The FEBS Journal

FEBS PRESS science publishing by scientists

# Structural basis of steroid binding and oxidation by the cytochrome P450 CYP109E1 from *Bacillus megaterium*

Ilona K. Jóźwik[1], Flora M. Kiss[2], Łukasz Gricman[3], Ammar Abdulmughni[2], Elisa Brill[2], Josef Zapp[4], Juergen Pleiss[3], Rita Bernhardt[2] and Andy-Mark W. H. Thunnissen[1]

1 Laboratory of Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, The Netherlands
2 Institute of Biochemistry, Saarland University, Saarbrücken, Germany
3 Institute of Technical Biochemistry, University of Stuttgart, Germany
4 Pharmaceutical Biology, Saarland University, Saarbrücken, Germany

**Keywords**
crystallography; cytochrome P450; steroid; structure-function; testosterone

**Correspondence**
R. Bernhardt, Institute of Biochemistry, Campus B2.2, 66123, Saarland University, Saarbrücken, Germany
Fax: +49 (0) 681 302 4739
Tel: +49 (0) 681 302 4241
E-mail: ritabern@mx.uni-saarland.de
and
A.-M. W. H. Thunnissen, University of Groningen, Laboratory of Biophysical Chemistry, Nijenborgh 7, 9747 AG Groningen, The Netherlands
Fax: +31 50 3634800
Tel: +31 50 3634380
E-mail: a.m.w.h.thunnissen@rug.nl

(Received 15 June 2016, revised 15 September 2016, accepted 27 September 2016)

doi:10.1111/febs.13911

Cytochrome P450 monooxygenases (P450s) are attractive enzymes for the pharmaceutical industry, in particular, for applications in steroidal drug synthesis. Here, we report a comprehensive functional and structural characterization of CYP109E1, a novel steroid-converting cytochrome P450 enzyme identified from the genome of *Bacillus megaterium* DSM319. *In vitro* and whole-cell *in vivo* turnover experiments, combined with binding assays, revealed that CYP109E1 is able to hydroxylate testosterone at position 16β. Related steroids with bulky substituents at carbon C17, like corticosterone, bind to the enzyme without being converted. High-resolution X-ray structures were solved of a steroid-free form of CYP109E1 and of complexes with testosterone and corticosterone. The structural analysis revealed a highly dynamic active site at the distal side of the heme, which is wide open in the absence of steroids, can bind four ordered corticosterone molecules simultaneously, and undergoes substantial narrowing upon binding of single steroid molecules. In the crystal structures, the single bound steroids adopt unproductive binding modes coordinating the heme-iron with their C3-keto oxygen. Molecular dynamics (MD) simulations suggest that the steroids may also bind in ~180° reversed orientations with the C16 carbon and C17-substituents pointing toward the heme, leading to productive binding of testosterone explaining the observed regio- and stereoselectivity. The X-ray structures and MD simulations further identify several residues with important roles in steroid binding and conversion, which could be confirmed by site-directed mutagenesis. Taken together, our results provide unique insights into the CYP109E1 activity, substrate specificity, and regio/stereoselectivity.

**Database**
The atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 5L90 (steroid-free CYP109E1), 5L91 (CYP109E1-COR4), 5L94 (CYP109E1-TES), and 5L92 (CYP109E1-COR).

**Enzymes**
Cytochrome P450 monooxygenase CYP109E1, EC 1.14.14.1, UniProt ID: D5DKI8, Adrenodoxin reductase EC 1.18.1.6.

**Abbreviations**
COR, corticosterone; P450, cytochrome P450 monooxygenase; SRS, substrate recognition site; TES, testosterone.

## Introduction

Cytochrome P450 monooxygenases (P450s) constitute a large superfamily of enzymes, found in all domains of life, which catalyze oxygen-mediated hydroxylation of a wide variety of aromatic and aliphatic compounds. In nature, these enzymes play essential roles in metabolic processes like steroid biosynthesis, fatty acid metabolism, or biotransformation of drugs and other xenobiotics. In the laboratory and biotechnological industry, they are considered as high potential biocatalysts, owing to their ability to selectively oxidize unreactive C-H bonds at mild conditions [1]. Their application toward cost-effective and environmentally friendly production of steroid derivatives is of particular interest, considering the wide use of such compounds as therapeutic agents [2,3]. Currently more than 300 steroid drugs are authorized, making them the most marketed group of products in the pharmaceutical industry [4]. The ability of P450s to perform regioselective and stereoselective hydroxylation of steroids has led to an ongoing search and characterization of new enzymes, in particular from prokaryotic sources, as these are more amenable to industrial application than their eukaryotic counterparts [5].

To efficiently handle P450s and to design improved variants for synthetic applications, it is crucial to understand the structure-function relationships governing their substrate specificity and regio- and stereoselectivity [6,7]. Although a wealth of functional and structural data are available for these enzymes, it has proven difficult to pinpoint the molecular determinants of their different specificities. Crystal structures of P450 enzymes from mammalian and bacterial sources have revealed a high conformational variability, despite their common overall fold [8–10]. In particular, a few characteristic, highly flexible regions around the P450 distal heme pocket (the BC loop, F and G helices together with the FG loop) are associated with substrate access, substrate binding and product release. Understanding substrate binding and conversion by P450s requires the analysis of crystallographic 'snapshots' of their open (substrate-free) and closed (substrate-bound) states, which frequently are not available for individual P450s. Combining findings from X-ray crystallography, computational docking, and MD simulations has significantly increased the understanding of mammalian and bacterial P450s [11,12]. Nevertheless, for new, soluble, prokaryotic P450s, structural characterization is indispensable for adapting them to specific biotechnological applications.

To date, only a few bacterial steroid-specific P450s have been functionally and structurally characterized. Recently published examples include CYP106A2 from *Bacillus megaterium* ATCC 13368 [13], CYP154C5 from *Nocardia farcinica* [14], CYP154C3 from *Streptomyces griseus* [15], and CYP109B1 from *Bacillus subtilis* [16]. Of these, CYP154C5 is of particular interest, as it exhibits an exceptionally high regio- and stereoselectivity to various pregnanes and androstanes, including pregnenolone, progesterone, testosterone, and androstenedione, yielding only 16α-hydroxylated steroids. Crystal structures of steroid-bound CYP154C5 reveal a narrow, nearly closed active site, mostly hydrophobic with two opposing polar regions, perfectly matching the size and polarity distribution of the steroids. The apolar/polar shape complementarity leads to a highly ordered binding mode of the steroids, with the α-face of their C16 carbon at a suitable distance from the heme-iron for allowing a radical attack by the reactive iron-oxo species (compound I) of the catalytic cycle, thus explaining the high regio- and stereoselectivity of the enzyme. Of the other two enzymes, CYP109B1 is of interest as it represents the first member of the CYP109 family of which a crystal structure has been determined. It was crystallized in an 'open' conformation with no ligands bound in the active site. Sequence alignment with different CYP109 family members, and members of related P450 families, suggested that variations in the so called BC loop, one of the common P450 regions involved in substrate binding, may primarily account for the diverse substrate specificities among these enzymes.

Recently, the genomes of *Bacillus megaterium* DSM319 [17] and ATCC 13368 were sequenced, allowing exploration of their cytochrome P450 complement. Its analysis has resulted in the identification and characterization of several novel steroid-converting P450s [18–21]. Here, we report the functional and structural properties of CYP109E1 from *B. megaterium* DSM319. Using *in vitro* and *in vivo* turnover experiments, we demonstrate that the enzyme converts testosterone to 16β-hydroxytestosterone with a high stereo- and regioselectivity. Crystal structures of CYP109E1 were determined both in steroid-free and steroid-bound states allowing a detailed analysis of the interactions and conformational changes associated with steroid binding. In addition, we employed MD simulations to test putative productive steroid-binding modes in the active site pocket of CYP109E1, leading to a better understanding of the structural determinants of the enzyme's activity.

## Results

### Identification and bioinformatic analysis of CYP109E1

The *cyp109e1* gene was identified in the genome of *B. megaterium* DSM319 by the same bioinformatic search strategy used previously for identifying *cyp106a1* [20]. A multiple amino acid sequence alignment of CYP109E1 (UniProtKB entry d5dki8) with its closest homologs is presented in Fig. 1. Classification of the enzyme into the CYP109 family was based on the conventional P450 nomenclature system [22]. Indeed, CYP109E1 shares close similarity with other characterized CYP109 family members, that is, CYP109B1 from *B. subtilis* strain 168 (47% sequence identity), CYP109A1 from *B. subtilis* strain W23 (43% sequence identity), and the three fatty acid-oxidizing proteins CYP109C1, CYP109C2, and CYP109D1 from *Sorangium cellulosum* strain So ce56 (39%, 41%, and 33% sequence identity, respectively) [16,23,24]. However, the phylogenetic tree (Fig. 2) indicates that CYP109E1 shares the closest similarity with CYP106A1 from the same organism (42% amino acid sequence identity). This close relationship, which has been noted also for other CYP109 and CYP106 family members from *Bacillus* species [25], may indicate that these enzymes convert similar substrates.

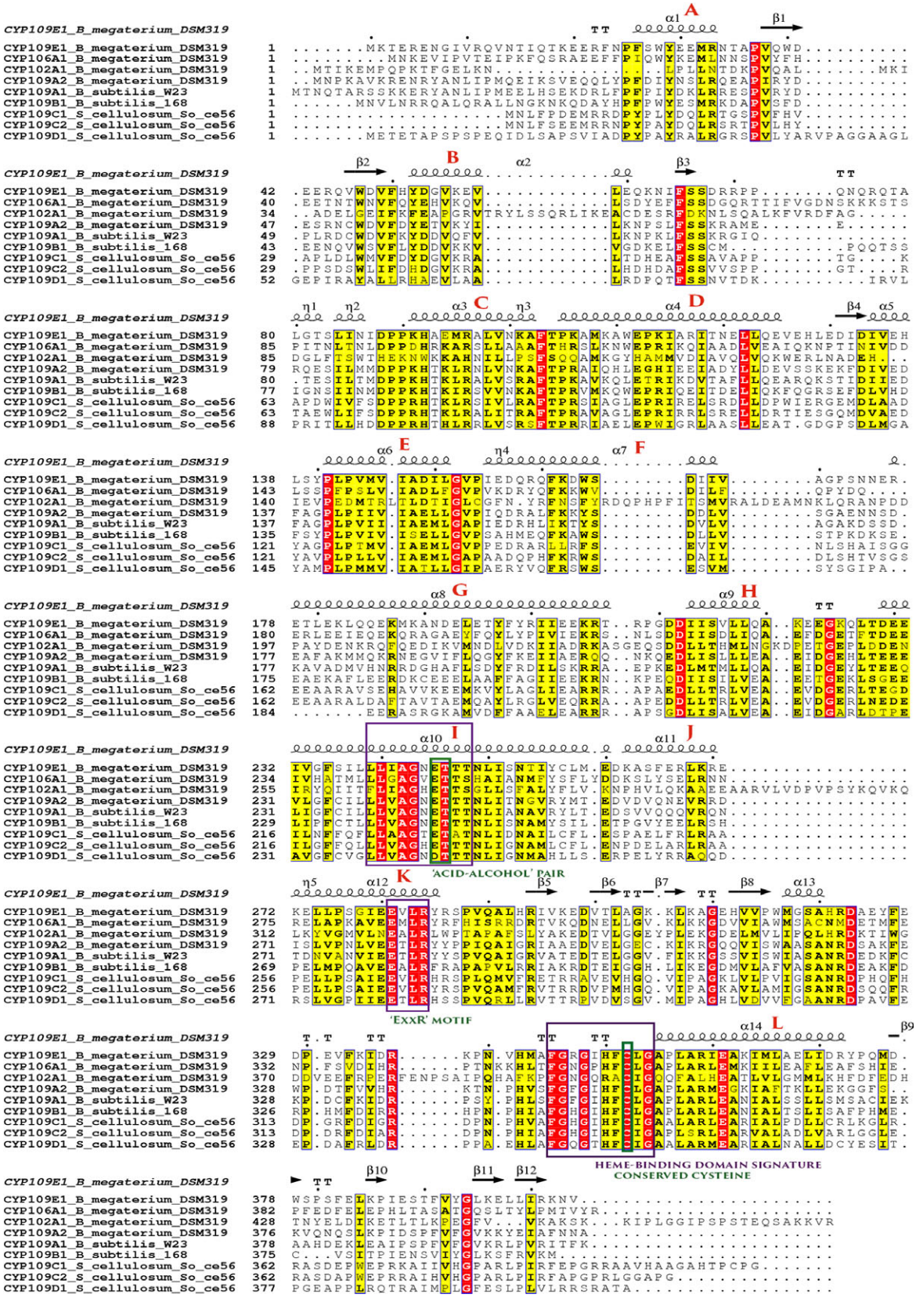
### Expression, purification, and spectroscopic characterization of CYP109E1

CYP109E1 was expressed in *E. coli* as a soluble protein, containing the gene-encoded residues 1–404, fused to a C-terminal polyhistidine tag. Purification was accomplished by employing a three-step purification strategy, including anion-exchange, size-exclusion, and mixed mode ion-exchange chromatography. Immobilized metal ion affinity chromatography was on purpose excluded from the purification protocol to eliminate the risk of imidazole affecting the functional assays and crystallization experiments. Typically, 25 mg of pure protein was obtained from 0.5 L of expression culture, with an estimated purity of > 95% as judged by SDS/PAGE. Analysis of UV-visible absorption and reduced CO-difference spectra of reduced CYP109E1 revealed all characteristic P450 peaks, confirming the structural integrity of the protein with correct incorporation of the heme cofactor. The purified protein thus obtained was used for subsequent functional and structural characterization.

### Substrate screening, conversion, and product identification

Considering the close phylogenetic distance between CYP109E1 and CYP106A1, it was proposed that steroids being substrates of CYP106A1 could also be suitable ligands for CYP109E1. The potential of CYP109E1 as a steroid hydroxylase was, therefore, tested in ligand-binding experiments using a library of 13 steroids (Table 1, see Table S2 for chemical structures). Six steroids induced a type I spectral shift upon titration to CYP109E1: androstenedione (1), corticosterone (4), deoxycorticosterone (5), dexamethasone (7), testosterone (11), and testosterone acetate (12), indicating displacement of the axial heme water and thus marking them as potential CYP109E1 substrates. However, several studies have shown that, on one hand, not all compounds inducing type I shift are substrates of P450s [19,26,27] while, on the other hand, substrates do not always induce a type I spectral shift [18]. Consequently, all 13 steroids were further subjected to an *in vitro* CYP109E1-dependent enzymatic conversion assay, using bovine adrenodoxin and adrenodoxin reductase ($Adx_{4–108}$ and AdR) as surrogate redox partners for the reconstitution of P450 activity. Most steroids were not converted by CYP109E1 (including all steroids which did not induce a type I spectral shift) or showed only very low to negligible conversion rates (< 10%). Only testosterone (11) showed substantial turnover by CYP109E1, and analysis of the reaction mixture by HPLC revealed one main (69%) and one minor (14%) product (Fig. 3), with a catalytic activity of 4.35 and 0.93 nmol product·(nmol P450)$^{-1}$·min$^{-1}$, respectively. Structure determination of these products by NMR spectroscopy required a substantial increase in their amounts, which could not easily be accomplished with the *in vitro* activity assay. This was the incentive for developing a *B. megaterium* whole-cell system overexpressing

**Fig. 1.** Multiple sequence alignment of CYP109E1 with P450s identified in *B. megaterium* DSM319 and other known CYP109 family members. Secondary structural elements are shown as in the substrate-free CYP109E1 crystal structure (helices labeled A–L). Conserved and similar residues are highlighted in red and yellow, respectively. Highly conserved, functionally relevant regions (central part of I helix, EXXR motif and heme-binding domain signature) are shown with violet frames. For all sequences shown, the UniProtKB accession numbers are the same as used in the phylogenetic tree (Figure 2) (for CYP102A1 only residues 1–472 of the heme domain are shown).


CYP109E1_B_megaterium_DSM319
CYP109E1_B_megaterium_DSM319
CYP106A1_B_megaterium_DSM319
CYP102A1_B_megaterium_DSM319
CYP102A2_B_megaterium_DSM319
CYP109A1_B_subtilis_W23
CYP109B1_B_subtilis_168
CYP109C1_S_cellulosum_So_ce56
CYP109C2_S_cellulosum_So_ce56
CYP109D1_S_cellulosum_So_ce56
'ACID-ALCOHOL' PAIR
'EXXR' MOTIF
HEME-BINDING DOMAIN SIGNATURE
CONSERVED CYSTEINE

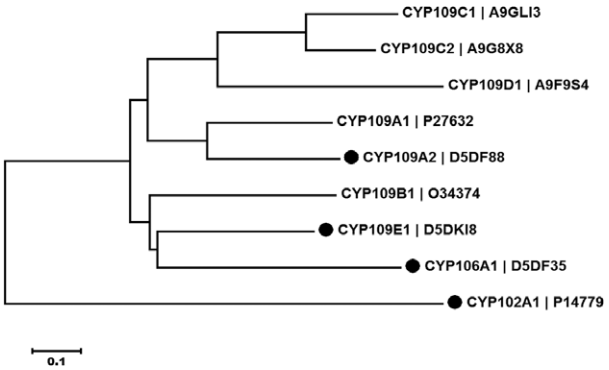



**Fig. 2.** Phylogenetic tree of the CYP109 family members and related P450s from *B. megaterium* DSM319. P450s from *B. megaterium* DSM319 are indicated with a closed circle (•). CYP109C1, CYP109C2, and CYP109D1 are from *Sorangium cellulosum* So ce56; CYP109A1 and CYP109B1 are from *B. subtilis* W23 and 168, respectively. The UniProtKB accession numbers are given next to the associated CYP names. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

CYP109E1, thus allowing the conversion of substrates under more optimal *in vivo* conditions and without the need of surrogate P450 redox partners. Testosterone turnover by CYP109E1 was successfully reproduced in the *B. megaterium* whole-cell system, enabling conversion of 26 mg of testosterone (300 μM) within 6 h to 19 mg and 5 mg of the main and minor product, respectively. Structure elucidation by NMR identified the main product as 16β-hydroxytestosterone and the side product as androstenedione (Table S3). The formation of androstenedione as a side product may be the result of a weak hydroxylation activity toward the C17 atom of testosterone, yielding a 17,17-gem-diol intermediate that subsequently dehydrates to the corresponding oxocarbon.

To investigate whether the differences in the conversion of compounds inducing a type I spectral shift are due to major differences in binding to CYP109E1, we compared the affinities of testosterone (converted substrate) and corticosterone (bound, but not converted). Quantitative analysis of the spectral shift titration data revealed that both compounds displayed very similar binding affinities to CYP109E1 with a dissociation constant ($K_D$) of 105 ± 10 μM and 91 ± 8 μM, respectively (Fig. 4). Thus, the highly selective activity of CYP109E1 toward testosterone, versus no conversion of corticosterone, is not correlated with a difference in steroid-binding affinity. In summary, the functional experiments revealed that CYP109E1 displays a rather narrow steroid-binding specificity, with relatively weak-binding affinities, and is able to hydroxylate testosterone to 16β-hydroxytestosterone with very high regio- and stereoselectivity.

**Table 1.** Steroid binding and *in vitro* conversion activity of CYP109E1 toward selected steroids. Chemical structures of the steroids are shown in supplementary Table S2. The + and – indicate positive and negative outcomes of what is stated in the column headers.

| Compound | Type I shift | *In vitro* conversion |
|---|---|---|
| Androstenedione (1) | + | + (< 10%) |
| Cortisol (2) | – | – |
| Cortisone (3) | – | – |
| Corticosterone (4) | + | – (< 2%) |
| Deoxycorticosterone (5) | + | + (< 7%) |
| 11-deoxycortisol (6) | – | – |
| Dexamethasone (7) | + | – |
| Prednisolone (8) | – | – |
| Prednisone (9) | – | – |
| Progesterone (10) | – | – |
| Testosterone (11) | + | + |
| Testosterone acetate (12) | + | + (< 10%) |
| 19-nortestosterone (13) | – | – |

## Overall structure of steroid-free CYP109E1

To evaluate the structural basis for the selectivity of steroid binding and conversion, crystallization of CYP109E1 in the presence and absence of a substrate was performed. Purified CYP109E1 was successfully crystallized in the absence of steroids, and its crystal structure was solved and refined to a 2.55 Å resolution (see Table 2 for the relevant data collection and refinement statistics). The CYP109E1 crystals contain two protein molecules in the asymmetric unit. Both polypeptide chains, and their associated heme cofactors, are well defined in the electron density, except for residues 1–20 and the C-terminal $His_6$-tags. Since the two crystallographically independent CYP109E1 molecules have nearly identical conformations [the root mean square deviation (RMSD) of Cα-backbone atoms is 0.34 Å for 374 aligned residues], we will restrict the description to only one of them. CYP109E1 adopts the characteristic triangular 'P450-fold' and contains 13 α-helices, two $3_{10}$ helices, and 10 β-strands (arranged in a five-stranded sheet, three-stranded sheet, and a β-finger, Fig. 5A). The overall fold of the P450 enzyme is well conserved: the closest structural homolog of CYP109E1 is substrate-free CYP109B1 from *B. subtilis* (PDB entry 4rm4, RMSD of Cα-backbone atoms of 1.19 Å for 347 common residues). A unique feature of the CYP109E1 structure is presented by its G helix (following nomenclature of secondary

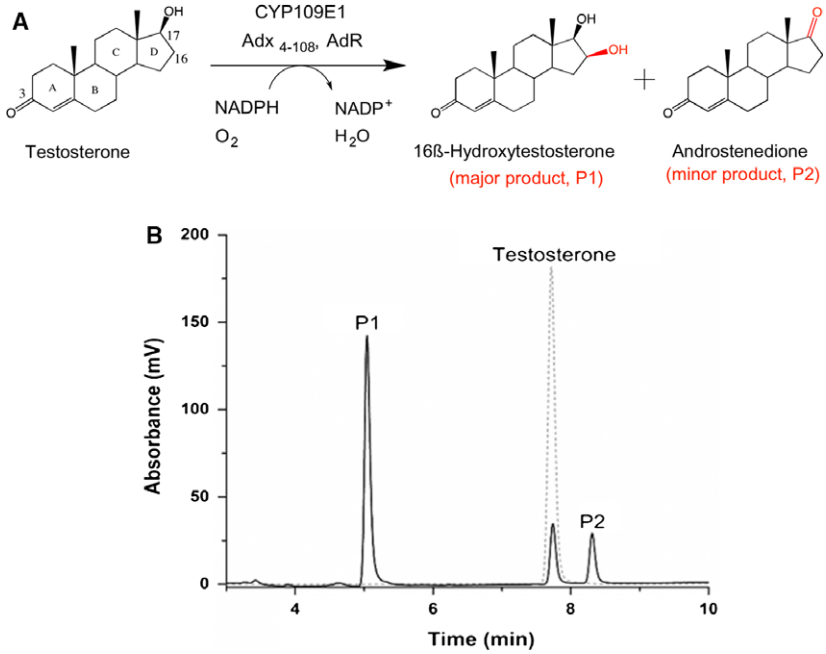



**Fig. 3.** *In vitro* conversion of testosterone by purified CYP109E1. (A) Schematic representation of the reaction and (B) HPLC chromatogram.

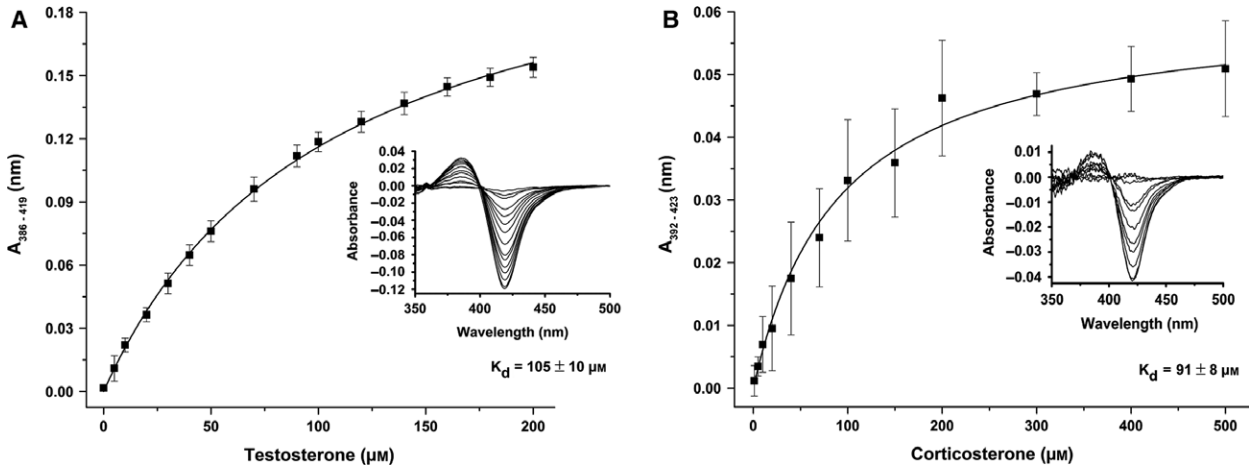



**Fig. 4.** Type I spectral shifts induced by steroid binding to CYP109E1 and the derived binding curves. (A) Binding curve of testosterone. (B) Binding curve of corticosterone. The spectral shifts upon titration of the steroids are shown as insets. The enzyme solution (10 µM, 50 mM potassium phosphate buffer, pH 7.4) was titrated with increasing amounts of substrates dissolved in DMSO. The peak-to-through absorbance differences were plotted against the increasing concentrations of the substrate. Plotted data points are the mean values from three independent measurements. Error bars represent the standard deviations. The curves were fitted by hyperbolic regression.

structural elements established by Poulos *et al.* [28]), which is significantly longer than in other bacterial P450 structures. A structural comparison with other P450s further reveals that steroid-free CYP109E1 adopts an open conformation with a spacious active site pocket at the distal side of the heme, which is freely accessible to solvent (Fig. 5B). The heme cofactor is bound between the I and L helices, with the thiolate group of Cys352 serving as the fifth ligand of the heme-iron. No electron density was observed for a water molecule coordinating the heme-iron as a sixth ligand. Instead, a small blob of electron density was visible at about 4 Å distance from the heme, but the density was of insufficient quality to allow identification of the bound ligand. The noncovalent interactions between CYP109E1 and its heme are highly similar as in other class I P450s. Stabilization is provided by van der Waals interactions between the heme core and several apolar protein side chains, and by ion pairs formed between

**Table 2.** Data collection, refinement, and model statistics of CYP109E1.

| | Substrate-free CYP109E1 | CYP109E1-COR4 | CYP109E1-TES | CYP109E1-COR |
|---|---|---|---|---|
| Conformation | Open | Open | Closed & open | Closed & open |
| PDB code | 5L90 | 5L91 | 5L94 | 5L92 |
| Model statistics | | | | |
| Monomers in the AU | 2 | 2 | 2 | 2 |
| Solvent content (%) | 61.7 | 60.2 | 49.1 | 49.1 |
| Ligands | n/a | Four corticosterone molecules/monomer | Testosterone (TES) (in chain A) | Corticosterone (COR), malonic acid (MLA) (in chain A) |
| Data collection | | | | |
| X-ray source (ESRF/ in-house) | ID29 | ID23-2 | In-house | ID23-1 |
| Wavelength (Å) | 1.04541 | 0.87261 | 1.5418 | 0.97241 |
| Resolution range (Å) | 60–2.55 (2.65–2.55)[a] | 49–2.20 (2.26–2.20) | 56–2.25 (2.32–2.25) | 48–2.10 (2.16–2.10) |
| Space group | P $3_2$ 2 1 | P $3_2$ 2 1 | P 1 $2_1$ 1 | P 1 $2_1$ 1 |
| Unit-cell parameters | | | | |
| a, b, c (Å) | 121.3, 121.3, 144.2 | 120.4, 120.4, 140.8 | 60.4, 134.9, 61.9 | 60.5, 135.7, 61.6 |
| α, β, γ, ° | 90, 90, 120 | 90, 90, 120 | 90, 113.8, 90 | 90, 114.4, 90 |
| Observed reflections | 293 843 (30 950) | 355 244 (26 012) | 293 196 (26 695) | 156 293 (12 345) |
| Unique reflections | 40 454 (4508) | 60 153 (4384) | 42 810 (3945) | 51 292 (12 345) |
| Multiplicity | 7.3 (6.9) | 5.9 (5.9) | 6.8 (6.8) | 3.0 (2.9) |
| Completeness (%) | 99.9 (99.9) | 99.7 (100) | 99.7 (99.1) | 97.4 (98.8) |
| $\langle I/\sigma(I)\rangle$ | 11.6 (2.6) | 15.9 (2.3) | 12.2 (2.7) | 23.3 (4.1) |
| $R_{merge}$ (%) | 9.6 (80.2) | 8.7 (66.9) | 14.2 (73.9) | 2.8 (22.2) |
| CC (1/2) (%) | 99.8 (84.8) | 99.8 (80.1) | 99.6 (83.7) | 99.9 (91.4) |
| Refinement | | | | |
| $R_{work}$ (%) | 21.44 | 18.48 | 21.90 | 20.09 |
| $R_{free}$ (%) | 25.88 | 23.21 | 26.28 | 24.31 |
| R.m.s.deviation, bond lengths (Å) | 0.017 | 0.014 | 0.014 | 0.014 |
| R.m.s.deviation, bond angles (°) | 1.679 | 1.322 | 1.252 | 1.299 |
| Average B-factors (Å)$^2$ | | | | |
| Overall | 73.65 | 46.81 | 40.16 | 42.65 |
| Protein | 73.96 | 46.49 | 40.45 | 42.92 |
| Heme | 48.29 | 28.30 | 27.39 | 28.22 |
| Testosterone (TES) | n/a | n/a | 35.29 | n/a |
| Corticosterone (COR) | n/a | n/a | n/a | 43.43 |
| COR-1 | n/a | 48.66 | n/a | n/a |
| COR-2 | n/a | 77.89 | n/a | n/a |
| COR-3 | n/a | 86.81 | n/a | n/a |
| COR-4 | n/a | 56.60 | n/a | n/a |
| Malonic acid (MLA) | n/a | n/a | n/a | 42.14 |
| Ramachandran plot statistics | | | | |
| Most favored (%) | 94.9 | 95.6 | 96.1 | 96.6 |
| Allowed regions (%) | 4.8 | 4.3 | 3.9 | 3.4 |
| Disallowed regions (%) | 0.3 | 0.1 | 0.00 | 0.00 |
| Molprobity overall score | 1.62 | 1.54 | 1.22 | 1.14 |

[a] Values in parentheses refer to data in the highest resolution shells.

the heme propionates and the side chains of highly conserved residues His92, Arg96, Arg294, and His350.

## Active site pocket of steroid-free CYP109E1

The geometry of the active site pocket in the open form of CYP109E1 was analyzed in more detail. As shown in Fig. 5B, it resembles a funnel, wide open at the entrance but becoming more constricted as it leads toward the heme. The structural regions surrounding the active site pocket are equivalent to the six substrate recognition sites commonly found in P450s [29]: the BC loop (substrate recognition site 1 or SRS1), parts of the F and G helices (SRS2 and SRS3, respectively), the central segment of the I helix (SRS4), the

region connecting helix K and strand β5 (SRS5) and the β9–β10 turn (SRS6). Close to the heme, the active site is very hydrophobic and its surface is roughly built of two rings of residues. The first ring (closest to the heme) contains residues Ile85 (BC loop); Leu238, Ala242, and Thr246 (I helix); Pro288, Val289, and Leu292 (K-β5 loop). Many of these residues are highly to moderately conserved among P450s, for example, Thr246, Ala242, or Val289, in agreement with common roles in catalysis and substrate binding [30]. The second ring contains residues Arg69, Leu80, and Asn86 (BC loop); Ile168 and Val169 (F helix); Ile241 and Glu245 (I helix); Phe391 and Val392 (β9–β10 turn). Residues Arg69, Asn86, and Glu245 form two polar-charged surface patches located at opposite sides of the active site pocket. Further away from the heme, near the entrance of the active site pocket, residues are mostly polar charged. Active site residues in the second ring, and residues at the entrance of the active site pocket in CYP109E1, show substantial divergence with respect to other P450s, including the testosterone-hydroxylating enzymes like CYP109B1 from *B. subtilis* or CYP154C5 from *N. farcinica*.

## Structures of CYP109E1 with single bound corticosterone or testosterone

To characterize how steroids interact with CYP109E1, crystal structures of CYP109E1 with bound corticosterone (type I shift, but no substrate) or testosterone (type I shift, substrate) (CYP109E1-COR and CYP109E1-TES) were obtained by cocrystallization and refined at 2.25–2.1 Å resolution (see Table 2 for details). Like the steroid-free crystals, the steroid-bound cocrystals contain two polypeptide chains per asymmetric unit (solvent content of 49%), yet they belong to a different space group and show a different crystal packing geometry. Only one of the two unique CYP109E1 protein molecules in each cocrystal has a bound steroid. The protein molecules lacking a bound steroid adopt an open conformation similar to that of steroid-free CYP109E1. No electron density is present for the BC loop, FG loop, and several residues at the end of the G helix, indicating that these regions are significantly disordered. In the protein molecules with a bound steroid, a remarkable narrowing and partial closure of the active site pocket is observed, allowing most substrate-binding regions to closely approach and interact with the steroids. The F and G helices, together with the FG loop, show the largest displacement, while the BC loop and β9–β10 turn show smaller readjustments (Fig. 6A). The central part of the BC loop (residues 71–75) is highly disordered and could not be resolved in the electron density map. The conformational change in the FG helices is accompanied by reorientations of the H helix and HI loop. In addition, a local widening of the I helix is observed at residues 242–246, creating a groove which allows close packing of the steroids near the heme. A similar groove in the I helix has also been observed in other

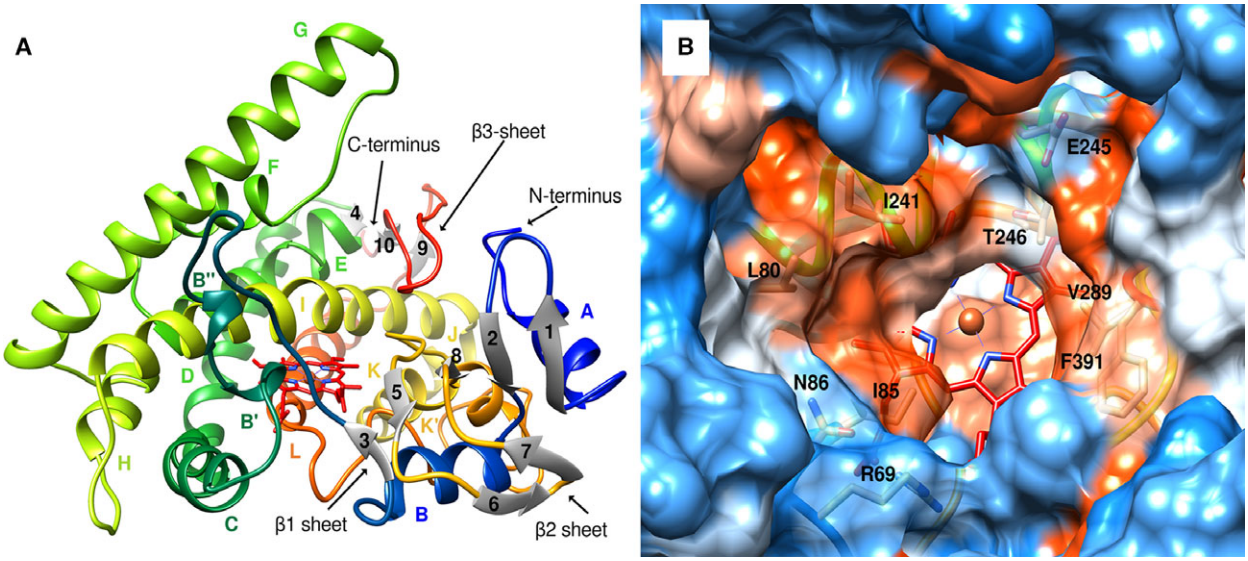



**Fig. 5.** Overall structure and the active site of CYP109E1. (A) Overall structure shown in ribbon representation, with rainbow coloring from blue N′ terminus to red C′ terminus. The heme cofactor is shown in red and beta strands in gray. Secondary structural elements are labeled following the common P450 nomenclature. (B) Active site pocket of substrate-free CYP109E1. Coloring apolar–polar, orange-blue.

substrate-bound bacterial P450 structures, and its substrate-induced conformation is believed to play an important role in oxygen activation and proton delivery [31]. Interestingly, four ordered water molecules are found at the interface of the I and E helix, two of which are located in the groove of the I helix (Fig. 7). The four waters form a continuous hydrogen-bonded network, which includes the main-chain carbonyls of Ala242 and Gly243, the side-chain hydroxyl groups of Thr246 and Thr247, and the main-chain amide of Thr248. Although the network is not directly connected to bulk solvent (the distance with the nearest water bound at the protein surface is 5.3 Å), it correlates well with the so called 'solvent channel' found in many P450s, located between the F, E, and I helices, which is believed to serve as a water access channel and/or proton delivery network [32].

Electron density of the bound steroids was of sufficient quality to allow a clear characterization of their binding modes (Fig. 6B, C). Both steroids show an unproductive binding mode: the corticosterone and testosterone molecules are bound in roughly perpendicular orientations relative to the heme plane, with their C3-keto oxygen atoms coordinating the heme-iron at the sixth axial position. The β-faces of the steroids are oriented toward the I helix, while the C17 substituents are pointing away from the heme toward the entrance of the active site pocket. The two steroids are mainly bound by van der Waals and hydrophobic interactions. Residues in the active site pocket making hydrophobic interactions with the steroids are Leu80 (BC loop, SRS1), Ile168, Val169 (F helix, SRS2), Leu238, Ile241, Ala242, Thr246 (I helix, SRS4), Val289, Ala291, His293 (K/β5 connecting region, SRS5), and Phe391, Val392 (β9–β10 turn, SRS6). No direct hydrogen bonds are observed between the protein and the two steroid molecules. In the structure with bound corticosterone, a malonic acid molecule is present in the active site forming van der Waals contacts with ring B of the steroid (Fig. 6C). The malonic acid molecule is further stabilized by hydrogen bonds with His293 (SRS5) and one of the heme-propionate groups, and by van der Waals contacts with residues Arg69, Ile85 (SRS1), Leu292 (SRS5), and Phe391 (SRS6). Its presence is most likely a crystallographic artifact, as malonic acid is the major component of tacsimate, the reagent used in the crystallization of CYP109E1-COR.

## Multiple corticosterone-bound CYP109E1 structure

An additional structure of corticosterone-bound CYP109E1 was obtained at 2.2 Å resolution by soaking the steroid into pregrown, steroid-free crystals (CYP109E1-COR4, Table 2). Interestingly, the corticosterone-soaked crystals display a water as the sixth axial ligand of the heme-iron and multiple steroid binding (four corticosterones, Fig. 8). For clarity, we abbreviate the ligands as COR-1, COR-2, COR-3, and COR-4. COR-1 is located closest to the heme-iron, while COR-4 is most distant. The protein molecules in CYP109E1-COR4 adopt an open state, with only a minor repositioning of the G helix toward the bound steroids. The limited conformational changes compared to steroid-free CYP109E1 are not surprising, since crystal packing interactions between the neighboring protein molecules lock the F and G helices, FG loop, and BC loop in an 'open' position (Fig. 8A). It is remarkable, however, that the open state allows ordered binding of multiple steroids.

The four steroids are bound in different orientations and show substantial intermolecular van der Waals contacts between their steroid rings. COR-1 is pointing with its bulky C17-substituent toward the heme and its C21-hydroxyl group forms a hydrogen bond with a water molecule coordinating the heme-iron (Fig. 8B). Additional interactions of this steroid at the active site pocket include two direct hydrogen bonds, between the C3-keto oxygen and the side chain of Lys187 (SRS3) and between the C11-hydroxyl group and the main-chain carbonyl oxygen of Ile241 (SRS4), two water-mediated hydrogen bonds, between the C11-hydroxyl group and the side chain of Glu245 and between the C21-hydroxyl group and the main-chain carbonyl group of Ala242 (SRS4), as well as several hydrophobic contacts with residues Ile85 (SRS1), Ile168, Val169 (SRS2), Thr246 (SRS4), Val289 (SRS5), and Val392 (SRS6). The other three corticosterones also make specific contacts with the protein (interactions described in detail in Table S4).

## Modeling of regio- and stereoselectivity of CYP109E1

In the CYP109E1-TES crystal structure, the shape of the electron density for the bound steroid suggested that testosterone may perhaps adopt an alternative, reversed binding mode, in which the C16 and C17 carbons are located close to the Fe atom (~4 Å), consistent with the observed activity of CYP109E1. Addition of such a putative productive binding mode in the refined CYP109E1-TES structure resulted in a poor fit to the observed electron density, thus, we assume that testosterone predominantly adopts the unproductive binding mode in the protein crystals. Therefore, to explore the molecular basis of its selective steroid conversion and regio/stereoselectivity, MD simulations of

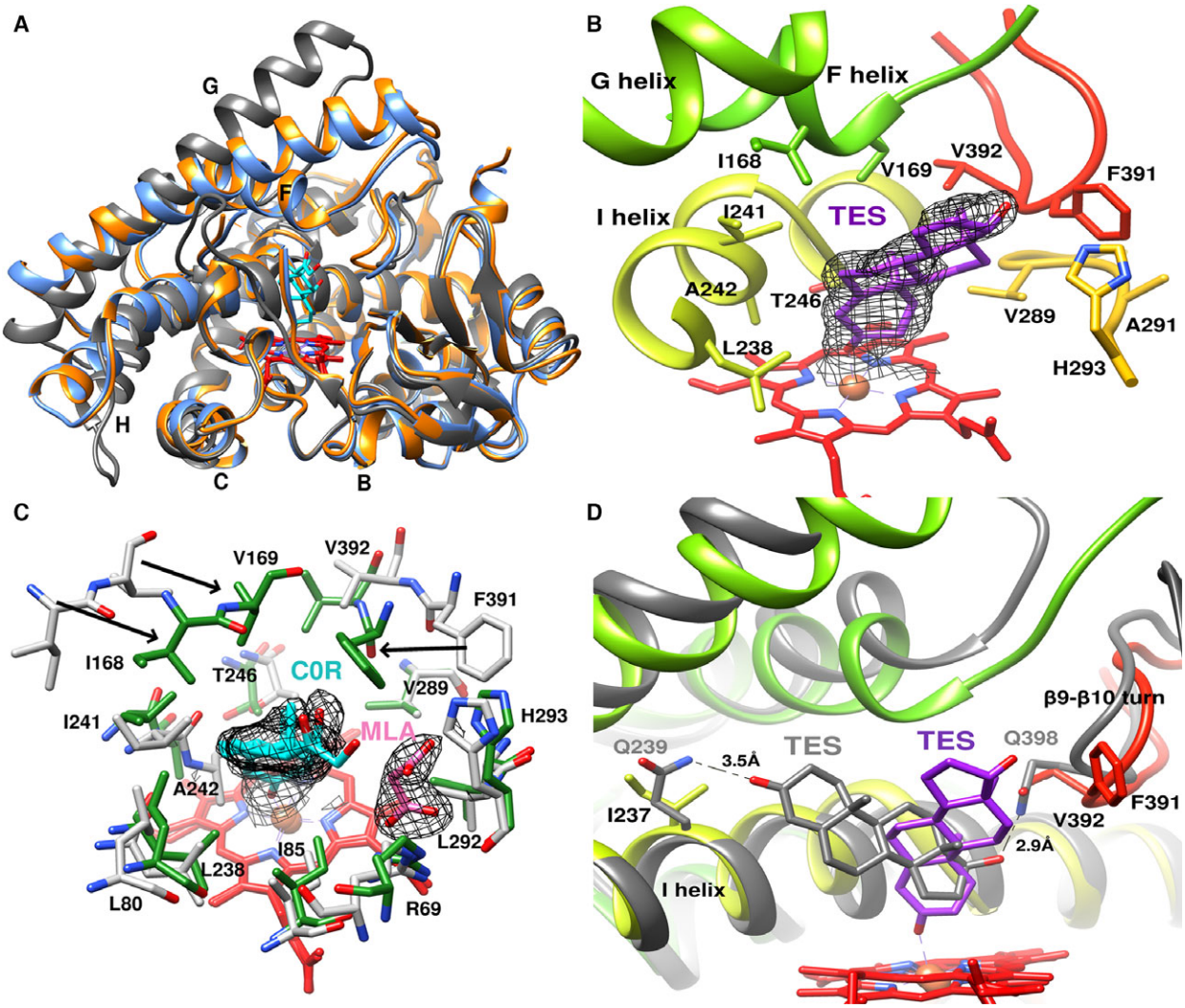



**Fig. 6.** Conformational changes and steroid-binding modes in CYP109E1. (A) Superposition of substrate-free CYP109E1 (gray), CYP109E1-TES (light blue), and CYP109E1-COR (orange) showing the open–closed conformational changes. COR is shown as cyan stick model. (B) Amino acid residues involved in testosterone binding (CYP109E1-TES) are shown and TES is shown as violet stick model. Black mesh is the composite omit $2F_o - F_c$ electron density map, calculated at 2.25 Å resolution and contoured at 1σ. (C) Residues binding single corticosterone and malonic acid in the active site of CYP109E1-COR (green) in comparison to substrate-free structure (gray), with COR shown in cyan and MLA in light pink. Black mesh is the composite omit $2F_o - F_c$ electron density map, calculated at 2.1 Å resolution and contoured at 1σ. (D) Comparison of testosterone binding modes observed in CYP109E1 (violet stick model) and CYP154C5 from *N. farcinica* (in gray, PDB entry 4j6d). Amino acid residues providing stabilizing interactions in CYP154C5 are shown and compared to their structural homologues in CYP109E1. Distances are given in ångstroms next to the black dashed lines.

the CYP109E1-oxoferryl species (compound I) in complex with corticosterone or testosterone were performed. Starting models for the MD simulations were based on the CYP109E1-TES structure in which the crystallographically observed testosterone molecule was replaced by docked steroids (corticosterone or testosterone) in reversed orientations, with the C16 carbon atom or C17 substituents oriented toward the heme-iron. Steroid conversion and regio/stereoselectivity of the CYP109E/steroid complexes were modeled by assigning each frame of the MD simulation as being in either a near-attack conformation or in a non-productive conformation, on the basis of distance and angle cutoffs, following previously published procedures [33–35]. The number of near-attack conformations was the highest for MD simulations with testosterone, where 16% of the simulation frames revealed near-attack conformations suggesting formation of 16β-hydroxytestosterone (pro-16β, 99%) or androstenedione (pro-17α, 1%). No pro-16α near-

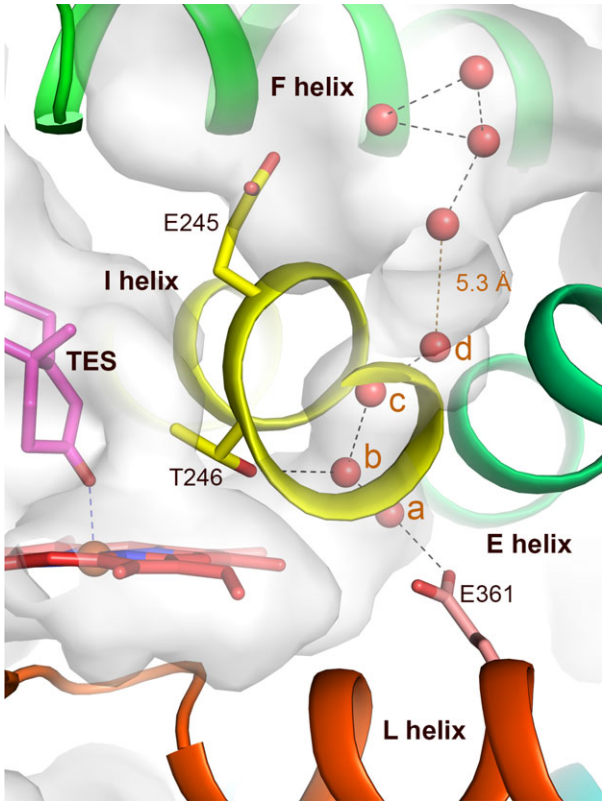



**Fig. 7.** Putative proton relay network in CYP109E1. Local distortion in the central part of I helix, upon binding of single testosterone or corticosterone (CYP109E1-TES is depicted here) allowed for binding of four ordered water molecules in the back of I helix (labeled a, b, c, and d). The putative proton delivery pathway from conserved Thr246 to bulk solvent is shown, involving water molecules found in the 'solvent channel' between the F, E and I helices. The solvent accessible surface is shown (calculated after removing all waters, testosterone and the heme from the structure). For clarity, only a selection of the hydrogen bonds of the waters with protein atoms are shown. Residue Glu245 does not participate in the hydrogen bond network.

attack conformations occurred in the simulations, in accordance with the observed stereoselectivity of the enzyme. Interactions of one of the most frequently visited testosterone-binding poses were analyzed in detail (Fig. 9). The testosterone molecule is positioned optimally for abstraction of the C16β-hydrogen, and forms hydrophobic interactions with the same residues as in the CYP109E1-TES crystal structure, that is, Leu80 (SRS1), Ile168 (SRS2), Ile241 and Thr246 (SRS4), Val289 (SRS5), Phe391 and Val392 (SRS6). In addition, the testosterone C3-keto oxygen is in hydrogen-bonding distance to Lys187 (SRS3). In the case of corticosterone, only 0.8% of the MD simulation frames showed near-attack conformations (suggesting generation of a 16α-hydroxylated product). The results are in concordance with the lack of experimentally observed corticosterone conversion by CYP109E1.

### Site-directed mutagenesis

To substantiate the crystallographic and modeling results, single alanine mutations were prepared of a few selected residues (Val169, Lys187, Ile241, Glu245, and Thr246) in the active site pocket of CYP109E1 and their effect on CYP109E1-catalyzed conversion of testosterone was analyzed using the *in vitro* activity assay (Table 3). Replacement of Val169 and Ile241 by alanine resulted in almost complete abolishment of 16β-hydroxytestosterone production, confirming the importance of these residues for productive steroid binding. The K187A mutation also caused a decrease in activity compared to the wild-type enzyme, but the effect is much smaller than for the V169 and I241 mutations. Thus, the hydrogen bond of the testosterone C3-keto group with the side chain of Lys187, as observed in the MD simulations, is not a crucial interaction for productive binding of testosterone. Interestingly, alanine mutations of Glu245 and Thr246 (the conserved 'acid-alcohol' pair) led to opposite effects on CYP109E1 activity toward testosterone. While the T246A mutation resulted in a drastic decrease in 16β-hydroxytestosterone production, the E245A mutation did not significantly affect the CYP109E1 activity. These results support the relevance of the water channel observed in the single steroid-bound CYP109E1 structures, and implicates a role for Thr246, but not Glu245, in proton delivery and oxygen activation within the active site.

## Discussion

Only a few bacterial P450s have been characterized that are able to hydroxylate testosterone at different positions in the steroid skeleton, with high regio- and stereoselectivity [36]. Recent examples include CYP109B1 from *B. subtilis* and CYP154C5 from *N. farcinica*, which produce 15β- and 16α-hydroxylated testosterone, respectively [37,38]. CYP109E1 from *B. megaterium* is the first example of a wild-type bacterial P450 showing highly selective 16β-hydroxylase activity toward testosterone. Production of 16β-hydroxytestosterone has previously been observed with specific P450 BM3 mutants, obtained by protein engineering, but these mutants do not display the same high level of regio- and stereoselectivity as CYP109E1 [39–41]. Similarly, various mammalian P450s are able to hydroxylate testosterone at position 16β, but they

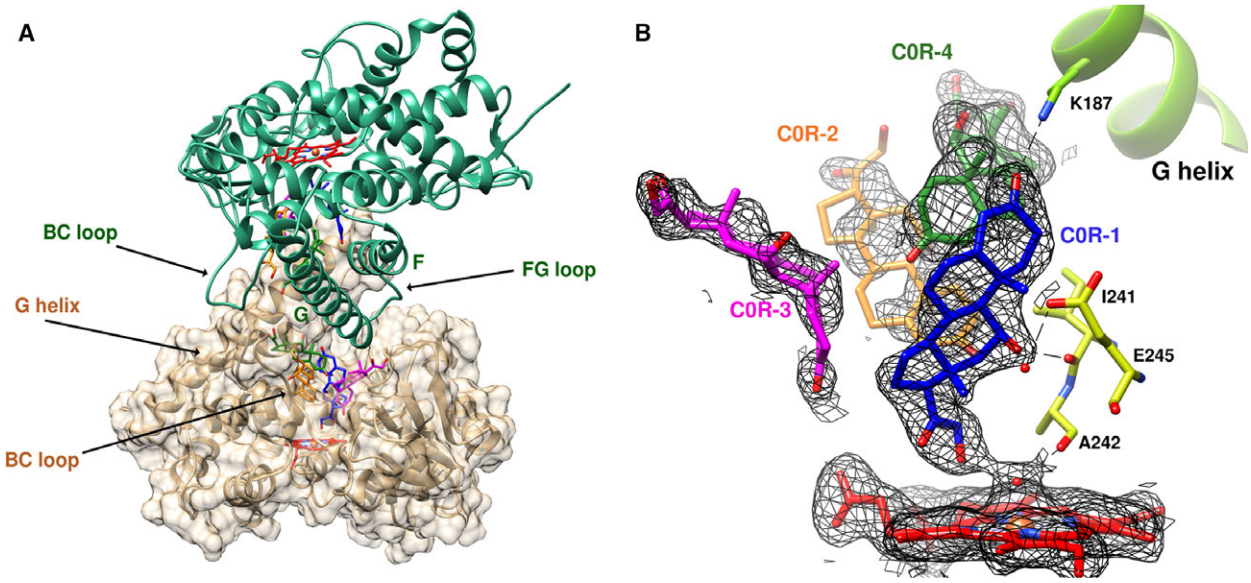



**Fig. 8.** Multiple corticosterone binding in CYP109E1. (A) Representation of intermolecular packing interactions in the crystal lattice of CYP109E1-COR4. The mobility of the F and G helices, FG and BC-loop is restricted, thus locking the CYP109E1 molecules in an open state. (B) Close up on multiple corticosterone binding orientations in CYP109E1 and recognition of COR-1. Hydrogen-bonding interactions of COR-1 are shown as black dashed lines. Black mesh is the composite omit $2F_o - F_c$ electron density map, calculated at 2.2 Å resolution and contoured at 1σ.

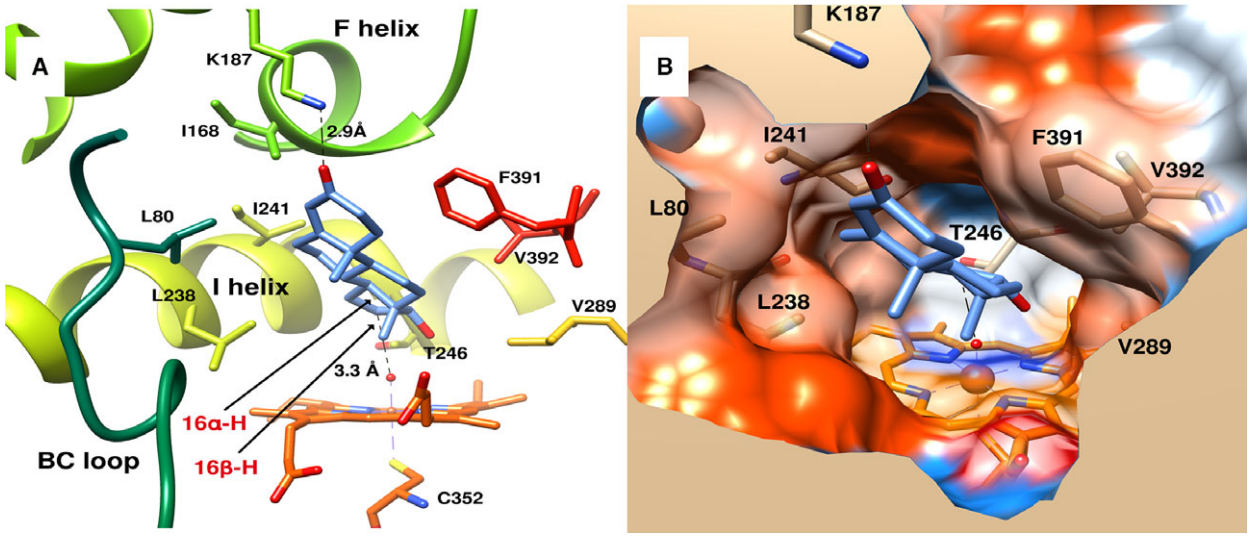



**Fig. 9.** Main productive testosterone binding mode in CYP109E1. (A) Representation of the most frequently visited testosterone conformation in the hydrophobic active site of CYP109E1 during molecular dynamics simulations. The C16 atom of testosterone is at an appropriate distance and angle from the oxoferryl oxygen atom to allow abstraction of its 16β-hydrogen, in accordance with the formation of the main observed turnover product, 16β-hydroxytestosterone. Testosterone is drawn in light blue-colored sticks and the hypothetical heme oxoferryl moiety (Fe(IV) = O or compound I) is in orange. The locations of the two C16-hydrogens targeted in predictions by MD simulations are shown by arrows. Distances are given in ångstroms next to the black dashed lines. (B) Surface representation of the same binding mode as depicted in panel A, coloring apolar–polar, orange-blue.

lack sufficient regio- and stereoselectivity. Low solubility and generally low expression levels of mammalian P450s further hinder their biotechnological use [42]. In contrast, the successful development of a CYP109E1-catalyzed whole-cell system for the production of 16β-hydroxylated testosterone, as reported here,

**Table 3.** Effect of introduced mutations on production of 16-β-hydroxytestosterone (16β-OH-TES) by CYP109E1 enzyme variants. The *in vitro* reactions were carried out in 50 mM potassium phosphate buffer with 2% glycerol, pH 7.4, at 30 °C for 30 min. Bovine $Adx_{4-108}$ and AdR were used as redox partners and 200 μM of substrate dissolved in DMSO was added. Shown are the mean values and standard deviations (±SD) of three independent measurements.

| CYP109E1 variant | Catalytic activity for the formation of 16β-OH-TES, nmol·(nmol P450)$^{-1}$·min$^{-1}$ |
|---|---|
| WT | 4.35 ± 0.02 |
| K187A | 1.43 ± 0.06 |
| V169A | 0.07 ± 0.02 |
| I241A | 0.01 ± 0.01 |
| E245A | 4.69 ± 0.04 |
| T246A | 0.12 ± 0.04 |

demonstrates the potential of this bacterial enzyme for biotechnological applications.

To further improve CYP109E1 for biotechnological purposes and to better understand its structure-function relationships, crystal structures of CYP109E1 with and without steroids were obtained, revealing interesting features related to steroid binding and protein conformational dynamics. The structures confirm the general view that P450s possess a highly dynamic active site, which exists in a primarily open state in the absence of bound substrate, but changes toward a more closed state when the substrate is bound. Unexpectedly, and for the first time, crystal structures have revealed a P450 distal pocket with either four ligand molecules or only a single ligand molecule bound in a distinct way. Arguably, the presence of four bound corticosterone molecules in the open active site pocket of CYP109E1 is a crystallographic artifact, as it is influenced by the high concentration of the steroid in the crystal-binding experiment. Furthermore, crystal-packing interactions in the CYP109E1-COR4 structure effectively lock CYP109E1 in an open conformation, resisting the conversion into the closed structure expected upon substrate binding. On the other hand, the crystallographically observed binding of multiple corticosterone steroids to the open form of CYP109E1 may represent a snapshot of the initial substrate recognition and binding events. Multiple substrate binding occurs in other P450s, and has been implicated in the mechanisms of homo- and heterotopic cooperativity observed in mammalian P450s enzymes [43,44]. There are no indications, however, for cooperative steroid binding by CYP109E1, as evident from the hyperbolic binding curves derived from the P450 spectral shift titration assays (no sigmoidal fit).

The single steroid-bound structures of CYP109E1 in the closed conformation, complemented with the MD simulations and site-directed mutagenesis results, provide clear insights on how this enzyme accomplishes selective 16β-hydroxylation of testosterone. The narrow shape of the active site pocket in the closed form of the enzyme, and the almost exclusively hydrophobic surface of the active site walls near the heme, restrict the binding modes of the steroids to an orientation in which their longitudinal axis is roughly perpendicular to the heme plane. Two main binding modes are possible for 3-oxo-$\Delta^4$-steroids like testosterone and corticosterone, with either the C3-keto oxygen down (toward heme) and C17-substituent up (away from heme) or with the C17-substituent down and C3-keto oxygen up. The first binding mode may have a possible inhibitory effect, by preventing oxygen binding to the heme-iron. Although it should be noted that coordination of the keto C3-group is less likely to occur with reduced iron in the ferrous state. The second binding mode may lead to 16β-hydroxylation for steroids carrying a small C17 substituent, like testosterone, as supported by the MD simulations which show a substantial number of pro-16β productive conformations. Bulky polar C17 substituents (as in corticosterone) most probably cause a steric hindrance, blocking an optimal approach of the C16 carbon to the heme-iron, and result in steroids being bound in unproductive conformations (as supported by our MD simulations with corticosterone). For testosterone, a third, minor productive binding mode positions the α-face of the C17 carbon close to the heme-iron, (in MD simulations a few pro-17α near-attack conformations were observed), so that a second hydroxylation at this position can occur followed by removal of a water molecule, explaining the production of androstenedione as a side product. A similar oxidation reaction has very recently been observed with CYP106A1 where an 11-oxidase activity toward 11-hydroxy steroids was demonstrated [19,45].

The molecular basis for the stereoselectivity of CYP109E1 toward 16β-hydroxylation of testosterone is further clarified by a structural comparison with CYP154C5, which converts testosterone to 16α-hydroxytestosterone. The shape and volume of the active site pockets in the testosterone-bound crystal structures of these two enzymes are significantly different, which is primarily due to the highly variable BC loop. This loop is much longer in CYP154C5 and functions as a lid that almost completely locks the active site pocket. A closure to this extent is not observed in the testosterone-bound CYP109E1 structure. Although near to the heme, the active site pocket in testosterone-bound CYP109E1 is more restricted

compared to CYP154C5, near its entrance, it is more open and accessible to solvent. The different shapes of the active site pockets are coupled to different orientations of the bound steroids relative to the heme plane: roughly perpendicular in the CYP109E1-TES structure, while more-or-less parallel in the CYP154C5-TES complex. While the perpendicular binding mode of testosterone is optimal for C-H abstraction from the β-face of the C16 carbon, the parallel binding orientation in CYP154C5-TES positions the α-face of the testosterone C16 carbon close to the heme-iron, explaining the difference in stereoselectivity of these enzymes [14]. The parallel binding orientation of testosterone in CYP154C5 is stabilized by interactions of the apolar steroid ring system with hydrophobic residues from two opposite regions (BC loop and I helix) in the active site pocket. The polar C17 substituent forms a hydrogen bond with Gln398 (from SRS6), while the C3 substituent is bound near to Gln239 in a solvent-accessible pocket. In CYP109E1, the equivalent residues to Gln239 and Gln398 of CYP154C5 are Ile237 and Val392, respectively (Fig. 6D). Lack of polar residues in the active site pocket near the heme thus prevents a similar binding orientation of testosterone in CYP109E1 as in CYP154C5. In addition, the side chains of Leu80 (BC loop) and Ile241 (I helix) in CYP109E1 pose a steric hindrance to the incoming steroid molecule, favoring a perpendicular over a parallel binding orientation. Thus, the difference in stereoselectivity between CYP109E1 and CYP154C5 toward C16-hydroxylation of testosterone is a result of the differences in the shape and the apolar/polar surface distribution of their active site pocket, leading to a different binding orientation of the steroid relative to the heme.

In conclusion, we have identified CYP109E1 as a novel steroid-hydroxylating cytochrome P450 enzyme from *Bacillus megaterium* DSM319, with the ability to selectively convert testosterone to 16β-hydroxytestosterone. Our combined structural, biochemical, and molecular modeling studies provided several insights into the molecular basis of steroid binding by CYP109E1 and of its regio- and stereoselectivity toward testosterone. First, binding of single steroid molecules like testosterone and corticosterone stabilizes a change in the active site pocket toward a more closed and narrow conformation. The steroid-induced structural changes include the local widening of the central I helix, which is coupled with the formation of a water channel believed to function as a water access channel and/or proton delivery network during catalysis. Secondly, the steroids may bind in two opposite orientations, with either their C3-keto oxygen or their C16 carbon and C17 substituent directed toward the heme. The first orientation leads to nonproductive binding. The second orientation results in productive binding for testosterone, and nicely explains the high selectivity toward 16β-hydroxylation. The larger C17 substituent of corticosterone, and the presence of its C21-hydroxyl group, prohibit productive binding of this steroid. Our results will facilitate future protein engineering experiments to improve this enzyme for biotechnological applications.

## Experimental procedures

### Materials

The steroid compounds used in this study were obtained from Sigma-Aldrich (Steinheim, Germany). All other chemicals were of highest grade available.

### Bioinformatics analysis

Identification of close homologs and comparison of protein sequences were performed using the Basic Local Alignment Search Tool (BLAST, NCBI). Multiple sequence alignment was done with Clustal Omega [46] and visualized with ESPript3 [47]. Evolutionary analysis was carried out using Molecular Evolutionary Genetics Analysis (MEGA) version 6.0 software [48]. The phylogenetic tree was constructed using the Neighbor-joining method [49] and the evolutionary distances were computed using the Poisson correction method [50].

### Cloning of wild-type enzyme

The gene encoding CYP109E1 (GenBank GeneID 9119265) was amplified by the PCR using genomic DNA of *B. megaterium* MS941, a mutant form of *B. megaterium,* derived from the DSM319 strain [51]. The PCR primers were designed (MWG-Biotech AG, Ebersberg, Germany) to introduce an *Nde*I restriction site at the 5′ end of the fragment and a *Kpn*I restriction site with a 6-histidine tag at the 3′ end. Following the amplification of the *cyp109e1* gene, the PCR product was cloned into the pCR4Bl-TOPO vector (Invitrogen, San Diego, CA, USA) and the vector was further digested with the restriction enzymes cutting at the above-mentioned sites and ligated into the expression vector pET17b (Merck Bioscience, Bad Soden, Germany), creating the pET17b_109E1 vector. The *cyp109e1* gene was further amplified by PCR from the previously constructed pET17b expression vector using PCR primers designed to include a *Spe*I restriction site at the 5′ end and a *Kpn*I restriction site at the 3′ end of the fragment. The resulting PCR product was subcloned in the pCR4-TOPO vector and digested with the corresponding restriction enzymes.

The fragment was then ligated to the previously linearized pSMF2.1 vector [52], yielding the pSMF2.1E construct. Sequences of the designed primers are given in supplementary Table S1. Sequences of all created vectors were verified by DNA sequencing, carried out by Eurofins-MWG (Ebersberg, Germany).

## Site-directed mutagenesis

The mutants of CYP109E1 were generated by the QuikChange site-directed mutagenesis method using the plasmid pET17b_109E1 as template and Phusion polymerase for DNA replication (Thermo Fisher Scientific GmbH, Dreireich, Germany). The reactions were performed in 50 μL, using a gradient cycler (PTC-200 DNA Engine cycler). Twenty cycles were carried out as follows: initial denaturation at 95 °C for 30 sec, denaturation at 95 °C for 30 sec, annealing at 58 °C for 30 sec, and extension at 72 °C for 4 min. The oligonucleotide primers for mutagenesis are shown in Table S1. Correct generation of the desired mutations was confirmed by DNA sequencing.

## Expression and purification

A 30-mL preculture of *E. coli* C43 (DE3) cells carrying the pET17b_109E1 vector was grown overnight in LB medium containing 100 $\mu g \cdot mL^{-1}$ ampicillin at 37 °C (150 rpm). This culture was used to inoculate a 1.2-L production culture, divided over four 2-L baffled flasks, in Terrific Broth (TB) medium containing 100 $\mu g \cdot mL^{-1}$ ampicillin. Cultivation was continued at 37 °C (150 rpm) until the $OD_{600}$ reached 0.5, after which 1 mM IPTG and 0.5 mM δ-aminolevulinic acid were added to start protein expression and support heme synthesis, respectively. After 24 h of incubation at 30 °C, 100 rpm, the cells were harvested by centrifugation (4500 ***g*** for 35 min) and the cell pellet was stored at −20 °C until further use. All purification steps were performed at 4 °C. For crystallization and spectral characterization of wild-type CYP109E1, a three-step purification procedure was applied, starting with resuspension of the cell pellet in 100 mL cold lysis buffer containing 50 mM Tris/HCl, pH 8.0, 1 mM EDTA, 20 mM NaCl, and 0.1 mM dithioerythritol, followed by the addition of 50 $\mu g \cdot mL^{-1}$ PMSF. The mixture was sonicated for 15 min (15 sec on, 15 sec off) on ice and, subsequently, the same amount of PMSF was added. Cell-free extract was obtained by ultracentrifugation at 40 000 ***g*** for 35 min at 4 °C. The supernatant containing CYP109E1 was loaded onto a 50-mL SOURCE 30Q anion-exchange column (GE Healthcare, Solingen, Germany) equilibrated with three column volumes of 20 mM Tris/HCl, pH 7.4, 0.1 mM dithioerythritol. The column was washed with the same buffer before elution of CYP109E1 with a linear gradient of 0–500 mM NaCl. Fractions with the highest $A_{418}/A_{280}$ ratio were combined and concentrated by ultrafiltration using a 30-kDa cutoff membrane (Amicon Ultra/Millipore). The protein concentrate was then manually loaded onto a Superdex 75 (200 mL) gel filtration column (GE Healthcare) and CYP109E1 was eluted with 50 mM potassium phosphate buffer, pH 7.4, 0.1 mM dithioerythritol. The fractions with the highest $A_{418}/A_{280}$ ratio were pooled, diluted 1:5 with 5 mM potassium phosphate buffer, pH 7.4 and 0.05 mM dithioerythritol, before loading them onto a hydroxyapatite column (50 mL, Bio-Rad, Hercules, CA, USA). The column was washed with 10 mM potassium phosphate buffer, pH 7.4, 0.1 mM dithioerythritol, and CYP109E1 was eluted with a buffer concentration gradient ranging from 10 to 100 mM. Fractions containing purified CYP109E1 with an $A_{418}/A_{280}$ ratio larger than 1.6 were collected, concentrated by ultrafiltration using a 30-kDa cutoff membrane, and stored at −80 °C after flash-freezing in liquid nitrogen. For *in vitro* conversion experiments, the wild-type protein and its mutants were purified with an alternative one-step purification procedure. The cell pellets were resuspended in 50 mM potassium phosphate buffer, pH 7.4, containing 300 mM NaCl and 20% glycerol. Then, PMSF was added to a final concentration of 1 mM and the suspension was sonicated for 15 min (15 sec on, 15 sec off) on ice. Cell-free extract was obtained by ultracentrifugation at 30 000 rpm for 30 min, at 4 °C. The supernatant was applied to immobilized metal ion affinity chromatography column (TALON, Takara Bio Europe, Saint-Germain-en-Laye, France) equilibrated with 50 mM potassium phosphate buffer, pH 7.4, containing 300 mM NaCl and 20% glycerol. The column was washed with 5 column volumes of the same buffer containing 20 mM imidazole, and the tagged protein was eluted with a buffer containing 150 mM imidazole.

## Enzyme analysis

The UV/Vis spectra of the purified protein were recorded using a double-beam spectrophotometer (UV-2101PC, Shimadzu, Japan) from 200 to 700 nm. CYP109E1 concentrations were determined by CO-difference spectroscopy of the reduced protein, following the method of Omura and Sato [53] and using an extinction coefficient of 91 $mM^{-1} \cdot cm^{-1}$. Protein purity was analyzed by SDS/PAGE.

## Difference spectroscopy and determination of dissociation constants ($K_D$)

Substrate-induced spin-shift states were studied via a double-beam spectrophotometer (UV-2101PC, Shimadzu, Japan), using tandem quartz cuvettes, according to the method of Schenkman and Jansson [54]. One of the cuvette chambers contained the purified CYP109E1 protein solution (10 μM) in 50 mM potassium phosphate buffer, pH 7.4, while the other chamber was filled with the corresponding

buffer only. The steroids were dissolved in DMSO (2.5–20 mM stock solutions) and the enzyme solution was titrated with increasing amounts of testosterone or corticosterone, in the range of 0–200 μM or until saturation was reached, while the spectrum was recorded between 350 and 500 nm. All titrations were carried out in triplicates. The data were analyzed by plotting the peak-to-trough differences ($\Delta A$: $A_{max}-A_{min}$) against the steroid concentrations. The subsequent hyperbolic fitting was performed using ORIGIN software (OriginLab Corporation, Northampton, MA, USA), and the equilibrium dissociation constants ($K_D$) were determined with a regression coefficient of $R^2 = 0.99$.

### *In vitro* substrate turnover

The *in vitro* turnover of tested steroids was performed in a final volume of 250 μL, using 50 mM potassium phosphate buffer with 2% glycerol, pH 7.4, at 30 °C for 30 min. The reconstituted system contained CYP109E1 (1 μM), bovine $Adx_{4-108}$ (20 μM), AdR (2 μM), a NADPH-regenerating system (1 mM $MgCl_2$, 5 mM glucose-6-phosphate, 1 U glucose-6-phosphate dehydrogenase) and 200 μM of the corresponding steroid dissolved in DMSO. The reactions were initiated by the addition of NADPH to a final concentration of 1 mM, then stopped and extracted twice with the addition of 250 μL of ethyl acetate. The samples were centrifuged (10 000 rpm, 10 min), the organic phases were combined and evaporated until complete dryness and, after resuspension, analyzed by HPLC. Since, the absorption properties of the products did not differ from the respective substrates, the product formation was calculated from the relative peak area (area %) of the HPLC chromatograms, dividing each respective product peak area by the sum of all peak areas.

### Whole-cell conversion with *Bacillus megaterium* MS941

The *in vivo* conversions were done with the *B. megaterium* strain MS941 [51]. The MS941 cells were transformed with the pSMF2.1E vector, using the polyethylene glycol-mediated protoplast transformation method [55]. For cultivation, a complex medium was used (24 $g \cdot L^{-1}$ yeast extract, 12 $g \cdot L^{-1}$ soytone, 2.31 $g \cdot L^{-1}$ $KH_2PO_4$, and 1.25 $g \cdot L^{-1}$ $KHPO_4$) supplemented with 10 $\mu g \cdot mL^{-1}$ tetracycline at 30 °C, 180 rpm. First, a 50-mL overnight culture was prepared, inoculated from a −80 °C glycerol stock of the transformed MS941 cells. This culture was then used to inoculate the main culture in a 300-mL baffled shake flask filled with 50 mL complex medium. The main culture was incubated, until the $OD_{578}$ reached 0.4, when the protein expression was induced by the addition of xylose at a final concentration of 5 $mg \cdot mL^{-1}$. After 24 h, the cultures were harvested by centrifugation (15 min, 10 000 ***g***, 4 °C) and the pellets resuspended in 50 mL of 50 mM potassium phosphate buffer, pH 7.4. The substrates, dissolved in DMSO, were added at a final concentration of 200 μM. The 500 μL samples were taken from the cultures at fixed time points, followed by extraction and HPLC analysis. To obtain sufficient product quantities for structural analysis by NMR spectroscopy (mg range), the volume of the main culture was increased to 750 mL (3 × 250 mL) in 2-L baffled flasks. After a 24-h expression period, the cultures were harvested by centrifugation, resuspended in 375 mL of 50 mM potassium phosphate buffer, pH 7.4, and the corresponding substrates were added at a final concentration of 300 μM. Following the conversion, the cultures were extracted twice with ethyl acetate, the organic phases were combined and evaporated (Rotavapor R-114; BÜCHI Labortechnik AG, Flawil, Switzerland). The purification and isolation of products was carried out by preparative HPLC.

### HPLC analysis and product isolation

The HPLC analyses were carried out with a Jasco system (Pu-980 HPLC pump, AS-950 sampler, UV-975 UV/visible detector, LG-980-02 gradient unit; Jasco, Gross-Umstadt, Germany), using a reversed-phase ec MN Nucleodur $C_{18}$ (3 μM, 4.0 × 125 mm) column (Macherey-Nagel, Bethlehem, PA, USA) kept at an oven temperature of 40 °C. The steroids were eluted using a gradient method, starting with a mobile phase consisting of acetonitrile:$H_2O$ in a ratio of 1:9, increasing it to 1:1. The flow rate was 1 $mL \cdot min^{-1}$ and the UV detection of the substrate and product was accomplished at 240 or 254 nm. For product isolation, the dissolved extracts were filtered (Acrodisc® PTFE syringe filter, 0.45 μm, PALL) and injected to a reversed-phase ec MN Nucleodur $C_{18}$ (5 μM, 8.0 × 250 mm) column (Macherey-Nagel, Bethlehem, PA, USA). The HPLC run was carried out analogously to the analytical HPLC conditions with an increased flow rate of 2.5 $mL \cdot min^{-1}$. Fractions were collected with an Advantec CHF122 SB fraction collector, combined, and evaporated on a rotary evaporator (Rotavapor R-114 from BÜCHI Labortechnik AG).

### NMR spectroscopy

The NMR spectra were recorded in deuterated chloroform ($CDCl_3$) with a Bruker DRX 500 or a Bruker Avance 500 NMR spectrometer at 300 K. The chemical shifts were relative to $CDCl_3$ at δ 24.7 ($^1H$ NMR) and 77.00 ($^{13}C$ NMR), using the standard δ notation in parts per million. The 1D NMR ($^1H$ and $^{13}C$ NMR, DEPT135) and the 2D NMR spectra (gs-HH-COSY, gs-NOESY, gs-HSQCED, and gs-HMBC) were recorded using the BRUKER pulse program library. All assignments were based on extensive NMR spectral evidence.

### Crystallization

After thawing, aliquots of purified CYP109E1 were buffer exchanged to 20 mM Tris/HCl, pH 8.0, 0.1 mM

dithioerythritol using a PD-10 desalting column (GE Healthcare) and concentrated to 40 mg·mL$^{-1}$. Screening for crystallization growth conditions was done at 293 K by the sitting-drop vapor-diffusion method, using 96-well crystallization plates, a Mosquito crystallization robot, (TTP LabTech, Melbourn, UK) and a few commercially available screens. Lead crystallization conditions were optimized manually in 24-well plates at 293 K. Initial crystals grew with a reservoir solution containing 25% poly(ethylene glycol) 3350, 0.1 M Bis-tris, pH 6.5, and 4% tacsimate reagent (pH 6.0) (Hampton Research, Aliso Viejo, CA, USA). Diffraction quality crystals were obtained by applying a streak-seeding protocol. First, protein drops were prepared by mixing 2 μL of concentrated CYP109E1 (40 mg·mL$^{-1}$) with an equal volume of reservoir solution, containing 20% poly(ethylene glycol) 3350, 0.1 M Bis-tris, pH 6.5, and 4% tacsimate reagent pH 6.0. Droplets were equilibrated against 500 μL of the reservoir solution for 1 h, and then streak-seeded. Red-colored, cubic-shaped crystals of CYP109E1 (space group $P3_221$) grew overnight and reached a final average size of approximately $0.15 \times 0.15 \times 0.15$ mm$^3$ in about 7 days. Corticosterone-bound CYP109E1 crystals were initially obtained by crystal soaking experiments. For this, corticosterone was added as a solid powder directly into crystallization drops containing native CYP109E1 crystals, followed by equilibration against the original crystallization solution for 3–4 weeks. Alternatively, cocrystallization screens were employed with corticosterone and testosterone (in a 1:5 molar ratio of protein:steroid, using stocks solutions of the steroids in DMSO), resulting in large cuboid-shaped crystals (space group $P2_1$) grown from 9% poly(ethylene glycol) 3350 and 8% tacsimate pH 5.0 (Hampton Research). Prior to data collection, crystals were briefly soaked in a cryoprotectant solution containing mother liquor, supplemented with 20% (v/v) glycerol. Cryoprotection of crystals obtained by cocrystallization was accomplished by raising the poly(ethylene glycol) 3350 concentration to 35%. Subsequently, all crystals were flash-cooled at 100 K in the cold nitrogen gas stream of the camera's cryostat.

## Data collection and structure determination

X-ray diffraction data were collected at the ID29, ID23-1, and ID23-2 beam lines of the European Synchrotron Radiation Facility (ESRF), Grenoble, all equipped with Pilatus detectors. Additional data were recorded using an in-house rotating anode X-ray source (Bruker MicroSTAR) and an image plate detector (*mar345*™). Single crystals of CYP109E1 were used to obtain diffraction datasets in the 2.55–2.1 Å resolution range. Reflections were indexed and integrated using iMosflm [56] or XDS [57], while scaling and merging of the data was done with AIMLESS from the CCP4 software suite [58]. The structure of native CYP109E1 was solved by molecular replacement with Phaser from the PHENIX suite [59], using the structure of HmtT from *Streptomyces himastatinicus* (PDB ID 4ggv, 39% sequence identity) as a search model. Two protein molecules were located in the asymmetric unit, consistent with Matthew coefficient calculations indicating a solvent content of 61%. Automatic model rebuilding using routines in PHENIX yielded approximately 80% of the polypeptide model. The model was completed by iterative cycles of model building with Coot [60] and structure refinement with Phenix.refine [61]. The structure of substrate-free CYP109E1 then served as a starting point to solve the CYP109E1 steroid-bound structures. At the final stages of refinement, water molecules were added to the structures based on peaks in the electron density maps and using strict interaction criteria. The quality of the refined protein models was validated using MolProbity [62].

## Structure analysis

Pairwise comparison of the obtained structures with other P450 structures and RMSD calculations were done with the PDBeFold engine [63] Protein–ligand interactions were analyzed by LigPlot+ [64]. Substrate recognition sites (SRS) in CYP109E1 were identified based on alignment with $P450_{cam}$ and the description provided by Gotoh [29]. The respective residue ranges in CYP109E1 are: SRS1 67–89, SRS2 165–171, SRS3 187–194, SRS4 230–249, SRS5 286–295, and SRS6 388–395.

## Molecular dynamics simulations

Molecular dynamics simulations were performed using the CYP109E1-TES crystal structure as a starting model, following the same procedure as previously described [33]. First, testosterone and all water molecules were removed from the crystal structure, and the heme and its cysteine ligand were replaced by a model of heme compound I covalently bound to cysteine. The tested ligand (testosterone or corticosterone) was then manually placed back in the binding pocket using PyMOL (Schrödinger, Cambridge, MA, USA). Testosterone was placed in 10 different starting orientations where the distance between the C16β hydrogen and the ferryl oxygen atom of heme compound I was less than 4 Å. Corticosterone was placed in starting orientations similar to the COR-1 steroid orientation in the CYP109E1-COR4 crystal structure. To ensure extensive conformational sampling, ten 100-ns MD simulations with different initial orientations of the steroid were performed for each CYP109E1-steroid system, and the AMBER03 force field [65] in GROMACS version 5.0.4 was used [66]. Force fields for testosterone and corticosterone were generated by use of the structures derived from the PubChem database [67] and subsequent energy minimization in YASARA (www.YASARA.org). The RESP partial charges of the molecules were calculated using the R.E.D webserver with

the RESP-A1B charge model [68]. The force field was built with the ANTECHAMBER module of AMBER 10 [69] and converted into the GROMACS topology format. The force field of the cysteine-heme compound I complex was used as previously described [70]. An octahedral water box of SPC/E water with periodic boundaries at least 1.2 nm from the protein was used. Simulations were run at 300 K and 1 bar. Pressure coupling was performed with a Parrinello–Rahman barostat [71]. The Nose–Hoover coupling scheme was used to maintain the temperature, with coupling constant of 0.5 ps [72]. Initial velocities were randomly assigned. The LINCS algorithm was applied to constrain all bonds containing hydrogen atoms [73]. Seventeen sodium counter ions were added to maintain the neutral charge of the systems. Long-range electrostatic interactions were treated by using the particle-mesh Ewald method [74]. Energy minimization was performed using the steepest descent method with positional restraints on protein heavy atoms and a maximum allowed force of 1000 $kJ \cdot mol^{-1} \cdot nm^{-1}$. A 2-fs time step was used and coordinates were saved every 1000 steps (2 ps). To determine near-attack conformations the distance between the 16α or 16β hydrogen and the ferryl oxygen of compound I ferryl oxygen was measured. In case of simulations with testosterone, additionally the distance between the 17α hydrogen and the ferryl oxygen was measured. Angles between the C16 or C17 carbon, their hydrogens and the ferryl oxygen were also measured. Based on the previously described cut-offs [33–35], ligand conformations were considered near to attack, if at least one of the measured distances was < 3.5 Å and the angle was 180 ± 45°. All other conformations were considered nonproductive. The number of frames in MD trajectories suggesting pro-16α, pro-16β, pro-17α, and nonproductive conformations were counted and expressed as percentages.

## Acknowledgements

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's 7th Framework Programme (FP7/2007–2013) under REA Grant Agreement 289217 (ITN P4FIFTY). We thank the beam-line scientists of ID29, ID23-1, and ID23-2 (ESRF, European Synchrotron Radiation Facility) for assistance. We thank the high performance computing center Stuttgart (HLRS) for their support and for supplying the computational resources.

## Author contributions

IKJ, FMK, LG, RB, and AMWHT designed the study. IKJ, FMK, LG, AA, EB, and JZ performed the experiments. IKJ, FMK, LG, EB, and AA analyzed the data. IKJ, FMK, LG, AA, JP, RB, and AMWHT wrote the manuscript. JP, RB, and AMWHT provided supervision.

## References

1 Bernhardt R & Urlacher VB (2014) Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Appl Microbiol Biotechnol* **98**, 6185–6203.
2 Bhatti HN & Khera RA (2012) Biological transformations of steroidal compounds: a review. *Steroids* **77**, 1267–1290.
3 Donova MV & Egorova OV (2012) Microbial steroid transformations: current state and prospects. *Appl Microbiol Biotechnol* **94**, 1423–1447.
4 Tong WY & Dong X (2009) Microbial biotransformation: recent developments on steroid drugs. *Recent Pat Biotechnol* **3**, 141–153.
5 Grogan G (2011) Cytochromes P450: exploiting diversity and enabling application as biocatalysts. *Curr Opin Chem Biol* **15**, 241–248.
6 Ost TW, Miles CS, Murdoch J, Cheung Y, Reid GA, Chapman SK & Munro AW (2000) Rational re-design of the substrate binding site of flavocytochrome P450 BM3. *FEBS Lett* **486**, 173–177.
7 Noble MA, Miles CS, Chapman SK, Lysek DA, MacKay AC, Reid GA, Hanzlik RP & Munro AW (1999) Roles of key active-site residues in flavocytochrome P450 BM3. *Biochem J* **339** (Pt 2), 371–379.
8 Denisov IG, Shih AY & Sligar SG (2012) Structural differences between soluble and membrane bound cytochrome P450s. *J Inorg Biochem* **108**, 150–158.
9 Johnson EF & Stout CD (2005) Structural diversity of human xenobiotic-metabolizing cytochrome P450 monooxygenases. *Biochem Biophys Res Commun* **338**, 331–336.
10 Podust LM & Sherman DH (2012) Diversity of P450 enzymes in the biosynthesis of natural products. *Nat Prod Rep* **29**, 1251–1266.
11 Otyepka M, Berka K & Anzenbacher P (2012) Is there a relationship between the substrate preferences and structural flexibility of cytochromes P450? *Curr Drug Metab* **13**, 130–142.
12 McLean KJ, Lafite P, Levy C, Cheesman MR, Mast N, Pikuleva IA, Leys D & Munro AW (2009) The structure of *Mycobacterium tuberculosis* CYP125: molecular basis for cholesterol binding in a P450 needed for host infection. *J Biol Chem* **284**, 35524–35533.
13 Janocha S, Carius Y, Hutter M, Lancaster CR & Bernhardt R (2016) Crystal structure of CYP106A2 in substrate-free and substrate-bound form. *ChemBioChem* **9**, 852–860.

14 Herzog K, Bracco P, Onoda A, Hayashi T, Hoffmann K & Schallmey A (2014) Enzyme-substrate complex structures of CYP154C5 shed light on its mode of highly selective steroid hydroxylation. *Acta Crystallogr D Biol Crystallogr* **70**, 2875–2889.
15 Makino T, Katsuyama Y, Otomatsu T, Misawa N & Ohnishi Y (2014) Regio- and stereospecific hydroxylation of various steroids at the 16alpha position of the D ring by the *Streptomyces griseus* cytochrome P450 CYP154C3. *Appl Environ Microbiol* **80**, 1371–1379.
16 Zhang A, Zhang T, Hall EA, Hutchinson S, Cryle MJ, Wong LL, Zhou W & Bell SG (2015) The crystal structure of the versatile cytochrome P450 enzyme CYP109B1 from *Bacillus subtilis*. *Mol BioSyst* **3**, 869–881.
17 Eppinger M, Bunk B, Johns MA, Edirisinghe JN, Kutumbaka KK, Koenig SS, Creasy HH, Rosovitz MJ, Riley DR, Daugherty S *et al*. (2011) Genome sequences of the biotechnologically important *Bacillus megaterium* strains QM B1551 and DSM319. *J Bacteriol* **193**, 4199–4213.
18 Schmitz D, Zapp J & Bernhardt R (2014) Steroid conversion with CYP106A2 – production of pharmaceutically interesting DHEA metabolites. *Microb Cell Fact* **13**, 81
19 Kiss FM, Khatri Y, Zapp J & Bernhardt R (2015) Identification of new substrates for the CYP106A1-mediated 11-oxidation and investigation of the reaction mechanism. *FEBS Lett* **589**, 2320–2326.
20 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-beta-boswellic acid (KBA). *Appl Microbiol Biotechnol* **98**, 1701–1717.
21 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.
22 Nelson DR (2009) The cytochrome P450 homepage. *Human Genomics* **4**, 59–65.
23 Khatri Y, Hannemann F, Ewen KM, Pistorius D, Perlova O, Kagawa N, Brachmann AO, Muller R & Bernhardt R (2010) The CYPome of *Sorangium cellulosum* So ce56 and identification of CYP109D1 as a new fatty acid hydroxylase. *Chem Biol* **17**, 1295–1305.
24 Khatri Y, Hannemann F, Girhard M, Kappl R, Meme A, Ringle M, Janocha S, Leize-Wagner E, Urlacher VB & Bernhardt R (2013) Novel family members of CYP109 from *Sorangium cellulosum* So ce56 exhibit characteristic biochemical and biophysical properties. *Biotechnol Appl Biochem* **60**, 18–29.
25 Furuya T, Shibata D & Kino K (2009) Phylogenetic analysis of *Bacillus* P450 monooxygenases and evaluation of their activity towards steroids. *Steroids* **74**, 906–912.
26 Khatri Y, Ringle M, Lisurek M, von Kries JP, Zapp J & Bernhardt R (2016) Substrate hunting for the myxobacterial CYP260A1 revealed new 1alpha-hydroxylated products from C-19 steroids. *ChemBioChem* **17**, 90–101.
27 Schmitz D, Zapp J & Bernhardt R (2012) Hydroxylation of the triterpenoid dipterocarpol with CYP106A2 from *Bacillus megaterium*. *FEBS J* **279**, 1663–1674.
28 Poulos TL, Finzel BC, Gunsalus IC, Wagner GC & Kraut J (1985) The 2.6-A crystal structure of *Pseudomonas putida* cytochrome P-450. *J Biol Chem* **260**, 16122–16130.
29 Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* **267**, 83–90.
30 Gricman Ł, Vogel C & Pleiss J (2015) Identification of universal selectivity-determining positions in cytochrome P450 monooxygenases by systematic sequence-based literature mining. *Proteins* **83**, 1593–1603.
31 Hamdane D, Zhang H & Hollenberg P (2008) Oxygen activation by cytochrome P450 monooxygenase. *Photosynth Res* **98**, 657–666.
32 Cojocaru V, Winn PJ & Wade RC (2007) The ins and outs of cytochrome P450s. *Biochim Biophys Acta* **1770**, 390–401.
33 Eichler A, Gricman L, Herter S, Kelly P, Turner N, Pleiss J & Flitsch S (2015) Enantioselective benzylic hydroxylation catalysed by P450 monooxygenases: characterisation of a P450cam mutant library and molecular modelling. *ChemBioChem* **5**, 426–432.
34 Paulsen MD & Ornstein RL (1992) Predicting the product specificity and coupling of cytochrome P450cam. *J Comput Aided Mol Des* **6**, 449–460.
35 Filipovic D, Paulsen MD, Loida PJ, Sligar SG & Ornstein RL (1992) Ethylbenzene hydroxylation by cytochrome P450cam. *Biochem Biophys Res Commun* **189**, 488–495.
36 Agematu H, Matsumoto N, Fujii Y, Kabumoto H, Doi S, Machida K, Ishikawa J & Arisawa A (2006) Hydroxylation of testosterone by bacterial cytochromes P450 using the *Escherichia coli* expression system. *Biosci Biotechnol Biochem* **70**, 307–311.
37 Girhard M, Klaus T, Khatri Y, Bernhardt R & Urlacher VB (2010) Characterization of the versatile monooxygenase CYP109B1 from *Bacillus subtilis*. *Appl Microbiol Biotechnol* **87**, 595–607.
38 Bracco P, Janssen DB & Schallmey A (2013) Selective steroid oxyfunctionalisation by CYP154C5, a bacterial cytochrome P450. *Microb Cell Fact* **12**, 95.

39 van Vugt-Lussenburg BM, Damsten MC, Maasdijk DM, Vermeulen NP & Commandeur JN (2006) Heterotropic and homotropic cooperativity by a drug-metabolising mutant of cytochrome P450 BM3. *Biochem Biophys Res Commun* **346**, 810–818.
40 Rea V, Kolkman AJ, Vottero E, Stronks EJ, Ampt KA, Honing M, Vermeulen NP, Wijmenga SS & Commandeur JN (2012) Active site substitution A82W improves the regioselectivity of steroid hydroxylation by cytochrome P450 BM3 mutants as rationalized by spin relaxation nuclear magnetic resonance studies. *Biochemistry* **51**, 750–760.
41 Venkataraman H, Beer SB, Bergen LA, Essen N, Geerke DP, Vermeulen NP & Commandeur JN (2012) A single active site mutation inverts stereoselectivity of 16-hydroxylation of testosterone catalyzed by engineered cytochrome P450 BM3. *ChemBioChem* **13**, 520–523.
42 Niwa T, Murayama N, Imagawa Y & Yamazaki H (2015) Regioselective hydroxylation of steroid hormones by human cytochromes P450. *Drug Metab Rev* **47**, 89–110.
43 Davydov DR & Halpert JR (2008) Allosteric P450 mechanisms: multiple binding sites, multiple conformers or both? *Expert Opin Drug Metab Toxicol* **4**, 1523–1535.
44 Zhao B, Lei L, Kagawa N, Sundaramoorthy M, Banerjee S, Nagy LD, Guengerich FP & Waterman MR (2012) Three-dimensional structure of steroid 21-hydroxylase (cytochrome P450 21A2) with two substrates reveals locations of disease-associated variants. *J Biol Chem* **287**, 10613–10622.
45 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.
46 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.
47 Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320–W324.
48 Tamura K, Stecher G, Peterson D, Filipski A & Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729.
49 Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
50 Zuckerlandl E & Pauling L (1965) Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins* (Bryson V & Vogel HJ, eds), pp. 97–166. Academic Press, Cambridge, MA.
51 Wittchen KD & Meinhardt F (1995) Inactivation of the major extracellular protease from *Bacillus megaterium* DSM319 by gene replacement. *Appl Microbiol Biotechnol* **42**, 871–877.
52 Bleif S, Hannemann F, Zapp J, Hartmann D, Jauch J & Bernhardt R (2012) A new *Bacillus megaterium* whole-cell catalyst for the hydroxylation of the pentacyclic triterpene 11-keto-beta-boswellic acid (KBA) based on a recombinant cytochrome P450 system. *Appl Microbiol Biotechnol* **93**, 1135–1146.
53 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.
54 Schenkman JB & Jansson I (1998) Spectral analyses of cytochromes P450. *Methods Mol Biol* **107**, 25–33.
55 Barg H, Malten M, Jahn M & Jahn D (2005) Protein and vitamin production in *Bacillus megaterium*. In *Microbial Processes and Products* (Barredo J, ed.), pp. 205–223. Humana Press, New York, NY.
56 Battye TG, Kontogiannis L, Johnson O, Powell HR & Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**, 271–281.
57 Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.
58 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242.
59 Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221.
60 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.
61 Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH & Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352–367.
62 Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21.
63 Krissinel E & Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256–2268.

64 Laskowski RA & Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**, 2778–2786.
65 Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T *et al.* (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* **24**, 1999–2012.
66 Hess B, Kutzner C, van der Spoel D & Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* **4**, 435–447.
67 Wang Y, Xiao J, Suzek TO, Zhang J, Wang J & Bryant SH (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res* **37**, W623–W633.
68 Vanquelef E, Simon S, Marquant G, Garcia E, Klimerak G, Delepine JC, Cieplak P & Dupradeau FY (2011) R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res* **39**, W511–W517.
69 Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W *et al.* (2008) *Amber 10*. University of California.
70 Seifert A, Tatzel S, Schmid RD & Pleiss J (2006) Multiple molecular dynamics simulations of human p450 monooxygenase CYP2C9: the molecular basis of substrate binding and regioselectivity toward warfarin. *Proteins* **64**, 147–155.
71 Parrinello M & Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* **52**, 7182–7190.
72 Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys* **52**, 255–268.
73 Hess B, Bekker H, Berendsen HJC & Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* **18**, 1463–1472.
74 Darden T, York D & Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* **98**, 10089–10092.

## Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

**Table S1.** Oligonucleotide primers used in this work.
**Table S2.** Chemical structures of compounds used for CYP109E1 substrate screening.
**Table S3.** Structural NMR data of 16β-hydroxytestosterone and androstendione in $CDCl_3$.
**Table S4.** Binding interactions of COR-2, COR-3, and COR-4 in CYP109E1-COR4 crystal structure.