Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Enzyme–substrate complex structures of CYP154C5 shed light on its mode of highly selective steroid hydroxylation

Konrad Herzog,[a,b]‡ Paula Bracco,[a]‡§ Akira Onoda,[c] Takashi Hayashi,[c] Kurt Hoffmann[b] and Anett Schallmey[a]*§

[a]Junior Professorship for Biocatalysis, Institute of Biotechnology, RWTH Aachen University, Worringerweg 3, 52074 Aachen, Germany, [b]Institute of Molecular Biotechnology, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany, and [c]Department of Applied Chemistry, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

‡ These authors contributed equally to the work.
§ Present address: Biocatalysis, Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Postbus 94720, 1090 GS Amsterdam, The Netherlands

Correspondence e-mail: a.schallmey@uva.nl

Received 14 May 2014
Accepted 25 August 2014

PDB references: CYP154C5, complex with progesterone, 4j6c; complex with androstenedione, 4jbt; complex with testosterone, 4j6d; complex with pregnenolone, 4j6b

CYP154C5 from *Nocardia farcinica* is a bacterial cytochrome P450 monooxygenase active on steroid molecules. The enzyme has recently been shown to exhibit exclusive regioselectivity and stereoselectivity in the conversion of various pregnans and androstans, yielding 16α-hydroxylated steroid products. This makes the enzyme an attractive candidate for industrial application in steroid hormone synthesis. Here, crystal structures of CYP154C5 in complex with four different steroid molecules were solved at resolutions of up to 1.9 Å. These are the first reported P450 structures from the CYP154 family in complex with a substrate. The active site of CYP154C5 forms a flattened hydrophobic channel with two opposing polar regions, perfectly resembling the size and polarity distribution of the steroids and thus resulting in highly specific steroid binding with K_d values in the range 10–100 n*M*. Key enzyme–substrate interactions were identified that accounted for the exclusive regioselectivity and stereoselectivity of the enzyme. Additionally, comparison of the four CYP154C5–steroid structures revealed distinct structural differences, explaining the observed variations in kinetic data obtained for this P450 with the steroids pregnenolone, dehydroepiandrosterone, progesterone, androstenedione, testosterone and nandrolone. This will facilitate the generation of variants with improved activity or altered selectivity in the future by means of protein engineering.

1. Introduction

Cytochrome P450 monooxygenases (CYPs) are highly interesting haem-containing enzymes as they are involved in many physiologically important reactions such as the detoxification of xenobiotics, the synthesis of secondary metabolites and drug metabolism (Ortiz de Montellano, 2005). Furthermore, they are promising biocatalysts owing to their remarkable ability to directly hydroxylate non-activated C atoms simply using molecular oxygen. Besides hydroxylation, CYPs have been shown to catalyze a range of chemically diverse reactions such as epoxidations, dealkylations, dehalogenations and heteroatom oxidations, amongst others (Bernhardt, 2006; Isin & Guengerich, 2007). The monooxygenases are found in eukaryotes and prokaryotes mainly as multi-component systems and require one or two redox partners for electron transfer from the cofactor NAD(P)H to the haem iron (Hannemann *et al.*, 2007).

An industrially very important application of cytochrome P450 monooxygenases is the regioselective and stereoselective hydroxylation of steroid compounds in the pharmaceutical industry for the synthesis of highly valuable steroid hormones such as glucocorticoids, mineralocorticoids and sexual hormones (Bureik & Bernhardt, 2007; Bernhardt, 2006).

Today, numerous eukaryotic CYPs are known to carry out this reaction owing to their natural involvement in steroid synthesis. In consequence, many crystal structures of eukaryotic P450 monooxygenases have been solved to date, some of them carrying a steroid molecule in the active site. In contrast, the crystal structures of only very few bacterial steroid-converting CYPs such as CYP125A1 (PDB entry 2x5w; Ouellet *et al.*, 2010) and CYP142A2 (PDB entry 2yoo; García-Fernández *et al.*, 2013) can be found in the PDB; both of these originate from mycobacteria and have a steroid bound in the active site. This observation is somewhat surprising as eukaryotic CYPs are typically membrane-bound while prokaryotic CYPs are soluble enzymes, rendering the latter more accessible for crystallization efforts. Owing to their easier expression and facilitated protein engineering for the generation of improved variants, the latter also seem to be more promising candidates for industrial application.

Very recently, a novel bacterial cytochrome P450 monooxygenase from *Nocardia farcinica*, CYP154C5, was identified to hydroxylate various pregnans and androstans with exceptionally high regioselectivity and stereoselectivity, yielding only the corresponding 16α-hydroxylated steroids as products (Bracco *et al.*, 2013). The enzyme belongs to the class I three-component P450 monooxygenases, requiring a ferredoxin reductase and a ferredoxin as redox partners for catalysis. Using putidaredoxin (Pdx) and putidaredoxin reductase (PdR) from *Pseudomonas putida* (Peterson *et al.*, 1990) as surrogate electron-transfer partners for CYP154C5, a highly efficient *Escherichia coli*-based whole-cell biocatalyst was generated for preparative-scale steroid hydroxylations, achieving productivities of up to 4 g l^{-1} d^{-1}. Thus, CYP154C5 constitutes a promising biocatalyst for the future production of 16α-hydroxylated steroids. Additionally, the enzyme could serve as a model for protein-engineering studies in order to generate variants with changed but still exclusive selectivity. Although CYP154C5 always exhibits the same high regioselectivity and stereoselectivity in the conversion of pregnenolone (**1**), dehydroepiandrosterone (**2**), progesterone (**3**), androstenedione (**4**), testosterone (**5**) and nandrolone (**6**) (Fig. 1), significant differences in reaction rates were observed which are likely to be explained on the basis of its crystal structure and structural insights into its steroid binding.

Of the CYP154 family of P450 monooxygenases, the crystal structures of two CYPs, namely CYP154A1 (PDB entry 1odo; Podust *et al.*, 2004) and CYP154C1 (PDB entry 1gwi; Podust *et al.*, 2003) both from *Streptomyces coelicolor* A3(2), have been reported to date, revealing significant differences in their active sites. The most intriguing finding in the active site of CYP154A1 is the 180° flip of the haem orientation compared with most other CYP structures. CYP154C1 was reported to

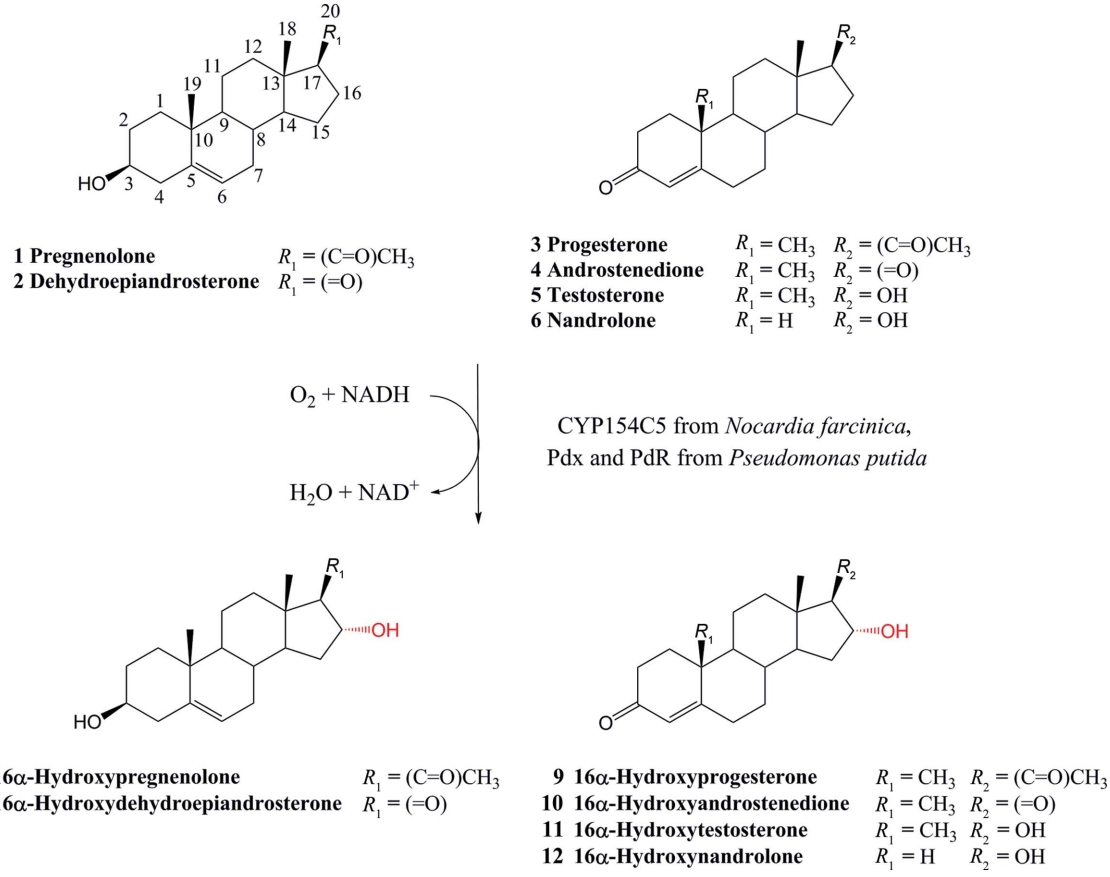


Figure 1
Bioconversion of steroids **1**–**6** by the three-component system composed of CYP154C5, Pdx and PdR. In all cases, 16α-monohydroxylated products (**7**–**12**) are exclusively obtained (Bracco *et al.*, 2013).

hydroxylate various macrolactones, while CYP154A1 catalyzes the cyclization of a dipentaenone molecule (Podust *et al.*, 2003; Cheng *et al.*, 2010). Both enzyme structures, however, were only obtained without substrate bound in their respective active sites, impeding the generation of a reliable CYP154C5 homology model to study enzyme–substrate interactions. Furthermore, although both enzymes belong to the same subfamily of P450 monooxygenases, CYP154C1 and CYP154C5 share only 54% sequence identity at the protein level. Hence, elucidation of the crystal structure of CYP154C5 with substrate bound in the active site is essential to identify the structural determinants for the high regioselectivity and stereoselectivity in steroid hydroxylations and to facilitate the future generation of CYP154C5 variants with improved activity or altered selectivity by protein engineering.

For this purpose, protein crystallization trials were performed, yielding structures with resolutions of up to 1.9 Å of four CYP154C5–steroid complexes with pregnenolone (**1**), progesterone (**3**), androstenedione (**4**) and testosterone (**5**) bound in the active site. Based on these crystal structures, the exclusive selectivity of CYP154C5 can be explained and likely reasons for the differences in steroid binding affinity as well as the conversion of **1** to **6** by this P450 are given.

2. Methods

2.1. General

All of the enzymes required for the molecular-biology work were purchased from New England Biolabs (NEB). For genetic manipulations the strain *Escherichia coli* TOP10 (Invitrogen) was used, while *E. coli* C43 (DE3) (Lucigen) was used for expressions. Plasmid pET-28a(+) was purchased from Novagen (EMD Biosciences). Plasmids pACYCcamA and pACYCcamB were used for recombinant expression of putidaredoxin reductase (CamA or PdR) and putidaredoxin (CamB or Pdx) from *Pseudomonas putida*, respectively, in *E. coli* (Schallmey *et al.*, 2011). Pdx and PdR were applied as redox partners of CYP154C5 in bioconversions. Plasmid pKNL031_M17 from a genomic library of *N. farcinica* containing the gene for CYP154C5 (nfa53110; GenBank NC_006361) was kindly provided by Jun Ishikawa (National Institute of Infectious Diseases, Japan).

Dehydroepiandrosterone (**2**), androstenedione (**4**) and nandrolone (**6**) were purchased from Steraloids Inc., USA. Pregnenolone (**1**), progesterone (**3**), testosterone (**5**), cholesterol, cytochrome *c* from horse heart, formate dehydrogenase from *Candida boidinii*, superoxide dismutase from bovine erythrocytes, catalase from bovine liver and hydroxypropyl β-cyclodextrin were purchased from Sigma–Aldrich. All other solvents, chemicals and media components necessary for the experiments were obtained from Applichem, Roth or Sigma–Aldrich.

2.2. Cloning and expression

Generation of plasmid pIT2cyp154C5 for the recombinant expression of CYP154C5 from *N. farcinica* IFM 10152 in *E. coli* has been described previously (Bracco *et al.*, 2013). Additionally, the *cyp154C5* gene was subcloned from pIT2cyp154C5 into pET-28a(+) vector using *Nde*I and *Hin*dIII restriction sites, resulting in plasmid pET28cyp154C5. Expression of CYP154C5 from the pET28cyp154C5 vector resulted in a fusion protein with an N-terminal His tag.

Expression of CYP154C5 without the N-terminal His tag for protein crystallization was carried out in a 15 l fermenter (Biostat ED, Sartorius) using strain *E. coli* C43 (DE3) (pIT2cyp154C5). A 200 ml Terrific broth (TB) preculture supplemented with tetracycline (10 µg ml^{-1}) was inoculated with 1%(*v*/*v*) overnight culture and incubated for 24 h at 37°C and 250 rev min^{-1}. Finally, a 15 l fermenter containing 9.8 l TB medium supplemented with tetracycline (10 µg ml^{-1}) and 0.1%(*v*/*v*) trace-element solution (0.5 g l^{-1} $CaCl_2.2H_2O$, 0.18 g l^{-1} $ZnSO_4.7H_2O$, 0.10 g l^{-1} $MnSO_4.H_2O$, 20.10 g l^{-1} Na_2EDTA, 16.70 g l^{-1} $FeCl_3.6H_2O$, 0.16 g l^{-1} $CuSO_4.5H_2O$ and 0.18 g l^{-1} $CoCl_2.6H_2O$) was inoculated with 2%(*v*/*v*) preculture. The pH was set to 7.0 and was adjusted during expression using solutions of 5 *M* KOH and H_3PO_4. Dissolved oxygen was measured with a pO_2 electrode. On reaching an OD_{600} of 1, the system was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG; 100 µg ml^{-1} final concentration), and δ-aminolevulinic acid (84 µg ml^{-1} final concentration) was added as a precursor for haem synthesis. Protein expression was carried out at 30°C and 300 rev min^{-1}. After 6 h of induction, glycerol feeding was started at 0.1 g min^{-1}. After 24 h of induction, the feeding was doubled (0.2 g min^{-1}). Finally, after 30 h of induction, a feeding rate of 0.4 g min^{-1} was set until the end of the fermentation process. After 48 h of expression, the cells were harvested by centrifugation (4420*g*, 4°C for 15 min) and washed once with 50 m*M* potassium phosphate buffer pH 7.4 before storing the pellets at −20°C.

For bioconversions using the three-component system consisting of CYP154C5, Pdx and PdR, all three proteins were expressed separately in *E. coli*. Expression of CYP154C5 containing an N-terminal His tag was performed in a 5 l fermenter (Biostat M, Sartorius) using *E. coli* C43 (DE3) (pET28cyp154C5) according to the protocol described above with the following modifications: kanamycin (50 µg ml^{-1} final concentration) was used as an antibiotic instead of tetracycline, IPTG at a final concentration of 200 µg ml^{-1} was added to induce protein expression and no glycerol feeding was applied. Expression of PdR and Pdx using *E. coli* C43 (DE3) (pACYCcamA) and *E. coli* C43 (DE3) (pACYCcamB), respectively, was performed in 2 l shake flasks. 200 ml TB medium supplemented with chloramphenicol (25 µg ml^{-1}) and 0.1%(*v*/*v*) trace-element solution were inoculated with 1%(*v*/*v*) overnight culture of the respective strain and incubated at 37°C and 250 rev min^{-1} until an OD_{600} of 1 was reached. Protein expression was then induced with IPTG (200 µg ml^{-1} final concentration). After 48 h of incubation at 30°C and 250 rev min^{-1}, the cells were recovered by centrifugation (4420*g*, 4°C for 15 min) and washed once with 50 m*M* potassium phosphate buffer pH 7.4 and the pellets were stored at −20°C.

2.3. Purification

2.3.1. Purification of CYP154C5. For the purification of CYP154C5 without the N-terminal His tag, *E. coli* C43(DE3) (pIT2cyp154C5) cells were resuspended in 20 m*M* MES buffer pH 6.5 containing 2 m*M* dithiothreitol (DTT). Typical protease inhibitors were added and cell disruption was carried out by initial lysozyme treatment followed by sonication on ice [using a Sonopuls sonicator (Bandelin) with MS73 sonotrode (Bandelin) at 60% power for five cycles of 30 s sonication and 30 s resting on ice]. The resulting cell lysate was centrifuged at 40 000*g* at 4°C for 45 min to remove all cell debris. Purification of CYP154C5 by liquid chromatography was performed on a Q Sepharose FF column (GE Healthcare) equilibrated with 20 m*M* MES buffer pH 6.5 containing 2 m*M* DTT and the column was rinsed once with the same buffer after loading the cell-free extract. For elution of the target protein, a linear gradient of 0–1 *M* KCl in 20 m*M* MES buffer pH 6.5 containing 2 m*M* DTT was applied. Fractions containing CYP154C5 were combined, desalted and loaded onto a Mono Q HR1010 column (GE Healthcare) with an elution profile as outlined above. Afterwards, the fractions containing CYP154C5 were combined and dialyzed overnight at 4°C against a 50-fold volume of 20 m*M* MES buffer pH 6.5 containing 2 m*M* DTT and 1 *M* ammonium sulfate using SERVAPOR dialysis tubing (SERVA) with a cutoff of 12–14 kDa. The dialysed protein solution was then loaded onto a TSK Phenyl column (Tosoh Bioscience) previously equilibrated with 20 m*M* MES buffer pH 6.5 containing 2 m*M* DTT and 1 *M* ammonium sulfate. The target protein was eluted with a linear gradient of 1–0 *M* ammonium sulfate in 20 m*M* MES buffer pH 6.5 containing 2 m*M* DTT. Finally, the collected fractions containing CYP154C5 were concentrated to 1 ml by ultrafiltration at 4000*g* at 4°C using an Amicon Ultra-15 centrifugal filter unit with a 30 kDa cutoff (Millipore). Subsequently, the concentrated solution was loaded onto a HiLoad 16/60 Superdex 75 gel-filtration column (GE Healthcare) equilibrated with 20 m*M* MES buffer pH 6.5 containing 100 m*M* KCl and 2 m*M* DTT and eluted isocratically. An SDS–PAGE gel of purified CYP154C5 after the final gel-filtration step is shown in Supplementary Fig. S1[1].

The purity of CYP154C5 was monitored by SDS–PAGE under nonreducing conditions with alkylation of the samples prior to electrophoresis by incubation of 2 µl protein solution in 10 µl 100 m*M* ammonium bicarbonate, 25 m*M* iodoacetamide for 30 min in the dark (Lane, 1978). Gels were washed in Milli-Q H_2O and stained with a sensitive Coomassie Brilliant Blue 250G solution (Kang *et al.*, 2002). For the purpose of quality analysis of the final CYP154C5 preparation, the total mass of CYP154C5 solution previously desalted with a C4 ZipTip (Millipore) was measured using a Q-TOF2 mass spectrometer (Micromass) and DLS measurements of 1 µl protein droplets under paraffin oil were carried out using an in-house light-scattering device (self-constructed) in order to validate the monodispersity of the sample.

Purification of CYP154C5 carrying an N-terminal His tag was performed by anion-exchange chromatography and subsequent immobilized metal-ion affinity chromatography (IMAC). *E. coli* C43(DE3) (pET28cyp154C5) cells were resuspended in 50 m*M* potassium phosphate buffer pH 7.4, and PMSF (phenylmethylsulfonyl fluoride; 17 µg ml^{-1} final concentration) was added as a protease inhibitor. Cell disruption was performed using a French press (EmulsiFlex Homogenizer, Avestin) with five cycles of 10.3 MPa. The cell debris was removed by centrifugation at 17 700*g* at 4°C for 45 min and the supernatant was filtered through 0.22 µm pore-size filters. The resulting cell-free extract was loaded onto a Q Sepharose FF column (GE Healthcare) pre-equilibrated with 50 m*M* potassium phosphate buffer pH 7.4. Protein was eluted with a linear gradient of 0–500 m*M* KCl in 50 m*M* potassium phosphate buffer pH 7.4. Fractions containing CYP154C5 were combined and loaded onto a HisTrap HP column (GE Healthcare) pre-equilibrated with 50 m*M* potassium phosphate buffer pH 7.4 containing 500 m*M* KCl. The column was then washed with 50 m*M* potassium phosphate buffer pH 7.4 containing 20 m*M* imidazole and 500 m*M* KCl in order to remove nonspecifically bound protein. Subsequently, CYP154C5 was eluted with 250 m*M* imidazole in 50 m*M* potassium phosphate buffer pH 7.4 containing 500 m*M* KCl. Fractions containing the P450 monooxygenase were combined and concentrated by ultrafiltration using Amicon centrifugal filters (Millipore) with 30 kDa cutoff. The concentrate was loaded onto a PD10 column (GE Healthcare) and was eluted with 50 m*M* potassium phosphate buffer pH 7.4 to desalt the protein solution. Additionally, dialysis in 50 m*M* potassium phosphate buffer pH 7.4 was performed to remove all traces of imidazole.

2.3.2. Purification of PdR and Pdx. For purification of PdR, *E. coli* C43 (DE3) (pACYCcamA) cells were resuspended in 50 m*M* potassium phosphate buffer pH 7.4 and cell disruption was performed using a French press (EmulsiFlex Homogenizer, Avestin) with five cycles at 10.3 MPa. Cell debris was removed by centrifugation at 17 700*g* at 4°C for 45 min and the supernatant was filtered through 0.22 µm pore-size filters. The resulting cell-free extract was loaded onto a Q Sepharose FF column (GE Healthcare) pre-equilibrated with 50 m*M* potassium phosphate buffer pH 7.4. Protein was eluted with a linear gradient of 0–500 m*M* KCl in 50 m*M* potassium phosphate buffer pH 7.4. PdR-containing fractions were combined, adjusted to 1.5 *M* KCl and loaded onto a Phenyl Sepharose FF column (GE Healthcare) pre-equilibrated with 50 m*M* potassium phosphate buffer pH 7.4 containing 1.5 *M* KCl. The target protein was eluted with a linear gradient of 1.5–0 *M* KCl in 50 m*M* potassium phosphate buffer pH 7.4. PdR-containing fractions were combined and concentrated by ultrafiltration using Amicon centrifugal filters (Millipore) with 30 kDa cutoff. The concentrate was loaded onto a PD10 column (GE Healthcare) and eluted with 50 m*M* potassium phosphate buffer pH 7.4 for desalting.

Analogously, purification of Pdx also started with an ion-exchange Q Sepharose FF column as described for the PdR purification. Pdx-containing fractions were combined and

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: CB5063).

higher molecular-weight proteins were removed by ultrafiltration using a 30 kDa cutoff membrane. The resulting Pdx-containing filtrate was subsequently concentrated using Amicon centrifugal filters (Millipore) with 10 kDa cutoff.

2.4. Crystallization and data collection

All crystallization experiments with CYP154C5 were carried out as sitting-drop vapour-diffusion experiments with 2.0 µl droplets and 250 µl reservoirs in crystallization plates from Taorad. Crystallization droplets were created by mixing 1.0 µl protein solution (at 40 mg ml^{-1}) and crystallization condition. During crystallization experiments, the six steroid hormones **1**–**6** were used. Stock solutions of hormones were prepared in 100% DMSO with concentrations of 50 m*M* for **1** and **4** and 100 m*M* for **2**, **3**, **5** and **6**. All solutions used for crystallization experiments were filtered (0.22 µm diameter) before use. Steroid hormones were added from the stock solutions to the pure protein solution prior to filtration. Diffracting crystals were obtained within a 'window' of crystallization conditions consisting of 0.2–0.3 *M* magnesium formate pH 6.0–6.5 in the presence of 0.5–2 equivalents of steroid substrate.

For X-ray crystallography, the loop-mounted crystals were first equilibrated for several seconds in solutions of 0.3 *M* magnesium formate pH 6.5 and then in solutions enriched with increasing concentrations of glycerol as a cryoprotective agent, and were finally flash-cooled at 100 K in a liquid-nitrogen stream. Near-complete X-ray data sets were collected using a Bruker FR591 rotating-anode X-ray generator and a MAR345dtb detector. Diffraction data were processed using *iMOSFLM* (Battye *et al.*, 2011), their symmetry was analyzed using *POINTLESS* and they were scaled with *SCALA* (Evans, 2006), both of which are from the *CCP*4 suite (Winn *et al.*, 2011).

2.5. Structure determination, refinement and modelling

The structures of CYP154C5 were solved by molecular replacement using *MOLREP* from *CCP*4 (Vagin & Teplyakov, 2010). For CYP154C5 in complex with **1**, the structure of CYP154C1 (PDB entry 1gwi; Podust *et al.*, 2003) was processed with *CHAINSAW* from *CCP*4 (Stein, 2008) and used as a model for molecular replacement. The fully solved CYP154C5–pregnenolone structure was then used as the molecular-replacement model for all of the other CYP154C5 structures. All structures were iteratively refined by manual inspection of the electron density with *Coot* (Emsley & Cowtan, 2004) and refinement with *REFMAC*5 (Murshudov *et al.*, 2011). Refinement to convergence was carried out with isotropic *B* factors and using TLS parameters initially computed using the *TLSMD* webserver (Painter & Merritt, 2006*a*,*b*). Alternative conformations were modelled for a number of side chains, while occupancies were adjusted manually for some atoms. Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 4j6b, 4j6c, 4j6d and 4jbt. A model for the substrate-free open CYP154C5 conformation was built based on the *A* chain of CYP154C1 (PDB entry 1gwi) by manually fitting the CYP154C5 residues and backbone into the electron density reported for 1gwi.

2.6. Enzyme assays

The CYP154C5 concentration was measured using CO-difference spectra (Omura & Sato, 1964) and was calculated based on the maximum absorbance of CO-bound CYP154C5 at 450 nm (ε_{450} = 91 mM^{-1} cm^{-1}). The concentrations of pure Pdx and PdR solutions were determined at 455 nm (ε_{455} = 10.4 mM^{-1} cm^{-1}) and 454 nm (ε_{454} = 10.0 mM^{-1} cm^{-1}), respectively (Gunsalus & Wagner, 1978). The activity of the electron-transfer components Pdx and PdR was determined by a cytochrome *c* reduction assay (Lacour *et al.*, 1998), monitoring the increase in absorbance at 550 nm (ε_{550} = 19.1 mM^{-1} cm^{-1}).

2.7. Substrate binding

Dissociation constants (K_d) of CYP154C5 for the six different steroids were determined by spectroscopic measurements upon titration of the purified enzyme with increasing steroid concentrations (Schenkman *et al.*, 1967). Purified CYP154C5 carrying an N-terminal His tag was diluted with 50 m*M* potassium phosphate buffer pH 7.4 in order to give a final enzyme concentration of 3 µ*M*. To this mixture, substrate was added at concentrations of 0–100 µ*M* from 0.1, 0.5 and 1 m*M* steroid stock solutions in 0.1%(*w*/*v*) hydroxypropyl-β-cyclodextrin. The absorbance spectra of all samples were measured using a Cary 50 spectrophotometer (Agilent) in the range 300–500 nm at 30°C. Difference spectra were obtained by subtracting the absorbance spectra of 3 µ*M* CYP154C5 in 50 m*M* potassium phosphate buffer pH 7.4 without steroid substrate but with the addition of an equivalent amount of buffer. Each substrate concentration was measured in duplicate. Dissociation constants for the different steroids were ultimately obtained by fitting the plots of the resulting differences in absorbance ($Abs_{396\ nm} - Abs_{419\ nm}$) against the applied substrate concentration using the tight binding equation (Williams & Morrison, 1979).

2.8. Differential scanning fluorimetry experiments

All differential scanning fluorimetry (DSF) experiments were carried out in a MyiQ2 RT-PCR thermal cycler (Bio-Rad) using film-sealed 96-well plates and a filter set to λ_{ex} = 485 nm and λ_{em} = 530 nm. Each single reaction had a volume of 50 µl and contained 0.15 mg ml^{-1} (~3.3 µ*M*) CYP154C5, 12 m*M* MES, 60 m*M* KCl, 1.2 m*M* DTT or tris(2-carboxyethyl)phosphine (TCEP) and 5× SYPRO Orange. Thermal unfolding data for CYP154C5 were recorded with steroid hormones **1**–**6** from a stepwise temperature gradient of 0.5°C steps from 25 to 95°C. The steroids were dissolved in 100% DMSO to obtain stock solutions of 200 µ*M* (substrates **2**, **3**, **5** and **6**) or 100 µ*M* (substrates **1** and **4**). Thermal unfolding data for CYP154C5 were collected at different molar equivalent (eq) concentrations of steroid compared with the CYP154C5 concentration, namely 0, 1, 2, 4 and 8 eq for substrates **2**, **3**, **5**

Table 1
X-ray crystallographic data-collection, refinement and model statistics.

	CYP154C5–pregnenolone (**1**)	CYP154C5–progesterone (**3**)	CYP154C5–testosterone (**5**)	CYP154C5–androstenedione (**4**)
PDB code	4j6b	4j6c	4j6d	4jbt
Data-collection statistics				
Space group	*H*3	*H*3	*H*3	*H*3
Unit-cell parameters (Å)	$a = b = 103.12$, $c = 217.83$	$a = b = 103.00$, $c = 217.86$	$a = b = 102.99$, $c = 218.42$	$a = b = 102.80$, $c = 217.52$
Wavelength (Å)	1.5418	1.5418	1.5418	1.5418
Resolution (Å)	82.63–2.20 (2.32–2.20)	82.55–1.90 (2.00–1.90)	82.60–2.40 (2.53–2.40)	82.393–2.200 (2.32–2.20)
$\langle I/\sigma(I)\rangle$	10.5 (3.3)	10.9 (4.1)	5.9 (2.9)	7.6 (2.4)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	99.8 (99.1)	93.0 (82.8)
Unique reflections	43848	67934	33730	40488
Average multiplicity	5.9 (5.7)	5.9 (5.7)	3.4 (3.3)	2.4 (2.2)
$R_{\rm merge}$	0.139 (0.559)	0.099 (0.382)	0.141 (0.391)	0.076 (0.339)
Refinement statistics				
Resolution (Å)	82.63–2.20 (2.26–2.20)	82.55–1.90 (1.95–1.90)	82.60–2.40 (2.46–2.40)	51.40–2.20 (2.26–2.20)
Reflections used	43846 (3025)	67930 (3448)	33701 (2348)	40105 (2359)
$R_{\rm work}$	0.156 (0.215)	0.192 (0.229)	0.180 (0.181)	0.183 (0.250)
$R_{\rm free}$	0.202 (0.245)	0.238 (0.264)	0.244 (0.269)	0.248 (0.329)
CC $(F_{\rm o} - F_{\rm c})_{\rm work}$	0.960	0.943	0.935	0.954
CC $(F_{\rm o} - F_{\rm c})_{\rm free}$	0.931	0.909	0.870	0.915
Average *B* factor (Å^2)	42.30	28.87	44.14	35.41
Model statistics				
Monomers in asymmetric unit	2	2	2	2
Nonsolvent atoms	6412	6489	6406	6413
Solvent atoms	490	694	375	321
R.m.s.d.$_{\rm ideal}$, bond lengths (Å)	0.010	0.013	0.011	0.008
R.m.s.d.$_{\rm ideal}$, bond angles (°)	1.658	1.607	1.415	1.256

and **6** and 0, 0.5, 1, 2 and 4 eq of substrates **1** and **4** owing to their lower solubility. The mentioned relative steroid concentrations correspond to absolute DMSO concentrations of 0, 1.6, 3.2, 6.4 and 12.8%(*v*/*v*), respectively.

2.9. Bioconversions

Bioconversions using purified CYP154C5 (with N-terminal His tag), Pdx and PdR were carried out to determine turnover numbers (TON; micromoles of substrate consumed per minute per micromole of CYP154C5) and coupling efficiencies (percentage of electrons from consumed cofactor NADH used for product formation) of the three-component system in steroid conversion. For TON determination, each reaction consisted of 3 µ*M* CYP154C5, 3 µ*M* PdR, 16 µ*M* Pdx, 0.5 U ml^{-1} formate dehydrogenase from *C. boidinii*, 150 m*M* sodium formate, 300 U ml^{-1} catalase from bovine liver, 50 µ*M* NADH and 2 m*M* of the respective steroid [5 m*M* stock in 3.6%(*w*/*v*) hydroxypropyl-β-cyclodextrin in water] in 50 m*M* potassium phosphate buffer pH 7.4. All bioconversions were carried out on a 5 ml scale for 8 h at 30°C and 250 rev min^{-1}. During bioconversions, 0.5 ml samples were taken every hour for subsequent HPLC and GC analysis. Each sample was extracted twice with ethyl acetate (300 and 250 µl) and once with chloroform (300 µl); the organic phases were combined, dried with sodium sulfate and the solvent was removed under reduced pressure. Each reaction was performed in duplicate. TON were calculated for the time period for which a linear increase in substrate consumption was observed (0–1.5 h for **3**, 0–2 h for **1** and **2**, 0–3 h for **4** and **5**, and 0–4 h for **6**).

For coupling efficiency determination, reactions of 1 ml total volume consisting of 0.4 µ*M* CYP154C5, 0.4 µ*M* PdR, 14 µ*M* Pdx, 600 U ml^{-1} catalase from bovine liver, 200 U ml^{-1} superoxide dismutase, 200 µ*M* NADH and 1 m*M* of the respective steroid [5 m*M* stock in 3.6%(*w*/*v*) hydroxypropyl-β-cyclodextrin in water] in 50 m*M* potassium phosphate buffer pH 7.4 were incubated at 25°C. Firstly, complete depletion of NADH was monitored spectrophotometrically at 340 nm (ε_{340} = 6.22 mM^{-1} cm^{-1}). Each 0.5 ml reaction was then extracted as described before and analyzed by GC or HPLC in order to determine steroid conversion for the calculation of coupling efficiencies. All reactions were performed in duplicate.

2.10. HPLC and GC analysis

Substrate consumption and product formation in bioconversions of steroids using CYP154C5 together with Pdx and PdR were measured by GC (steroids **1** and **2**) and HPLC (steroids **3**–**6**). In the case of **1** and **2** the solid residues after reaction workup were redissolved in chloroform containing 30 m*M* cholesterol as an internal standard and were analyzed on a GC2010 gas chromatograph (Shimadzu) equipped with an OPTIMA 17 ms column (Macherey Nagel, 0.25 mm × 30 m) with a linear gradient starting at 250°C and heating to 300°C at 10°C min^{-1}. The injector and detector temperature were set to 350 and 300°C, respectively. Substrates and products were detected by a flame-ionization detector (FID). Substrates **1** and **2** eluted with retention times of 9.69 and 7.94 min, respectively, whereas their products (**7** and **8**) eluted at 9.27 and 10.18 min, respectively. The dried residues of bioconversions with steroids **3**–**6** were dissolved in acetonitrile:water (60:40) and injected onto an Ultra Fast Liquid Chromatograph (UFLC; Prominence, Shimadzu) equipped with a C18 reversed-phase column (CS Chromatographie Service, 250 × 4.5 mm) at 50°C. A mixture of acetonitrile and water (60:40) was used as the mobile phase with a flow rate of

1.2 ml min^{-1}. Substrates and their respective products were detected by UV absorbance at 242 nm. The substrates eluted at 10.70 min (**3**), 6.34 min (**4**), 6.08 min (**5**) and 5.51 min (**6**), whereas the products were detected at 4.23 min (**9**), 3.93 min (**10**), 4.47 min (**11**) and 3.29 min (**12**). In all cases, conversions were calculated based on substrate consumption.

3. Results and discussion

3.1. Overall structure of CYP154C5

Initial crystallization trials with substrate-free CYP154C5 yielded no crystals. Upon the introduction of steroid substrates into the crystallization screening process, diffracting crystals were obtained for CYP154C5 with all six steroids mentioned above. However, data sets of publishable quality could only be obtained for the complexes of CYP154C5 with pregnenolone (**1**), progesterone (**3**), androstenedione (**4**) and testosterone (**5**). The structure of the CYP154C5–pregnenolone complex (PDB entry 4j6b) could be solved successfully by molecular-replacement phasing with CYP154C1 from *S. coelicolor* (PDB entry 1gwi; Podust *et al.*, 2003), thus providing a perfect molecular-replacement model that allowed the structure solution of the other CYP154C5–steroid complexes. The statistics of data collection and structure refinement are given in Table 1.

The overall structure of CYP154C5 is characterized by the helix-rich cytochrome P450 monooxygenase fold, as shown in Fig. 2 for the *A* chain of the CYP154C5–progesterone complex. At the bottom of the active site, a haem *b* molecule is bound to the typically conserved residues His99, Arg103, Arg297, Tyr320, His355 and Cys357. Clear and distinct electron density was observed for all co-crystallized steroid substrates, as shown exemplarily for the structures with progesterone and androstenedione (Fig. 3), thus allowing an accurate and unambiguous determination of substrate positions and orientations within the active site of CYP154C5. The active-site-bound steroid molecule is surrounded by a narrow channel with a small surface opening. Despite the strong hydrophobicity of the steroids, the channel opening is not extensively surrounded by hydrophobic patches that could facilitate initial substrate surface association and subsequent entrance of the active-site channel by the steroid. Instead, the crystallographically observed substrate-channel opening is formed by Arg77, Asp86, Val87, Asp88, Ile193, Asp194 and His197, which exhibit primarily polar or charged side chains.

The four structures published here, as well as the unpublished structures of the CYP154C5–steroid complexes with dehydroepiandrosterone (**2**) and nandrolone (**6**), exhibit an asymmetric unit formed of two CYP154C5 molecules. Interestingly, the two molecules in the asymmetric unit are covalently linked *via* a Cys4(*A*)–Cys4(*B*) disulfide bond and are surrounded by adjacent bivalent metal-coordination sites forming crystal contacts with neighbouring asymmetric units, as visible in the CYP154C5–testosterone structure (Fig. 4). The covalent disulfide interaction seems to be crucial for successful crystallization, as crystallization trials with the His-tagged variant of CYP154C5 failed. This was demonstrated by determining the evolution of the second virial coefficient over time using static light-scattering experiments in microdroplets

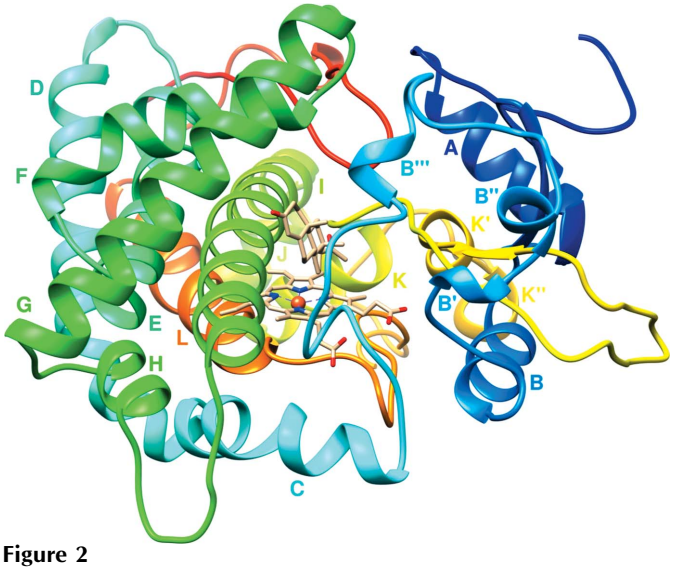


Figure 2
Overall structure of CYP154C5 in complex with **3** as a ribbon model with P450 helix numbering.

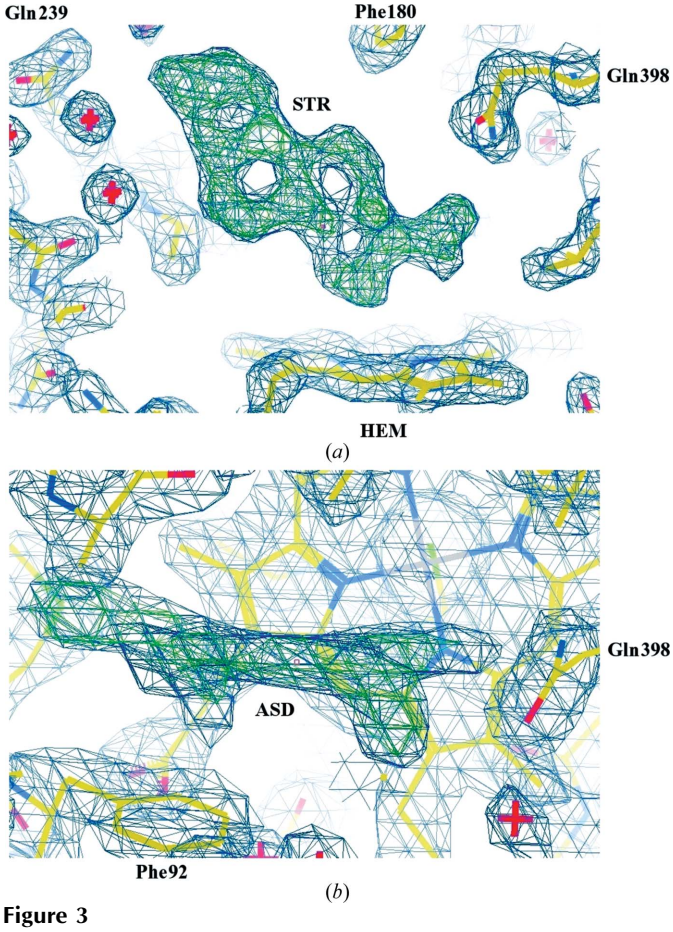


Figure 3
Electron densities of steroid substrates bound to the active site of CYP154C5 with (*a*) progesterone (**3**) and (*b*) androstenedione (**4**) ($2F_o - F_c$ map in blue, $F_o - F_c$ map in green).

Table 2
R.m.s.d. values in Å calculated for the superpositions of selected CYP154C5 chains (italics) and of selected CYP154C5 active sites (bold).

Letters in parentheses denote the superposed chains. The active-site superpositions consist of Met84, Phe92, Phe179, Gln239, Ala240, Ala243, Ala244, Thr248, Val291, Leu294, Gln398 and the haem complex.

	CYP154C5– pregnenolone (**1**)	CYP154C5–progesterone (**3**)	CYP154C5–androstenedione (**4**)	CYP154C–testosterone (**5**)
CYP154C5–pregnenolone (**1**)	*0.175 (A–B)*	*0.270(A–A)*	*0.338 (A–A)*	*0.548 (A–A)*
	0.397 (A–*B*)	**0.129 (A–A)**	**0.291 (A–A)**	**0.338 (A–A)**
CYP154C5–progesterone (**3**)	*0.280 (B–B)*	*0.147 (A–B)*	*0.407 (A–A)*	*0.588 (A–A)*
	0.133 (*B*–*B*)	**0.401 (A–*B*)**	**0.284 (A–A)**	**0.339 (A–A)**
CYP154C5–androstenedione (**4**)	*0.316 (B–B)*	*0.386 (B–B)*	*0.161 (A–B)*	*0.525 (B–B)*
	0.273 (*B*–*B*)	**0.259 (*B*–*B*)**	**0.399 (A–*B*)**	**0.261 (*B*–*B*)**
CYP154C5–testosterone (**5**)	*0.373 (B–B)*	*0.460 (B–B)*	*0.591 (A–A)*	*0.489 (A–B)*
	0.349 (*B*–*B*)	**0.349 (*B*–*B*)**	**0.246 (A–A)**	**0.400 (A–*B*)**

(in-house-built system; data not shown). With Cys4 being located on a highly flexible N-terminal loop (Fig. 2, ribbon model; dark blue), the dimerization constrains the opportunity for the loop to undergo thermal motion and thus strongly increases the overall stability of CYP154C5. Although electron density for the disulfide bond could only be observed in the CYP154C5 structure with **5** as well as in the unpublished structures with **2** and **6**, it can be assumed that this disulfide bond is also present in the other CYP154C5 structures. The N-terminal loop region, however, exhibits a relatively high remaining flexibility, as shown by relatively high *B* factors from Cys4 to Asp8. Owing to those high *B* factors and the comparably poor observed electron density in this region, the bivalent metal ion in the adjacent coordination site cannot be determined with absolute certainty from our recorded data sets. The amino-acid composition and coordination geometry of the metal-binding site might indicate a typical Zn^{2+}-coordination site (Harding, 2001). However, no anomalous scattering signal at the Zn *K* X-ray absorption edge could be observed for any of the reported structures at a data-collection wavelength of 1.54 Å, while all of the structures exhibited strong anomalous density for the haem iron (data not shown). Furthermore, Zn^{2+} was not added during protein purification and crystallization, in contrast to the high Mg^{2+} concentrations that were present during crystallization. Hence, it is more likely that Mg^{2+} was incorporated into the metal-binding site upon enzyme dimerization and crystallization. Nevertheless, it cannot be excluded that Zn^{2+}, if present in sufficient amounts, might be the metal ion that is preferentially incorporated in the metal-binding site.

As shown by the r.m.s.d. values in Table 2, structural differences between the CYP154C5–steroid complexes are small. Except for the 'self-alignments' (*A*–*B*, red) the active-site r.m.s.d.s are generally lower than the overall r.m.s.d.s, showing structural conservation of the residues that are directly involved in ligand binding. The higher active-site r.m.s.d.s for the 'self-alignments' can be explained by chain-specific structural differences in the asymmetric unit being required for the formation of crystal contacts. Comparative analysis of the different enzyme–substrate complexes should therefore only be carried out for the same chains. The overall and active-site r.m.s.d. between the complexes with **1** and **3** are much smaller at the protein level than those between the complexes with **1** or **3** and the other two steroids. This could be explained by an adaption of CYP154C5 to the larger steroids **1** and **3** which share an acetyl substituent at position C17, while substrates **4** and **5** bear a smaller C17 oxygen substituent.

3.2. Active-site architecture

The co-crystallized steroids share an overall structure of two peripheral polar groups (hydroxyl or carbonyl groups) and the typical strongly hydrophobic steroid core (Fig. 1). This architecture is completely resembled by the active site of CYP154C5, forming a cavity of two oppositely located polar regions that define a flat central hydrophobic 'tube' (Fig. 5). The replication of the polarity distribution of the steroid and the flat substrate-channel shape allow an initial orientation of the steroid molecule within the active site. Ultimately, these steroids are bound in the active site so that only C16 faces the haem iron with sufficient proximity to allow attack by the highly reactive Fe(IV)=O complex (compound I) during catalysis (Denisov *et al.*, 2005). This mode of binding explains the exclusive regioselectivity of this enzyme that had been observed previously in bioconversions with steroids **1**–**6**. Additionally, the high stereoselectivity of CYP154C5 can be

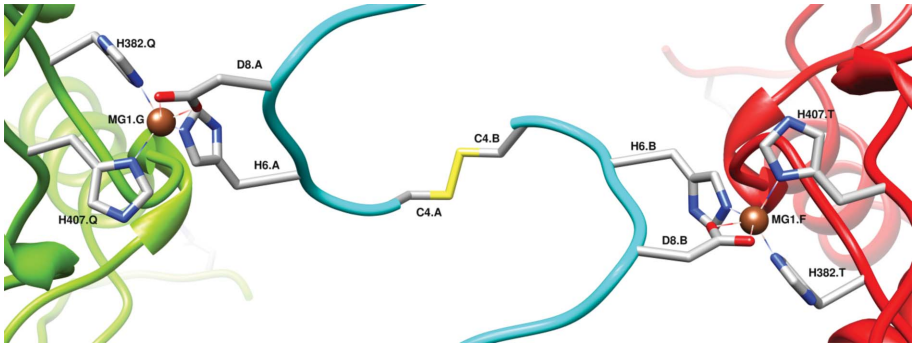
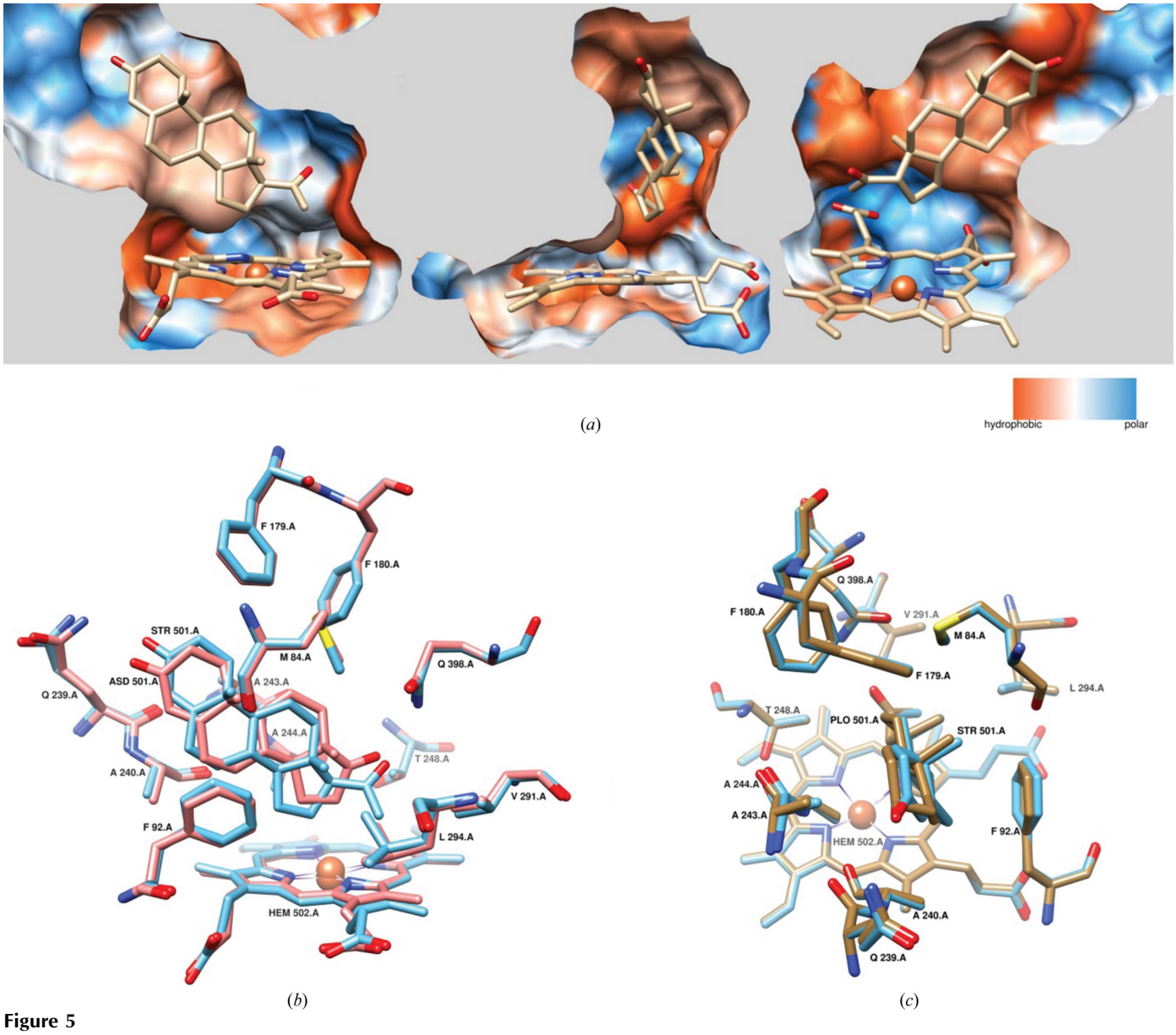


Figure 4
Disulfide dimerization site in the CYP154C5–testosterone structure. Flexible N-terminal loops (blue) are connected *via* Cys4. Adjacent Mg^{2+}-coordination sites form crystal contacts with other neighbouring CYP154C5 molecules in the unit cell.

deduced from the relative orientation of the flat hydrophobic tube relative to the haem. Substrate binding is not centred above the haem iron (Fig. 5); instead, the steroid molecule is laterally shifted towards the haem carboxyl groups. Consequently, the catalytic haem-bound oxygen radical can attack the steroid C16 only from the α-direction in such a way that CYP154C5 exclusively catalyzes α-hydroxylation of the investigated steroids.

The steroid-contacting active-site residues of CYP154C5 are visualized in Fig. 6. Several aromatic and aliphatic residues, namely Val87, Met84, Phe92, Phe179, Phe180, Ala240, Ala243, Ala244, Val291 and Leu294, form the flat hydrophobic tube flanking the substrate and positioning it *via* steric and hydrophobic interactions. Possibly the most important residue for achieving the observed high regioselectivity of CYP154C5 in steroid hydroxylation seems to be Gln398, as it forms a hydrogen bond to the C17-substituent oxygen of the steroid, thus stabilizing the substrate in a position that allows C16 hydroxylation. The configuration of Gln398 is stabilized *via* a network of hydrogen bonds with Thr397 and several water molecules bound to the backbone peptides of the loop/β-strand region between α-helices K and K′ (Fig. 2, yellow). On the opposite site of Gln398, an additional polar surface is established by Gln239. Although this residue is not directly involved in enzyme–substrate interaction as observed by X-ray crystallography, it might play an important role in the dynamic process of steroid binding in the active site as it generates a polar match for the steroid O3. At least for the two larger steroids **1** and **3**, this polar region contributes to substrate binding *via* water-mediated hydrogen-bonding



(*a*) (*b*) (*c*)

Figure 5
Active site of CYP154C5. (*a*) Three perspectives of the active site with **3** and haem *b*. (*b*) Superposition of the active-site residues surrounding **3** (cyan) or **4** (magenta). (*c*) Superposition of the active-site residues surrounding **3** (cyan) or **1** (brown). The surface of the active site is calculated from 4j6c and is coloured by hydrophobicity.

networks between Gln239 and O3. Although those waters are not visible (and hence not well defined) in the other two structures, it can be assumed that water molecules are present there as well as this polar region forms the substrate entrance and is therefore in contact with the aqueous solvent surrounding the protein. The absence of ordered water molecules in this region in the structures with **4** and **5** might be a consequence of the lower resolution of 4jbt and 4j6d, respectively.

As previously indicated by the active-site r.m.s.d. values in Table 2, as well as demonstrated by the selected superpositions depicted in Figs. 5(*b*) and 5(*c*), the conformation of the CYP154C5 active site is strongly conserved. Nevertheless, a major difference in the steroid binding orientation can be observed between the C17-acetyl-substituted steroids **1** and **3** in comparison to the C17-hydroxyl/carbonyl-substituted steroids **4** and **5**. The superposition of the active site of CYP154C5 with **3** and **4** (Fig. 5*b*) illustrates that the positions and orientations of the steroid-surrounding amino acids are highly similar, while the steroid orientation is significantly changed depending on the C17 substituent. Owing to the internal hydrogen-bond stabilization of the Gln398 configuration in CYP154C5 and the limited number of allowed rotamers, its side chain cannot easily be flipped away to account for bulkier substrates. Thus, the larger acetyl substituent in **1** and **3** induces the rotation of these steroids around an axis located approximately along the C10—C19 bond while maintaining the crucial Gln398–steroid hydrogen bond. Despite this rotation, the Fe–C16 distance for the C17-acetylated steroids (Table 3) is not significantly changed compared with those bearing a C17-hydroxyl/carbonyl substitution which allows the conversion of both types of substrates.

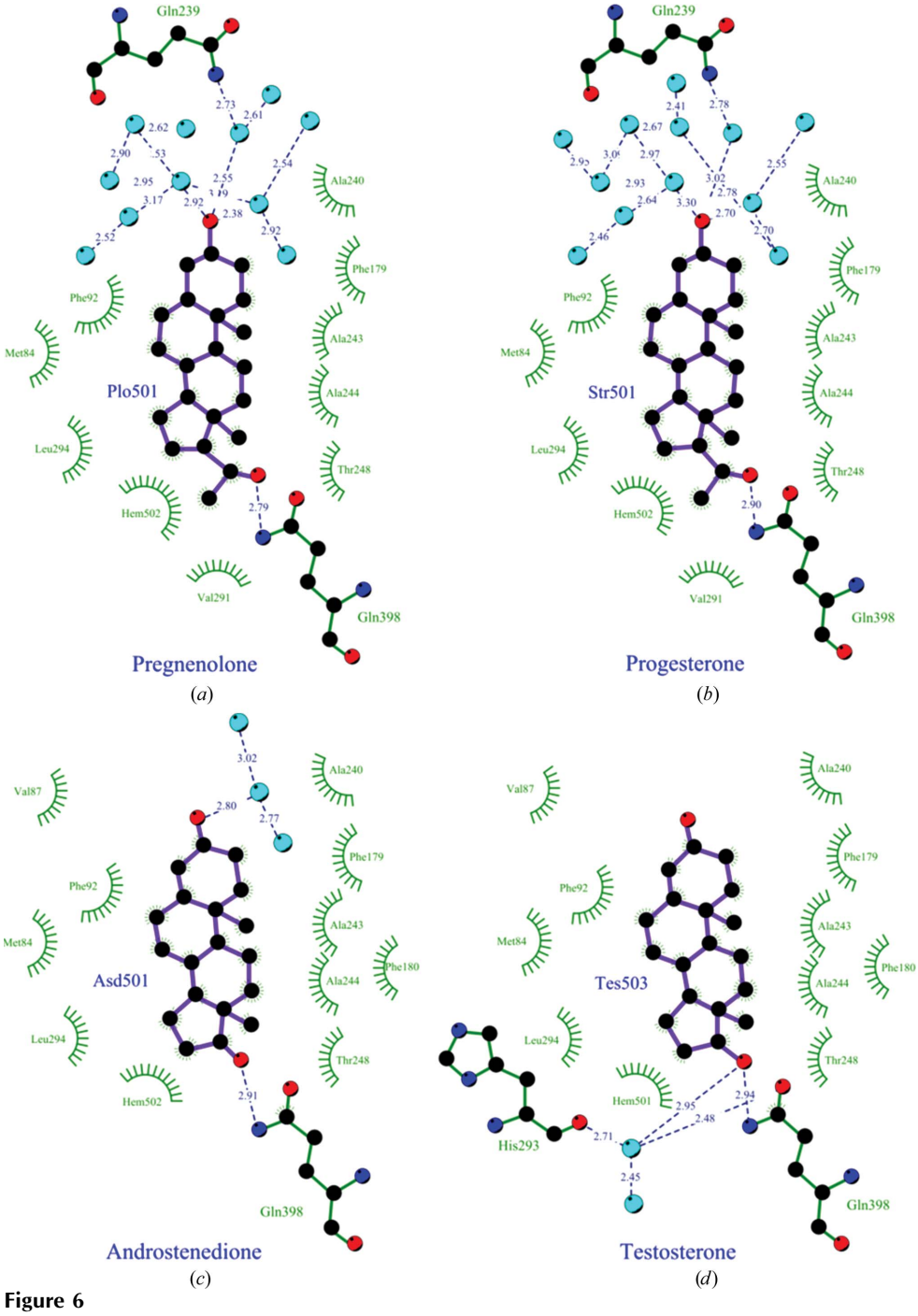


Figure 6
Schematic two-dimensional representation of the active-site architecture of CYP154C5 in complex with (*a*) **1**, (*b*) **3**, (*c*) **4** and (*d*) **5**. Hydrogen bonds are represented by blue dashed lines with the respective distances in Å. Green circle segments indicate hydrophobic interactions.

3.3. CYP154C5–steroid interactions

In order to allow a thorough analysis of the small CYP154C5–steroid interaction differences, additional experiments were carried out to determine the dissociation constants (K_d) of CYP154C5 for each steroid and to elucidate possible ligand-induced enzyme stabilization *via* protein-unfolding temperatures (T_m) determined by differential scanning fluorimetry (DSF; Pantoliano *et al.*, 2001; Senisterra & Finerty, 2009). The K_d determination for P450 monooxygenases is based on an optically measurable spin shift of

Table 3
Turnover numbers (TON) and coupling efficiencies as well as characteristic active-site distances of CYP154C5 with steroids **1**–**6**.

Kinetic data were determined using Pdx and PdR as redox partners for CYP154C5. Respective NADH oxidation rates for this three-component system are given in Supplementary Table S1.

Substrate	TON (min^{-1})	Coupling efficiency (%)	Average Fe–C16 distance (Å)	Average Gln398–O17/20 distance (Å)
1	4.47 ± 0.36	68 ± 1	4.31	2.93
2	5.10 ± 0.11	86 ± 5	4.33	2.89
3	5.70 ± 0.22	84 ± 3	4.51	2.84
4	3.14 ± 0.06	83 ± 1	4.22	2.94
5	1.55 ± 0.23	62 ± 5	4.88	2.91
6	1.33 ± 0.10	40 ± 7	4.56	3.21

the haem iron upon substrate binding. In the absence of substrates, P450 enzymes usually exhibit a low-spin haem iron. Substrate addition shifts the iron to a high-spin state, leading to an absorbance change (in type I spectral changes, usually from 420 to roughly 390 nm) owing to the displacement of a water molecule by the substrate at the sixth coordination position of the haem iron (Schenkman *et al.*, 1967). In particular, CYP154C5 exhibits a steroid concentration-dependent shift from 419 to 386 nm in the presence of the six tested steroid substrates (Supplementary Fig. S2). All CYP154C5–steroid dissociation constants using His-tagged P450 were obtained by fitting the peak-to-trough difference for increasing substrate concentration to the quadratic tight binding equation (Morrison, 1969; Williams & Morrison, 1979). Strikingly, all K_d values appear to be remarkably small, in the range of only 10–100 n*M* (Fig. 7). This tight binding is likely to be a consequence of steroid hydrophobicity, which leads, upon their solvation in aqueous environments, to a thermodynamically unfavourable entropy reduction of the steroid-surrounding water molecules. Active-site binding of the substrate releases these low-entropic waters from the solvation shell of the steroid, thus increasing the overall entropy of the system. In addition, substrate binding also displaces low-entropic water molecules from the hydrophobic CYP154C5 active site. As a consequence of the large hydrophobic interaction interfaces, the entropic energy gain can be assumed to exceed that from hydrogen bonding, rendering the overall process of steroid binding to CYP154C5 predominantly driven by entropy. Similar low K_d values have also been reported very recently for CYP154C3 from *Streptomyces griseus* for the binding of various steroid molecules (Makino *et al.*, 2014).

DSF measurements with non-His-tagged P450 revealed multiple melting events for CYP154C5–substrate mixtures, which could be assigned to a monomeric and a dimeric CYP154C5 species by combinatorial cross-validation with additional experimental results (Supplementary Fig. S3). The dimeric CYP154C5 can be assumed to be the covalently linked disulfide dimer that has been found in the crystal structures (Fig. 4). Dimerization *via* disulfide-bond formation seems to be enforced by DMSO, which has been described to be a mild oxidant for thiol compounds (Wallace, 1964; Wallace & Mahon, 1964, 1965; Epstein & Sweat, 1967; Tam *et al.*, 1991) and which has been used in crystallization trials as well as in DSF measurements for dissolving the steroids. Generally, the dimeric form was found to be much more stable than the monomeric CYP154C5 in all DSF experiments (Fig. 7), thus confirming the importance of fixing the N-terminal loop for successful CYP154C5 crystallization. However, in the reducing environment of the cytosol it can be assumed that only the monomeric form of CYP154C5 exists and that the dimeric form is likely to play no physiological role.

For all investigated steroids, a substrate-binding-induced thermal stabilization of CYP154C5 was observed which, interestingly, is more pronounced for the dimeric compared with the monomeric P450 (Fig. 7). This effect can be attributed to a structural plasticity that has been described for several bacterial CYPs, which preferentially exhibit an open conformation in the absence and a closed conformation in the presence of substrate in the active site (Sherman *et al.*, 2006; Savino *et al.*, 2009; Lee *et al.*, 2010; Montemiglio *et al.*, 2010; Yasutake *et al.*, 2010; Yang *et al.*, 2011; Stoll *et al.*, 2012). This conformational transition is mediated *via* helices F and G and the B–C loop region, with at least one of these being flipped out upon opening of the active site. Indeed, all of the CYP154C5–steroid crystal structures reported in this paper exhibit the closed conformation. As the surface opening of the active-site channel is too small to allow steroid entry, an open

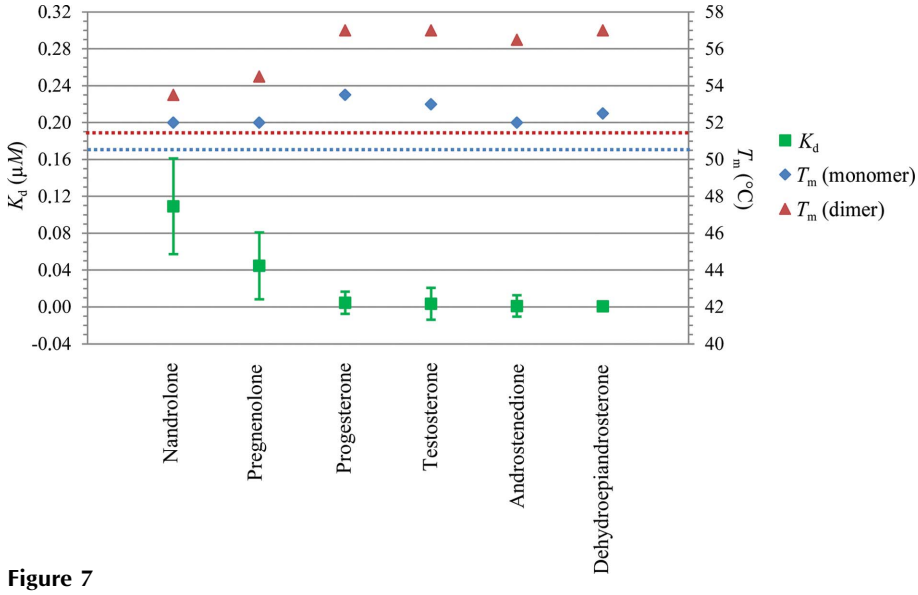


Figure 7
K_d and T_m values for CYP154C5–steroid complexes. K_d values were determined at 30°C for the monomeric His-tagged CYP154C5. K_d error bars show the 95% parameter confidence intervals from the curve fitting. T_m data points were measured at $c_{steroid} = c_{CYP154C5}$. Horizontal dotted lines represent the T_m values of substrate-free CYP154C5.

CYP154C5 conformation will be required for successful steroid binding in the active site. CYP154C1 (PDB entry 1gwi), which is to date the closest published structural relative to CYP154C5, exhibits an open conformation in the absence of substrate (Podust *et al.*, 2003) and can be used as general model for the substrate-free CYP154C5 structure, especially as the active-site residues of the enzymes are quite conserved. In the 1gwi-based CYP154C5 open-conformation model the active site is exposed to the solvent and thus allows binding of steroid substrate (Fig. 8 and Supplementary Fig. S4). However, as stated above, this exposure is thermodynamically unfavourable especially owing to the strong hydrophobicity of the active site of the enzyme, which in consequence also reduces the thermostability of the enzyme. Upon steroid binding to the active site, the CYP154C5 core is compacted by the displacement of low-entropic water molecules by increasing hydrophobic interactions between the enzyme and substrate and is accompanied by the active-site closure that buries hydrophobic active-site residues from the aqueous solvent (Supplementary Fig. S4). Consequently, the thermal stability of the enzyme–substrate complex is increased, as reflected by the DSF-derived T_m values.

A clear steroid-dependent trend for substrate affinity and thermostabilization is observed, showing that **1** and **6** bind more weakly than the other four steroids in the active site of CYP154C5 (Fig. 7). This weaker binding of **6** can be attributed to the missing C19, which reduces the surface of the molecule and increases its water solubility by at least 33% (∼200 µ*M* for nandrolone *versus* ∼20–150 µ*M* for the other steroids, as calculated by *ALOGPS* 2.1; Yasutake *et al.*, 2010; Yang *et al.*, 2011). Consequently, the hydrophobic effect as a main contribution towards the CYP154C5–steroid interaction is weakened which, in consequence, results in a higher K_d and a reduced thermal stability owing to the less tight hydrophobic protein core packing. Binding of **6** is furthermore weakened by a significantly longer Gln398 hydrogen bond (Table 3). In contrast to **6**, the water solubility of **1** is much lower (∼40 µ*M*), which is also reflected by its lower K_d value. However, the K_d for **1** is still higher than those of the four remaining steroids, and the T_m values of monomeric and dimeric CYP154C5 in complex with **1** are also almost as low as those for **6**. This indicates a rather weak interaction of enzyme and **1** that cannot be attributed to a reduced hydrophobic effect but must be caused by other specific interactions.

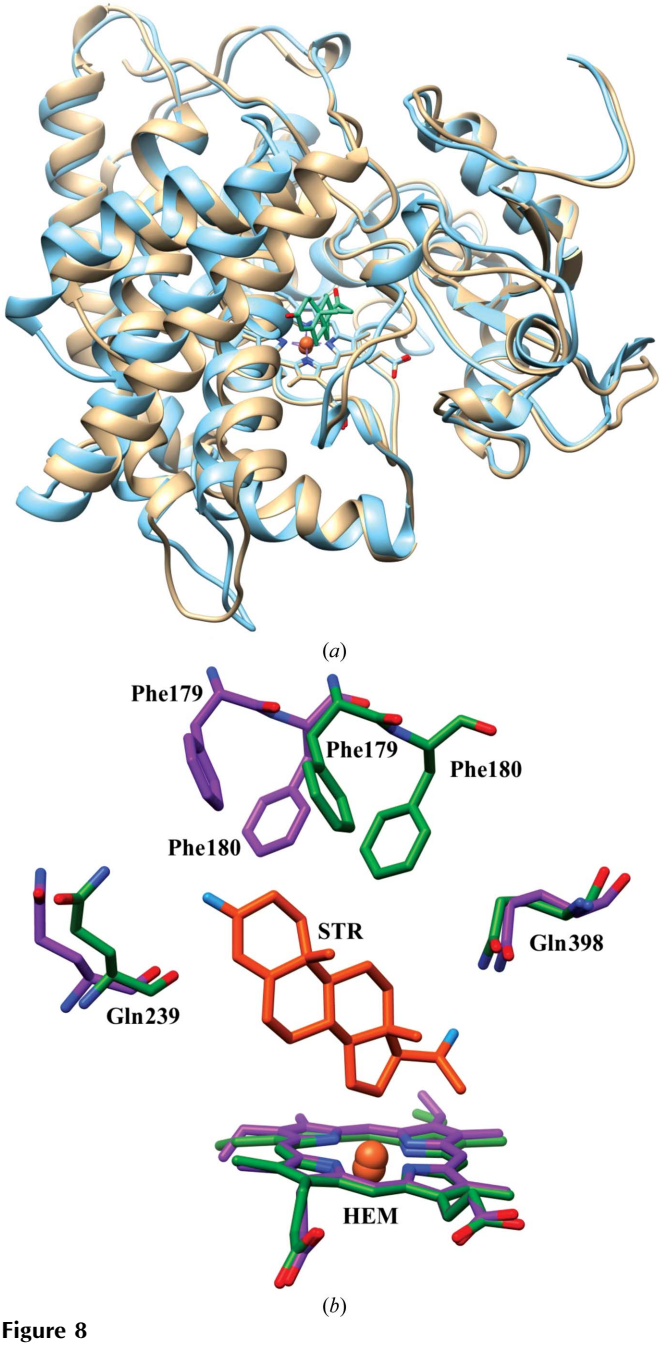


Figure 8
Superposition of the closed, substrate-bound conformation of CYP154C5 as well as its open-conformation model based on the structure of CYP154C1 (PDB entry 1gwi). (*a*) Superposition of the overall structure of CYP154C5 in closed (beige) and open (cyan) conformations. (*b*) Superposition of important active-site residues that move during active-site closure. Residues in the open conformation are shown in magenta and residues in the closed conformation are shown in green.

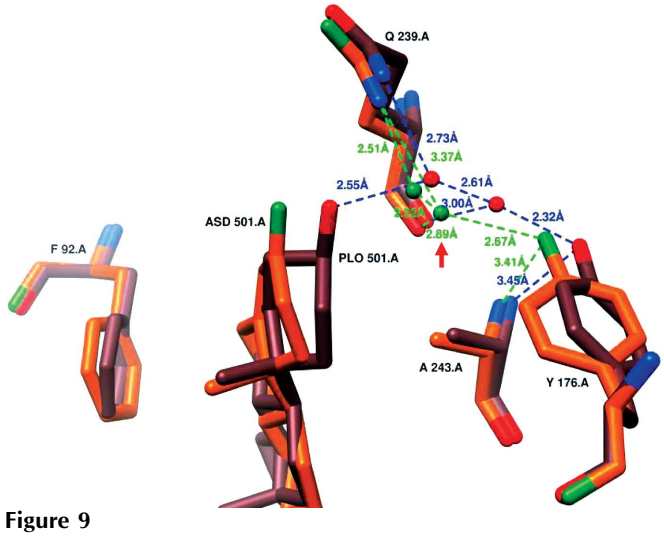


Figure 9
Superposition of the CYP154C5 *A* chains with **1** (PLO; brown) and **4** (ASD; orange) and the hydrogen-bond network between the steroid, Tyr176, Gln239 and Ala243. The red arrow marks the conserved water molecule.

Fig. 9 shows the superposition of pairwise interactions for steroids **1** and **4** with the residues Phe92 and the more distant Tyr176 which seem to be responsible for the weaker binding of **1**. Phe92 mediates important hydrophobic interactions between CYP154C5 and the steroid core and can be therefore assumed to contribute significantly to the steroid-binding process and to the forces that stabilize the CYP154C5–steroid complex. In the case of **1** the Phe92 phenyl ring is rotated clockwise probably by π-electron repulsion and steric interactions. This misalignment can be assumed to reduce the strength of local hydrophobic interactions and thus contributes to the higher K_d value for **1** compared with the structurally similar compound **3**. Furthermore, Tyr176 belongs to helix F and mediates a hydrogen-bonding network with the large active-site helix I *via* Gln239 and Ala243. As shown in Fig. 9, the displacement of Tyr176 in the CYP154C5–pregnenolone complex changes the local hydrogen-bonding network, especially by displacing the water molecule marked with a red arrow. Hence, an additional water molecule is required to contact the Gln239 side chain, thus weakening the interactions of Gln239 and Tyr176. The importance of the Tyr176 interactions becomes obvious when the substrate-induced transition of CYP154C5 from an opened to a closed conformation is considered. As mentioned above, the active-site closure is mediated *via* the movement of helices F and G, which can be assumed to act according to a zipping mechanism. These interactions with the bound steroid substrate pull helices F and G down to close the active site. Thereby, Tyr176 plays an important role as it connects helices F and I. Consequently, it must be assumed that this connection is weakened upon binding of **1** as the equilibrium of the open–closed reaction is slightly shifted towards the open conformation. This hypothesis is also supported by the comparably low T_m value for the CYP154C5–pregnenolone complex, indicating a less compact protein core packing and increased flexibility of helices F and G.

3.4. Steroid hydroxylation kinetics

To gain further insight into the catalytic efficiency of CYP154C5, turnover numbers (TON) and coupling efficiencies in the conversion of steroid substrates **1**–**6** were determined. Measurements were carried out using purified P450 together with purified redox partners Pdx and PdR from *P. putida*, since the natural redox-partner system of CYP154C5 is not known. Depending on the steroid substrate, the TON range between 1.3 and 5.7 min^{-1} (Table 3, Supplementary Fig. S5). Except for **6**, these values are significantly higher than the TON obtained earlier using a whole-cell biocatalyst containing CYP154C5, Pdx and PdR (Bracco *et al.*, 2013), although the volumetric activity of the electron-transfer components Pdx and PdR in the purified enzyme system was twofold lower compared with the whole-cell system. Hence, steroid uptake into the bacterial cells indeed seems to be limiting the overall efficiency of the whole-cell biocatalyst. On the other hand, formate dehydrogenase from *C. boidinii* together with sodium formate as a cosubstrate were used for NADH cofactor regeneration during TON determination with the purified three-component system. This setup might be more efficient than NADH supply by the metabolism of the cell in the whole-cell biocatalyst upon glucose addition.

Furthermore, the TON determined for **5** and **6** are, in agreement with earlier results (Bracco *et al.*, 2013), a factor of two to four lower than for the other steroids (Table 3). With the crystal structures in hand, these low turnover numbers might now be rationalized by an additional hydrogen-bonding network between the steroid O17, His293 and Gln398 (Fig. 6) which was only observed in the structures with **5** and **6**. Here, the addition of another hydroxyl group at C16 might result in an extra hydrogen bond, which will in turn stabilize this product in the active site and will consequently hamper product release. Additionally, the about 10% longer Fe–C16 distance might also contribute to a reduced conversion rate for **5** (Table 3). For substrate **6**, the lower TON might also be influenced by the weaker substrate binding, as indicated by a rather high K_d value.

In the case of coupling efficiencies, the values obtained for the three-component system with steroids **1**–**6** are all below 100%, revealing a significant loss of electrons from the cofactor NADH. However, taking into account that surrogate redox partners were used for CYP154C5, the obtained coupling efficiencies are still rather high for substrates **2**–**4**, with values above 80%. The coupling efficiency was drastically reduced only for the conversion of **6**. As coupling efficiencies are affected by the efficiency of electron transfer from the cofactor NADH to the P450 haem *via* the redox partners PdR and Pdx, optimal interaction of the different protein components is essential for high coupling efficiencies. For CYP154C5, the interaction with Pdx is of special interest as the interaction of the natural electron-transfer partners Pdx and PdR is assumed to be optimal: P450cam together with Pdx and PdR exhibits a coupling efficiency of 100% in the

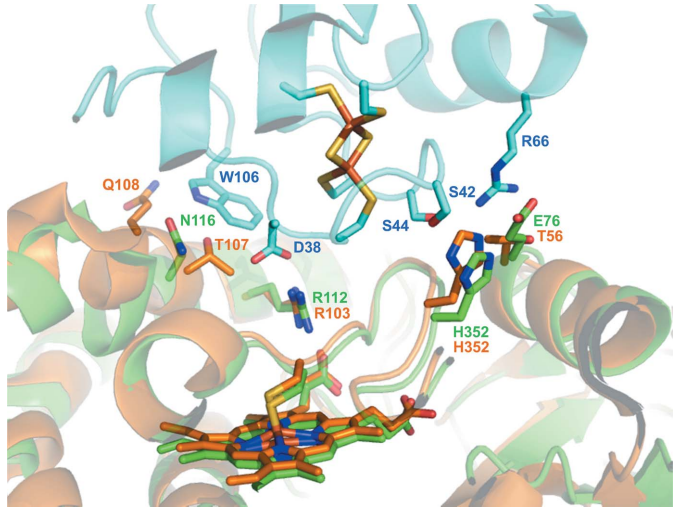


Figure 10
Superposition of CYP154C5 (orange) with P450cam (green) in complex with Pdx (blue). Residues involved in the interaction of P450cam and Pdx (PDB entry 4jx1) as well as their equivalents in CYP154C5 (4j6b) are highlighted.

conversion of its natural substrate D-camphor (Hayashi *et al.*, 2009). Recently, X-ray crystallographic structures for P450cam–putidaredoxin (Pdx) complexes were reported by Tripathi *et al.* (2013) and can serve as a valuable model in analyzing the interactions between CYP154C5 and Pdx. Fig. 10 provides comparative insights into the interactions of P450cam and Pdx (PDB entry 4jx1) as well as the interactions of CYP154C5 and Pdx based on a structural alignment of the CYP154C5–progesterone *A* chain and 4jx1. In general, the ferredoxin-contacting surface of both CYPs is mainly positively charged. Despite the relatively large interaction area, some key residues can be identified from the interaction map of P450cam and Pdx. The most important CYP residue is Arg112 (Arg103 in CYP154C5), which forms an 'electron bridge' with Asp38 of Pdx during electron transfer (Koga *et al.*, 1993). The conformation of this residue is highly conserved in P450cam and CYP154C5, explaining why Pdx is also able to deliver electrons to the haem iron of CYP154C5. Other specific side-chain interactions reported to hold the two redox partners together by hydrogen bonding are not completely conserved in CYP154C5. As an example, the interaction with Pdx Trp106 is mediated by Asn116 in P450cam, while this is probably mediated through Thr107 or Gln108 in CYP154C5. Furthermore, Glu76 of P450cam is matched by Thr56 in CYP154C5, which probably breaks the hydrogen bond to Pdx Arg66 as a consequence of the change in residue size and charge, and thus reduces the interaction between the proteins. In consequence, the interacting surfaces of CYP154C5 and Pdx are less optimal compared with the P450cam–Pdx interaction, explaining the generally reduced (below 100%) coupling efficiencies.

Structural alignments of the four CYP154C5 structures revealed only negligible differences in the CYP154C5–Pdx interacting regions (data not shown). Hence, the steroid-dependent differences in coupling efficiencies cannot be attributed to substrate-induced changes of the Pdx-binding interface in the CYP154C5 structures. Instead, the obtained coupling efficiencies follow the trend of the determined K_d values of CYP154C5 for steroids **1**–**6**. The only exception here is steroid **5**, for which a reduced coupling efficiency was observed although the measured K_d was similar to the values obtained for steroids **2**–**4**, resulting in significantly higher coupling efficiencies. Again, the significantly longer Fe–C16 distance observed in the CYP154C5–testosterone structure might provide a plausible explanation here. This haem–testosterone distance might be too long for highly efficient radical attack at the testosterone C16 by the reactive iron-oxo species (compound I) of the catalytic cycle (Denisov *et al.*, 2005). This might slow down the rate of C16 hydroxylation and thereby increase the time for and probability of uncoupling by the collapse of compound I. However, further studies, such as determination of the electron-transfer rates between Pdx and CYP154C5 as a function of the employed steroid substrate, would be necessary to better understand the substrate-dependent changes in coupling efficiencies.

In summary, we were able to shed light on the structural determinants of the high regioselectivity and stereoselectivity of CYP154C5 in steroid hydroxylation by solving the crystal structure of this P450 in complex with four different steroid molecules. A very recently published paper on CYP154C3 (Makino *et al.*, 2014), an enzyme that also hydroxylates various steroids exclusively at the 16α position, supports our structural findings as most of the crucial active-site residues in CYP154C5 are conserved in CYP154C3 (data not shown). Despite the fact that crystal structures of the homologues CYP154C1 and CYP154A1 have been published previously (Podust *et al.*, 2003, 2004), this is the first report of the structure of a P450 from family CYP154 with substrate bound in the active site and thus could serve as a model for studying the closed, substrate-bound conformations of other CYP154 family members. Although the identified structural differences between the four CYP154C5–steroid complexes are rather subtle, they help us to rationalize the steroid-dependent differences in the biochemical as well as the biocatalytic data obtained for this P450. Overall, this will eventually allow us to fine-tune the activity and selectivity of the enzyme by protein engineering in the future in order to, for example, enhance its applicability in steroid hormone synthesis.

This project was financially supported by the German Research Foundation (DFG) within the national Excellence Initiative funding scheme to promote science and research at German universities. Additionally, the work was supported by the European Regional Development Fund (ERDF) and the European Union ('Die Europäische Kommission investiert in Ihre Zukunft'). Furthermore, this work was also partly supported by the JSPS Japanese–German Graduate Externship and DFG through the International Research Training Group 'Selectivity in Chemo- and Biocatalysis (SeleCa)'. PB expresses her special thanks to the SeleCa program for financial support of her stay at Osaka University. Furthermore, we thank Dr Heinrich Delbrück (Institute of Molecular Biotechnology, RWTH Aachen University) for helpful discussion during protein crystallization and structure determination. Molecular graphics were produced and analyses were performed with the *UCSF Chimera* package. *Chimera* was developed by the Resource for Biocomputing, Visualization and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

References

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Bernhardt, R. (2006). *J. Biotechnol.* **124**, 128–145.
Bracco, P., Janssen, D. B. & Schallmey, A. (2013). *Microb. Cell Fact.* **12**, 95.
Bureik, M. & Bernhardt, R. (2007). *Modern Biooxidation*, edited by R. D. Schmid & V. B. Urlacher, pp. 155–176. Weinheim: Wiley–VCH Verlag.
Cheng, Q., Lamb, D. C., Kelly, S. L., Lei, L. & Guengerich, F. P. (2010). *J. Am. Chem. Soc.* **132**, 15173–15175.
Denisov, I. G., Makris, T. M., Sligar, S. G. & Schlichting, I. (2005). *Chem. Rev.* **105**, 2253–2277.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Epstein, W. W. & Sweat, F. W. (1967). *Chem. Rev.* **67**, 247–260.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

García-Fernández, E., Frank, D. J., Galán, B., Kells, P. M., Podust, L. M., García, J. L. & Ortiz de Montellano, P. R. (2013). *Environ. Microbiol.* **15**, 2342–2359.
Gunsalus, I. C. & Wagner, G. C. (1978). *Methods Enzymol.* **52**, 166–188.
Hannemann, F., Bichet, A., Ewen, K. M. & Bernhardt, R. (2007). *Biochim. Biophys. Acta*, **1770**, 330–344.
Harding, M. M. (2001). *Acta Cryst.* D**57**, 401–411.
Hayashi, T., Harada, K., Sakurai, K., Shimada, H. & Hirota, S. (2009). *J. Am. Chem. Soc.* **131**, 1398–1400.
Isin, E. M. & Guengerich, F. P. (2007). *Biochim. Biophys. Acta*, **1770**, 314–329.
Kang, D., Gho, Y., Suh, M. & Kang, C. (2002). *Bull. Korean Chem. Soc.* **23**, 1511–1512.
Koga, H., Sagara, Y., Yaoi, T., Tsujimura, M., Nakamura, K., Sekimizu, K., Makino, R., Shimada, H., Ishimura, Y., Yura, K., Go, M., Ikeguchi, M. & Horiuchi, T. (1993). *FEBS Lett.* **331**, 109–113.
Lacour, T., Achstetter, T. & Dumas, B. (1998). *J. Biol. Chem.* **273**, 23984–23992.
Lane, L. C. (1978). *Anal. Biochem.* **86**, 655–664.
Lee, Y.-T., Wilson, R. F., Rupniewski, I. & Goodin, D. B. (2010). *Biochemistry*, **49**, 3412–3419.
Makino, T., Katsuyama, Y., Otomatsu, T., Misawa, N. & Ohnishi, Y. (2014). *Appl. Environ. Microbiol.* **80**, 1371–1379.
Montemiglio, L. C., Gianni, S., Vallone, B. & Savino, C. (2010). *Biochemistry*, **49**, 9199–9206.
Morrison, J. (1969). *Biochim. Biophys. Acta*, **185**, 269–286.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Omura, T. & Sato, R. (1964). *J. Biol. Chem.* **239**, 2370–2378.
Ortiz de Montellano, P. (2005). *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd ed. New York: Kluwer/Plenum.
Ouellet, H., Guan, S., Johnston, J. B., Chow, E. D., Kells, P. M., Burlingame, A. L., Cox, J. S., Podust, L. M. & Ortiz de Montellano, P. R. (2010). *Mol. Microbiol.* **77**, 730–742.
Painter, J. & Merritt, E. A. (2006*a*). *Acta Cryst.* D**62**, 439–450.
Painter, J. & Merritt, E. A. (2006*b*). *J. Appl. Cryst.* **39**, 109–111.
Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P. & Salemme, F. R. (2001). *J. Biomol. Screen.* **6**, 429–440.
Peterson, J. A., Lorence, M. C. & Amarneh, B. (1990). *J. Biol. Chem.* **265**, 6066–6073.
Podust, L. M., Bach, H., Kim, Y., Lamb, D. C., Arase, M., Sherman, D. H., Kelly, S. L. & Waterman, M. R. (2004). *Protein Sci.* **13**, 255–268.
Podust, L. M., Kim, Y., Arase, M., Neely, B. A., Beck, B. J., Bach, H., Sherman, D. H., Lamb, D. C., Kelly, S. L. & Waterman, M. R. (2003). *J. Biol. Chem.* **278**, 12214–12221.
Savino, C., Montemiglio, L. C., Sciara, G., Miele, A. E., Kendrew, S. G., Jemth, P., Gianni, S. & Vallone, B. (2009). *J. Biol. Chem.* **284**, 29170–29179.
Schallmey, A., den Besten, G., Teune, I. G., Kembaren, R. F. & Janssen, D. B. (2011). *Appl. Microbiol. Biotechnol.* **89**, 1475–1485.
Schenkman, J. B., Remmer, H. & Estabrook, R. W. (1967). *Mol. Pharmacol.* **3**, 113–123.
Senisterra, G. A. & Finerty, P. J. Jr (2009). *Mol. Biosyst.* **5**, 217–223.
Sherman, D. H., Li, S., Yermalitskaya, L. V., Kim, Y., Smith, J. A., Waterman, M. R. & Podust, L. M. (2006). *J. Biol. Chem.* **281**, 26289–26297.
Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.
Stoll, S., Lee, Y.-T., Zhang, M., Wilson, R. F., Britt, R. D. & Goodin, D. B. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 12888–12893.
Tam, J. P., Wu, C. R., Liu, W. & Zhang, J. W. (1991). *J. Am. Chem. Soc.* **113**, 6657–6662.
Tripathi, S., Li, H. & Poulos, T. L. (2013). *Science*, **340**, 1227–1230.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wallace, T. J. (1964). *J. Am. Chem. Soc.* **86**, 2018–2021.
Wallace, T. J. & Mahon, J. J. (1964). *J. Am. Chem. Soc.* **86**, 4099–4103.
Wallace, T. J. & Mahon, J. J. (1965). *J. Org. Chem.* **30**, 1502–1506.
Williams, J. W. & Morrison, J. F. (1979). *Methods Enzymol.* **63**, 437–467.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yang, W., Bell, S. G., Wang, H., Zhou, W., Bartlam, M., Wong, L.-L. & Rao, Z. (2011). *Biochem. J.* **433**, 85–93.
Yasutake, Y., Fujii, Y., Nishioka, T., Cheon, W.-K., Arisawa, A. & Tamura, T. (2010). *J. Biol. Chem.* **285**, 31193–31201.