 displayed the canonical fold of the helix-rich heme-dependent CYP monooxygenase, composed of 15 α-helices and 8 β-strands with tightly bound heme molecules (Fig. 3). A structural homology search using the DALI server [36] showed that *Ba*CYP106A2 has high structural similarities with CYP106A2 from *B. megaterium* (PDB code 4YT3) [19], CYP109E1 from *B. megaterium* (PDB code 5L91) [11], EryK from *Saccharopolyspora erythraea* (PDB code 2WIO) [37], CYP109B1 from *Bacillus subtilis* (PDB code 4RM4) [38], and CYP epoxidase from *Streptomyces antibioticus* (PDB code 4XE3) [39] (Table 2).

Active Site of *Ba*CYP106A2

In the *Ba*CYP106A2 structure, the heme molecule is located between the α12 helix and α17 helix and interacts with several conserved residues. In detail, the O2D of heme interacts with NH1 of Arg100 and NE of His96. The O1D of heme makes a hydrogen bond with ND1 of His353. O1A and O2A of heme interact with NH2 and NH1 of Arg296, respectively. The central iron is coordinated with the side chain of Cys355 (2.46 Å) on the proximal side. The putative ligand-binding site of *Ba*CYP106A2 is composed of Val82, Leu239, Ala243, Phe289, Ile292, Leu294, and Ala395, forming a hydrophobic pocket (Fig. 4). Structural superposition of ligand-free *Ba*CYP106A2 and *Bm*CYP106A2 (PDB code 4YT3) showed a root-mean-square deviation value of 0.833 Å for

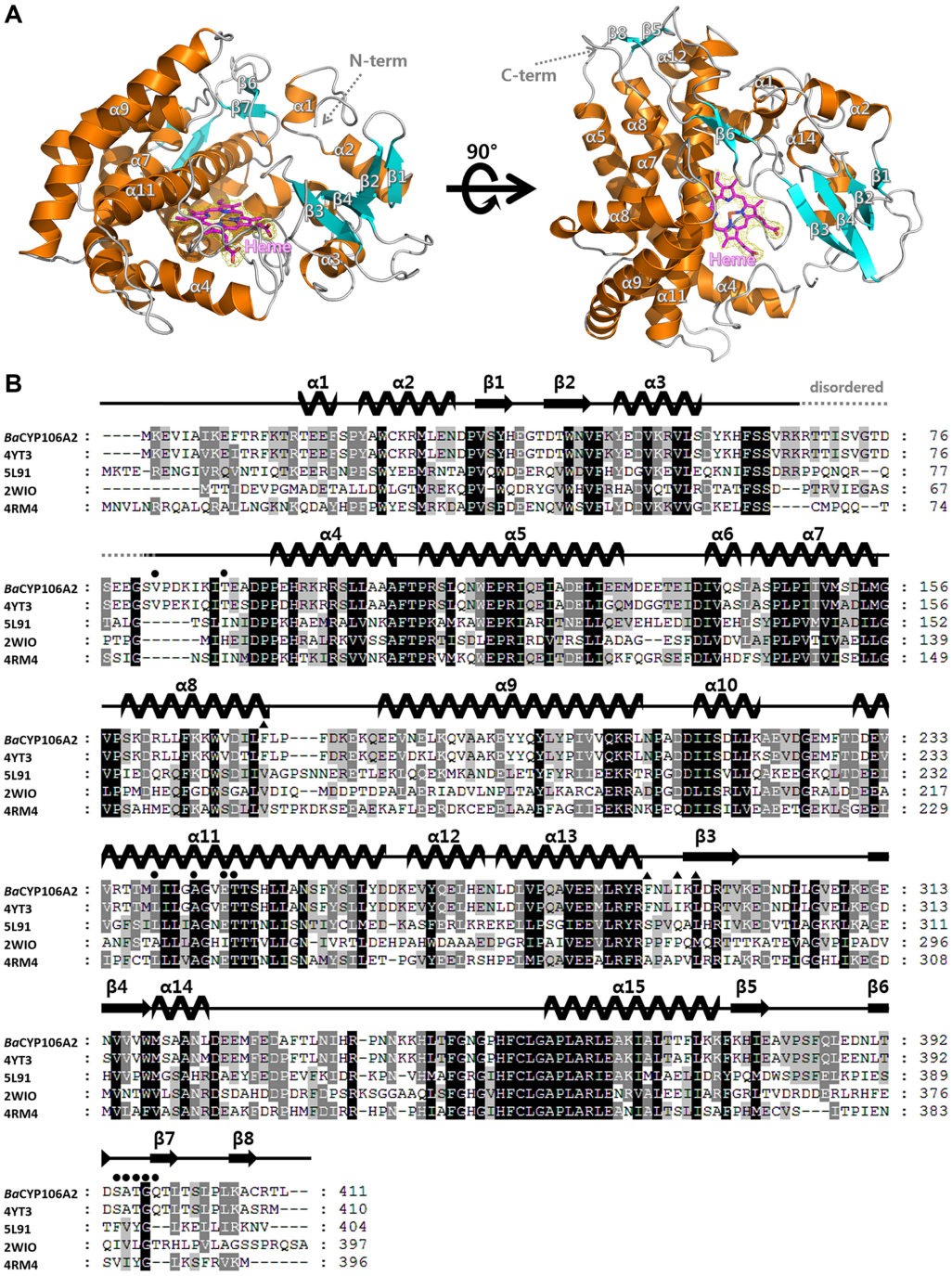

Fig. 3. Crystal structure of cytochrome P450 (*Ba*CYP106A2) from *Bacillus* sp. PAMC 23377 and multiple sequence alignment with other CYPs.

(**A**) Ribbon diagram showing the overall structure of the apo form of *Ba*CYP106A2. The orientation of the image on the right is rotated clockwise at 90° along the Y-axis relative to that on the left. The N- and C-termini are indicated. The bound heme molecule is labeled and represented by a magenta stick model. (**B**) Multiple sequence alignment between *Ba*CYP106A2 and representative structural homologous sequences used in a DALI search. The secondary structures obtained from the crystal structure of *Ba*CYP106A2 are shown above the aligned sequences, which include CYP106A2 from *B. megaterium* (PDB code 4YT3; UniProtKB code Q06069), CYP109E1 from *B. megaterium* (PDB code 5L91; UniProtKB code D5DKI8), EryK from *Saccharopolyspora erythraea* (PDB code 2WIO; UniProtKB code P48635), and CYP109B1 from *B. subtilis* (PDB code 4RM4; UniProtKB code O34374). The residues constituting putative substrate-binding pockets in *Ba*CYP106A2 are indicated by black circles. The Phe173 residue of *Bm*CYP106A2 is indicated by black triangles.

Table 1. Data collection and refinement statistics.

Dataset	*Ba*CYP106A2
X-ray source	PAL 5C beam line
Space group	$P6_122$
Wavelength (Å)	0.9795
Resolution (Å)	50.00–2.70 (2.75–2.70)
Total reflections	170,911
Unique reflections	14,704 (716)
Average I/σ (I)	40.2 (10.9)
R_{merge} [a]	0.123 (0.590)
Redundancy	11.6 (12.7)
Completeness (%)	99.4 (99.9)
Refinement	
Resolution range (Å)	48.68–2.70 (2.77–2.70)
No. of reflections of the working set	13,914 (1026)
No. of reflections of the test set	714 (43)
No. of amino acid residues	392
No. of water molecules	24
R_{cryst} [b]	0.191 (0.195)
R_{free} [c]	0.258 (0.351)
R.m.s. bond length (Å)	1.657
R.m.s. bond length (°)	0.0126
Average B value (Å^2) (protein)	58.703
Average B value (Å^2) (solvent)	41.971
PDB accession code	5XNT

[a] $R_{merge} = \Sigma |<I> - I| / \Sigma <I>$.

[b] $R_{cryst} = \Sigma ||Fo| - |Fc|| / \Sigma |Fo|$.

[c] R_{free} calculated with 5% of all reflections excluded from refinement stages using high-resolution data.

Values in parentheses refer to the highest resolution shells.

332 C_α atoms [19]. Although the two structures have high sequence similarity, there are clear structural differences in the α8–α9 and β6–β7 loop regions (Fig. 5A).

In the *Bm*CYP106A2 structure, the α8–α9 loop region is partially disordered, and the corresponding region displays high average B factors in *Ba*CYP106A2 (Fig. 5B). Structural comparisons between *Ba*CYP106A2 and abietic acid-complexed *Bm*CYP106A2 (PDB code 5IKI) showed the largest structural difference in the α8–α9 loop region by ligand binding (Fig. 5C) [19].

In the ligand-free *Ba*CYP106A2 structure, the α8–α9 loop region stretches outward from the substrate-binding site. However, in the abietic acid-complexed *Bm*CYP106A2 structure, the α8–α9 loop extends inward of the substrate-binding site. Notably, the Phe173 residue of *Bm*CYP106A2, located in the α8–α9 loop region, has hydrophobic interaction with abietic acid. This interaction likely results in re-orientations of the α8 helix and a consequent conformational change of the α8–α9 loop region. Similar changes at these regions have been observed in other CYPs. Another example is the conformational change found in the crystal structure of CYP105P2 from *Streptomyces peucetius* (PDB code 5IT1; 30% sequence identity with *Ba*CYP106A2) [31]. Upon binding of the biphenyl derivative, the α8 region loop of CYP105P2 went through distinctive conformational changes (Fig. 5D). These results confirmed that despite differences in the ligand types and locations of ligand binding, these CYP regions can be intrinsically mobile and altered by ligand interactions. In conclusion, establishment of our ligand-free *Ba*CYP106A2 structure enabled identification of the conformational changes that are likely responsible for the open-to-closed transition of the α8–α9 loop region during ligand binding in CYP106A2.

Table 2. Selected structural homologs of *Ba*CYP106A2 from a DALI search (DALI-Lite server[a]).

Protein	PDB code	DALI score	Ligand	Sequence % ID with *Ba*CYP106A2 (aligned residue number/total residue number)	Reference
CYP106A2 from *Bacillus megaterium*	4YT3	59.6	NL[b]	94% (385/388)	[19]
CYP109E1 from *Bacillus megaterium*	5L91	45.6	Corticosterone	41% (377/391)	[11]
Cytochrome P450 Eryk from *Streptomyces erythraea*	2WIO	43.1	NL[b]	31% (363/393)	[37]
CYP109B1 from *Bacillus subtilis*	4RM4	42.8	NL[b]	42% (357/364)	[38]
Cytochrome P450 epoxidase from *Streptomyces antibioticus*	4XE3	42.8	Clotrimazole	31% (370/394)	[39]
Cytochrome P450 MycG F286A mutant from *Micromonospora griseorubida*	3ZSN	42.7	Mycianmicin IV	32% (368/393)	[40]
Cytochrome P450 Vdh from *Pseudonocardia autotrophica*	3A4G	42.6	PEG 6000	31% (370/395)	[41]

[a] DaliLite ver. 3 server; http://ekhidna.biocenter.helsinki.fi/dali_server/.

[b] NL, no ligand molecule is bound.

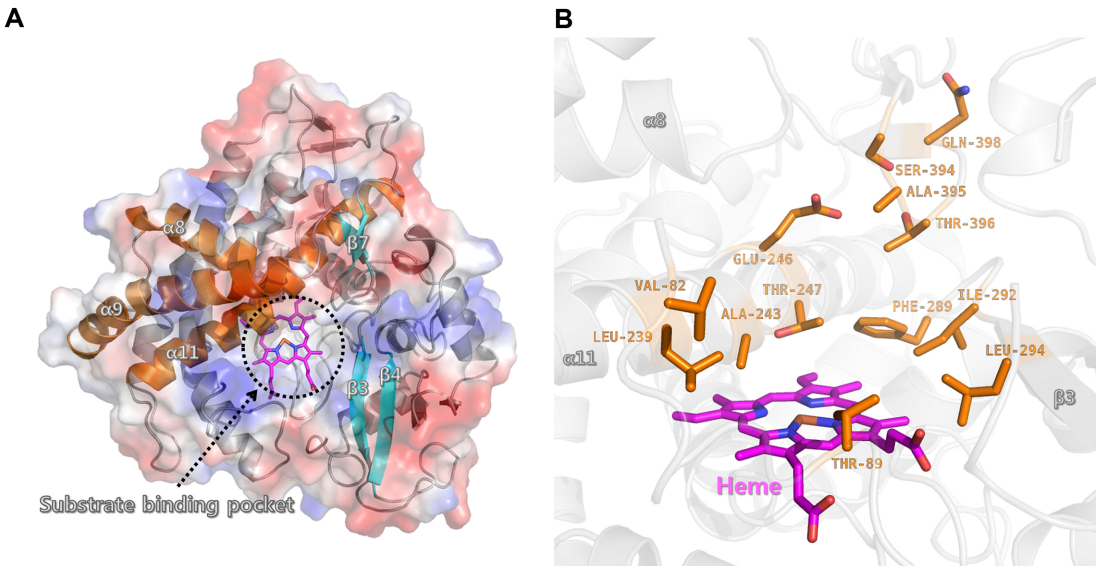

Fig. 4. Putative substrate-binding site of cytochrome P450 (*Ba*CYP106A2) from *Bacillus* sp. PAMC 23377.
(**A**) Semi-transparent electrostatic surface view of *Ba*CYP106A2. The putative substrate-binding pocket in *Ba*CYP106A2 is shown within the dotted circle, whereas the heme molecule is located at the bottom of this pocket. Notably, the putative substrate-binding pocket has hydrophobic surface electrostatic properties. (**B**) Close-up view of the putative substrate-binding pocket of *Ba*CYP106A2. The hydrophobic substrate pocket is generated by the Val82, Leu239, Ala243, Phe289, Ile292, Leu294, and Ala395 residues.

HPLC and LC-MS Analysis for In Vitro CYP106A2 Hydroxylase Activity

4-Androstenedione and nandrolone have been reported as steroidal substrates when bovine adrenodoxin and adrenodoxin reductase were used for supplying electrons to CYP from NAD(P)H [16]. CYP106A1, another steroid hydroxylating CYP showing 65% sequence similarity with *Bm*CYP106A2, is proposed as a potential CYP with industrial potential [10, 16]. Some studies have also demonstrated activity using electron transfer partners such as FDX and FDR or new electron transfer systems present in the bacterium itself [9, 42]. Therefore, in this study, we used FDX and FDR in an in vitro assay. Specifically, we sought to identify the function of *Ba*CYP106A2, which is similar to that of *Bm*CYP106A2, to determine whether there is enzymatic activity. For this purpose, we performed an in vitro assay with HPLC and LC-MS. HPLC analysis showed almost 100% conversion of 4-androstenedione as a substrate, where product P1 was found as a major peak along with some other minor product peaks. Analysis of the reaction mixture by LC-MS showed P1 with a mass value of m/z^+ $[M+H]^+$ 303.1958, for which the calculated mass for the molecular formula $C_{19}H_{27}O_3^+$ was 303.1960 (Fig. 6A). The S peak represents the actual structure and the P1 peak is a predicted structure based on the fact that 4-androstenedione is mainly hydroxylated at the 15β position [16]. Interestingly, other subfamily enzymes exhibited different regioselectivities toward 4-androstenedione. For instance, using adrenodoxin and adrenodoxin reductase as redox partners, CYP106A2 converted almost 100% of the substrate into two main products, 15β-hydroxyandrostenedione and 7β,15β-dihydroxyandrostenedione, after a 30-min reaction. In contrast, CYP106A1 produced hydroxylated metabolites at positions 6β and 7β [16].

Similarly, nandrolone also showed almost 100% conversion, where the major product P1 and some minor products peaks were found. P1 was monohydroxylated at m/z^+ $[M+H]^+$ 291.1975, for which the calculated mass for the molecular formula $C_{18}H_{27}O_3^+$ was 291.1960 (Fig. 6B). The position of hydroxylation for nandrolone has not been reported previously, and thus the possible position of its hydroxylation can now be suggested based on our results.

Conclusively, steroids have diverse properties with several potential applications, including for human health, such as anti-inflammatory and anticancer effects, and are therefore important resources in the pharmaceutical industry. From an industrial point of this view, it is inefficient to change the substituent with traditional synthetic methods. Therefore, there has been extensive research effort on the use of CYPs for conducting more effective reactions [43, 44]. However, the role of CYPs as steroid hydroxylases has been rarely investigated; two such CYPs, CYP106A1 and CYP106A2,

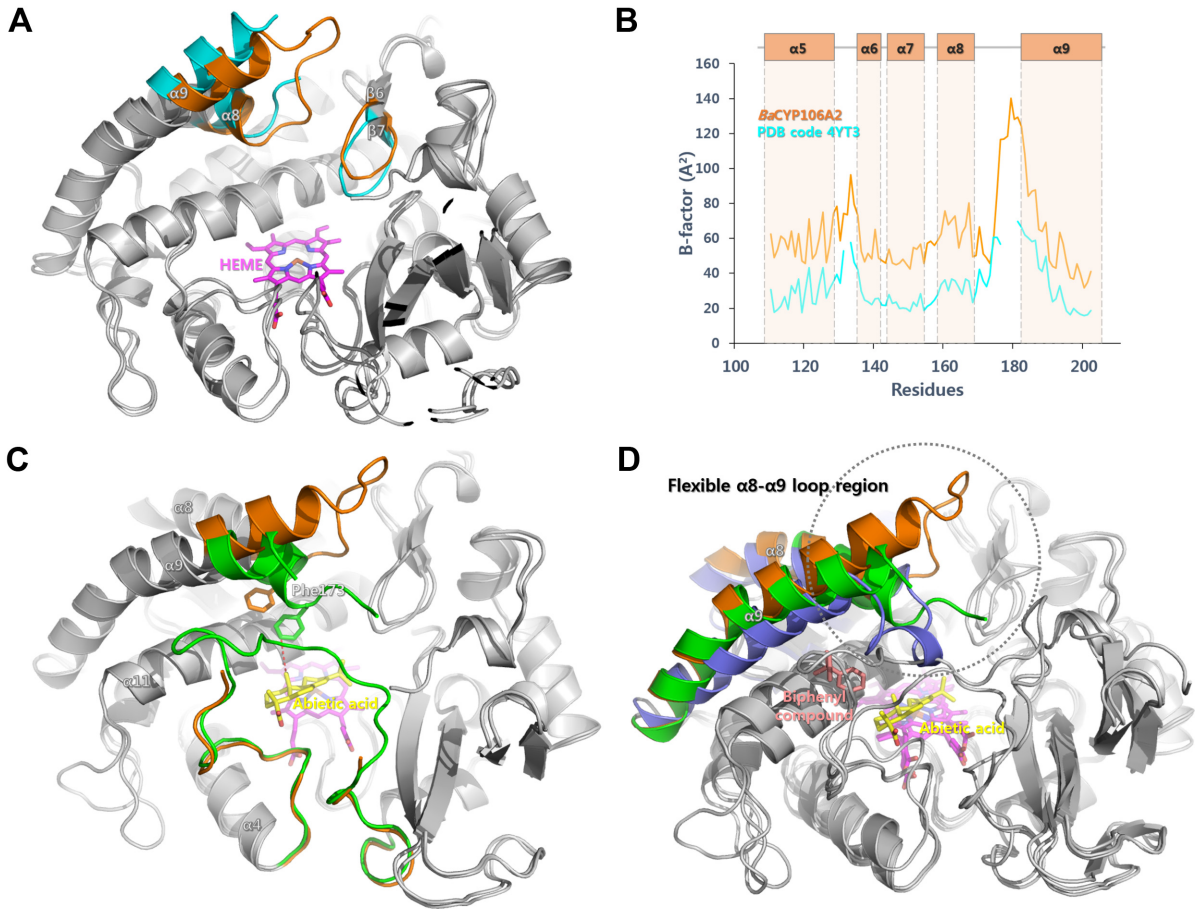

Fig. 5. Structures of the active site of cytochrome P450s from *Bacillus* sp. PAMC 23377 (*Ba*CYP106A2) and *B. megaterium* (*Bm*CYP106A2).

Conformational changes of CYP106A2 induced by ligand binding (**A**) Structure superposition of ligand-free *Ba*CYP106A2 (orange) and *Bm*CYP106A2 (PDB code 4YT3; cyan) revealed that the most significant structural difference exists in the loop region between α8 and α9. Notably, a part of the α8–α9 loop region in the crystal structure of *Bm*CYP106A2 is disordered. (**B**) Plot of crystallographic B-factor values against residue number for ligand-free *Ba*CYP106A2 (orange) and *Bm*CYP106A2 (PDB code 4YT3; cyan). (**C**) Structural comparison between apo *Ba*CYP106A2 (orange) and abietic acid-bound *Bm*CYP106A2 (PDB code 5IKI; green) at the active site. The bound abietic acid in *Bm*CYP106A2 is indicated by a yellow stick model. (**D**) Structural comparisons of apo *Ba*CYP106A2 (orange), abietic acid-bound *Bm*CYP106A2 (PDB code 5IKI; green), and biphenyl compound-bound CYP105P2 from *S. peucetius* (PDB code 5IT1; blue) showing that the α8–α9 loop region is intrinsically flexible and has a common conformational change induced by ligand binding.

have been identified in the same CYP106A family [16]. In the present study, we first identified *Ba*CYP106A2 from *Bacillus* sp. PAMC 23377 from the Arctic Ocean. The gene was overexpressed, and the purified soluble protein was used for a substrate-binding assay and in vitro assay. Our results demonstrated that the newly identified *Ba*CYP106A2 has hydroxylase activities similar to those of the previously reported enzyme *Bm*CYP106A2. The crystal structure of *Ba*CYP106A2 was also determined, and structural movements in the active site were detected in comparison with *Bm*CYP106A2. Therefore, in future work, we aim to carry out an in vitro assay with other substrates and to determine the position of their hydroxylation. Our structural analysis can help to select further substrates, and the potent activity of CYP106A2 can be applied to biotechnology applications.

Acknowledgments

We would like to thank the staff at the X-ray Core Facility of the Korea Basic Science Institute (KBSI; Ochang, Korea) and BL-5C of the Pohang Accelerator Laboratory (Pohang, Korea) for their kind help in data collection. This

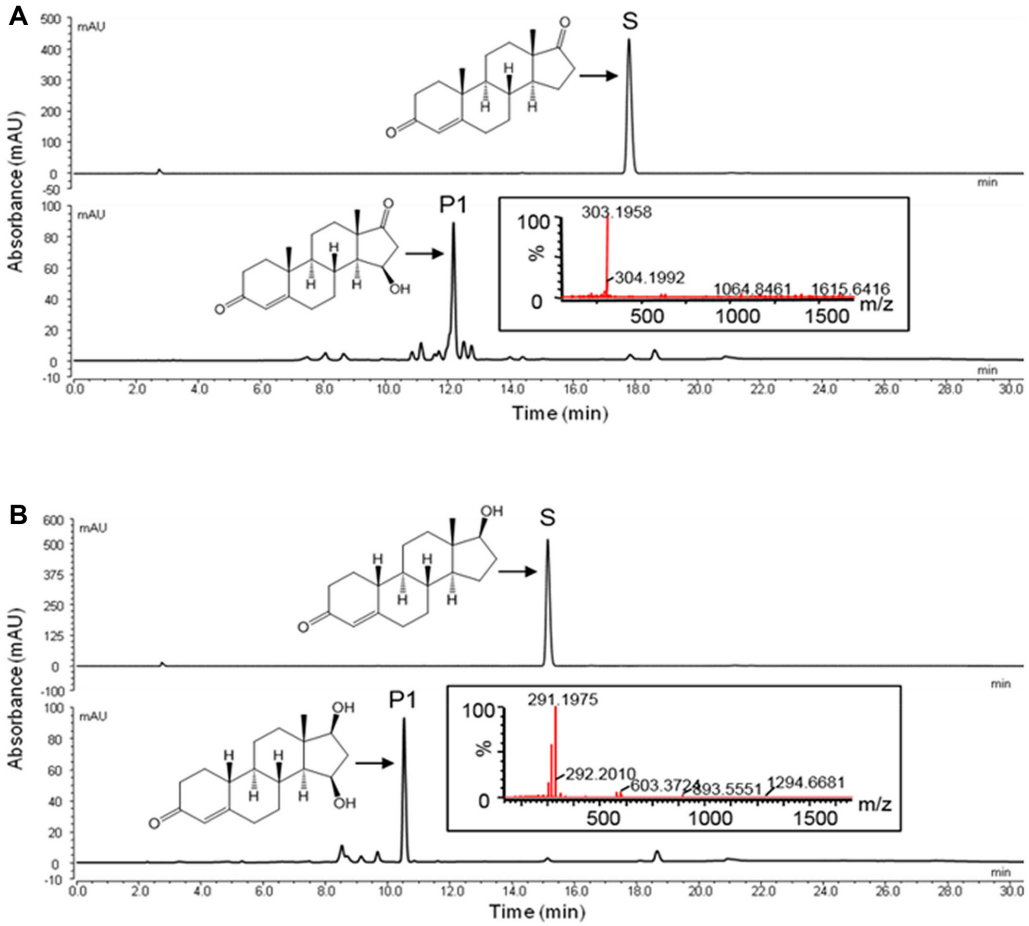

Fig. 6. HPLC chromatogram and LC-MS analysis in the in vitro assay.
P1 is the peak of the major product identified as a monohydroxylated product. S peak, substrate; P1 peak, main product. (**A**) 4-Androstenedione substrate. P1 was analyzed by LC-MS at m/z^+ $[M+H]^+$ 303.1958. (**B**) Nandrolone substrate. P1 was analyzed by LC-MS at m/z^+ $[M+H]^+$ 291.1975. These reactions were carried out by adding 25 μg of ferredoxin, 0.1 U of ferredoxin reductase, and 10 μM CYP106A2 to 50 mM potassium phosphate buffer (pH 7.4).

work was supported by the Polar Genomics 101 Project: Genome Analysis of Polar Organisms and Establishment of Application Platform (PE17080) funded by the Korea Polar Research Institute. In addition, this work was carried out with the support of the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2016R1D1A3B03933814).

Conflict of Interest

The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; and in the decision to publish the results.

References

1. Bernhardt R. 1996. Cytochrome P450: structure, function, and generation of reactive oxygen species. *Rev. Physiol. Biochem. Pharmacol.* **127:** 137-221.
2. Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, *et al*. 1996. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* **6:** 1-42.
3. Urlacher VB, Eiben S. 2006. Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends Biotechnol.* **24:** 324-330.
4. Milhim M, Gerber A, Neunzig J, Hannemann F, Bernhardt R. 2016. A novel NADPH-dependent flavoprotein reductase from *Bacillus megaterium* acts as an efficient cytochrome P450 reductase. *J. Biotechnol*. **231:** 83-94.

5. Ortiz de Montellano PR. 2005. Cytochrome P450 Structure, Mechanism, and Biochemistry, 3rd Ed. Kluwer Academic/ Plenum Publishers, New York, USA.
6. Zhang A, Zhang T, Hall EA, Hutchinson S, Cryle MJ, Wong LL, *et al*. 2015. The crystal structure of the versatile cytochrome P450 enzyme CYP109B1 *Bacillus subtilis*. *Mol. Biosyst.* **11:** 869-881.
7. Holland HL. 1999. Recent advances in applied and mechanistic aspects of the enzymatic hydroxylation of steroids by whole-cell biocatalysts. *Steroids* **64:** 178-186.
8. Bureik M, Lisurek M, Bernhardt R. 2002. The human steroid hydroxlases CYP1B1 and CYP11B2. *Biol. Chem.* **383:** 1537-1551.
9. Lee GY, Kim DH, Kim D, Ahn T, Yun CH. 2015. Functional characterization of steroid hydroxylase CYP106A1 derived from *Bacillus megaterium*. *Arch. Pharm. Res.* **38:** 98-107.
10. Berg A, Gustafsson JA, Ingelman-Sundberg M. 1976. Characterization of a cytochrome P-450-dependent steroid hydroxylase system present in *Bacillus megaterium*. *J. Biol. Chem.* **251:** 2831-2838.
11. Jozwik IK, Kiss FM, Gricman L, Abdulmughni A, Brill E, Zapp J, *et al*. 2016. Structural basis of steroid binding and oxidation by the cytochrome P450 CYP109E1 from *Bacillus megaterium*. *FEBS J*. **283:** 4128-4148.
12. Makino T, Katsuyama Y, Otomatsu T, Misawa N, Ohnishi Y. 2014. Regio- and stereospecific hydroxylation of various steroids at the 16α position of the D ring by the *Streptomyces griseus* cytochrome P450 CYP154C3. *Appl. Environ. Microbiol.* **80:** 1371-1379.
13. Bracco P, Janssen DB, Schallmey A. 2013. Selective steroid oxyfunctionalisation by CYP154C5, a bacterial cytochrome P450. *Microb. Cell Fact*. **12:** 95.
14. Khatri Y, Ringle M, Lisurek M, von Kries JP, Zapp J, Bernhardt R. 2016. Substrate hunting for the myxobacterial CYP260A1 revealed new 1α-hydroxylated products from C-19 steroids. *Chembiochem* **17:** 90-101.
15. Salamanca-Pinzon SG, Khatri Y, Carius Y, Keller L, Muller R, Lancaster CR, *et al*. 2016. Structure-function analysis for the hydroxylation of Δ4 C21-steroids by the myxobacterial CYP260B1. *FEBS Lett*. **590:** 1838-1851.
16. Kiss FM, Schmitz D, Zapp J, Dier TK, Volmer DA, Bernhardt R. 2015. Comparison of CYP106A1 and CYP106A2 from *Bacillus megaterium* – identification of a novel 11-oxidase activity. *Appl. Microbiol. Biotechnol.* **99:** 8495-8514.
17. Nguyen KT, Virus C, Gunnewich N, Hannemann F, Bernhardt R. 2012. Changing the regioselectivity of a P450 from C15 to C11 hydroxylation of progesterone. *Chembiochem* **13:** 1161-1166.
18. Nikolaus J, Nguyen KT, Virus C, Riehm JL, Hutter M, Bernhardt R. 2017. Engineering of CYP106A2 for steroid 9α- and 6β-hydroxylation. *Steroids* **120:** 41-48.
19. Janocha S, Carius Y, Hutter M, Lancaster CR, Bernhardt R. 2016. Crystal structure of CYP106A2 in substrate-free and substrate-bound form. *Chembiochem* **17:** 852-860.
20. Omura T, Sato R. 1964. The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *J. Biol. Chem.* **239:** 2370-2378.
21. Johnston JB, Ouellet H, Ortiz de Montellano PR. 2010. Functional redundancy of steroid C_{26}-monooxygenase activity in *Mycobacterium tuberculosis* revealed by biochemical and genetic analyses. *J. Biol. Chem.* **285:** 36352-36360.
22. Minor W, Otwinowski Z. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276:** 307-326.
23. Vagin A, Teplyakov A. 1997. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30:** 1022-1025.
24. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, *et al*. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67:** 235-242.
25. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, *et al*. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67:** 355-367.
26. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, *et al*. 2010. PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66:** 213-221.
27. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, *et al*. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66:** 12-21.
28. DeLano WL. 2002. The PyMOL molecular graphics system. *CCP4 Newsletter on Protein Crystallography* **40:** 82-92.
29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al*. 2006. The Protein Data Bank, 1999, pp. 675-684. *International Tables for Crystallography*, Vol. F. John Wiley & Sons, IN.
30. Rauschenbach R, Isernhagen M, Noeske-Jungblut C, Boidol W, Siewert G. 1993. Cloning sequencing and expression of the gene for cytochrome P450meg, the steroid-15 beta-monooxygenase from *Bacillus megaterium* ATCC 13368. *Mol. Gen. Genet*. **241:** 170-176.
31. Lee CW, Lee JH, Rimal H, Park H, Lee JH, Oh TJ. 2016. Crystal structure of cytochrome P450 (CYP105P2) from *Streptomyces peucetius* and its conformational changes in response to substrate binding. *Int. J. Mol. Sci.* **17:** 813.
32. Lee CW, Yu SC, Lee JH, Park SH, Park H, Oh TJ, *et al*. 2016. Crystal structure of a putative cytochrome P450 alkane hydroxylase (CYP153D17) from *Sphingomonas* sp. PAMC 26605 and its conformational substrate binding. *Int. J. Mol. Sci.* **17:** 2067.
33. Schenkman JB, Sligar SG, Cinti DL. 1981. Substrate interaction with cytochrome P-450. *Pharmacol. Ther.* **12:** 43-71.
34. Brill E. 2013. Identifizierung und charakterisierung neuer cytochrom P450 systeme aus *Bacillus megaterium* DSM319 (Doctoral dissertation).

35. Schmitz D, Zapp J, Bernhardt R. 2014. Steroid conversion with CYP106A2 – production of pharmaceutically interesting DHEA metabolites. *Microb. Cell Fact.* **13:** 81.
36. Holm L. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38:** W545-W549.
37. Savino C, Montemiglio LC, Sciara G, Miele AE, Kendrew SG, Jemth P, *et al*. 2009. Investigating the structural plasticity of a cytochrome P450: three-dimensional structures of P450 EryK and binding to its physiological substrate. *J. Biol. Chem.* **284:** 29170-29179.
38. Zhang A, Zhang T, Hall EA, Hutchinson S, Cryle MJ, Wong LL, *et al*. 2015. The crystal structure of the versatile cytochrome P450 enzyme CYP109B1 from *Bacillus subtilis*. *Mol. Biosyst.* **11:** 869-881.
39. Montemiglio LC, Parisi G, Scaglione A, Sciara G, Savino C, Vallone B. 2016. Functional analysis and crystallographic structure of clotrimazole bound OleP, a cytochrome P450 epoxidase from *Streptomyces antibioticus* involved in oleandomycin biosynthesis. *Biochim. Biophys. Acta* **1860:** 465-475.
40. Li S, Tietz DR, Rutaganira FU, Kells PM, Anzai Y, Kato F, *et al*. 2012. Substrate recognition by the multifunctional cytochrome P450 MycG in mycinamicin hydroxylation and epoxidation reactions. *J. Biol. Chem.* **287:** 37880-37890.
41. Yasutake Y, Fujii Y, Nishioka T, Cheon WK, Arisawa A, Tamura T. 2010. Structural evidence for enhancement of sequential vitamin D3 hydroxylation activities by directed evolution of cytochrome P450 vitamin D3 hydroxylase. *J. Biol. Chem.* **285:** 31193-31201.
42. Brill E, Hannemann F, Zapp J, Bruning G, Jauch J, Bernhardt R. 2014. A new cytochrome P450 system from *Bacillus megaterium* DSM319 for the hydroxylation of 11-keto-β-boswellic acid (KBA). *Appl. Microbiol. Biotechnol.* **98:** 1701-1717.
43. Donova MV, Egorova OV. 2012. Microbial steroid transformations: current state and prospects. *Appl. Microbiol. Biotechnol.* **94:** 1423-1447.
44. Wang R, Sui P, Hou X, Cao T, Jia L, Lu F, *et al*. 2017. Cloning and identification of a novel steroid 1α-hydroxylase gene from *Absidia coerulea*. *J. Steroid Biochem. Mol. Biol.* DOI: 10.1016/j.jsbmb.2017.04.006.