





RESEARCH LETTER

Structural comparison of the cytochrome P450 enzymes CYP106A1 and CYP106A2 provides insight into their differences in steroid conversion

 Yvonne Carius¹ , Michael Hutter² , Flora Kiss³, Rita Bernhardt³  and C. Roy D. Lancaster¹ 

1 Department of Structural Biology, Faculty of Medicine, Center of Human and Molecular Biology (ZHMB), Saarland University, Homburg, Germany

2 Centre for Bioinformatics, Saarland University, Saarbrücken, Germany

3 Institute of Biochemistry, Saarland University, Saarbrücken, Germany

Correspondence

R. Bernhardt, Institute of Biochemistry, Saarland University, 66123 Saarbrücken, Germany

Tel: +49 681 302 4241

E-mail: ritabern@mx.uni-saarland.de

C. R. D. Lancaster, Department of Structural Biology, Faculty of Medicine, Center of Human and Molecular Biology (ZHMB), Saarland University, Building 60, 66421 Homburg, Germany

Tel: +49 6841 16 26235

E-mail: roy.lancaster@structural-biology.eu

(Received 19 August 2022, revised 13 September 2022, accepted 14 September 2022, available online 6 October 2022)

doi:10.1002/1873-3468.14502

Edited by Dietmar J. Manstein

Understanding the structural basis of the selectivity of steroid hydroxylation requires detailed structural and functional investigations on various steroid hydroxylases with different selectivities, such as the bacterial cytochrome P450 enzymes. Here, the crystal structure of the cytochrome P450 CYP106A1 from *Priestia megaterium* was solved. CYP106A1 exhibits a rare additional structural motif of a cytochrome P450, a sixth β -sheet. The protein was found in different unusual conformations corresponding to both open and closed forms even when crystallized without any known substrate. The structural comparison of CYP106A1 with the previously investigated CYP106A2, including docking studies for both isoforms with the substrate cortisol, reveals a completely different orientation of the steroid molecule in the active sites. This distinction convincingly explains the experimentally observed differences in substrate conversion and product formation by the two enzymes.

Keywords: crystal structure; Cytochrome P450; docking; oxidoreductase; steroid conversion

Cytochromes P450 (P450s) are a superfamily of haem containing monooxygenases found in prokaryotes, eukaryotes and even in viruses. They are involved in numerous biological processes, such as the biosynthesis of steroid hormones and human xenobiotic metabolism [1]. They catalyse a large variety of reactions (e.g. hydroxylations, dehydrogenations, epoxidations, dealkylations, oxidation etc.) converting a broad range of substrates (e.g. steroids, terpenes, fatty acids etc.) which make them versatile biocatalysts for industrial processes [2,3].

Beside antibiotics, the production of steroids has become an important field in pharmaceutical industry

[4]. Steroids play an important role in many living systems and regulate cellular processes such as sexual differentiation, inflammation, signal transduction but also the fluidity of membranes. Steroid derivatives exhibit higher biological activity and stability than their non-hydroxylated forms and are therefore more preferable for industrial applications and as biomedicine [5]. Bacterial cytochromes P450 show highest potential as biocatalysts for steroid hydroxylation and can be expressed to high levels in bacterial expression systems [6] or yeast [7]. A disadvantage is the insufficient regio- and/or stereoselectivities of these enzymes or low activity which can be solved by rational protein engineering [8].

Abbreviations

P450, cytochrome P450; PDB, Protein Data Bank; r.m.s.d, root mean square deviation.

The gram-positive soil-bacterium *Priestia megaterium* (formerly *Bacillus megaterium*) has been a valuable industrial organism for decades and is widely used as a production strain for enzymes in biotechnological application [9,10]. But besides this, the most interesting proteins present in *P. megaterium* are P450s with novel functions such as the self-sufficient fatty acid hydroxylase P450 BM3 (CYP102A1) [11,12], the vitamin D hydroxylases CYP109E1 and CYP109A2 [13,14], the recently functionally characterized CYP106A1 (P450 BM1) [15–18] and CYP106A2 (P450meg) involved in steroid, di- and triterpene conversions of biotechnological relevance [19–25].

While CYP106A2 was extensively investigated and is well characterized since decades [26–30], including the publication of three crystal structures with and without substrate, respectively [31,32], significantly less was known about its closest homolog, CYP106A1. The CYP106 subfamily members share 63% amino acid identity, a good indication for catalytic similarity. The potential of CYP106A1 for steroidal drug or drug metabolite production was thus probed by the conversion of a focused steroid library [17]. The catalysts were found to have a similar binding behaviour towards the selected substrates but showed different product patterns and activities. Although both enzymes bound the steroids examined with similar affinities, CYP106A2 produced one or two main products, whereas the conversion with CYP106A1 yielded a higher amount of additional side products. CYP106A2 also exhibited higher conversion velocities in all transformations except for 11 β -hydroxysteroids (cortisol, corticosterone). While the CYP106A1 enzyme enabled the dehydrogenation of cortisol and corticosterone, converting these 11 β -hydroxysteroids into their 11-keto forms, CYP106A2 performed the same reaction only with corticosterone [18]. Since, such an 11-oxo-steroid formation is uncommon in cytochrome P450 catalysis, a follow-up study investigated the underlying mechanism while identifying further CYP106A1 substrates [17]. The structural background for the differences in substrate conversion between CYP106A1 and CYP106A2 is not yet known but would pave the way for a deeper understanding of their steroid conversion and provide a basis for defining new possibilities for a rationally designed enzyme activity and specificity.

Here, we describe the crystal structure of CYP106A1 at 1.7 Å resolution. This structure, accompanied by computational substrate docking studies using cortisol, provides the basis for understanding the differences between CYP106A1 and CYP106A2 concerning their substrate conversion and product formation.

Materials and methods

Protein purification and crystallization

CYP106A1 was expressed and purified as described previously [15]. The protein was produced in *Escherichia coli* strain C43(DE3) and purified by immobilized metal ion affinity chromatography using TALON™ resin (Takara Bio USA, Inc., Mountain View, CA, USA) followed by size-exclusion chromatography. After purification, the protein was concentrated to 1 mM. CYP106A1 crystals were obtained in 96-well plates with the sitting-drop vapour diffusion method at 291K using the automated crystallization facility at the Department of Structural Biology [33]. Equal amounts of protein solution (1 mM and 670 μ M protein concentration) and reservoir solution were mixed and equilibrated against the reservoir solution. First reddish cube- and triangle-shaped protein crystals appeared after 30 days and were further optimized in 24-well plates with the hanging-drop vapour diffusion method using 670 μ M protein concentration (Fig. 2A). The best crystals diffracting to 1.7 Å were obtained in 0.3 M LiSO₄ and 25% PEG 3350.

Data collection, structure determination and refinement

The protein crystals were transferred to a buffer consisting of 0.3 M LiSO₄ supplemented with 30% PEG 3350 and 10% PEG 400 for cryoprotection and flash-cooled in liquid nitrogen. After initial diffraction experiments at the Department of Structural Biology home source (Oxford Diffraction Nova system), X-ray diffraction data were collected at 100 K at beamline ID23-1 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) [34]. Data were processed with IMOSFLM [35] and scaled with SCALA [36] from the CCP4 software package [37]. The CYP106A1 structure was solved by molecular replacement with MOLREP [38] using the structure of CYP106A2 from *Priestia megaterium* ATCC 13368 (PDB entry 4YT3) [31] as a search model. The program COOT [39] was used for manual rebuilding and completion of the model and refinement was performed using REFMAC5 [40]. All data collections and refinement statistics are summarized in Table 1. The coordinates and associated structure factors have been deposited at the Protein Data Bank under (PDB entry 7ZZL). Graphical representations of the structural model were created using PYMOL [41].

Computational methods

Docking of cortisol into CYP106A1 and CYP106A2

For better comparison, the crystal structure of CYP106A1 (chain D) was superimposed onto that of CYP106A2 (PDB

Table 1. Data collection and refinement statistics.

PDB entry	7ZZL
Beamline	ID23-2
Space group	P1
Unit cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> [Å]	63.69, 82.87, 84.86
$\alpha/\beta/\gamma$ [°]	95.36/90/90
Wavelength [Å]	0.97916
Resolution of data ^a	25.92–1.70 (1.79–1.70)
No. of observations ^a	553 189 (76487)
No. of unique reflections ^a	168 988 (23968)
Completeness [%] ^a	88.9 (86.3)
Redundancy ^a	3.3 (3.2)
$\langle I/\sigma(I) \rangle^a$	7.05 (2.11)
R_{merge} [%] ^a	6.6 (30.7)
R_{meas} [%] ^{a,b}	7.9 (36.7)
$R_{\text{p.i.m.}}$ [%] ^{a,c}	4.3 (19.9)
CC(1/2) ^a	0.996 (0.920)
Wilson <i>B</i> factor [Å ²]	18.6
Refinement	
$R_{\text{cryst}}^d/R_{\text{free}}^e$ [%]	16.9/20.6
No. of molecules in the asymmetric unit	4
Residues included in the model (total no. of protein atoms)	1560 (14031)
Water molecules (belonging to chain A/B/C/D)	1027
Ligands	Protoporphyrin IX containing Fe, sulphate, imidazole, polyethylene glycol, dimethyl sulfoxide, cobalt
Overall <i>B</i> factor [Å ²]	22.75
<i>B</i> Factor for protein chain (A/B/C/D)	23.27/22.86/25.15/25.02
<i>B</i> Factors for ligands/ions [Å ²]	
Waters	30.92
Haem (A/B/C/D)	12.75/11.52/15.39/14.54
Ramachandran outliers [%]	
Favoured	90.0
Allowed	10
Outliers	0
r.m.s.d. for bond lengths [Å]	0.011
r.m.s.d. for bond angles [°]	1.674

^aValues in parentheses are calculated for the highest resolution shell.

$$^b R_{\text{meas}} = \sum_{\mathbf{h}} \left(\frac{n_{\mathbf{h}}}{n_{\mathbf{h}} - 1} \right) \sum_i |I_{\mathbf{h}i} - \langle I_{\mathbf{h}} \rangle| / \sum_{\mathbf{h}} \sum_i \langle I_{\mathbf{h}} \rangle$$

$$^c R_{\text{p.i.m.}} = \sum_{\mathbf{h}} \left(\frac{1}{n_{\mathbf{h}} - 1} \right) \sum_i |I_{\mathbf{h}i} - \langle I_{\mathbf{h}} \rangle| / \sum_{\mathbf{h}} \sum_i \langle I_{\mathbf{h}} \rangle$$

^d $R_{\text{cryst}} = 100 \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$; ^eCalculation of R_{free} was performed analogous to R_{cryst} excluding 5% of randomly chosen reflexes.

entry code 5IKI, chain B) [31]. Missing loops (FG loop) and stretches traced to polyalanine residues (aa232–237) due to missing electron density were modelled using the SWISS-pdb viewer (version 4.0.1) [42] as published earlier [43,44]. Both structures were prepared for subsequent docking using AUTODOCKTOOLS (Windows version 1.5.6r3) [45]

assigning Kollman charges to the protein part and Gasteiger–Marsili charges to the HEM moiety, as described previously [44]. Preparation of cortisol was carried out following the same procedure as for similar ligands [44]. Protonation states of the histidine residues were assigned by visual inspection to optimize their hydrogen-bonding network. For both receptors, the grid box had the same dimensions (46 by 54 by 52 points) and was centred above the HEM moiety. Default values of AUTODOCK (Version 4.2) were used, except that the number of Lamarckian genetic algorithm docking runs was increased to 250, in order to capture all relevant docking poses [45].

Protein structures alignment and ligand interaction analysis

Sequence alignments were done with the program CLUSTAL OMEGA [46]. The sequence alignments with the secondary structure elements were performed with the ESPRIT 3.0 server [47]. The protein ligand analysis of the docked substrates was done with LIGPLOT+ [48].

Results

Overall structure of CYP106A1

The crystal structure of full-length CYP106A1 was solved at 1.7 Å resolution in the space group P1 with four molecules in the asymmetric unit. As a model, the crystal structure of CYP106A2 from *Priestia megaterium* ATCC 13368 (PDB entry 4YT3) [31] with a sequence identity of 63% and 76% similarity was used for the molecular replacement. The protein structure displays the typical triangular shape and the P450 fold [49,50] with 14 α -helices and 9 β -strands in chains A and B but 13 α -helices and 11 β -strands in chains C and D respectively (Figs 1, 5 and S1). Gaps without visible electron density are found in all chains at the N terminus (first 3 to 4 amino acids), the BC loop (aa73–85 in chain A, aa73–86 in chain B, aa76–87 in chain C and D) and the C-terminal hexa-histidine tag. Additional breaks are only seen in chain A and B in the F helix (aa164 + 165) and the FG loop (aa175–186 in chain B, aa175–182 in chain A). Structural alignment of the four molecules reveals an equal conformation of chains A and B (Fig. 2B). Chain D displays prominent differences to chain A/B in the orientation of the F and G helices (deviation of 13.2 Å between the C α -atoms of Glu184, angle of 19.2°) with their connecting FG loop, the first part of the I-helix (deviation of 2.1 Å between the C α -atoms of Asp231, angle of 6.3°) and the beginning of the BC loop (Fig. 2C). Chain C varies slightly in the orientation of the F helix and the G helix from chain D. Parts of the BC loop are not

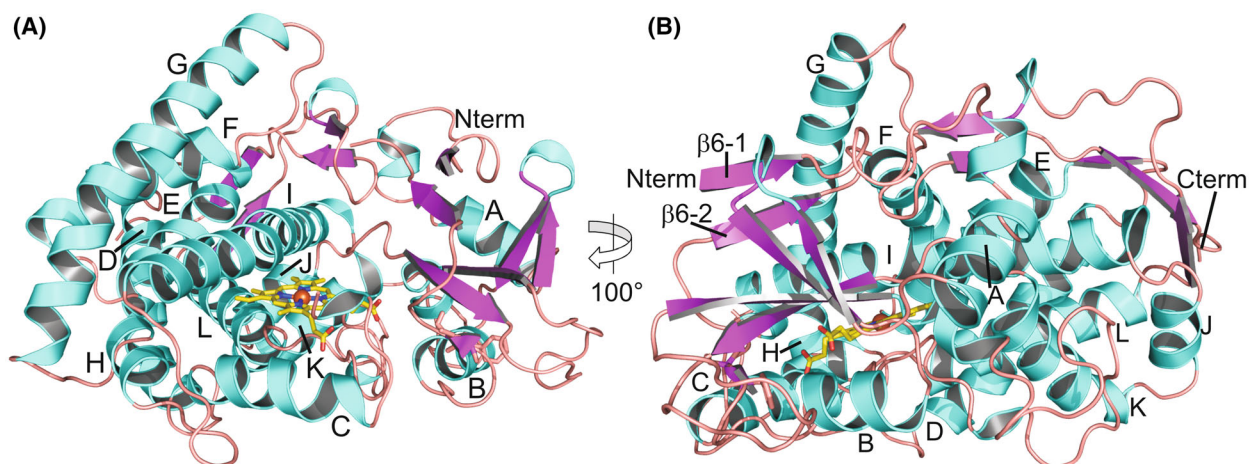


Fig. 1. Overall structure of CYP106A1 (chain D) in front view (A) and side view (B). The secondary-structure elements are shown as ribbons with helices in cyan and β -strands in magenta. The nomenclature of the secondary-structure elements is according to [62]. The haem cofactor is displayed in stick presentation (C-atoms coloured in yellow, O-atoms in red, N-atoms in blue, Fe-atom as an orange sphere).

visible in all molecules which is not unusual for CYP450 enzymes crystallized in substrate-free form, whereas the electron density of the F helix in chain A/B is less defined and the FG loop is not visible, in chain C/D this region is clearly observable. Chain D provides the most complete electron density especially around the active site and the substrate recognition sites and, therefore, was used for the structure analysis and the docking experiments.

CYP106A1 displays the β -sheets 1, 3 and 4, but lacks β -sheets 2 and 5. Surprisingly, two β -strands were found in chain C and D at the N terminus and in the BC loop (numbered β 6-1 and β 6-2) building together an additional unusual β -sheet (Figs 1 and 5). In chains A and B, an additional alpha-helix (A') similar as in CYP106A2 is found at the N terminus instead of β 6-1 (Fig. S1).

The protein was crystallized without any substrate. In the active sites of all four molecules, a water molecule binds as the sixth ligand to the haem-iron. In chain B and C, a polyethylene molecule coordinated via two water molecules was found in the active site which results from the crystallization conditions. Further ligands were among other sulphate ions, polyethylene glycol and imidazole bound to the C-terminal histidine tags.

The active site of CYP106A1 is enclosed by the dominant I helix with substrate recognition site 4 (SRS4) and residues belonging to SRS1, 5 and 6. The propionate side chains of the haem cofactor form hydrogen bonds to His97, Arg101, Arg297 and His345 similar to CYP106A2. Unusual is an additional hydrogen bond to Arg295 which is replaced by leucine in

CYP106A2. Cys356 serves as the fifth ligand of the bound iron.

Structural comparison of CYP106A1 with other P450 structures using the PDBfold server [51] identified CYP106A2 from *Priestia megaterium* (PDB entries 4YT3 and 5IKI) [31] and from *Bacillus* sp. PAMC 23377 (PDB entry 5XNT) [32] as the most similar structural homologs. Whereas chains A and B of CYP106A1 are most similar to the substrate-bound CYP106A2 (PDB entry 5IKI, chain B), chain C and D of CYP106A1 are more comparable to the substrate-free structure of CYP106A2 (PDB entry 4YT3; Fig. 3 A,B). Pairwise alignment of the CYP106A1 chain D with CYP106A2 structures using the DALI server [52] showed a r.m.s. d. of 1.2 Å for the open form of CYP106A2 (4YT3 and chain A of 5IKI) and 1.7 Å for the closed form with bound substrate (5IKI, chain B) respectively. Chain A of CYP106A1 showed a r.m.s. d. of 1.2 Å for the open form of CYP106A2 (4YT3 and chain A of 5IKI) and 1.5 Å for the closed form with bound substrate (5IKI, chain B) respectively.

Further structurally homologous enzymes with lower sequence identity were also assessed and found to be surprisingly different concerning chains A/B and C/D. Chains A and B are more similar to CYP109B1 from *Priestia megaterium* (PDB entry 4RM4, r.m.s.d. 1.62, [53]) and CYP109E1 from *Bacillus subtilis* (PDB entry 5L92, r.m.s.d. 1.61, [54]) respectively. Chains C and D resemble more CYP105P1 from *Streptomyces avermitilis* (PDB entry 3E5K, r.m.s.d. 2.07 [55]) and MoxA from *Nonomuraea recticatena* (PDB entry 2Z36, r.m.s.d. 1.86, [56]) which belongs also to CYP105 family.

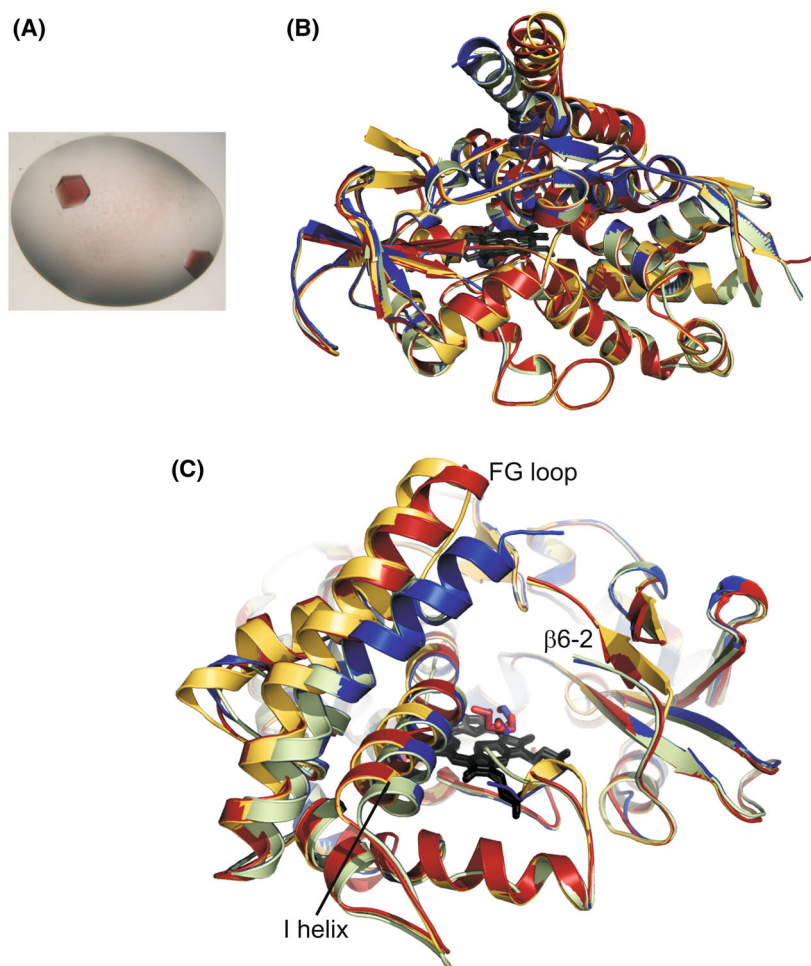


Fig. 2. Reddish crystals of CYP106A1 (A) and structural alignment of the four chains found in the asymmetric unit (B) and the main differences in detail (C). Chains A and B (light green/blue) as well as C and D (dark red/orange) exhibited the most similarity. The main differences between A/B and C/D are in the orientation of the F and G helices, the I helix and the BC loop with an additional β -sheet (β -6). The two PEG molecules found in the active site of chains B and C are shown as stick models with corresponding colours (C).

Docking results

To gain further knowledge and understand the reason behind the different product patterns of CYP106A1 and CYP106A2 despite similar substrate binding, docking simulations with cortisol were performed.

The major differences between CYP106A1 and CYP106A2 affecting the binding pocket are found in the region SRS6. Especially the two amino acids, alanine and threonine, lie closer to the HEM moiety in CYP106A2. Decisive for the binding conformation of cortisol are, however, the different residues in position 293–295 in region SRS5. These comprise two arginines in CYP106A1 that are exchanged to lysine and leucine in the A2 form (Fig. 3C). The energetically most favoured docking conformation shows that the 17 β -OH as well as the 22-OH group of cortisol can form hydrogen bonds to amino acids Ser293, Arg294 and Arg295 in this region of CYP106A1, thus orienting the 11-OH group towards the HEM-iron (4.12 Å

distance), thereby enabling oxidation to the keto functionality (Figs 4A,C and S2). A possible side product of this conformation is 1 β -OH-cortisol (4.43 Å distance to the iron). In the docking pose ranked second, which is, however, most often adopted in the docking runs, the 15 β position is closest to the iron (3.58 Å). In CYP106A2, a completely different binding position of cortisol is found with hydrogen bonds to amino acids in the SRS1 region to amino acids Ile 88, Thr89 and Glu90. Here, hydroxylation is expected to take place in positions 6 and 7 (Figs 4B,D and S2).

Discussion

To obtain a deeper insight and understanding for the selectivity of steroid hydroxylation, structural investigation of close relatives of P450s with overlapping substrate specificity but distinct differences concerning the selectivity of hydroxylation are of great help. Fortunately, we were able during the past years to identify,

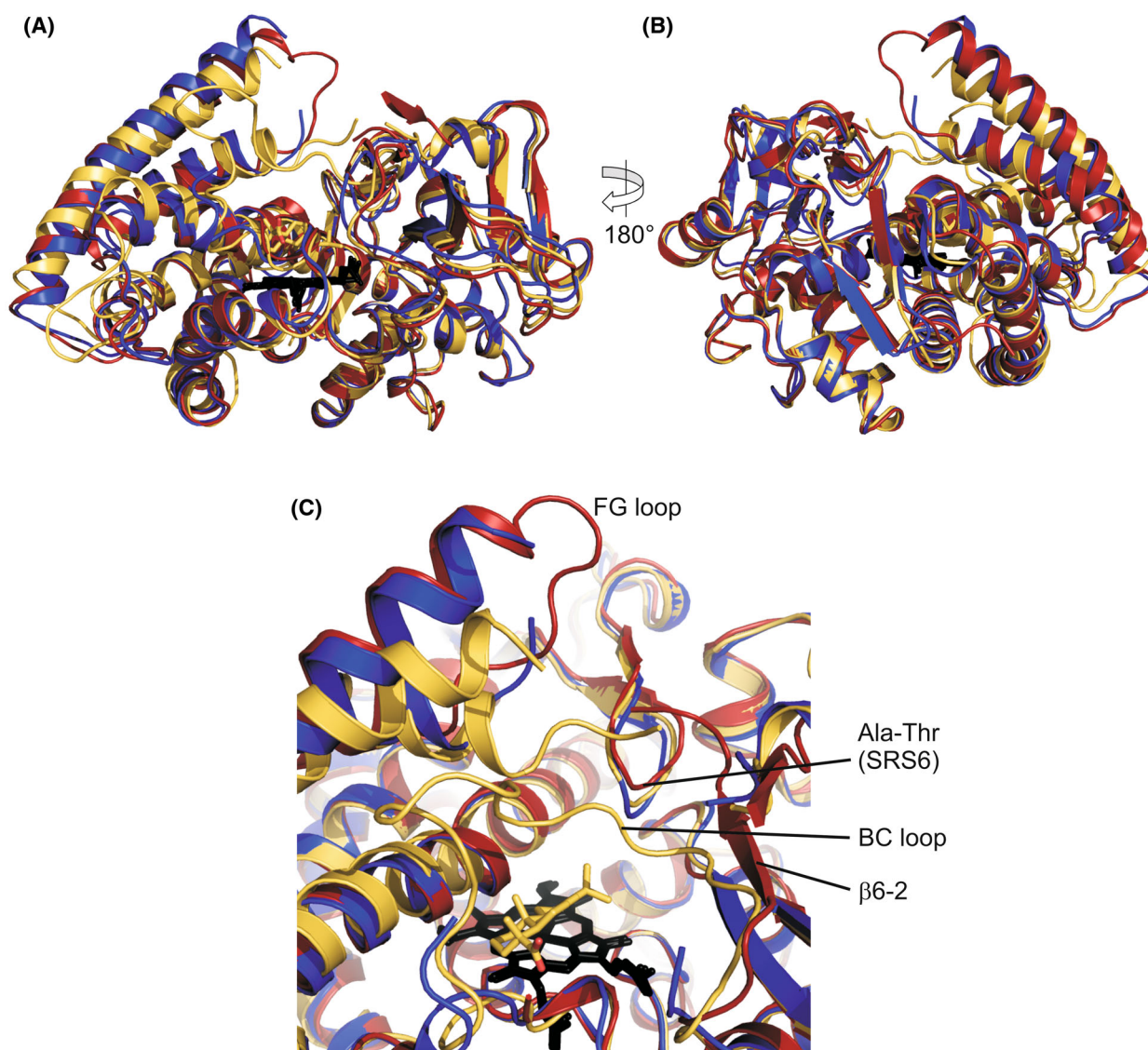


Fig. 3. Structural alignment of CYP106A1 (red, chain D), substrate-free CYP106A2 (blue, PDB entry 4YT3) and CYP106A2 with bound substrate (orange, PDB entry 5IKI, chain B) (A, B) and their active sites in detail (C). Major differences are in the location of the F and G helices and the intermediate FG loop. The active sites of CYP106A1 and A2 differ mainly in SRS6 and the BC loop with an additional β -strand in CYP106A1 (β 6-2). The substrate abietic acid in the active site of CYP106A2 is represented as a stick model with corresponding colours.

clone, express and characterize two members of the CYP106A subfamily, CYP106A1 and CYP106A2, and solve the X-ray structure of one of them, CYP106A2 [31].

The previously unknown crystal structure of CYP106A1 from *Priestia megaterium* DSM319 was now solved and is described here. The four molecules are found to be in different conformations. The structural comparison with other P450 enzymes points towards an open conformation of chains C and D and a closed conformation in chains A and B. Since CYP106A1 was crystallized without any substrate, this

finding is rather unexpected. The PEG molecule as a potential substrate mimic can be excluded since it was present in both conformations. Based on a large number of available P450 structures, it is known that a rearrangement of the BC loop, the FG loop and the adjacent helices from an open conformation to a closed one occurs if a substrate is bound. Protein structures showing an open and closed conformation in absence of substrate were also described for PikC [57], EryK [58] and P450cam [59]. For EryK, an equilibrium of two conformations was suggested instead of an induced-fit rearrangement, while for P450cam a

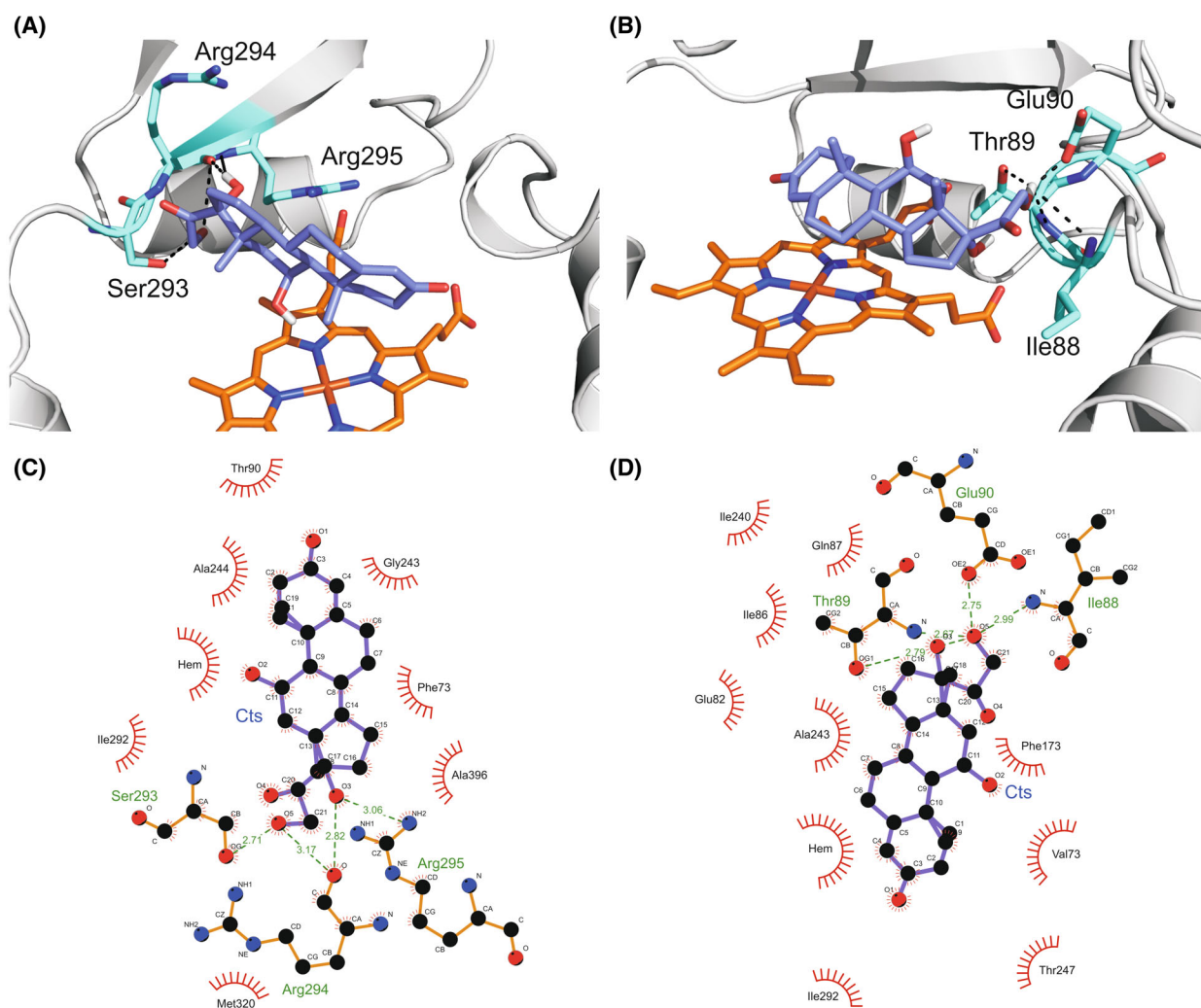


Fig. 4. Docking studies of CYP106A1 and CYP106A2 with cortisol and corresponding ligplots of ligand interactions. Docking of cholesterol in the active site of CYP106A1 (A) and CYP106A2 (B). The protein structure is represented in light grey. Residues involved in substrate binding are coloured in cyan and represented as sticks. Bonds involving carbon atoms are coloured in purple (substrate) and orange (haem cofactor) respectively. Hydrogen bonds are indicated in black dashed lines. The energetically most favoured docking conformations of cortisol in CYP106A1 (A) and CYP106A2 (B) show that the different hydrogen-bond pattern determines the orientation. In CYP106A1, the experimentally observed reduction of the 11 β -OH group to the keto functionality takes place, yielding cortisone. The ligplots of CYP106A1 (C) and CYP106A2 (D) show the hydrophobic and polar contacts between cortisol and the protein. Hydrogen bonds are drawn in green, residues involved in hydrophobic contacts are half-circled in orange.

three-step model of open, closed and intermediate state was proposed [60]. The four molecules in CYP106A1 differ mainly in the location of the F and G helices, in the flexibility of the FG loop and the additional β -sheet in the BC loop found in chains C and D. Since chain C is a partial outlier and might be an incomplete open form, the hypothesis of a multi-step model is reasonable.

The unusual sixth β -strand in the BC loop of chain C/D was also described for CYP105P1 from *Streptomyces avermitilis* but only in the substrate-free form

[55]. This β -strand in CYP105P1 is included in β -sheet 1 as a sixth strand while in CYP106A1 a new β -sheet is built with the N terminus. Due to this unusual secondary structure element, the BC loop is more rigid with a divergence of 6.3 Å between the C α -atoms of CYP106A1-Ile72 and CYP106A2-Ile71 in substrate-free form providing together with the bending of the N-terminal part of the I-helix a larger substrate access channel for bulky tethering substrates in CYP106A1. In CYP106A1, the highest deviation is 4.6 Å for residue Ile72 between chains A/B and C/D and due to the

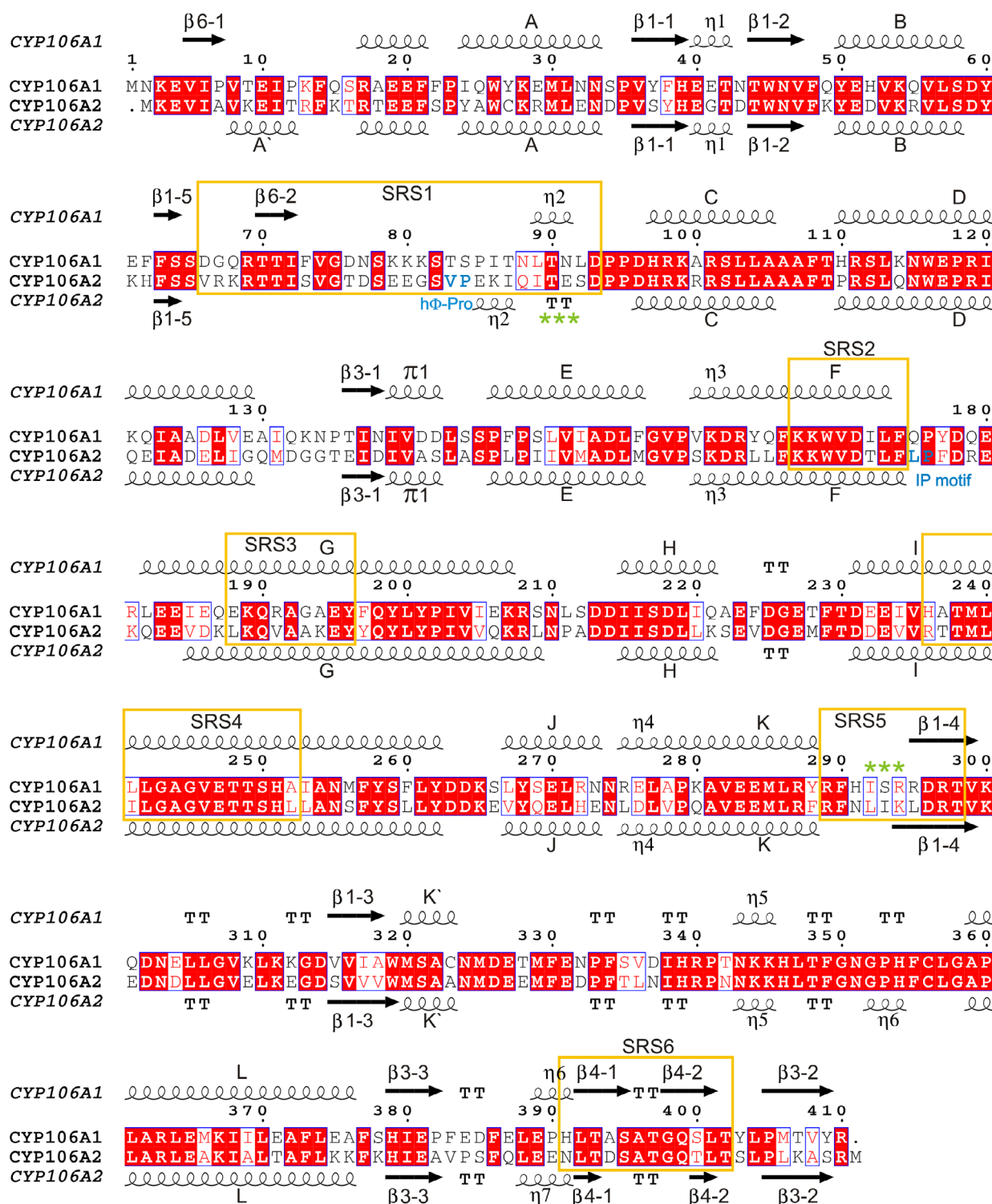


Fig. 5. Sequence alignment between CYP106A1 (chain D) and CYP106A2 from *Priestia megaterium* performed with the program CLUSTAL OMEGA [46]. The final figure with the secondary-structure elements was prepared with the server ESPRIT 3.0 [47] using chain D of CYP106A1 (this work) and chain B of CYP106A2 (PDB entry 5IKI). The residues highlighted in blue belong to the hΦ-pro and the IP motifs in CYP106A2. The orange frames highlight the substrate recognition sites 1–6. The green asterisks indicate the amino acids involved in cortisol binding in the active sites. While SRS1, SRS4, SRS5, and SRS6 contour the catalytic site, SRS 2 and 3 are involved in forming the substrate access channel [62,63].

more closed conformation the β -strand might not be formed in chain A/B. Since none of these two β -strands are involved in crystal packaging, a potential crystallization artefact can be excluded.

CYP106A1 shows the highest structural similarity to CYP106A2. While chain A/B resembles more the closed CYP106A2 form, chain C/D is more similar to the open CYP106A2 form. The main deviation is in the FG loop, not in the I helix. Therefore, the structural differences cannot be the reason behind the different substrate conversion. Substrate specificity and/or regio- and stereoselectivity in P450s often correlates with special sequence motifs. One is the h Φ -Pro motif, in which a hydrophobic residue (h Φ = Phe, Ile, Val, Leu) precedes a proline at the beginning of the B' helix in the SRS1 region. Another motif is the IP-motif in the FG loop. Instead of Ile, hydrophobic residues like Leu or Phe are also found in some P450s. Enzymes lacking the h Φ -Pro motif in the B–C loop contain the IP motif initiating the F' helix in the F–G loop [61]. Interestingly, in CYP106A1, none of these motifs are present, whereas in CYP106A2, perhaps quite uniquely, both motifs are found (Figs 5 and S1). In the h Φ -Pro motif of CYP106A2, Val82 precedes Pro83, while in CYP106A1 the hydrophobic residue is replaced by the polar Thr83. In the FG loop of CYP106A2, Leu174 precedes Pro175, whereas in CYP106A1 a glutamine (Gln175) is present. These differences in the motifs, the particular structural features like the additional β -sheet and the higher flexibility of the substrate gate in CYP106A1 even in a substrate-free state explain why CYP106A1 is less selective than CYP106A2 in the conversion of C18-, C19-, C20- or C21-steroids [18].

Even though most steroid hydroxylations took place at the B- and D-rings, the transformation of cortisol and corticosterone revealed unexpected results. CYP106A1 demonstrated 11-oxidase activity, while CYP106A2 did so only with corticosterone (yet still producing the 15 β -hydroxy derivative as a main metabolite) [18]. To understand this new transformation and the underlying differences between the subfamily members, cortisol docking studies were performed with both enzymes. We found a completely different location of the cortisol molecule over the haem cofactor. In CYP106A1, the substrate forms hydrogen bonds to amino acids in SRS5, while in CYP106A2 cortisol is fixed over the haem *via* residues in SRS1. The location of cortisol in the active site of CYP106A1 enables the conversion to mainly cortisone, whereas in case of CYP106A2, a different product pattern is expected (e.g. hydroxylation at 6 β , 7 β positions). Thus, the structure of CYP106A1 obtained,

together with the docking studies performed, provides a conclusive explanation of the experimentally observed differences in the substrate conversion of cortisol leading to the formation of cortisone in the case of CYP106A1 but not for CYP106A2.

Conclusion

The crystal structure of CYP106A1 from *Priestia megaterium* is presented here and compared with the CYP106A2 investigated previously. CYP106A1 displays a rare additional β -sheet and was crystallized in different unusual conformations corresponding to both open and closed forms. Docking studies of both isoforms with the substrate cortisol result in completely different orientation of the steroid molecule in the active sites, explaining the distinction of substrate conversion and product formation observed experimentally. Taken together, these data provide valuable insights into the structural basis for differences in the specificity of substrate hydroxylation by both enzymes and gives a basis for the engineering of CYP106A1 to create a more potent enzyme for biotechnological use, e.g., for the production of the important anti-inflammatory compound cortisone.

Acknowledgements

The structural biology instruments in Homburg were supported by DFG infrastructure grants INST 256/275-1 FUGG and INST 256/299-1 FUGG to CRDL. We thank the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities, and we would like to thank Benoit Maillot for assistance in using beamline ID23-1 and the Frankfurt-Saarland Block Allocation group (BAG) MX1635. This work was supported by the Staatskanzlei des Saarlandes, Landesforschungsförderungsprogramm des Saarlandes (LFFP) 11/02 and 15/04 to CRDL. Open Access funding enabled and organized by Projekt DEAL. Open Access funding enabled and organized by Projekt DEAL.

Author contributions

YC performed crystallization experiments, data collection and processing, solved and refined the structures, did the structural analysis and wrote the main part of the manuscript; FK purified the proteins and revised the manuscript; MH did the docking experiments; RB designed the study and revised the manuscript; CRDL coordinated structural analysis, edited and revised the manuscript.

Data accessibility

The data that support the findings of this study are openly available in the protein database (<https://www.rcsb.org/>) under entry number 7ZZL.

References

- Bernhardt R. Cytochromes P450 as versatile biocatalysts. *J Biotechnol.* 2006;**124**:128–45.
- Urlacher VB, Girhard M. Cytochrome P450 monooxygenases in biotechnology and synthetic biology. *Trends Biotechnol.* 2019;**37**:882–97.
- Di Nardo G, Gilardi G. Natural compounds as pharmaceuticals: the key role of cytochromes P450 reactivity. *Trends Biochem Sci.* 2020;**45**:511–25.
- Donova M. Microbial steroid production technologies: current trends and prospects. *Microorganisms.* 2022;**10**:53.
- Szaleniec M, Wojtkiewicz AM, Bernhardt R, Borowski T, Donova M. Bacterial steroid hydroxylases: enzyme classes, their functions and comparison of their catalytic mechanisms. *Appl Microbiol Biotechnol.* 2018;**102**:8153–71.
- Yogo Y, Yasuda K, Takita T, Yasukawa K, Iwai Y, Nishikawa M, et al. Metabolism of non-steroidal anti-inflammatory drugs (NSAIDs) by *Streptomyces griseolus* CYP105A1 and its variants. *Drug Metab Pharmacokinet.* 2022;**45**:100455.
- Szczębara FM, Chandelier C, Villeret C, Masurel A, Bourot S, Dupont C, et al. Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nat Biotechnol.* 2003;**21**:143–9.
- Zhang X, Peng Y, Zhao J, Li Q, Yu X, Acevedo-Rocha CG, et al. Bacterial cytochrome P450-catalyzed regio- and stereoselective steroid hydroxylation enabled by directed evolution and rational design. *Bioresour Bioprocess.* 2020;**7**:1–18.
- Vary PS, Biedendieck R, Fuerch T, Meinhardt F, Rohde M, Deckwer WD, et al. *Bacillus megaterium*-from simple soil bacterium to industrial protein production host. *Appl Microbiol Biotechnol.* 2007;**76**:957–67.
- Korneli C, David F, Biedendieck R, Jahn D, Wittmann C. Getting the big beast to work-systems biotechnology of *Bacillus megaterium* for novel high-value proteins. *J Biotechnol.* 2013;**163**:87–96.
- Narhi LO, Wen LP, Fulco AJ. Characterization of the protein expressed in *Escherichia coli* by a recombinant plasmid containing the bacillus megaterium cytochrome P-450BM-3 gene. *Mol Cell Biochem.* 1988;**79**:63–71.
- Whitehouse CJC, Bell SG, Wong LL. P450BM3 (CYP102A1): connecting the dots. *Chem Soc Rev.* 2012;**41**:1218–60.
- Abdulgugni A, Jóźwik IK, Putkaradze N, Brill E, Zapp J, Thunnissen AMWH, et al. Characterization of cytochrome P450 CYP109E1 from *Bacillus megaterium* as a novel vitamin D3 hydroxylase. *J Biotechnol.* 2017;**243**:38–47.
- Abdulgugni A, Jóźwik IK, Brill E, Hannemann F, Thunnissen AMWH, Bernhardt R. Biochemical and structural characterization of CYP109A2, a vitamin D3 25-hydroxylase from *Bacillus megaterium*. *FEBS J.* 2017;**284**:3881–94.
- Brill E, Hannemann F, Zapp J, Brüning G, Jauch J, Bernhardt R. A new cytochrome P450 system from *Bacillus megaterium* DSM319 for the hydroxylation of 11-keto- β -boswellic acid (KBA). *Appl Microbiol Biotechnol.* 2014;**98**:1703–17.
- Lee GY, Kim DH, Kim D, Ahn T, Yun CH. Functional characterization of steroid hydroxylase CYP106A1 derived from *Bacillus megaterium*. *Arch Pharm Res.* 2015;**38**:98–107.
- Kiss FM, Khatri Y, Zapp J, Bernhardt R. Identification of new substrates for the CYP106A1-mediated 11-oxidation and investigation of the reaction mechanism. *FEBS Lett.* 2015;**589**:2320–6.
- Kiss FM, Schmitz D, Zapp J, Dier TKF, Volmer DA, Bernhardt R. Comparison of CYP106A1 and CYP106A2 from *Bacillus megaterium* – identification of a novel 11-oxidase activity. *Appl Microbiol Biotechnol.* 2015;**99**:8495–514.
- Berg A, Gustafsson JA, Ingelman Sundberg M, Carlstrom K. Characterization of a cytochrome P 450 dependent steroid hydroxylase system present in *Bacillus megaterium*. *J Biol Chem.* 1976;**251**:2831–8.
- Simgen B, Contzen J, Schwarzer R, Bernhardt R, Jung C. Substrate binding to 15 β -hydroxylase (CYP106A2) probed by FT infrared spectroscopic studies of the iron ligand CO stretch vibration. *Biochem Biophys Res Commun.* 2000;**269**:737–42.
- Lisurek M, Kang MJ, Hartmann RW, Bernhardt R. Identification of monohydroxy progesterones produced by CYP106A2 using comparative HPLC and electrospray ionisation collision-induced dissociation mass spectrometry. *Biochem Biophys Res Commun.* 2004;**319**:677–82.
- Nguyen KT, Virus C, Günnewich N, Hannemann F, Bernhardt R. Changing the regioselectivity of a P450 from C15 to C11 hydroxylation of progesterone. *Chembiochem.* 2012;**13**:1161–6.
- Bleif S, Hannemann F, Lisurek M, Von Kries JP, Zapp J, Dietzen M, et al. Identification of CYP106A2 as a regioselective allylic bacterial diterpene hydroxylase. *Chembiochem.* 2011;**12**:576–82.
- Bleif S, Hannemann F, Zapp J, Hartmann D, Jauch J, Bernhardt R. A new *Bacillus megaterium* whole-cell catalyst for the hydroxylation of the pentacyclic triterpene 11-keto- β -boswellic acid (KBA) based on a

- recombinant cytochrome P450 system. *Appl Microbiol Biotechnol.* 2012;**93**:1135–46.
- 25 Schmitz D, Zapp J, Bernhardt R. Hydroxylation of the triterpenoid dipterocarpol with CYP106A2 from *Bacillus megaterium*. *FEBS J.* 2012;**279**:1663–74.
- 26 Berg A, Ingelman-Sundberg M, Gustafsson JA. Purification and characterization of cytochrome P-450meg. *J Biol Chem.* 1979;**254**:5264–71.
- 27 Schmitz D, Zapp J, Bernhardt R. Steroid conversion with CYP106A2 – production of pharmaceutically interesting DHEA metabolites. *Microb Cell Fact.* 2014;**13**:81.
- 28 Schmitz D, Janocha S, Kiss FM, Bernhardt R. CYP106A2—a versatile biocatalyst with high potential for biotechnological production of selectively hydroxylated steroid and terpenoid compounds. *Biochim Biophys Acta Proteins Proteomics.* 2018;**1866**:11–22.
- 29 Putkaradze N, Kiss FM, Schmitz D, Zapp J, Hutter MC, Bernhardt R. Biotransformation of prednisone and dexamethasone by cytochrome P450 based systems – identification of new potential drug candidates. *J Biotechnol.* 2017;**242**:101–10.
- 30 Zehentgruber D, Hannemann F, Bleif S, Bernhardt R, Lütz S. Towards preparative scale steroid hydroxylation with cytochrome P450 monooxygenase CYP106A2. *Chembiochem.* 2010;**11**:713–21.
- 31 Janocha S, Carius Y, Hutter M, Lancaster CRD, Bernhardt R. Crystal structure of CYP106A2 in substrate-free and substrate-bound form. *Chembiochem.* 2016;**17**:852–60.
- 32 Kim KH, Lee CW, Dangi B, Park SH, Park H, Oh TJ, et al. Crystal structure and functional characterization of a cytochrome P450 (BaCYP106A2) from *Bacillus* sp. PAMC 23377. *J Microbiol Biotechnol.* 2017;**27**:1472–82.
- 33 Müller FG, Lancaster CRD. Crystallization of membrane proteins. *Methods Mol Biol.* 2013;**1033**:67–83.
- 34 Nurizzo D, Mairs T, Guijarro M, Rey V, Meyer J, Fajardo P, et al. The ID23-1 structural biology beamline at the ESRF. *J Synchrotron Radiat.* 2006;**13**:227–38.
- 35 Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr.* 2011;**67**:271–81.
- 36 Evans P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr.* 2006;**62**:72–82.
- 37 Potterton E, Briggs P, Turkenburg M, Dodson E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr.* 2003;**59**:1131–7.
- 38 Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. *J Appl Cryst.* 1997;**30**:1022–5.
- 39 Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* 2010;**66**:486–501.
- 40 Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* 1997;**53**:240–55.
- 41 DeLano WL. The PyMOL molecular graphics system. Palo Alto, CA: DeLano Scientific LLC; 2006. <http://www.pymol.org>
- 42 Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.* 1997;**18**:2714–23.
- 43 Sagadin T, Riehm J, Putkaradze N, Hutter MC, Bernhardt R. Novel approach to improve progesterone hydroxylation selectivity by CYP106A2 via rational design of adrenodoxin binding. *FEBS J.* 2019;**286**:1240–9.
- 44 Nikolaus J, Nguyen KT, Virus C, Riehm JL, Hutter M, Bernhardt R. Engineering of CYP106A2 for steroid 9 α - and 6 β -hydroxylation. *Steroids.* 2017;**120**:41–8.
- 45 Allouche A. Software news and updates Gabedit — a graphical user Interface for computational chemistry Softwares. *J Comput Chem.* 2012;**32**:174–82.
- 46 Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019;**47**:W636–41.
- 47 Gouet P, Courcelle E, Stuart DI, Métoz F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics.* 1999;**15**:305–8.
- 48 Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model.* 2011;**51**:2778–86.
- 49 Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure.* 1995;**3**:41–62.
- 50 Peterson JA, Graham SE. A close family resemblance: the importance of structure in understanding cytochromes P450. *Structure.* 1998;**6**:1079–85.
- 51 Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr.* 2004;**60**:2256–68.
- 52 Holm L. Using Dali for protein structure comparison. *Methods Mol Biol.* 2020;**2112**:29–42.
- 53 Zhang A, Zhang T, Hall EA, Hutchinson S, Cryle MJ, Wong LL, et al. The crystal structure of the versatile cytochrome P450 enzyme CYP109B1 from *Bacillus subtilis*. *Mol Biosyst.* 2015;**11**:869–81.
- 54 Józwiak IK, Kiss FM, Grieman Ł, Abdulmughni A, Brill E, Zapp J, et al. Structural basis of steroid binding and oxidation by the cytochrome P450 CYP109E1 from *Bacillus megaterium*. *FEBS J.* 2016;**283**:4128–48.

- 55 Xu LH, Fushinobu S, Ikeda H, Wakagi T, Shoun H. Crystal structures of cytochrome P450 105P1 from *Streptomyces avermitilis*: conformational flexibility and histidine ligation state. *J Bacteriol.* 2009;**191**:1211–9.
- 56 Yasutake Y, Imoto N, Fujii Y, Fujii T, Arisawa A, Tamura T. Crystal structure of cytochrome P450 MoxA from *Nonomuraea recticatena* (CYP105). *Biochem Biophys Res Commun.* 2007;**361**:876–82.
- 57 Sherman DH, Li S, Yermalitskaya LV, Kim Y, Smith JA, Waterman MR, et al. The structural basis for substrate anchoring, active site selectivity, and product formation by P450 PikC from *Streptomyces venezuelae*. *J Biol Chem.* 2006;**281**:26289–97.
- 58 Savino C, Montemiglio LC, Sciara G, Miele AE, Kendrew SG, Jemth P, et al. Investigating the structural plasticity of a cytochrome P450. Three dimensional structures of P450 EryK and binding to its physiological substrate. *J Biol Chem.* 2009;**284**:29170–9.
- 59 Lee YT, Wilson RF, Rupniewski I, Goodin DB. P450cam visits an open conformation in the absence of substrate. *Biochemistry.* 2010;**49**:3412–9.
- 60 Lee YT, Glazer EC, Wilson RF, Stout CD, Goodin DB. Three clusters of conformational states in P450cam reveal a multistep pathway for closing of the substrate access channel. *Biochemistry.* 2011;**50**:693–703.
- 61 Pochapsky TC, Kazanis S, Dang M. Conformational plasticity and structure/function relationships in cytochromes P450. *Antioxid Redox Signal.* 2010;**13**:1273–96.
- 62 Gotoh O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem.* 1992;**267**:83–90.
- 63 Schuler MA, Berenbaum MR. Structure and function of cytochrome P450S in insect adaptation to natural and synthetic toxins: insights gained from molecular modeling. *J Chem Ecol.* 2013;**39**:1232–45.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Sequence alignment between CYP106A1 (chain B) and CYP106A2 from *Priestia megaterium* performed with the program CLUSTAL OMEGA [46].

Fig. S2. The respective steroids and their identified conversion products by the CYP106A1 and CYP106A2 enzymes.