

University of Groningen

Regio- and stereoselective steroid hydroxylation by CYP109A2 from *Bacillus megaterium* explored by X-ray crystallography and computational modeling

Jóźwik, Ilona K; Bombino, Elvira; Abdulmughni, Ammar; Hartz, Philip; Rozeboom, Henriette J; Wijma, Hein J; Kappl, Reinhard; Janssen, Dick B; Bernhardt, Rita; Thunnissen, Andy-Mark W H

Published in:
The FEBS Journal

DOI:
[10.1111/febs.16906](https://doi.org/10.1111/febs.16906)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Jóźwik, I. K., Bombino, E., Abdulmughni, A., Hartz, P., Rozeboom, H. J., Wijma, H. J., Kappl, R., Janssen, D. B., Bernhardt, R., & Thunnissen, A-M. W. H. (2023). Regio- and stereoselective steroid hydroxylation by CYP109A2 from *Bacillus megaterium* explored by X-ray crystallography and computational modeling. *The FEBS Journal*, 290(20), 5016-5035. <https://doi.org/10.1111/febs.16906>

Copyright





Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Regio- and stereoselective steroid hydroxylation by CYP109A2 from *Bacillus megaterium* explored by X-ray crystallography and computational modeling

Ilona K. Jóźwik¹ , Elvira Bombino¹, Ammar Abdulmughni², Philip Hartz² , Henriette J. Rozeboom¹, Hein J. Wijma¹, Reinhard Kappl³, Dick B. Janssen¹ , Rita Bernhardt² and Andy-Mark W. H. Thunnissen¹ 

¹ Biotransformation and Biocatalysis, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, The Netherlands

² Department of Biochemistry, Saarland University, Saarbrücken, Germany

³ Department of Biophysics, CIPMM, School of Medicine, Saarland University, Saarbrücken, Germany

Keywords

Cytochrome P450; hydroxylation; near-attack conformations; steroids; X-ray crystallography

Correspondence

A.-M. W. H. Thunnissen, Biotransformation and Biocatalysis, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Tel: +31 50 363 4380

E-mail: a.m.w.h.thunnissen@rug.nl

(Received 12 December 2022, revised 31 May 2023, accepted 16 June 2023)

doi:10.1111/febs.16906

The P450 monooxygenase CYP109A2 from *Bacillus megaterium* DSM319 was previously found to convert vitamin D3 (VD3) to 25-hydroxyvitamin D3. Here, we show that this enzyme is also able to convert testosterone in a highly regio- and stereoselective manner to 16 β -hydroxytestosterone. To reveal the structural determinants governing the regio- and stereoselective steroid hydroxylation reactions catalyzed by CYP109A2, two crystal structures of CYP109A2 were solved in similar closed conformations, one revealing a bound testosterone in the active site pocket, albeit at a nonproductive site away from the heme-iron. To examine whether the closed crystal structures nevertheless correspond to a reactive conformation of CYP109A2, docking and molecular dynamics (MD) simulations were performed with testosterone and vitamin D3 (VD3) present in the active site. These MD simulations were analyzed for catalytically productive conformations, the relative occurrences of which were in agreement with the experimentally determined stereoselectivities if the predicted stability of each carbon-hydrogen bond was taken into account. Overall, the first-time determination and analysis of the catalytically relevant 3D conformation of CYP109A2 will allow for future small molecule ligand screening *in silico*, as well as enabling site-directed mutagenesis toward improved enzymatic properties of this enzyme.

Introduction

Cytochrome P450 monooxygenases (P450s or CYPs) are versatile heme-containing enzymes, capable of oxidizing C-H bonds in a wide variety of substrates. They occur in most organisms, where they play essential roles in the biosynthesis of natural endogenous compounds, like steroids, vitamins and lipids, as well as in the metabolism and biodegradation of xenobiotics

[1,2]. Their ability to catalyze the regio- and stereospecific oxidation of nonfunctionalized hydrocarbons has led to a considerable interest in utilizing P450s as biocatalysts for the production of fine chemicals, including pharmaceuticals [3–5]. In particular, P450s' ability to hydroxylate steroids in a highly regio- and stereoselective manner makes them attractive as recombinant

Abbreviations

16 β (OH)TES, 16 β (OH)-hydroxytestosterone; 25(OH)VD3, 25-hydroxyvitamin D3; MD, molecular dynamics; NAC, near-attack conformation; P450, CYP, cytochrome P450 monooxygenase; RMSD, root-mean-square deviation; RP-HPLC, reversed-phase high-performance liquid chromatography; SRS, substrate recognition site; TES, testosterone; VD3, vitamin D3.

biocatalysts, allowing the production of high-value steroidal drugs and their precursors for the pharmaceutical industry [6,7]. Microbial P450s displaying steroid-converting activities are particularly attractive for biocatalytic applications, as they can be overexpressed in higher amounts in soluble form, and are usually more active and stable than their counterparts from higher eukaryotic organisms [8]. In addition, microbial P450s can be efficiently integrated in whole-cell biocatalyst systems, which are the preferred biotechnological tool for large-scale production of fine and bulk chemicals [9,10]. For these reasons, various microbial P450s, in particular from bacteria, capable of hydroxylating steroids with different regio- and stereoselectivity have been documented [11–13]. Biocatalytic and structural characterization has allowed for a better understanding of the sequence/structure relationships of these enzymes, while the use of various protein engineering strategies has been successful in altering or improving enzymatic properties like substrate specificity, regio/stereoselectivity, catalytic activity, and thermostability [3,14–16]. Nevertheless, full characterization of new and already known bacterial steroid-converting P450s remains of high importance to increase the toolbox for synthesis of functionalized steroidal drugs and to provide a better basis for the rational redesign of these enzymes toward industrial applications [17–19].

The bacterial strain *Bacillus megaterium* DSM319 [20] contains three different P450s with wild-type steroid hydroxylation activities, that is, CYP106A1, CYP109E1, and CYP109A2. The latter two enzymes, which share a sequence identity of 45%, have both been subject to catalytic and structural analysis [21–26]. CYP109E1 was found to convert testosterone (TES) to 16 β -hydroxytestosterone (16 β -OH-TES) with high regio- and stereoselectivity, but is also able to hydroxylate cholesterol and vitamin D3 (VD3) at carbons C24 and C25, albeit with low regioselectivity [22,24,26]. In addition, the enzyme displays highly regioselective hydroxylase and epoxidase activities with a selection of statin and terpene substrates [25]. Crystal structures of CYP109E1 revealed a dynamic active site that is wide open in the substrate-free state, but closes when bound to its substrate TES [24]. The TES-bound crystal structure of CYP109E1 displays the steroid in a nonproductive orientation with the C3-carbonyl group facing the heme-iron. Using docking and molecular dynamics (MD) simulations, we previously demonstrated that the steroid can also bind in a reversed orientation, positioning its C16-H β hydrogen in close proximity to the heme-iron and in line with the observed 16 β -hydroxylation activity.

Less functional and structural data are available for CYP109A2. Like CYP109E1, CYP109A2 is able to hydroxylate VD3, but with a higher regioselectivity, resulting in the formation of C25-hydroxylated VD3 (25-OH-VD3) as the main product [21,23]. However, it is unknown whether CYP109A2 also displays hydroxylating activities with other steroids. A crystal structure is available for CYP109A2 [21], but only in a substrate-free open conformation, which limits its suitability for evaluating the role of active site residues in steroid binding.

Here, we report a comprehensive structure–function analysis of CYP109A2 focusing on its catalytic steroid hydroxylation activity. Using an *in vitro* activity assay, we show that the enzyme is able to convert TES to 16 β -OH-TES, confirming its close relationship to CYP109E1. In addition, we determined two crystal structures of CYP109A2 in similar closed conformations. To establish whether these crystal structures display catalytically relevant conformations, we performed docking of TES and VD3, followed by MD simulations and quantification of near-attack conformations (NACs) of the steroids in the active site. The results of this combined experimental and computational analysis allowed a clear assessment of the residues and active site regions in CYP109A2 associated with selective and productive steroid binding.

Results

CYP109A2-dependent conversion of TES to 16 β -OH-TES

To explore whether CYP109A2 has a broader substrate scope beyond VD3, we tested various steroids for conversion by CYP109A2, using an *in vitro* assay containing the purified P450 enzyme, supplemented with adrenodoxin (Adx) and adrenodoxin reductase (AdR) as surrogate redox partners (Fig. 1). With this assay, a CYP109A2-dependent conversion of the androgen steroids TES, 17 α -methyl-TES and androstenedione was observed, whereas the estrogen steroids estrone and estradiol were not converted. CYP109A2 was able to convert each androgen with high selectivity into one major product. By using authentic standards and comparing the HPLC reaction profile with that of CYP109E1, it was established that CYP109A2 converts TES to 16 β -OH-TES (Fig. 2). The conversion products for 17 α -methyl-TES and androstenedione could not be unambiguously identified due to lack of appropriate authentic standards and insufficient amounts of produced material for NMR analysis. Unlike CYP109E1, CYP109A2 did not yield typical

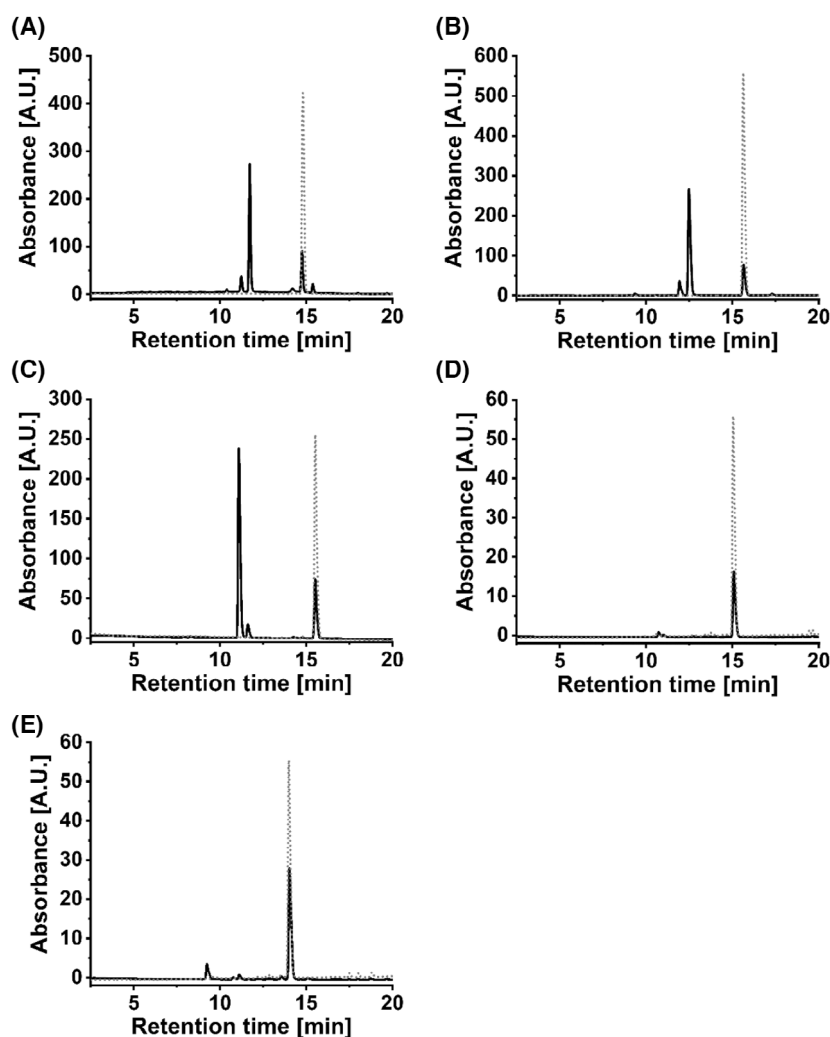


Fig. 1. Substrate screening of CYP109A2. Reversed-phase HPLC chromatograms for five steroid substrates after their addition to the *in vitro* CYP109A2/AdR/Adx assay system (black lines). (A) testosterone, (B) 17(α)-methyltestosterone, (C) androstenedione, (D) estrone and (E) estradiol. The gray dotted lines represent the injected pure standards of the corresponding substrates.

type I binding spectra upon binding of TES or VD3; thus K_d values could not be determined. Nevertheless, the common bioconversion activity toward TES and VD3 emphasizes the close functional relationships of these enzymes.

Crystal structures of ligand-bound CYP109A2 in a closed conformation

To establish the structural basis of its regio- and stereoselective hydroxylation activity with TES, we co-crystallized CYP109A2 in the presence of the steroid and determined two crystal structures of CYP109A2 in a closed conformation (Fig. 3, see Table 1 for relevant crystallographic statistics). Previous co-crystallization

experiments of CYP109A2 carried out with VD3 did not result in a steroid-bound structure and captured the enzyme in an open conformation [21]. Now, for the first time, we could confirm that CYP109A2 is able to close its active site, mainly via a movement of its F-G helices, like observed in other P450s including CYP109E1. Surprisingly, neither structure contains TES in a productive binding mode.

One of the crystal structures, determined at 2.10 Å resolution, revealed two bound ligands in the active site, which we believe to be fatty acids, based on the shape of their associated electron densities and the inferred binding interactions (Fig. 3A,C, Fig. S1, PDB entry 8ABR). One fatty acid molecule is bound close to the heme and was modeled as hexanoic acid,

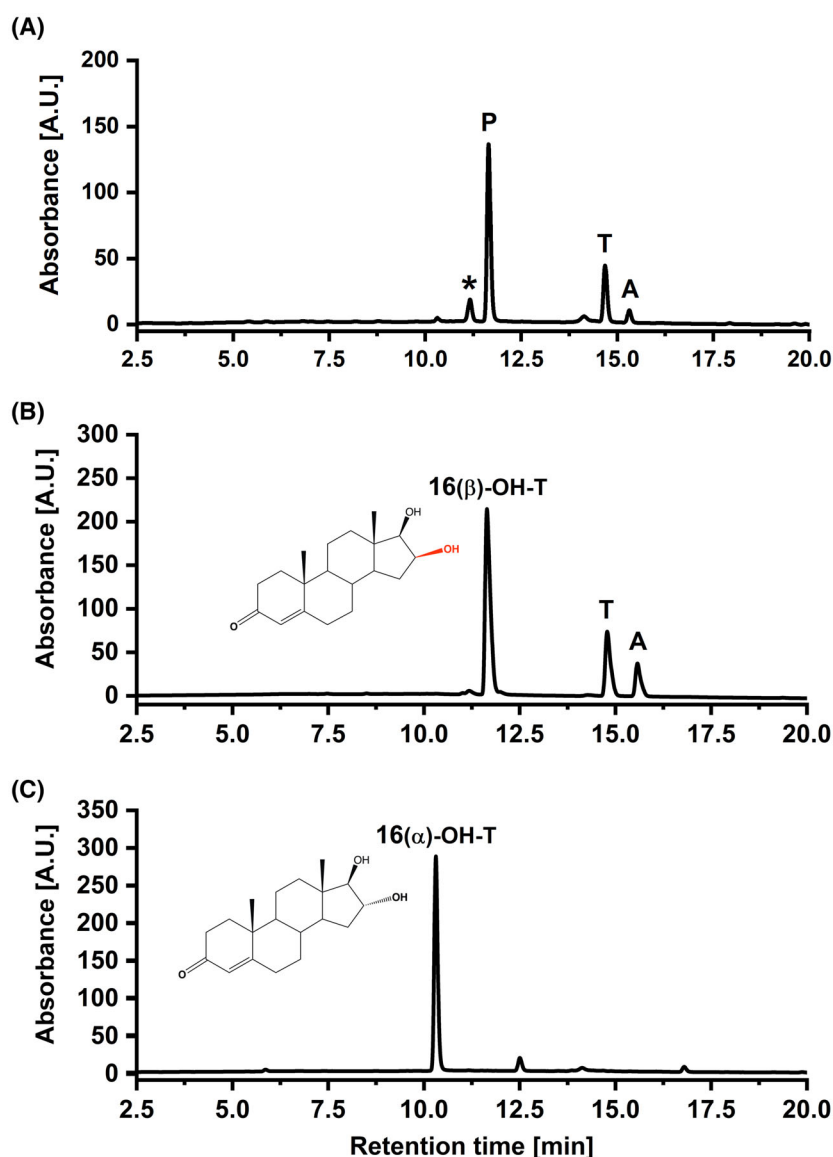
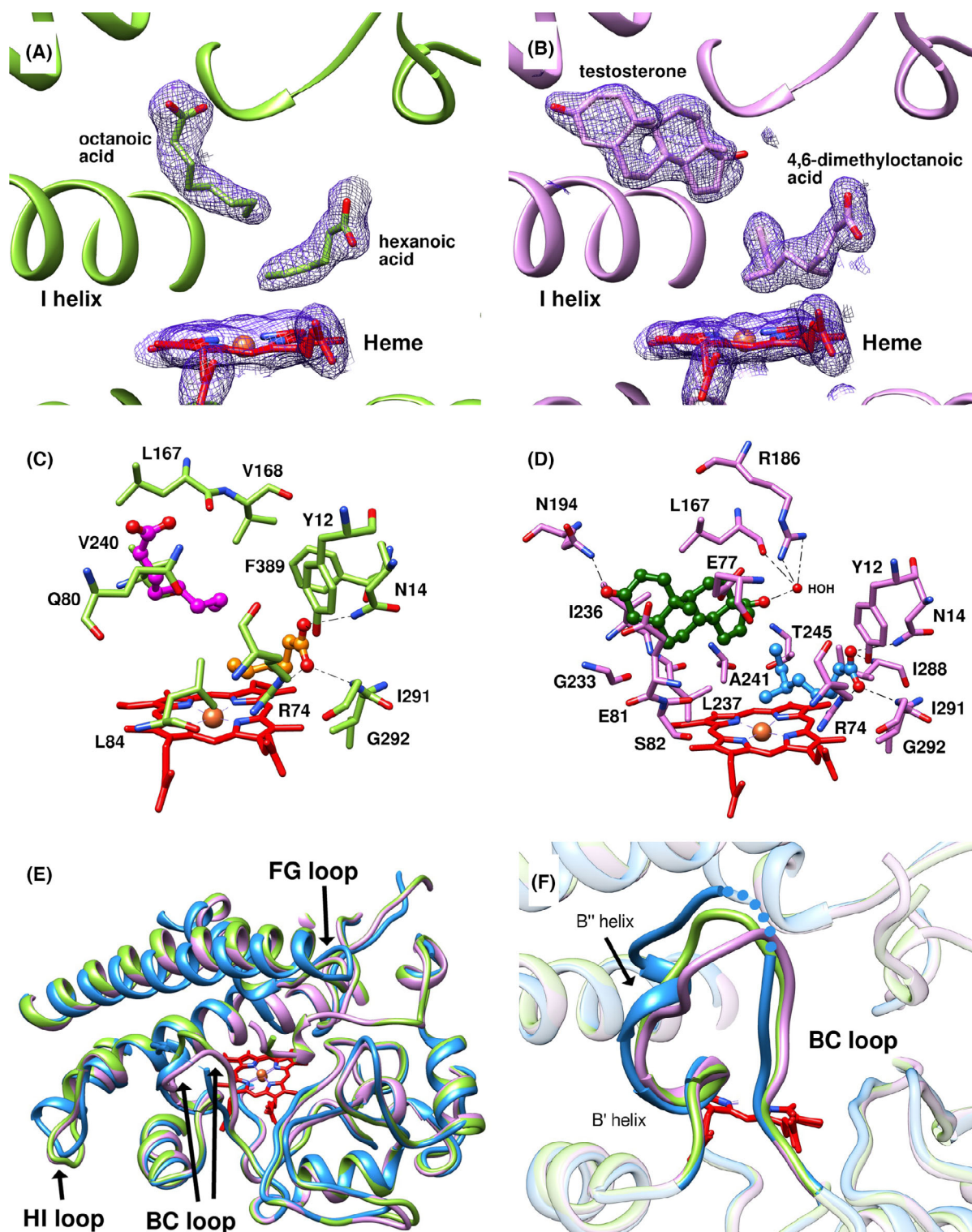


Fig. 2. CYP109A2 converts TES to 16(β)-OH-TES. (A) A representative reversed-phase HPLC chromatogram of the CYP109A2-catalyzed conversion of TES, showing the presence of residual TES (T), a main product (P), identified as 16(β)-OH-TES, one side product (A), identified as androstenedione, and one unknown residual side product (*). (B) Reversed-phase HPLC chromatogram of CYP109E1-catalyzed conversion of TES to 16(β)-OH-TES serving as authentic standard. (C) Reversed-phase HPLC chromatogram of 16(α)-OH-TES. The compound 16(α)-OH-TES was obtained from prof. Annet Schallmeyer (Technische Universität Braunschweig) and was prepared as described [27].

while the other, more distantly bound fatty acid was modeled as octanoic acid. It should be noted, however, that the exact nature of the fatty acids, regarding length and shape of their acyl chains, could not be unambiguously deduced from the electron density maps. The fatty acid modeled as hexanoic acid forms a salt bridge between its carboxylate group and the side chain of Arg74 [B-C loop or substrate recognition site 1 (SRS1, identification and nomenclature of conserved P450 SRSs follows Gotoh [28]). In addition,

the carboxylate group forms hydrogen bonds with the backbone amide of Gly292 (SRS5) and the side chains of Tyr12 and Asn14. The acyl chain closely packs against one of the propionate moieties of the heme, while making additional hydrophobic contacts with the side chains of Ile291 (SRS5) and Phe389 (SRS6). The terminal carbon atom of the fatty acid tail approaches the heme-iron at a distance of 4.6 Å. The carboxylate of the fatty acid modeled as octanoic acid points toward the guanidinium group of Arg186 (G-



helix/SRS3), while the acyl chain makes hydrophobic contacts with residues from the I-helix (Val240/SRS4), the F-G loop (Leu167 and Val168/SRS2) and the B-C loop (Gln80 and Leu84/SRS1). Since no fatty acids or

related compounds were added to the buffers used for protein purification and crystallization, the bound ligands must originate from the *Escherichia coli* cells used as expression host. Furthermore, their affinities

Fig. 3. Crystal structures of ligand-bound CYP109A2 in a closed conformation. (A) Active site view of the 2.10 Å CYP109A2 crystal structure, showing the proposed bound ligands and their associated electron densities (unbiased composite omit 2Fo-Fc electron density map, contoured at 1σ). (B) Similar active site view of the 1.75 Å CYP109A2 crystal structure. (C) and (D) Protein-ligand interactions are presented for structures corresponding to panels (A) and (B), respectively. Hydrogen bonds are indicated by black dashed lines. More details are provided in Fig. S1. (E) Overlay of the 2.10 Å (green) and 1.75 Å (pink) CYP109A2 structures with the closed TES-bound CYP109E1 structure (blue, PDB entry 5L94). Substrates are not shown. Substantial conformational differences are confined mainly to the BC- and HI-loop regions. (F) Zoomed-in view on the BC-loop region presented in panel (E). Same coloring scheme is applied. The blue dots indicate the unresolved part of the BC-loop in the CYP109E1 crystal structure (residues 72–74).

Table 1. Crystallographic data collection and refinement statistics.

	CYP109A2 (fatty acid-bound)	CYP109A2 (TES/fatty acid-bound)
PDB entry	8ABR	8ABS
Data collection		
Beamline	I03 (DLS)	ID30A-1 (ESRF)
Wavelength (Å)	0.9763	0.9655
Resolution range (Å)	94–2.10 (2.16–2.10)	49–1.75 (1.78–1.75)
Space group	P 4 ₃ 2 ₁ 1	P 4 ₃ 2 ₁ 1
Unit-cell parameters, a, b, c (Å)	78.67, 78.67, 188.02	78.03, 78.03, 187.86
Observed reflections	922 831 (77 808)	335 918 (18 620)
Unique reflections	35 461 (2830)	59 408 (3215)
Multiplicity	26.0 (27.5)	5.7 (5.8)
Completeness (%)	100.0 (100.0)	99.9 (100.0)
<I/σ(I)>	27.6 (1.5)	13.9 (1.0)
R _{merge}	0.065 (2.935)	0.053 (1.797)
R _{p.i.m.}	0.013 (0.567)	0.025 (0.813)
CC _{1/2} (%)	1.000 (0.695)	0.999 (0.491)
Refinement		
R _{work} , R _{free} (%)	19.2, 22.0	18.7, 20.5
RMSD bond lengths (Å), bond angles (°)	0.009, 0.9	0.007, 0.9
Average B-factors (Å) ²		
Protein, ligands, solvent	65.7, 69.9, 60.0	45.5, 50.6, 46.0
Ramachandran (%)		
Most favored, allowed, outliers	96.9, 3.1, 0.0	97.2, 2.8, 0.0
Rotamer outliers (%)	1.15	0.29
Clashscore	3.87	2.62
Molprobrity overall score	1.15	1.18

Values in parentheses refer to the highest resolution shell.

must be strong enough to resist displacement by TES, which was present at 100 μM concentration in all buffers used during protein purification.

Treatment of the purified protein with Lipidex, a fatty acid binding resin commonly used to remove fatty acids and lipids from proteins [29], followed by the addition of TES prior to crystallization, resulted in a second crystal structure determined at 1.75 Å resolution (PDB entry 8ABS). In this second structure, the

distantly positioned fatty acid is replaced by a TES molecule (Fig. 3B,D). The ligand close to the heme was not displaced, and, based on the more detailed 1.75 Å electron density map, was modeled as 4,6-dimethyloctanoic acid. The two branching methyl groups in the acyl chain of 4,6-dimethyloctanoic acid fit snugly into two small protruding electron densities and form favorable apolar contacts with the protein. We are not sure, however, of the identity of the ligand and we cannot exclude that its electron density is an averaged density resulting from binding a mixture of different fatty acids, or from binding a fatty acid in different conformations. Disordered binding may also account for the different ligand electron density features observed in the 2.1 Å crystal structure, although the same ligand should be bound. It appeared unfeasible to experimentally identify the presumed bound fatty acids using mass spectrometry, nor did we investigate whether fatty acids are converted by CYP109A2. The bound TES displays well-defined electron density and is mainly stabilized via hydrophobic contacts of its carbon ring skeleton with a similar set of residues as used for the binding of the distant fatty acid in the 2.10 Å crystal structure (Fig. 3B,D). In addition, TES forms a hydrogen bond between its C3-keto oxygen and the side chain of Gln194 (G-helix/SRS3), while water-mediated hydrogen bonds are observed between the C17-hydroxyl group of TES, the side chain of Arg186 (G-helix/SRS3) and the main chain carbonyl of Leu167 (F-helix/SRS2). The two CYP109A2 crystal structures display highly similar overall structures, with a Cα backbone root-mean-square deviation (RMSD) of 0.56 Å (Fig. 3E). The only difference is observed for residues 77–81 in the BC-loop/SRS1 (Fig. 3F), which flipped out of the active site pocket in the 1.75 Å structure in order to accommodate the bound TES (the Cα backbone RMSD drops to 0.17 Å if these residues are not included in the superposition). The flexibility of the BC-loop in CYP109A2 is further evident from the relatively high B-factors for residues 77–81 in both crystal structures and is in line with the behavior of the equivalent loop in crystal structures of other P450s, including CYP109E1. Subsequent analysis

showed the closed conformation of the two CYP109A2 crystal structures to be very similar to the conformation of TES-bound CYP109E1 (PDB entry 5L94) with backbone RMSDs of 2.0 Å for 375 common C α atoms and large deviations restricted mainly to the BC and HI loops (Fig. 3E,F). The newly obtained crystal structures of CYP109A2 thus point to a close relationship of the two CYP109 enzymes from *B. megaterium* DSM319, in line with their amino acid sequence identity and highly similar substrate specificities.

Analysis of mutant CYP109A2-R74V

To destabilize binding of the fatty acid close to the heme, and thereby increase the chance to crystallographically capture a productively bound TES molecule, we decided to mutate Arg74 in the BC-loop/SRS1 of CYP109A2 to a valine. We anticipated that such a mutation would not only interfere with fatty acid binding but also lead to a more hydrophobic active site pocket, thus favoring steroid binding as observed previously in other bacterial P450s [30,31]. Surprisingly, the CYP109A2-R74V mutant completely lost the ability to convert TES in the *in vitro* assay using Adx/AdR as redox partners. The UV–visible spectrum of the purified R74V mutant showed a typical Soret peak at around 420 nm, which like with the wild-type enzyme shifted to 450 nm after reduction with sodium dithionite and saturation with carbon monoxide (CO), typical for a properly folded heme-containing P450 enzyme (Fig. 4A,B). In addition, we determined CO-difference spectra after incubating the protein with dithionite, and with NADPH and the BmCPR-Fdx2 redox protein pair from *B. megaterium* (Fig. 4C), a system which previously was demonstrated to successfully supply electrons to wild-type CYP109A2 [21]. While the mutant was readily reduced by sodium dithionite, the CO-difference spectra revealed a failure in reducing the mutant with NADPH and the BmCPR-Fdx2 redox protein pair. This indicates that the R74V mutation somehow disrupts the ferredoxin-heme electron transfer, by destabilizing binding of the redox partner to CYP109A2, by affecting the electron transfer pathway to the heme within the protein, or by increasing the reorganization energy for electron transfer [32]. Lowering of the redox potential of the heme by the R74V mutation may be another explanation, but this possibility was excluded by electron paramagnetic resonance spectroscopy measurements which showed that wild-type CYP109A2 and mutant R74V exhibited equal redox potentials of –420 mV. Furthermore, the redox potential was not changed upon binding with TES or VD3 (data

not shown). No further experiments were conducted to understand the role of Arg74 in electron transfer, as this would fall beyond the scope of the current study.

Molecular docking of TES and VD3

Although the two new crystal structures of CYP109A2 did not reveal a productively bound TES in the active site, they may still represent a catalytically relevant conformational state for binding and conversion of steroid substrates. To examine this possibility, molecular docking was performed with TES and VD3 using AutoDock4 [33]. The 2.10 Å crystal structure (PDB entry 8ABR) was selected for the docking calculations, since we reasoned that the inward BC-loop conformation would allow tighter binding of the steroids in the active site. Prior to docking, the crystallographic ligands were removed from the structure and the heme-iron was converted to the Compound I state, the reactive intermediate with which the substrate first reacts in the P450 catalytic cycle [34]. Docking poses were clustered using an RMSD cutoff of 1 Å. Subsequently, these were analyzed for catalytically favorable substrate binding modes or NACs, by examining the distances and orientations of the target steroid C-H groups relative to the ferryl oxygen of Compound I using the strict set of NAC criteria as described in the caption of Fig. 5.

As a result, one cluster of highly similar poses of TES was scored as potentially catalytically productive, with the best pose having the C16-H β hydrogen approaching the heme ferryl oxygen at 2.9 Å distance (Fig. 6A). The approach of the ferryl oxygen by the C16-H β hydrogen was restricted by the adjacent C17-OH group of TES, which was also positioned at a close distance from the ferryl oxygen, potentially allowing the formation of a hydrogen bond. We hypothesized that a C17-OH hydrogen bond to the heme ferryl oxygen may serve to keep the substrate in a suitable orientation for the selective abstraction of the C16-H β over the abstraction of the C16-H α hydrogen.

For VD3, 12 differently docked poses displayed NACs for abstraction of the hydrogen at carbon C25 (according to the strict criteria), reflecting the larger flexibility of this substrate as compared to TES (Fig. 6B displays one of the poses, NAC criteria are reported in Fig. 5). Compared with the C16-H β hydrogen in TES, the C25-H hydrogen in VD3 docked closer to the Compound I oxygen, within a 1.8–2.3 Å distance range. All 12 docked poses that were in a NAC for VD3 were selected for MD simulations, while for TES only the single NAC pose (according to strict criteria) was selected.

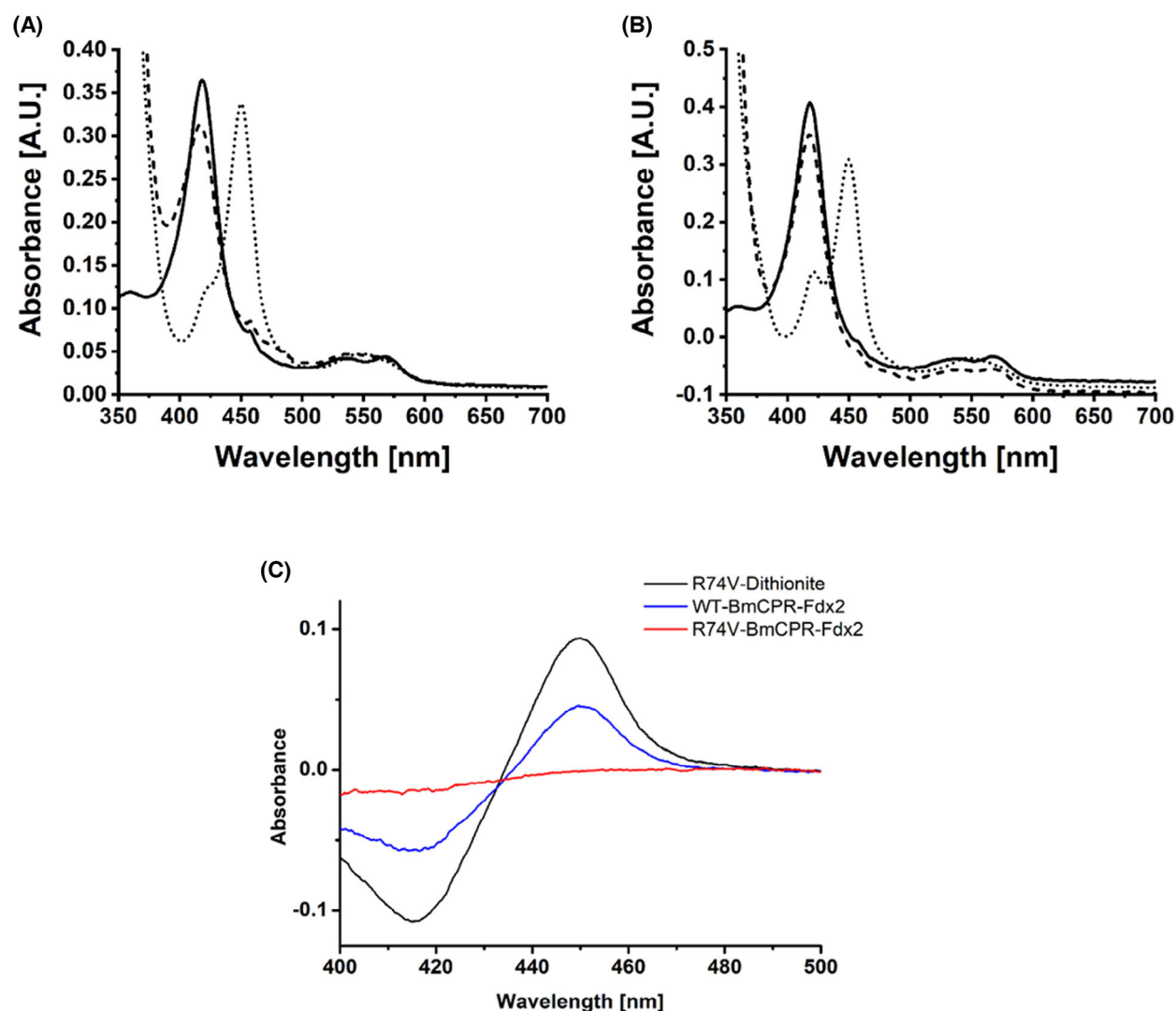


Fig. 4. Spectral characterization of CYP109A2 and its R74V variant. (A) UV-vis spectra of wild-type CYP109A2. (B) UV-vis spectrum of CYP109A2-R74V. Characteristic absorption spectra are shown for both enzymes in the oxidized state (solid line), dithionite-reduced state (dashed line) and dithionite-reduced CO-bound state (dotted line). (C) CO-difference spectra of wild-type CYP109A2 and CYP109A2-R74V obtained after reconstitution with bacterial cytochrome P450 reductase/ferredoxin pair BmCPR-Fdx2. The CO-difference spectrum obtained with dithionite-reduced CYP109A2-R74V is shown for comparison.

MD simulations and NAC analysis

To more thoroughly evaluate the catalytic relevance of the closed CYP109A2 crystal structure, the poses selected from dockings for TES and VD3 were used as starting points for MD simulations. For each of the 13 poses, five independently seeded MD simulations of 20 ns were run giving a total of 100 ns of simulation time per pose. The last 10 ns of each trajectory was analyzed for the frequency of occurrence of strict NACs, following published procedures [35,39]. Because NAC frequencies were analyzed on the fly, every 100 fs, the reported fractions are the average of

500 000 observations ($= 5 \times 10\,000\,000/100$). Even with strict NAC criteria (Fig. 5), the MD trajectories revealed a significant number of frames displaying NACs for both TES and VD3, including NACs consistent with the formation of the experimentally established main reaction products (pro-C16 β for TES, pro-C25 for VD3; Table 2, Fig. 7).

For TES, three out of the five independent 20 ns MD trajectories showed a stable binding of the steroid at a close distance from the heme, with an average distance $d(\text{O}-\text{H}16\beta)$ of ≈ 3 Å and a pro-16 β average NAC frequency of $\approx 0.2\%$. In addition to the C16-H β

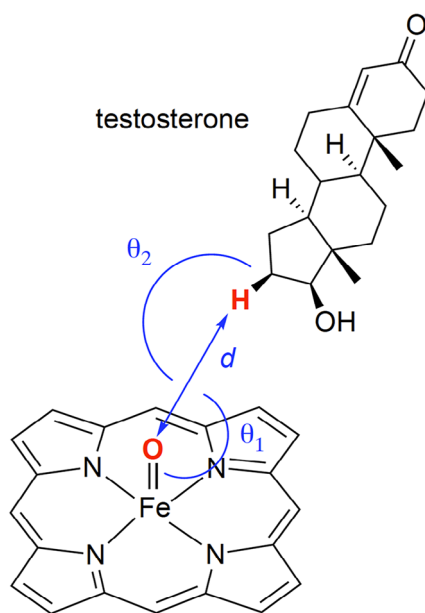


Fig. 5. NAC definitions for TES hydroxylation. Two sets of NAC criteria were used in this study. Our own set of strict NAC criteria required angles to be within roughly 20° of the transition state of the reaction and distances between reacting atoms that are closer than their van der Waals contact distance [35]. The second, more typical set of relaxed NAC criteria was the same as reported by Li *et al.* [36]. Like in most sets of published NAC criteria for P450s, the allowed ranges of angles in the second set are wider and the allowed distances significantly exceed the van der Waals contact distances (see also for instance [37]). For a conformation to be counted as a NAC with our strict criteria, it had to simultaneously meet $d < 2.72 \text{ \AA}$; $100^\circ < \theta_1 < 140^\circ$; and $\theta_2 > 140^\circ$. These criteria are based on the modeled transition state for the reaction of an alkane with compound I, in which θ_1 was 122.3° and θ_2 was 163.6° [38]. The more typical P450 NAC criteria in the second set required $2.0 \text{ \AA} < d < 3.5 \text{ \AA}$; $\theta_1 > 120^\circ$; without requirements for θ_2 .

hydrogen, significant NAC frequencies were observed for the C19-hydrogens of TES. However, due to the higher stability of the latter C-H bonds, one expects the C19-hydrogens to be relatively insensitive to hydroxylation (*vide infra*). No NACs were identified for C16-H α abstraction, consistent with stereoselective conversion to 16 β -OH-TES.

With VD3 as substrate, the MD runs of CYP109A2 revealed a more disordered binding. Four of the 12 VD3 poses did not yield reactive conformations for C25-H hydrogen abstraction, due to the instability of the initial docked enzyme–substrate complex. The MD runs starting from the other VD3 poses resulted in significant occurrences of NACs for C25-H hydrogen abstraction, especially with poses 7, 8, and 9 (Table 2, Fig. 8).

The complexes with TES and pose 7 of VD3, which displayed the highest NAC frequencies for hydroxylation at C25, were subjected to longer and more MD

simulations. A total of 20 replicas of 300 ns were analyzed, giving in total 6 μs of simulation time. Of each trajectory, the last 270 ns was analyzed on the fly, every 2 ps, for NACs giving a total of 2.7×10^6 geometry observations per complex ($= 20 \times 270\,000/2$). Here, analyses were done both with strict and with more typical NAC criteria (Fig. 5). For CYP109A2-VD3 (pose 7) a final average of 0.085% of NACs for C25-H hydroxylation was obtained using strict criteria, with 11 out of 20 simulations showing NAC conformations for at least some time (Fig. 9A). When using more typical criteria, 17 out of 20 simulations showed reactive conformations for CYP109A2-VD3 (Fig. 9B) and the total average was a 100-fold higher, at 6.3% NACs. When using strict criteria, for CYP109A2-testosterone 10 out of the 20 replicas did produce NACs for C16-H β hydroxylation (Fig. 9), to an average of 0.13%. With typical criteria, 19 out of the 20 replicas displayed NACs, often 100% of the time (Fig. 9B), giving an average NAC percentage of 68% for C16-H β hydroxylation. Thus, with strict NAC criteria the observed NAC percentages are small but still significant as the 0.085 and 0.13% correspond to $2\text{--}3 \times 10^5$ observed conformations in a NAC. With more typical NAC criteria, even overwhelmingly high NAC percentages were observed, of 6% and 68%. This all confirmed that these substrates can be bound in reactive conformations in the obtained 8ABR structure.

The r.m.s.d from the X-ray structure was lower for the averaged trajectory than for the individual trajectories (Fig. 9C). This agrees with the MD trajectories sampling different parts of the conformations represented by the X-ray structure [40]. The r.m.s.f. also revealed that the polypeptide maintained a stable conformation throughout the simulations, except for the N terminus, residues 74–81 in the BC-loop and residues 295–308 in the β -subdomain (Fig. 9D), which constitute the more flexible parts of the protein. The flexibility of these regions is consistent with the relatively high B-factors for these residues in the crystal structures.

During both $5 \times 20 \text{ ns}$ and $20 \times 300 \text{ ns}$ MD simulations, also several other VD3 and TES carbon-hydrogens displayed NACs (Table 2, Fig. 7). This does not mean that the NAC analysis predicted a poor regio- and stereoselectivity for these substrates, since also the differences in intrinsic stability of the C-H bonds that need to be broken determine where hydroxylation will actually occur. Therefore, using density functional theory (B3LYP with a 6-31G(d) basis set) we calculated the bond dissociation enthalpies of the C-H bonds in TES and VD3 (Table 3). These calculations predicted that breaking the C16-hydrogen bond of TES

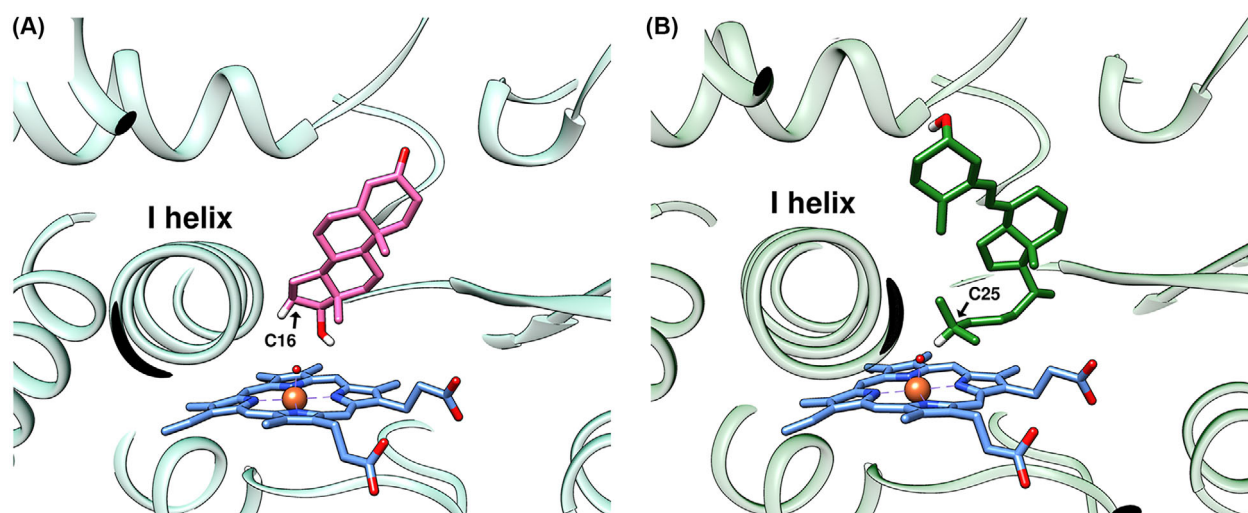


Fig. 6. Selected docked poses of TES and VD3 displaying near-attack conformations (NACs). (A) Docked binding mode of TES (pink) consistent with C16 β -hydroxylation. (B) Docked binding mode of VD3 (green) consistent with C25 hydroxylation. The heme (blue) is shown in its Compound I state. For clarity, only the polar hydrogen atoms and the hydrogens in TES and VD3 that are experimentally observed to be replaced by hydroxyl groups are shown. The pose shown for VD3 is one of the 12 different poses showing a reactive conformation for C25 hydroxylation.

Table 2. NAC frequencies (%) for C-hydrogens in TES and VD3.

TES	C16-H α	C16-H β	C18-H1	C18-H2	C18-H3	C15-H4			
Pose 1	0.000	0.087	1.1	0.97	0.36	7.2			
VD3	C23-H1	C24-H1	C25-H	C26-H1	C26-H2	C26-H3	C27-H1	C27-H2	
Pose 1	1.08	0.028	0.001	1.2	1.1	1.4	0.016	0.005	
Pose 2	0.000	0.000	0.000	0.054	0.053	0.059	0.039	0.030	
Pose 3	0.000	6.3	0.008	1.6	1.4	1.4	1.4	1.3	
Pose 4	0.000	0.001	0.009	1.4	1.3	1.4	3.7	3.6	
Pose 5	0.000	0.000	0.001	0.097	0.11	0.092	0.53	0.57	
Pose 6	0.000	0.000	0.006	2.3	2.3	2.4	0.21	0.17	
Pose 7	0.005	0.001	0.34	1.9	2.2	2.0	0.94	0.90	
Pose 8	0.93	0.17	0.014	0.76	0.80	1.15	0.72	0.70	
Pose 9	0.000	1.6	0.038	2.3	1.6	2.0	2.6	2.1	
Pose 10	0.000	0.000	0.000	0.11	0.13	0.10	0.04	0.06	
Pose 11	0.000	0.024	0.000	5.7	5.1	5.5	2.3	2.4	
Pose 12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	

The NAC frequencies were determined during 20 ns MD simulations of CYP109A2 using the strict criteria (caption of Fig. 5). The position of the hydrogens in the two steroids are shown in Fig. 7. For VD3, the results generated with all 12 initial poses are shown. Four of these poses (poses 2, 10, 11, and 12) did not result in any significant NAC formation at the C25 position. Poses 7, 8, and 9 generated the highest NAC frequencies for C25-H abstraction.

and the C25-hydrogen bond of VD3 will require 2–10 kcal mol⁻¹ less energy than the C-H hydrogen abstractions at all other positions for which NACs were observed. Previously, the bond dissociation energy of TES was calculated using the same method with a similar basis set [36]. In this study, not only quantum mechanics were used to distinguish C16-H β and C16-

H α hydroxylation, but also MD simulations were applied which show that for C16-H α no NACs occurred (Table 2). Furthermore, the current basis set was proven to be highly precise for calculating relative bond dissociation energies [41]. Thus, the modeling indicates that the experimentally observed regio- and stereoselectivity is explained by the formation of

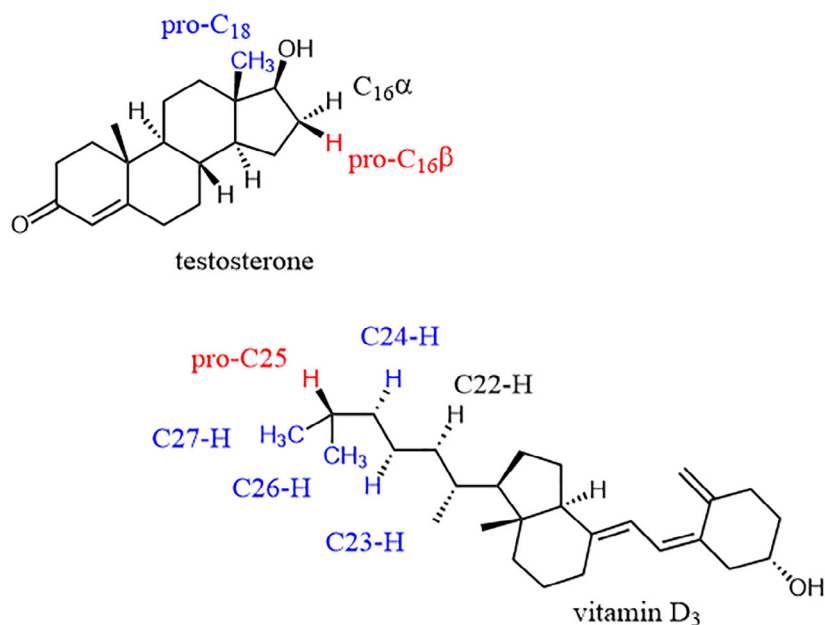


Fig. 7. TES and VD3 C-H hydrogens adopting reactive conformations in the MD simulations. Hydrogens colored in blue and red were found to form NACs along the 20 ns MD simulations (Table 2). The hydrogens colored in red are experimentally observed to be replaced by hydroxyl groups. Hydrogen C22-H (VD3) and C16 α (TES) do not adopt reactive conformations but are included for clarity.

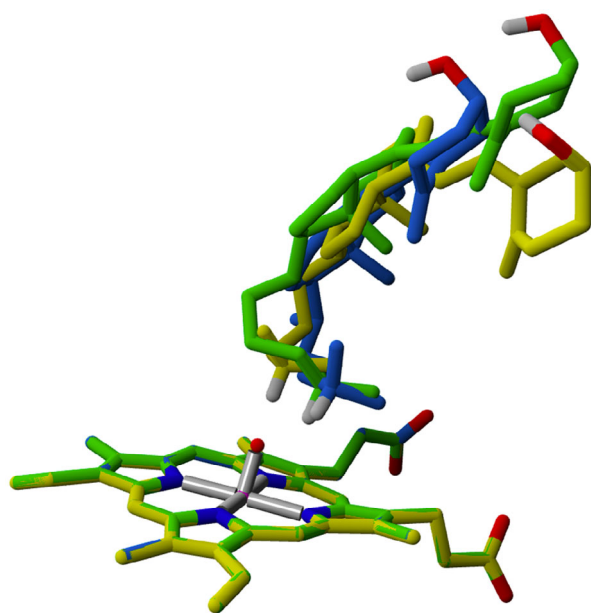


Fig. 8. Poses 7, 8, and 9 of vitamin D3. Initial positions of vitamin D3 pose 7, 8, and 9, which gave the highest number of NACs frequency for C25-H hydroxylation during 20 ns MD simulations, are shown. Pose 7, 8, and 9 are reported in yellow, green, and blue, respectively. C25-H hydrogen is shown in white.

catalytically productive complexes for a subset of the substrate's hydrogen atoms, of which those with the weakest carbon-hydrogen bonds will actually react.

Key residues stabilizing the productive TES and VD3 binding modes

Average structures along the MD simulations were computed and analyzed to identify key residues that stabilize the productive binding modes of TES and VD3 (Fig. S2). Despite differences in the way the two substrates interact at the active site pocket of CYP109A2, a similar core of residues was identified that keeps the steroids in a correct position for catalysis. This core includes Tyr12, Asn14, Glu78, Leu84, Met85, Arg186, Leu237, Ile288, and Phe389. Most of these residues form favorable hydrophobic van der Waals contacts with the steroids. Asn14, Glu78, and Arg186 seem to play a dual role as they contribute to the formation of hydrophobic contacts, as well as serve as hydrogen bonding partners for stabilizing the C3 polar substituent of the steroids (ketone group in TES, hydroxyl group in VD3). Two additional residues, Ala241 and Thr245, are important for the productive binding of TES. The interactions with these residues occur near Compound I and stabilize NAC formation for the conversion of TES to 16(β)-OH-TES. Arg74 does not play a direct role in productive TES binding, but, together with the highly conserved Arg293, is important for stabilizing the position of the heme in the active site, and consequently that of Compound I, via salt bridges formed with the heme's propionate groups. In addition, as in the docking experiments, the MD analysis reveals a persistent hydrogen

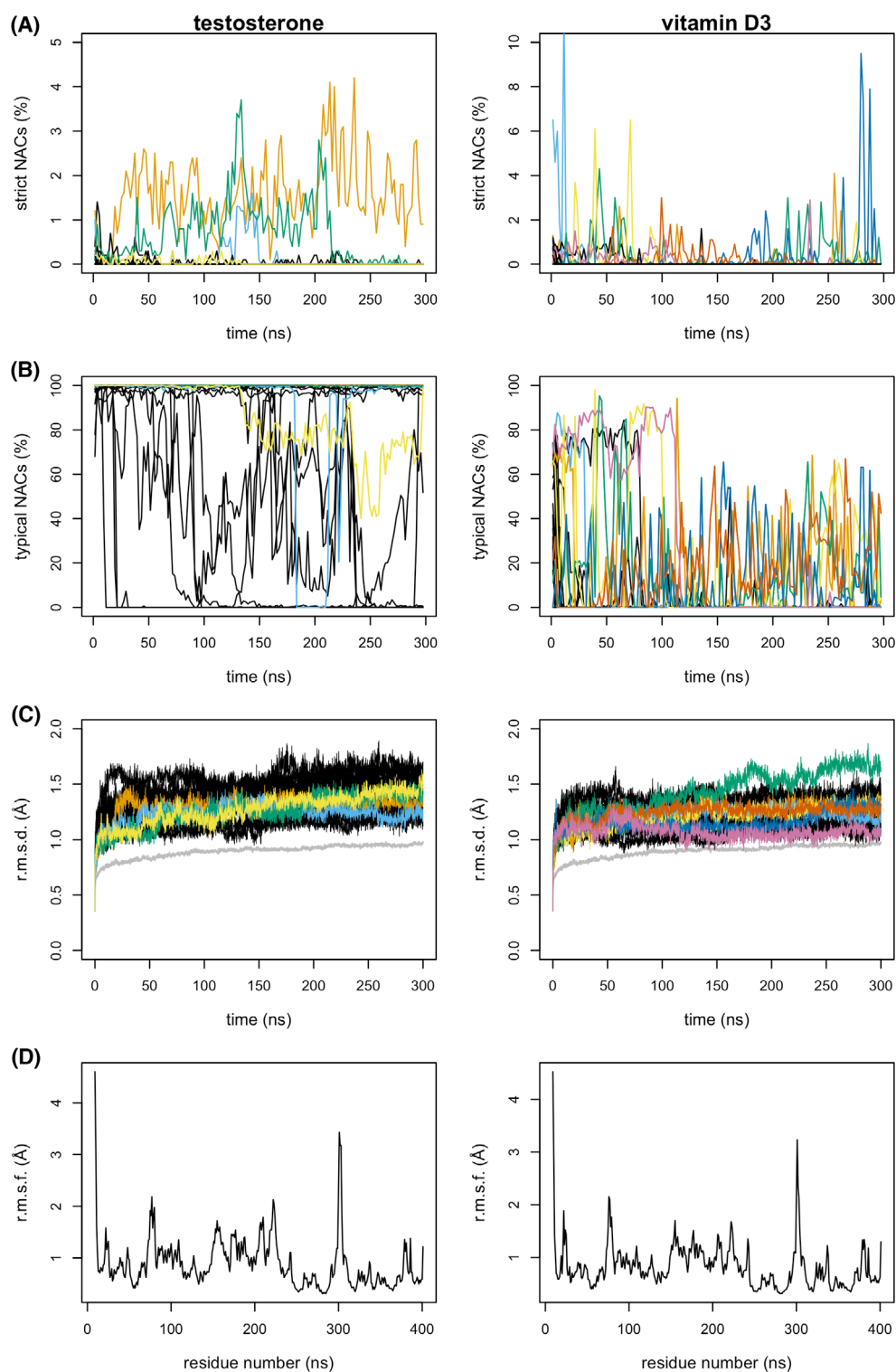


Fig. 9. CYP109A2 during the 20×300 ns MD simulations with TES and VD3. In panels (A–C), the colored traces belong to trajectories that gave relatively high NACs with strict criteria. (A) NACs percentages using strict criteria. (B) NACs percentages using more typical criteria. (C) Root-mean-square deviation for enzyme–TES and enzyme–VD3 during 300 ns simulations showing enzyme–substrate complexes stability. In gray is the average RMSD over all 20 trajectories. (D) RMSF plots revealing relatively large flexibilities for residues at the N terminus, in the BC-loop (residues 74–81) and in the β -subdomain (residues 295–308) for both enzyme–substrate complexes.

Table 3. Calculated dissociation energies (enthalpies) of C-H bonds in TES and VD3.

TES	C16-H	C15-H	C19-H			
Total (kcal mol ⁻¹)	93.9	95.4	99.0			
VD3	C25-H	C22-H	C23-H	C24-H	C26-H	C27-H
Total (kcal mol ⁻¹)	92.4	95.1	94.3	95.6	100.7	100.3

The calculations were performed at 298 K and only included those C-H bonds of which the hydrogen formed NACs during MD simulations.

bond formed between the C17-OH of TES and the heme ferryl oxygen, further indicating that this interaction may be important for optimal positioning of the substrate for the selective abstraction of the C16-H β .

Discussion

Cytochrome P450 monooxygenase CYP109A2 from *B. megaterium* DSM319 was previously shown to convert VD3 to 25-hydroxyvitamin D3 (25(OH)VD3) [21,23]. Our current study reveals that this enzyme is also able to convert TES to 16 β -OH-TES in a highly regio- and stereoselective manner. Thus, CYP109A2 and CYP109E1 from *B. megaterium* DSM319 share the same activities, although the latter enzyme appears to have a broader substrate scope, as it can also convert cholesterol, statin derivatives and a number of terpene substrates [25,26]. To reveal the structural determinants in CYP109A2 governing substrate specificity and product regio- and stereoselectivity, we solved two structures of the enzyme obtained from crystals grown in the presence of TES. Surprisingly, in the crystal structures a compound different than TES is bound close to the heme-iron, which, based on our interpretation of the associated electron density and the inferred binding interactions at the active site, could be a medium chain fatty acid, although the exact nature of the fatty acid remains unclear. The putative fatty acid must have been picked up by CYP109A2 during heterologous expression in *E. coli* or cell lysis, as no such compound was present in the buffers used for protein purification and crystallization. Subjecting the protein to Lipidex resin did not remove the heme-bound ligand from the active site, while it was successful in allowing the second, remotely bound ligand to be replaced by TES. This suggests that the putative heme-bound fatty acid binds with high affinity. We did not further investigate the nature of the heme-bound ligand in the crystal structures, as our interest was toward understanding how CYP109A2 binds TES and VD3. However, in future studies it may be worthwhile to establish whether CYP109A2 can convert fatty acids, as such an activity could point to a

possible physiological role of the enzyme in *B. megaterium* DSM319.

An additional surprising finding is the loss of steroid hydroxylation activity observed for CYP109A2 mutant R74V. In the wild-type crystal structures, Arg74 has a role in binding the putative fatty acid ligand. However, our modeling experiments did not reveal an obvious role of this residue in binding TES or VD3. In fact, the loss in activity of the mutant seems to originate from hampered electron transfer between CYP109A2 and its ferredoxin partner. We do not understand how the R74V mutation affects the electron transfer pathway. Given its position at the active site (BC-loop), a direct role of Arg74 in electron transfer seems unlikely. Presumably, the mutation has an indirect effect on the thermodynamics or kinetics of the electron transfer, caused by altered side chain interactions, a change in the chemico-physical environment around the heme or a partial dislocation of the heme. Establishing the exact nature of such effects requires further investigations.

Most importantly, our study emphasizes the strength of steroid docking, combined with MD and quantum mechanical calculations of C-H bond dissociation energies, to assess the functional relevance of P450 crystal structures and predict catalytically relevant steroid binding modes. Previous studies with CYP109E1 and CYP154C5 already showed how scoring NACs along MD simulation trajectories can be successfully applied to study and predict regio- and stereoselectivity in steroid conversion [24,35]. Using a similar approach, we were able to reproduce the experimentally determined regio- and stereoselectivities of CYP109A2 in converting two very different steroids, one being rigid and relatively small (TES), the other being rather flexible and long (VD3). These results demonstrate that the protein conformational state in the closed crystal structures of CYP109A2 is relevant for modeling catalytically relevant substrate binding modes. The docked binding modes of TES and VD3 in the active site of CYP109A2 and the average structures along the MD simulations further reveal how CYP109A2 uses a common core of six apolar and three polar/charged amino acid residues to stabilize the steroids. The overall features of the

steroid binding modes and the specific roles in steroid binding of some of the residues are similar as in related P450 enzymes [42]. This is particularly true for residues Leu84 and Ile288, which are located at opposite sides of the CYP109A2 active site in close vicinity of the heme. In various P450 enzymes, including CYP109E1, CYP260A1, and P450BM3, residues at these positions in the heme/substrate binding pocket have crucial activity and selectivity-determining roles [18,23,24,43–45].

The initial 20 ns MD simulations employed in this study for 13 enzyme–substrate complexes were long enough to identify productive binding modes which were subsequently confirmed in longer 300 ns MD simulations with 20 replicas. Even with strict NAC criteria, there were significant numbers of NACs along the 300 ns simulations. With more typically used NAC criteria, the fractions were extremely high in case of the TES, at >50%. These MD simulations might seem satisfactorily long as they are 4 orders of magnitude longer than the 100 ps MD simulations that were feasible two decades ago [46]. They do, however, not necessarily provide more information since essentially the time scales remain very short compared with the time scales on which enzyme conformational changes occur. The current sub-microsecond time scales are still >9 orders of magnitude short from relatively slow enzyme conformational changes, which occur on time scales of minutes to hours (e.g., [47]) to even days. Full sampling of enzyme conformational changes is therefore still far out of reach of accurate atomistic MD simulations, especially with single trajectories. The main sampling here was due to the use of multiple independent trajectories, which allows to observe the relatively rare conformations that provide NACs with even the strictest criteria (Fig. 9). These multitrajectory MD results do support that the crystallographically observed conformation is catalytically competent.

In conclusion, our complementary approach using X-ray crystallography, steroid modeling, and MD simulations combined with NAC analysis allowed us to obtain relevant new insights how two steroids are selectively converted by CYP109A2. We expect that our newly determined closed crystal structures of CYP109A2 may serve as valuable templates for future ‘*in-silico*’ screening for new substrates or for altering substrate selectivities by protein engineering.

Materials and methods

Chemicals, strains, and plasmids

Escherichia coli C43(DE3) cells and expression vector pET17b used for recombinant gene expression were obtained

Regio- and stereoselective steroid hydroxylation by CYP109A2

from Novagen (Darmstadt, Germany). Delta-aminolevulinic acid (δ -ALA) and isopropyl- β -D-thiogalactopyranoside (IPTG) were purchased from Carbolution Chemicals GmbH (Saarbrücken, Germany). NADPH was from Gerbu (Wieblingen, Germany). Glucose-6-phosphate and glucose-6-phosphate dehydrogenase (G6P-DH) were purchased from Roche (Mannheim, Germany). Bacterial media were obtained from Becton Dickinson (Heidelberg, Germany). All other chemicals were from standard sources and of highest purity available.

Recombinant protein expression and purification

CYP109A2 was overproduced in *E. coli* C43 (DE3) cells as a C-terminal His-tagged protein, using a pET17b-based expression vector, following previously published protocols [21]. Protein purification was initiated by resuspending the CYP109A2-containing cell pellet obtained from a 0.5-L-scale fermentation in 50 mL lysis buffer (50 mM potassium phosphate, pH 7.4, supplemented with 0.1 mM dithioerythritol), followed by sonication on ice for 15 min (15 s pulse on, 15 s pulse off, 13% amplitude). Cell debris was removed from the lysate by centrifugation at 30 000 *g* and 4 °C for 30 min. The supernatant was applied to a Protino Ni-NTA Agarose column (Macherey-Nagel) equilibrated with lysis buffer. Following a three-step wash with 50 mL lysis buffer containing 10, 20 and 40 mM imidazole, elution was carried out with lysis buffer containing 300 mM imidazole. The CYP109A2-containing elution fractions were pooled and subsequently concentrated using a 30 kDa cut-off Amicon Ultra Centrifugal Filter Unit (Merck Millipore). To prepare protein for crystallization, the purification procedure was repeated using buffers supplemented with 100 μ M TES. In addition, size exclusion chromatography was performed as a final step, using a HiLoad Superdex 75 column (GE Healthcare) with a running buffer containing 25 mM Tris/HCl, pH 7.4, 2% (v/v) glycerol and 100 μ M TES. Fractions with the highest protein purity (A_{420}/A_{280}) were combined, and the protein concentration was adjusted to ~850 μ M using a centrifugal filter unit.

Recombinant production and purification of the electron transfer proteins Adx₄₋₁₀₈ and AdR as well as Fdx2 and BmCPR was carried out as described elsewhere [48–51]. The protein purities of Adx₄₋₁₀₈ (A_{414}/A_{280}), AdR (A_{450}/A_{280}), Fdx2 (A_{390}/A_{280}), and BmCPR (A_{456}/A_{280}) were in good agreement with those from previous reports. Final protein samples were flash frozen with liquid nitrogen and stored at –80 °C until further use.

Site-directed mutagenesis and purification of CYP109A2-R74V

The R74V mutation in CYP109A2 was introduced by site-directed mutagenesis using the QuikChange method. The

whole-length pET17b carrying the CYP109A2 gene was amplified by PCR using Phusion DNA polymerase (New England BioLabs, Ipswich, UK) and the following oligonucleotide primers (Eurofins Genomics GmbH, Ebersberg, Germany): forward primer: 5'-CATCCAAAGTAGCAATGGAAGAACGCCAAG and reverse primer: 5'-CTTCCATTGCTACTTTGGATGAAAAAAGCG. The PCR reaction was performed in 50 μ L total volume using a thermocycler (PTC-200 DNA Engine cycler). The following PCR program was used: initial denaturation at 98 °C for 30 s, followed by 18 cycles of 98 °C, 30 s (denaturation), 60 °C, 50 s (annealing) and 72 °C, 4 min (extension). Correct generation of the desired mutation was confirmed by DNA sequencing, carried out by Eurofins Genomics GmbH (Ebersberg, Germany). The mutated plasmid was used to transform *E. coli* C43 (DE3) cells. Expression and purification of CYP109A2-R74V were carried out as for wild-type CYP109A2.

Spectral characterization

Concentrations of the electron transfer proteins were calculated based on the molar extinction coefficients of $\epsilon_{414} = 9.8 \text{ mM}^{-1} \text{ cm}^{-1}$ for Adx [52], $\epsilon_{450} = 10.9 \text{ mM}^{-1} \text{ cm}^{-1}$ for AdR [53], $\epsilon_{390} = 6.67 \text{ mM}^{-1} \text{ cm}^{-1}$ for Fdx2 [50], and $\epsilon_{456} = 21.0 \text{ mM}^{-1} \text{ cm}^{-1}$ for BmCPR [54].

Concentrations of the purified CYP109A2 proteins were estimated by CO-difference spectral assays using a molar extinction coefficient of $\epsilon_{450-490} = 91 \text{ mM}^{-1} \text{ cm}^{-1}$, as described elsewhere [55]. Electron transfer in the absence of substrate was investigated by the addition of dithionite as reducing agent, or by reconstituting the CYP109A2 proteins with diflavin reductase BmCPR and ferredoxin Fdx2 as electron transfer proteins in the presence of NADPH, followed by recording of CO-difference spectra as described previously [56]. All spectra were recorded with a double-beam spectrophotometer (UV2101PC; Shimadzu, Japan).

In vitro reactions

The *in vitro* conversion of tested steroids was performed in a final volume of 200 μ L, using 50 mM potassium phosphate buffer, pH 7.4, supplemented with 2% glycerol, at 30 °C for 30 min. The reconstituted system contained CYP109A2 (1 μ M), truncated bovine adrenodoxin (Adx₄₋₁₀₈, 20 μ M), bovine adrenodoxin reductase (AdR, 2 μ M), a NADPH-regenerating system (1 mM MgCl₂, 5 mM glucose-6-phosphate, 1 U glucose-6-phosphate dehydrogenase), and 200 μ M of the corresponding steroid dissolved in ethanol. Reactions were initiated by the addition of NADPH at a final concentration of 1 mM, then stopped, and extracted twice with the addition of 200 μ L of ethyl acetate. Samples were centrifuged (10 000 *g*, 10 min), and the organic phases were combined and evaporated until complete dryness and, after dissolving in acetonitrile, analyzed by HPLC.

Reversed-phase high-performance liquid chromatography (RP-HPLC)

Reversed-phase high-performance liquid chromatography (RP-HPLC) analysis was performed on a Jasco system (PU-980 HPLC pump, AS-950 sampler, UV-975 UV/visible detector, LG-980-02 gradient unit; Jasco, Gross-Umstadt, Germany) equipped with a Nucleodur 100–5 C18 column (125 \times 4 mm; Macherey-Nagel, Düren, Germany). The column temperature was adjusted to 40 °C. Samples were dissolved in 200 μ L acetonitrile. TES and its metabolites were eluted from the column using a gradient method, starting with a mobile phase ratio of 10% acetonitrile and increasing it to 100% acetonitrile over 30 min. UV detection of all steroids was carried out at a wavelength of 240 nm, except for estradiol and estrone which were detected at 280 nm. The absorption properties of the products did not differ from the substrate and, therefore, 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.

X-ray crystallography

Frozen samples of CYP109A2 (40 mg/mL) purified in the presence of 0.1 mM TES were thawed and either directly used for crystallization screening or first subjected to a treatment for removing the bound fatty acids. Removal of bound fatty acids from the protein was performed with Lipidex 1000 resin (hydroxyalkoxypropyl-Dextran, Sigma-Aldrich, Saint-Louis, MO, USA), following a procedure adapted from previously published protocols [29,57]. Briefly, 500 μ L Lipidex slurry was pipetted into an MC Ultrafree centrifuge filter cup (0.22 μ m filter, 500 μ L), spun-down at 10 000 *g*, resuspended with 50 mM citrate buffer, pH 5.5 (buffer A), and spun-down again. The last two steps were repeated. A Vivaspin 10 kDa ultracentrifugation device (500 μ L) was filled with 25 μ L CYP109A2 sample, diluted with buffer A to 500 μ L, centrifuged at 10 000 *g*, and washed/centrifuged again with buffer A. Retained CYP109A2 was diluted with buffer A to a total volume of 350 μ L. This solution was added to the washed Lipidex slurry in the MC filter cup and gently mixed. The cup was incubated overnight at 4 °C. Then, the cup was centrifuged (10 000 *g*) and the red-colored filtrate set aside. The filter cup was washed with 500 μ L buffer and centrifuged; filtrates were combined. The resulting protein sample was concentrated to ~33 mg/mL with 0.1 mM TES added in buffer A.

Initial crystallization conditions were identified using the sitting-drop vapor diffusion method, with the help of a Mosquito dispensing robot (SPT LabTech, Melbourne, UK) and applying several commercially available high-throughput crystallization screens. Drops were dispensed at a volume of 200 nL using a 1 : 1 protein:reservoir drop ratio, and the plates were incubated at a stable temperature

of 21 °C. Crystallization conditions were optimized with the aid of a Dragonfly crystal robot (SPT LabTech). The finally obtained crystallization solution contained 0.4–0.55 M ammonium sulfate, 1.05–1.2 M lithium sulfate, and 0.1 M sodium citrate buffer, pH 5.5. Crystals of untreated CYP109A2, and of protein treated with Lipidex, grew at identical conditions.

Prior to X-ray data collection, the crystals were soaked for few seconds in a mother liquor solution supplemented with 30% (v/v) glycerol to provide cryo-protection and then flash-cooled by immersion in liquid nitrogen. Diffraction data were collected at 100 K at the ID30A-3 beamline at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, and at the I03 beamline at the Diamond Light Source in Oxfordshire, UK. Data indexing, scaling, and merging were performed using the programs XDS [58] and Aimless [59] from the CCP4 software suite [60]. Initial phases and structures were determined by molecular replacement using the ‘open’ structure of CYP109A2 (PDB entry 5OFQ) as a search model. Molecular replacement was carried out with Phaser [61] within the Phenix software suite [62]. The structures were optimized via alternated rounds of model building and refinement using Coot [63] and Phenix.refine [64]. Waters and bound ligands were added at the final stages of model building and refinement, and the geometry of the final models was validated with MolProbity [65]. Coordinates and structure factors have been deposited in the Protein Data Bank with entry codes 8ABR (2.10 Å structure, two bound fatty acids) and 8ABS (1.75 Å structure, TES/fatty acid-bound).

Sequence and structure analysis

Protein sequences were compared using the Basic Local Alignment Search Tool (BLAST, NCBI). Pairwise structural alignments were done using PDBFold [66]. UCSF Chimera [67] was used for structure analysis and figure preparation. Substrate recognition sites (SRS) in CYP109A2 were identified by alignment with P450cam (WP_032492633) and conform to the classification according to Gotoh [28]: SRS1: 71–87 (BC-loop), SRS2: 163–169 (C'-terminal part of F-helix), SRS3: 185–192 (N'-terminal part of G-helix), SRS4: 229–247 (central part of I-helix), SRS5: 284–293 (K-β5 connection), and SRS6: 384–392 (β9–β10 turn).

TES and VD3 docking

Molecular docking of the steroids in the 2.10 Å crystal structure of closed CYP109A2 was carried out with the YASARA Structure package version 20.4.24 (YASARA Biosciences, Vienna, Austria). Prior to the docking, the two bound fatty acids were removed from the active site and the heme-iron was converted to Compound I following published specifications [35,68]. A structure of VD3 was obtained from the Cambridge structural database [69]. The

structures of TES and VD3 were optimized in YASARA using PM3 [70]. The docking simulations were performed using the AutoDock4 routines implemented in YASARA with 9990 runs of 25 000 and 250 000 energy evaluations per docking run, for the rigid TES and the flexible VD3, respectively. Docking poses were analyzed for being in a potentially reactive pose for the known regio- and stereospecificity by measuring the angles and distances that can be used to classify a pose as a NAC (Fig. 5). Only those docked orientations were selected for MD simulations that were in a NAC according to strict criteria and had a shorter catalytic distance *d* (Fig. 7) than similarly docked orientations of the substrate, namely those within an RMSD of 1 Å.

Molecular dynamics simulations

Molecular dynamics simulations were performed with YASARA, using the Yamber3 force field [71]. Long-range electrostatics were calculated using the particle mesh Ewald method [72]. Hydrogen atom positions were constrained to allow for a longer simulation time step of 2.0 fs [73–75]. Atomic point charges of TES and VD3 were calculated using the AM1-BCC method [76]. The octahedral simulation cell was extended 7.5 Å around the protein, and periodic boundary conditions were applied. During the MD simulations, the heme was maintained in the Compound I state, as generated for the docking calculations. For each of the selected docking poses, 5–20 independent MD simulations with periodic boundary conditions were run for 20–300 ns, giving total simulation times of 100 ns to 6 μs. After energy minimization, a gradual warm-up from 5 to 298 K was done in 30 ps. At this start of each individual simulation trajectory, the atom velocities were assigned randomly, following a Maxwell-Boltzmann distribution. Subsequently, an equilibration of 9.97 or 30 ns was performed for each enzyme–substrate complex. During the following production phases of 10–270 ns, NACs were analyzed on the fly, every 100 fs for the 22 ns MD simulations and every 2 ps for the 300 ns simulations. Snapshots were recorded every 50 ps for the 20 ns MD simulations and every 100 ps for the 300 ns MD simulations.

Bond dissociation energy calculations

Bond dissociation enthalpies of relevant C-H bonds in TES and VD3 were calculated under Gaussian09. The B3LYP density functional methods with 6-31G(d) basis set were used, based on previous literature [41].

Author contributions

IKJ, RB, and A-MWHT conceived the studies. All authors contributed to data analysis. IKJ and A-MWHT analyzed the X-ray crystallographic data.

HJR performed the X-ray crystallographic experiments. EB and HJW performed the docking and MD simulations. AA, PH and RK performed the biocatalytic and bioanalysis experiments. AA and PH purified the proteins and performed the site-directed mutagenesis experiments. IKJ, EB, HJW, and A-MWHT wrote the paper. DBJ and RB assisted with writing the paper. All authors read and approved the paper.

Acknowledgements

This research has received funding from the European Union's 7th Framework Programme (Marie Curie Actions-ITN P4FIFTY) under GA No. 289217 and the European Union's Horizon 2020 Programme (Marie Curie Actions-ITN ES-Cat) under GA No. 722610, which supported I.K.J. and E.B., respectively. We would like to thank Prof. Anett Schallmeyer (Technical University of Braunschweig, Germany) for kindly providing the 16(α)-hydroxytestosterone authentic standard for our analysis. We acknowledge the European Synchrotron Radiation Facility (ESRF) and the Diamond Light Source (DLS) for provision of synchrotron radiation facilities, and we would like to thank the beamline scientists for assistance in using ESRF beamline MASSIF-1 and DLS beamline I03. The Center of Information Technology at the University of Groningen is acknowledged for the use of the Peregrine and Hábrók High-Performance Computing clusters.

Peer review

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/febs.16906>.

Data availability

The structural data that support this study are openly available in the PDB under entry codes [8ABR](#) and [8ABS](#). Additional data that support this study can be found in the Supporting Information or are available from the corresponding author upon reasonable request.

References

- Bernhardt R (2006) Cytochromes P450 as versatile biocatalysts. *J Biotechnol* **124**, 128–145.
- Podust LM & Sherman DH (2012) Diversity of P450 enzymes in the biosynthesis of natural products. *Nat Prod Rep* **29**, 1251–1266.
- Bernhardt R & Urlacher VB (2014) Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Appl Microbiol Biotechnol* **98**, 6185–6203.
- Girvan HM & Munro AW (2016) Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. *Curr Opin Chem Biol* **31**, 136–145.
- Zhang L & Wang Q (2022) Harnessing P450 enzyme for biotechnology and synthetic biology. *Chembiochem* **23**, e202100439.
- Bhatti HN & Khera RA (2012) Biological transformations of steroidal compounds: a review. *Steroids* **77**, 1267–1290.
- Donova MV & Egorova OV (2012) Microbial steroid transformations: current state and prospects. *Appl Microbiol Biotechnol* **94**, 1423–1447.
- Donova M (2021) Microbial steroid production technologies: current trends and prospects. *Microorganisms* **10**, 53.
- Urlacher VB & Girhard M (2019) Cytochrome P450 monooxygenases in biotechnology and synthetic biology. *Trends Biotechnol* **37**, 882–897.
- Hu B, Zhao X, Wang E, Zhou J, Li J, Chen J & Du G (2023) Efficient heterologous expression of cytochrome P450 enzymes in microorganisms for the biosynthesis of natural products. *Crit Rev Biotechnol* **43**, 227–241.
- Agematu H, Matsumoto N, Fujii Y, Kabumoto H, Doi S, Machida K *et al.* (2006) Hydroxylation of testosterone by bacterial cytochromes P450 using the *Escherichia coli* expression system. *Biosci Biotechnol Biochem* **70**, 307–311.
- Szaleniec M, Wojtkiewicz AM, Bernhardt R, Borowski T & Donova M (2018) Bacterial steroid hydroxylases: enzyme classes, their functions and comparison of their catalytic mechanisms. *Appl Microbiol Biotechnol* **102**, 8153–8171.
- di Nardo G & Gilardi G (2020) Natural compounds as pharmaceuticals: the key role of cytochromes P450 reactivity. *Trends Biochem Sci* **45**, 511–525.
- Kille S, Zilly FE, Acevedo JP & Reetz MT (2011) Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat Chem* **3**, 738–743.
- McIntosh JA, Farwell CC & Arnold FH (2014) Expanding P450 catalytic reaction space through evolution and engineering. *Curr Opin Chem Biol* **19**, 126–134.
- 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.
- Hartz P, Strohmaier SJ, El-Gayar BM, Abdulmughni A, Hutter MC, Hannemann F *et al.* (2021) Resurrection and characterization of ancestral CYP11A1 enzymes. *FEBS J* **288**, 6510–6527.

- 18 Khatri Y, Jóźwik IK, Ringle M, Ionescu IA, Litzenburger M, Hutter MC, Thunnissen AMWH & Bernhardt R (2018) Structure-based engineering of steroidogenic CYP260A1 for stereo- and regioselective hydroxylation of progesterone. *ACS Chem Biol* **13**, 1021–1028.
- 19 Nguyen KT, Virus C, Gunnewich N, Hannemann F & Bernhardt R (2012) Changing the regioselectivity of a P450 from C15 to C11 hydroxylation of progesterone. *Chembiochem* **13**, 1161–1166.
- 20 Eppinger M, Bunk B, Johns MA, Edirisinghe JN, Kutumbaka KK, Koenig SS *et al.* (2011) Genome sequences of the biotechnologically important *Bacillus megaterium* strains QM B1551 and DSM319. *J Bacteriol* **193**, 4199–4213.
- 21 Abdulmughni A, Jóźwik IK, Brill E, Hannemann F, Thunnissen AMWH & Bernhardt R (2017) Biochemical and structural characterization of CYP109A2, a vitamin D3 25-hydroxylase from *Bacillus megaterium*. *FEBS J* **284**, 3881–3894.
- 22 Abdulmughni A, Jóźwik IK, Putkaradze N, Brill E, Zapp J, Thunnissen AMWH, Hannemann F & Bernhardt R (2017) Characterization of cytochrome P450 CYP109E1 from *Bacillus megaterium* as a novel vitamin D3 hydroxylase. *J Biotechnol* **243**, 38–47.
- 23 Abdulmughni A, Erichsen B, Hensel J, Hannemann F & Bernhardt R (2021) Improvement of the 25-hydroxyvitamin D3 production in a CYP109A2-expressing *Bacillus megaterium* system. *J Biotechnol* **325**, 355–359.
- 24 Jóźwik IK, Kiss FM, Gricman L, Abdulmughni A, Brill E, Zapp J, Pleiss J, Bernhardt R & Thunnissen AMWH (2016) Structural basis of steroid binding and oxidation by the cytochrome P450 CYP109E1 from *Bacillus megaterium*. *FEBS J* **283**, 4128–4148.
- 25 Putkaradze N, Litzenburger M, Abdulmughni A, Milhim M, Brill E, Hannemann F & Bernhardt R (2017) CYP109E1 is a novel versatile statin and terpene oxidase from *Bacillus megaterium*. *Appl Microbiol Biotechnol* **101**, 8379–8393.
- 26 Putkaradze N, Litzenburger M, Hutter MC & Bernhardt R (2019) CYP109E1 from *Bacillus megaterium* acts as a 24- and 25-hydroxylase for cholesterol. *Chembiochem* **20**, 655–658.
- 27 Bracco P, Janssen DB & Schallmey A (2013) Selective steroid oxyfunctionalisation by CYP154C5, a bacterial cytochrome P450. *Microb Cell Fact* **12**, 95.
- 28 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.
- 29 Oldham NJ, Krieger J, Breer H & Svatos A (2001) Detection and removal of an artefact fatty acid from the binding site of recombinant *Bombyx mori* pheromone-binding protein. *Chem Senses* **26**, 529–531.
- 30 Ma B, Wang Q, Ikeda H, Zhang C & Xu LH (2019) Hydroxylation of steroids by a microbial substrate-promiscuous P450 cytochrome (CYP105D7): key arginine residues for rational design. *Appl Environ Microbiol* **85**, e01530-19.
- 31 Sugimoto H, Shinkyo R, Hayashi K, Yoneda S, Yamada M, Kamakura M, Ikushiro SI, Shiro Y & Sakaki T (2008) Crystal structure of CYP105A1 (P450SU-1) in complex with 1 α ,25-dihydroxyvitamin D3. *Biochemistry* **47**, 4017–4027.
- 32 Marcus RA & Sutin N (1985) Electron transfers in chemistry and biology. *Biochim Biophys Acta* **811**, 265–322.
- 33 Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS & Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* **30**, 2785–2791.
- 34 Rittle J & Green MT (2010) Cytochrome P450 compound I: capture, characterization, and C-H bond activation kinetics. *Science* **330**, 933–937.
- 35 Bracco P, Wijma HJ, Nicolai B, Buitrago JAR, Klunemann T, Vila A *et al.* (2021) CYP154C5 regioselectivity in steroid hydroxylation explored by substrate modifications and protein engineering. *Chembiochem* **22**, 1099–1110.
- 36 Li J, Tang Y, Li W & Tu Y (2020) Mechanistic insights into the regio- and stereoselectivities of testosterone and dihydrotestosterone hydroxylation catalyzed by CYP3A4 and CYP19A1. *Chemistry* **26**, 6214–6223.
- 37 Eichler A, Gricman L, Herter S, Kelly PP, Turner NJ, Pleiss J & Flitsch SL (2016) Enantioselective benzylic hydroxylation catalysed by P450 monooxygenases: characterisation of a P450cam mutant library and molecular modelling. *Chembiochem* **17**, 426–432.
- 38 Shaik S, Lai W, Chen H & Wang Y (2010) The valence bond way: reactivity patterns of cytochrome P450 enzymes and synthetic analogs. *Acc Chem Res* **43**, 1154–1165.
- 39 Ashworth MA, Bombino E, de Jong RM, Wijma HJ, Janssen DB, McLean KJ & Munro AW (2022) Computation-aided engineering of cytochrome P450 for the production of pravastatin. *ACS Catal* **12**, 15028–15044.
- 40 Caves LS, Evanseck JD & Karplus M (1998) Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci* **7**, 649–666.
- 41 Xiu-Juan Q, Yong F, Lei L & Qing-Xiang G (2005) Assessment of performance of G3B3 and CBS-QB3 methods in calculation of bond dissociation energies. *Chin J Chem* **23**, 194–199.
- 42 Gricman L, 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.

- 43 Li QS, Ogawa J, Schmid RD & Shimizu S (2001) Residue size at position 87 of cytochrome P450 BM-3 determines its stereoselectivity in propylbenzene and 3-chlorostyrene oxidation. *FEBS Lett* **508**, 249–252.
- 44 Seifert A & Pleiss J (2009) Identification of selectivity-determining residues in cytochrome P450 monooxygenases: a systematic analysis of the substrate recognition site 5. *Proteins* **74**, 1028–1035.
- 45 Seifert A, Vomund S, Grohmann K, Kriening S, Urlacher VB, Laschat S & Pleiss J (2009) Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity. *Chembiochem* **10**, 853–861.
- 46 Chen K, Hirst J, Camba R, Bonagura CA, Stout CD, Burgess BK & Armstrong FA (2000) Atomically defined mechanism for proton transfer to a buried redox centre in a protein. *Nature* **405**, 814–817.
- 47 Otten R, Padua RAP, Bunzel HA, Nguyen V, Pitsawong W, Patterson M *et al.* (2020) How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science* **370**, 1442–1446.
- 48 Uhlmann H, Beckert V, Schwarz D & Bernhardt R (1992) Expression of bovine adrenodoxin in *E. coli* and site-directed mutagenesis of 2 Fe-2S cluster ligands. *Biochem Biophys Res Commun* **188**, 1131–1138.
- 49 Sagara Y, Wada A, Takata Y, Waterman MR, Sekimizu K & Horiuchi T (1993) Direct expression of adrenodoxin reductase in *Escherichia coli* and the functional characterization. *Biol Pharm Bull* **16**, 627–630.
- 50 Brill E, Hannemann F, Zapp J, Brüning G, Jauch J & Bernhardt R (2014) A new cytochrome P450 system from *Bacillus megaterium* DSM319 for the hydroxylation of 11-keto- β -boswellic acid (KBA). *Appl Microbiol Biotechnol* **98**, 1701–1717.
- 51 Milhim M, Gerber A, Neunzig J, Hannemann F & Bernhardt R (2016) A novel NADPH-dependent flavoprotein reductase from *Bacillus megaterium* acts as an efficient cytochrome P450 reductase. *J Biotechnol* **231**, 83–94.
- 52 Huang JJ & Kimura T (1973) Studies on adrenal steroid hydroxylases. Oxidation-reduction properties of adrenal iron-sulfur protein (adrenodoxin). *Biochemistry* **12**, 406–409.
- 53 Chu JW & Kimura T (1973) Studies on adrenal steroid hydroxylases. Molecular and catalytic properties of adrenodoxin reductase (a flavoprotein). *J Biol Chem* **248**, 2089–2094.
- 54 Vermilion JL & Coon MJ (1978) Purified liver microsomal NADPH-cytochrome P-450 reductase. Spectral characterization of oxidation-reduction states. *J Biol Chem* **253**, 2694–2704.
- 55 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.
- 56 Milhim M, Putkaradze N, Abdulmughni A, Kern F, Hartz P & Bernhardt R (2016) Identification of a new plasmid-encoded cytochrome P450 CYP107DY1 from *Bacillus megaterium* with a catalytic activity towards mevastatin. *J Biotechnol* **240**, 68–75.
- 57 Glatz JF & Veerkamp JH (1983) Removal of fatty acids from serum albumin by Lipidex 1000 chromatography. *J Biochem Biophys Methods* **8**, 57–61.
- 58 Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.
- 59 Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–1214.
- 60 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–422.
- 61 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Cryst* **40**, 658–674.
- 62 Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N *et al.* (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221.
- 63 Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501.
- 64 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.
- 65 Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21.
- 66 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.
- 67 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF chimera – a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612.
- 68 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.
- 69 Toan T, DeLuca HF & Dahl LF (1976) Solid-state conformations of vitamin D3. *J Org Chem* **41**, 3476–3478.

- 70 Stewart JJ (1990) MOPAC: a semiempirical molecular orbital program. *J Comput Aided Mol des* **4**, 1–105.
- 71 Krieger E, Darden T, Nabuurs SB, Finkelstein A & Vriend G (2004) Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins* **57**, 678–683.
- 72 Essmann U, Perera L & Berkowitz ML (1995) A smooth particle mesh Ewald method. *J Chem Phys* **103**, 8577–8593.
- 73 Krieger E & Vriend G (2015) New ways to boost molecular dynamics simulations. *J Comput Chem* **36**, 996–1007.
- 74 Hess B, Bekker H, Berendsen HJC & Fraaije JGEM (1998) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* **18**, 1463–1472.
- 75 Miyamoto S & Kollman PE (1992) Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* **13**, 952–962.
- 76 Jakalian A, Jack DB & Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **23**, 1623–1641.

Supporting information

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

Fig. S1. Ligand–protein interactions in the CYP109A2 crystal structures.

Fig. S2. CYP109A2-steroid interactions identified in reactive enzyme-substrate complexes from MD simulations.