The crystal structure of P450-TT heme-domain provides the first structural insights into the versatile class VII P450s

DOI: 10.1016/j.bbrc.2018.05.014

Citation for published version (APA):

Published in: Biochemical and Biophysical Research Communications

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
The crystal structure of P450-TT heme-domain provides the first structural insights into the versatile class VII P450s

Michele Tavanti a, Joanne L. Porter a, Colin W. Levy a, J. Rubén Gómez Castellanos b, 1, Sabine L. Flitsch a, **, Nicholas J. Turner a, *

a Manchester Institute of Biotechnology (MIB), School of Chemistry, The University of Manchester, 131 Princess Street, M1 7DN, Manchester, United Kingdom
b Department of Biology and Biotechnology “Lazzaro Spallanzani”, University of Pavia, Via Ferrata 9, 27100, Pavia, Italy

ARTICLE INFO

Article history:
Received 27 April 2018
Accepted 2 May 2018
Available online 17 May 2018

Keywords:
CYP
Biocatalysis
C-H activation
Crystallography
Hydroxylation

ABSTRACT

The first crystal structure of a class VII P450, CYP116B46 from Tepidiphilus thermophilus, has been solved at 1.9 Å resolution. The structure reveals overall conservation of the P450-fold and a water conduit around the I-helix. Active site residues have been identified and sequence comparisons have been made with other class VII enzymes. A structure similarity search demonstrated that the P450-TT structure is similar to enzymes capable of oxy-functionalization of fatty acids, terpenes, macrolides, steroids and statins. The insight gained from solving this structure will provide a guideline for future engineering and modelling studies on this catalytically promiscuous class of enzymes.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Cytochrome P450 monoxygenases (P450s or CYPs) are a widespread class of heme-containing enzymes capable of catalyzing a diverse array of oxidation reactions with important roles in biosynthetic pathways and xenobiotic metabolism [1]. Since these reactions often occur with high regio- and stereoselectivity, P450s have recently been targeted for synthetic applications [2]. The most common and synthetically attractive reaction catalyzed by these enzymes is the monooxygenation of unactivated C-H bonds under mild conditions using molecular oxygen as the sole oxidant. In order to carry out this reaction, two NAD(P)H-derived electrons are shuttled by redox partners for reductive activation of dioxygen. Two protonation steps are necessary to generate the ferryl-oxo compound I which represents the catalyst for oxygen insertion [3].

Cytochrome P450s can be classified according to the electron transport system exploited to complete their catalytic cycle [4]. Most of these P450 systems require separate redox proteins to function. However, P450 classes have been discovered in which the redox partners are fused to the heme domain of the enzyme, creating multicomponent, self-sufficient systems. These are attractive biocatalysts, as the need for identification and expression of separate redox partners is obviated.

Structural data available for P450 enzymes led to the conclusion that an overall-P450 fold is conserved and large conformational rearrangements are sometimes observed upon ligand binding [5]. Even though structural-based standard numbering systems have certainly improved the quality of P450 sequence alignments [6], the process of accurately converting sequence information into a 3D structure is complicated by the low sequence identity among P450s belonging to different families (10–30%) and by the diversity within the active site [7].

Class VII P450s represent a unique class of fused enzymes in which electrons are transferred by a phthalate family oxygenase (PFOR)-like reductase domain containing a FMN and a 2Fe–2S cluster [8]. These self-sufficient enzymes have recently received attention for their remarkable substrate promiscuity, with further efforts made to engineer them through directed evolution for higher dealkylation activity, to improve the catalytic performance in the hydroxylation of tetralin derivatives or to carry out anti-Markovnikov oxidation of styrenes [9–11]. Moreover, benefitting from the rapid expansion of available CYP sequences, we have recently reported the discovery of a panel of new class VII P450s...
with desirable biocatalytic properties such as good expression levels, thermal stability and diverse substrate scope [12,13]. In particular, CYP116B46 from T. thermophilus displayed not only better expression levels but also enhanced thermal stability when compared to the previously characterized CYP116B2 (P450-RhF).

Since the discovery of the first class VII P450-RhF in 2002, several more have been reported and thoroughly characterized biochemically [14–19]. However, as yet no physiological role has been determined for these enzymes with the exception of only CYP116B5 [18]. Moreover, X-ray crystal structures for the heme domain of these proteins are not available to date. In order to provide a link between the sequence and the 3D structure of this valuable class of enzymes, we have solved the first X-ray structure of the heme domain of a class VII P450. As thermally stable enzymes tend to generate protein crystals better suited for X-ray diffraction [20], we focused on CYP116B46 (UniProt A0A0K6ITW2) to get a first insight into the structure of this class of self-sufficient enzymes.

2. Material and methods

For full experimental details please refer to the Supporting Information.

2.1. Protein expression and purification

Protein expression was carried out as reported previously [12]. For protein purification, cells were resuspended in 50 mM HEPES buffer, 300 mM NaCl, 20 mM imidazole, pH 8.0 (buffer A) to a wet cell concentration of 200 mg/mL. After ultrasonication and ultra-centrifugation, immobilized metal ion affinity chromatography (IMAC) was employed as a first purification step using Ni-agarose resin (Qiagen) in a gravity flow column following the manufacturer instructions. Anion-exchange chromatography was carried out using a 6-ml RESOURCE Q column (GE Healthcare) on a AKTA Pure system (GE Healthcare). Size-exclusion chromatography was performed in 50 mM HEPES, 300 mM NaCl, pH 8 on a AKTA Pure system equipped with a HiLoad 16/600 Superdex 200 pg column. Enzyme elution was monitored at 280 nm and red fractions displaying at least 50% maximal absorbance of the main peak at 280 nm were analyzed by SDS-PAGE (Fig. S1), collected, concentrated and desalted in 50 mM HEPES pH 8 using a PD-10 column. Prior to crystallization trials, protein solutions were concentrated to 20 mg/mL, as determined by NanoDrop (Thermo Scientific), using ε280 = 67755 M⁻¹ cm⁻¹ and stored overnight at 4 °C.

2.2. Crystallization, data collection and structure determination

Crystals of CYP116B46 (P450-TT) were obtained by mixing 200 mL of protein with 200 mL of condition A12 (0.1 M MES 6.5 22% v/v PEG Smear Broad) from the BCS screen (Molecular Dimensions Ltd., Newmarket, UK). Drops were incubated at 4 °C for 24 h prior to inspection. Single crystals suitable for data collection were cryo protected with the addition of 15% PEG 200 prior to plunge cooling in liquid nitrogen.

Data were collected from a single cryo protected crystal of P450-TT at beamline I04-1 (Diamond Light Source). All data were indexed, scaled and subsequently integrated with Xia2. Structure determination was initially performed by molecular replacement in Phaser using a search model derived from the previously solved cytochrome P450 structure (5GWE). A combination of automated and manual rebuilding and refinement in Phenix and COOT were used to produce a complete model. Validation with both Molprobity and PDB_REDO were integrated into the iterative rebuild process. Complete data collection and refinement statistics are available in Table S1. The atomic coordinates and structure factors of P450-TT structure have been deposited in the RCSB Protein Data Bank under the accession code 6GII.

2.3. Sequence analysis

Sequence alignments were performed with ClustalW [21]. The output files were then used to run ESPrint [22] (for the characterized class VII P450s). Conservation analyses were carried out using the 3D™ system [subfamily K9GPR5, https://www.bio-prodikt. j [23]. The 3D™ system provides multiple sequence alignments of protein superfamilies based on multiple structure alignments, thereby generating a 3D numbering scheme (3DM standard numbering). When a 3D standard numbering could not be assigned, class VII sequences were aligned using the CYPED database [24]. Amino acid composition of characterized class VII P450 heme domains were computed starting from ClustalW alignments using MEGA7 [25].

2.4. Structural analysis

Structures were visualized and figures generated with PyMOL (the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). The CAVER plug-in in PyMOL [26] was employed using standard parameters and a probe radius of 1.4 Å to identify paths leading to the buried active site. Structural superposition were performed in COOT [27] and the output files used to generate structure-based sequence alignments in Chimera (Match- > Align, residue-residue cutoff distance of 5 Å).

3. Results and discussion

3.1. Overall structure

The substrate-free crystal structure of CYP116B46 (P450-TT) was determined at 1.9 Å resolution from residue 28 to 443 by molecular replacement using the recently solved structure of CYP288A2 (PDB 5GWE) as search model [28].

Overall, P450-TT adopts the triangular P450 fold composed of 19 α-helices and 10 β-strands divided into 3 sheets embedding the heme (Fig. 1A, the structural nomenclature of Hasemann et al. was adopted where possible and secondary structural elements annotated with ESPrint) [29]. Close inspection of the protein surface does not reveal an open tunnel leading to the cofactor. Analysis of the structure using the CAVER 3.0 tool enabled identification of a putative tunnel –14 Å long and with a bottleneck of ~1.7 Å (Fig. 1B). This tunnel is mainly lined by hydrophobic residues (Table S2) but its opening appears too small to accommodate organic molecules (for comparison, methane radius is 1.9 Å) [30]. Therefore, opening motions might be needed to accommodate the diverse range of substrates accepted by this class VII P450 [12].

The heme-iron is hexacoordinated, with Cys385 acting as the proximal ligand and water as the sixth ligand (Fig. 1C). The proximal ligand is part of the so-called "cysteine-pocket", which includes Phe378, Gly379, Tyr380, Gly381, Gln384, Leu386, Gly387 and Arg388. The K-helix bears the sequence 307IPXAXEECLR316 in which we can recognize the well described EXXR motif followed by FPRXEECL (the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). The CAVER plug-in in PyMOL [26] was employed using standard parameters and a probe radius of 1.4 Å to identify paths leading to the cofactor. Analysis of the structure using the CAVER 3.0 tool enabled identification of a putative tunnel ~14 Å long and with a bottleneck of ~1.7 Å (Fig. 1B). This tunnel is mainly lined by hydrophobic residues (Table S2) but its opening appears too small to accommodate organic molecules (for comparison, methane radius is 1.9 Å) [30]. Therefore, opening motions might be needed to accommodate the diverse range of substrates accepted by this class VII P450 [12].

The heme propionates are either coordinated to the protein chain directly through charged residues or indirectly through water, allowing the insertion of the heme cofactor in the interior of the molecule (Fig. 1D). In particular, the A-ring propionate interacts with Nε of His125, and Arg129 side chain, while the D-ring propionate hydrogen bonds with the main chain amide group of Tyr380,
the carbonyl group of the same residue and the main chain amide group of His383 through a water molecule, the side chain of Arg326 and finally with Val118, Trp324 and Arg326 through a second water molecule.

As in other P450 structures, the region of the I-helix between Ala274 and Thr279 shows a disruption of the regular α-helix hydrogen-bonding pattern (Fig. S3) [31]. Moreover an extensive chain of water molecules hydrogen bonding with Val273, Ala274, Ala275 (on the “back” of the I-helix) and Glu277 can be observed, the latter being the farthest from the heme and closest in distance to the bulk solvent. A similar water channel connecting the active site to the bulk solvent has been observed in other P450 structures [32].

3.2. Active site

The X-ray structure of P450-TT reveals that the active site is buried within the protein core, with 18 residues organized in what could be described as a tiered system above the heme (Fig. 2) [33]. The six residues of tier 1 (Val118, Ala274, Thr278, Pro320, Val321 and Trp324) are located just above the heme, with Ala274 and Thr278 involved in the I-helix distortion and Trp324 interacting with the D-ring propionate. Four residues form tier 2 (Val91, Thr116, Ala270 and Phe424) and lie in a ring above the tier 1 residues. Ala270 is the only residue located in the I-helix, while the other residues come from neighboring loops. Leu92, Phe206, Val273 and Glu277 comprise tier 3, which includes the conserved acidic residue (Glu277). Finally, tier 4 is entirely composed by residues from the Fα0-helix (His201, Thr202, Asn204, Ala205) and delineate a lid over the active site. Overall, the cone-shaped active site is dominated by apolar and aromatic residues, with charged residues only populating the upper tiers.

Guided by the insights gained from the first class VII P450 crystal structure, we proceeded to analyze local sequence similarities with other class VII P450s. The full heme-domain alignment of characterized class VII P450s can be found in the Supporting Information (Fig. S4). Active site residues and substrate channels are crucial factors determining P450 substrate-selectivity; identification of these residues is important for the purposes of drug design and for tailoring enzymes for biocatalysis. In order to carry out a sequence conservation analysis at putative substrate recognition sites, 96 class VII sequences (subfamily K9GPR5, including
uncharacterized proteins) were analyzed systematically using the 3DM™ system. This type of analysis is particularly useful when consensus mutagenesis is employed for protein engineering [34,35]. As shown in Table S3, amino acids found in P450-TT active site are well conserved throughout class VII P450s (occurrence higher than 87%), the exception being Val321 (70% conserved), which is often substituted by isoleucine. The same systematic analysis was carried out to evaluate the conservation of residues lining the putative substrate access channel. Most of these residues are conserved throughout class VII P450s (Table S2). Notable exceptions are observed between helix-B and helix-C (Phe113, Asn114) and in the I-helix (Ile271 and Ile272).

At this stage, the absence of general trends in active site diversity does not allow to draw general conclusions on substrate scope and product distributions for these P450s.

3.3. Structural comparison to other bacterial P450s

A DALI structural similarity search was carried out in order to gain more functional insights based on previously solved P450 crystal structures [36]. As shown in Table S4, P450-TT showed the highest structural similarity to P450-CreJ (PDB 5XJN, root-mean square deviation (RMSD) = 1.9 Å), an enzyme involved in the p-cresol biodegradation pathway via phosphorylated intermediates in Corynebacterium glutamicum [28]. As discussed in the Introduction, the P450 fold is conserved despite low degrees of sequence identity (30% sequence identity) between CYP families (Fig. S5, A). The structural superposition is poor at both N- and C-termini and most significant deviations in the structural alignment can be observed for βB and neighboring loops and the whole region from αF to the N-terminal region of αG (Fig. S5, B). These regions are generally recognized as substrate recognition sites in P450s [37].

Moreover, critical residues for substrate binding in P450-CreJ are not found in the corresponding position of P450-TT structure (Fig. S6).

Finally, additional similar structures include P450s acting on fatty acids, bulky substrates, such as vitamin D3, steroids, terpenes, statins, cholesterol and macrolides (see Table S4 for further information).

In summary, crystallization and structure determination of CYP116B46 (P450-TT) heme domain revealed that the features associated with the P450-fold are well conserved, including the cysteine-pocket, propionate coordination and the I-helix hydrogen-bonding pattern disruption. A comparative analysis of P450-TT structure demonstrated structural similarity to P450s involved in the oxidation of fatty acids and bulky substrates such as steroids, terpenes, statins, cholesterol and macrolides. As such, future investigations into substrate scope of class VII P450s might be extended to bulkier substrates, such as those accepted by human P450s. Eventually, this could reveal that the versatile framework of these biocatalysts can be exploited to generate enzymes with human-like activities. To conclude, the crystal structure of CYP116B46 provides the first glimpse into the structure of class VII P450s and provides a framework to inform engineering studies and improve our ability to build 3D models for this class of enzymes.

Acknowledgements

The research leading to these results received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613849 (BIOOX) and from the European Union’s Horizon 2020 Programme for research and innovation actions H2020-LEIT BIO-2014-1 under grant agreement no. 635734 (ROBOX). Additionally, we would like to thank Prof. Andrea Mattevi for his precious advice during structure determination and Mrs. Michiyo Sakuma for invaluable technical support. We thank Diamond Light Source for access to beamline i04-1 (MX17773-4) that contributed to the results presented here.

Any statement made herein reflects only the views of the authors. The European Union is not liable for any use that may be made of the information contained herein.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.bbrc.2018.05.014.

Transparency document

Transparency document related to this article can be found online at https://doi.org/10.1016/j.bbrc.2018.05.014.

References


[25] V.A.P. Martins Dos Santos, P.J. Schaap, 3DM: systematic analysis of hetero- 

