The crystal structure of the ring-hydroxylating dioxygenase from Sphingomonas CHY-1

Jean Jakoncic, Yves Jouanneau, Christine Meyer, V. Stojanoff

The crystal structure of the ring-hydroxylating dioxygenase from *Sphingomonas* CHY-1

Jean Jakoncic[1], Yves Jouanneau[2], Christine Meyer[2], Vivian Stojanoff[1]

[1]Brookhaven National Laboratory, National Synchrotron Light Source, Upton, NY 11973, USA. [2]Laboratoire de Biochimie et Biophysique des Systèmes Intégrés, CEA, DSV, DRDC and CNRS UMR 5092, CEA-Grenoble, F-38054 Grenoble Cedex 9, France.

running title: Structure of the *Sphingomonas* CHY-1 dioxygenase

subdivision: Structural Biology

Corresponding author:

Vivian Stojanoff

Brookhaven National Laboratory

Upton NY 11973 US

Tel. : 1 631 344 8375; Fax : 1 631 344 3238

Email : vivian.stojanoff@gmail.gov

Abbreviations: ring-hydroxylating dioxygenase (RHD), polycyclic aromatic hydrocarbons (PAHs)

Keywords: crystal structure, Rieske non-heme iron oxygenase, heavy molecular weight polycyclic aromatic hydrocarbons, bioremediation.

Summary

The ring-hydroxylating dioxygenase (RHD) from *Sphingomonas* CHY-I is remarkable due to its ability to initiate the oxidation of a wide range of polycyclic aromatic hydrocarbons (PAHs), including PAHs containing 4 and 5 fused rings, known pollutants for their toxic nature. Although the terminal oxygenase from CHY-1 exhibits limited sequence similarity with well characterized RHDs from the naphthalene dioxygenase family, the crystal structure determined to 1.85 Å by molecular replacement revealed the enzyme to share the same global $\alpha_3\beta_3$ structural pattern. The catalytic domain distinguishes itself from other bacterial non-heme Rieske iron oxygenases by a substantially larger hydrophobic substrate binding pocket, the largest ever reported for this type of enzyme. While residues in the proximal region close to the mononuclear Fe atom are conserved, the central region of the catalytic pocket is shaped mainly by the side chains of three amino acids, Phe 350, Phe 404 and Leu 356, which contribute to the rather uniform trapezoidal shape of the pocket. Two flexible loops, LI and LII, exposed to the solvent seem to control the substrate access to the catalytic pocket and control the pocket length. Compared to other naphthalene dioxygenases residues Leu 223 and Leu 226, on loop LI, are moved towards the solvent, thus elongating the catalytic pocket by at least 2 Å. An 11 Å long water channel extends from the interface between the α and β subunits to the catalytic site. The comparison of these structure with other known oxygenases suggests that the broad substrate specificity presented by the CHY-1 oxygenase is primarily due to the large size and particular topology of its catalytic pocket and provided the basis for the study of its reaction mechanism.

Polycyclic Aromatic Hydrocarbons (PAHs) are considered major environmental pollutants due to their cytotoxic, mutagenic or carcinogenic character. High molecular weight PAHs containing four or more fused benzene rings, are of particular concern as they are more resistant to biodegradation by microorganisms. Several bacteria, algae and fungi able to degrade PAHs have been described [1,2] but only a few have been shown to mineralize 4 and 5 ring PAHs [3,4,5,6,7]. Recently, a *Sphingomonas* strain was isolated for its ability to grow on chrysene [7]. In this strain, a single ring-hydroxylating dioxygenase (RHD) was found to catalyze the oxidation of a broad range of PAHs [8,9]. The dioxygenase has been purified and characterized as a three-component enzyme consisting of a NAD(P)H-dependent reductase, a [2Fe-2S] ferredoxin, and a terminal oxygenase, PhnI. This dioxygenase exhibited unique substrate specificity, as it could oxidize half of the 16 PAHs considered to be major pollutants by the US Environmental Protection Agency. Remarkably, the enzyme was found to be active on the 4-ring chrysene and benz[a]anthracene, and on the 5-ring benzo[a]pyrene, whereas none of the RHDs isolated so far were able to attack these high molecular weight PAHs. Sequence comparison of the oxygenase components of well characterized RHDs (Fig. 1) indicated that PhnI is most closely related to enzymes described as naphthalene dioxygenases [10].

To date the structures of seven RHD terminal oxygenases have been reported, including that of the naphthalene dioxygenases from *Pseudomonas sp.* strain NCIB9816-4 (NDO-$O_{9816\text{-}4}$) [11,12,13] and *Rhodococcus sp.* strain NCIMB12038 (NDO-O_{12038}) [14], the nitrobenzene dioxygenase from *Comamonas sp.* strain JS765 (NBDO-O_{JS765}) [15], the biphenyl dioxygenase from *Rhodococcus sp.* strain RHA1 (BPDO-O_{RHA1}) [16], the cumene dioxygenase from *P. fluorescens* strain IP01 (CDO-O_{IP01}) [17], the 2-oxoquinoline 8-monooxygenase from *P. putida* strain 86 (OMO-O_{86}) [18] and the carbazole-1-9 α-dioxygenase from *P. resinovorans* strain CA10 (CARDO-O_{CA10}) [19]. Except for OMO-O_{86} and CARDO-O_{CA10}, which were found to be homotrimers consisting of α-subunits only, all other enzymes exhibited a $\alpha_3\beta_3$ quaternary structure. The α subunit contains a hydrophobic pocket with a mononuclear Fe(II) center that

serves as substrate binding site. As found for all dioxygenases the iron atom is coordinated by a conserved 2-His-1-carboxylate triad [20], and is located about 12 Å from the [2Fe-2S] Rieske cluster of the adjacent α subunit.

Here we report the crystal structure of the terminal oxygenase component from *Sphingomonas sp.* strain CHY1, PhnI, in a substrate-free form. This is the first crystal structure of a terminal oxygenase that can catalyze the oxidation of a broad range of PAHs including four and five ring PAHs. Based on this structure it is inferred that the broad specificity of this RHD is due to the large size and specific topology of its hydrophobic substrate-binding pocket.

Results and Discussion

Overall Structure

The PhnI crystal structure was determined by molecular replacement using the α subunit structure from naphthalene dioxygenase NDO-$O_{9816\text{-}4}$ [11] and the β subunit from cumene dioxygenase CDO-O_{IP01} [17] as search model. The crystallographic model determined to 1.85 Å resolution was refined to yield an R-factor of 19.7 % and R_{free}-factor of 23.6 % (5 % of the reflections were used for the cross validation calculation), shown in Table 1. Consistent with biochemical analysis [9], the PhnI crystal structure can be described by an $\alpha_3\beta_3$ type heterohexamer (Fig. 2) with a 454 amino acid long α subunit and a 174 amino acid long β subunit[1]. In addition to the six polypeptidic chains, the final model contained three mononuclear iron atoms, three [2Fe-2S] Rieske clusters and 1096 water molecules. The electron density for one of the α subunits (chain A) was considerably better than that found for the other two subunits (chains C and E) while the electron density for the three β subunits (chains B, D, and F) was found to be equivalent. Residues located in flexible regions of the protein where no electron density was observed were not included in the final model. These residues include the four initial amino acids of all three β subunits, the C-termini of the α subunits, and loop regions

[1] Residues in different subunits will be designated as, aaa_u ijk, where u stands for the α or β subunit, aaa is the three letter residue denomination and ijk is the residue number.

located in the vicinity of the catalytic site. Five water molecules were found to be in direct contact with the catalytic Fe atoms. Over 88.8% of the residues were found in the most favorable region of the Ramachandran plot; all of the eleven outliers were located on β-turns in the α subunits and present well defined electron density except for Leu_{α} 238.

Like other members of the naphthalene dioxygenase family, PhnI presents a mushroom-like shape [11], 75 Å in height, with the three α subunits forming the cap (100 Å in diameter) and the three β subunits forming the stem (50 Å in diameter). Each αβ heterodimer is related to the other by a non-crystallographic three-fold symmetry axis (Fig. 2). No significant structural differences were observed between the three αβ heterodimers (average rmsd: 0.26 Å), Fig. 3. The overall <*B*> factor was slightly higher for chains C (32 $Å^2$) and E (34 $Å^2$) than for chain A (22 $Å^2$), indicating a higher dynamical disorder, and about the same for the three β subunits (25 $Å^2$). Overall, the crystal structure of PhnI is very similar to that of other RHDs (Fig. 4); the αβ heterodimers rmsd between alpha carbon chains being 1.2 Å between PhnI and NDO-$O_{9816\text{-}4}$ and 1.5 Å between PhnI and BPDO-O_{RHA1}. The description that follows is based on the structure of the αβ heterodimer formed by chains A and B.

β-subunit

The PhnI β subunit forms a funnel shaped conical cavity that contains in its core a twisted six stranded β-sheet surrounded by four α-helices, a short coil at the N terminal region (residues 5 to 10) and an extended loop (residues Pro_{β} 49 to Ala_{β} 69). The C-terminal coil and the third and fourth α-helices (ba3, ba4)[2] form the 20 Å entrance to the funnel. Together with the extended loop, which extends 20 Å from the center of the funnel, they form the base of the β subunit (Fig. 3). The last four residues in the C-terminal coil (residues 171 - 174) are deeply anchored inside the core of the conical shaped funnel by a

[2] secondary structure nomenclature: uvxi, where u=a,b stands for α or β subunit, v=r,c represents the Rieske or the catalytic domain of the α subunit and is absent when the structure is related to the β subunit, x=a,b stands for α-helix or β-strand, i=1,2,3,etc. represents the order following the sequence

hydrogen bond network with strictly conserved arginine residues among RHDs (residues 126, 140 and 156 in PhnI). Residues in the core region, mostly those located in the β−sheet, are mainly involved in interactions between neighboring β subunits, while the α−helices are located mostly on the outer part of the stem in contact with the solvent.

In spite of low amino acid sequence identity between the β subunits of related RHDs, the PhnI β subunit shares the global structural pattern (Fig. 4) with 24 to 35 % identical residues and main chain C_α rmsd ranging between 1.0 and 1.1 Å. The most significant structural difference between RHDs β subunits is observed in the extended loop region. In this region the PhnI secondary structure is closest to the CDO-O_{IPO1} and BPDO-O_{RHA1} structures (Fig. 4). Recently it has been suggested that the β subunit can play different roles in the various RHDs dioxygenases [31].

α-subunit

The α subunit, is composed of two domains: the *Rieske domain* with the [2Fe-2S] cluster (residues 38 to 156) and the *catalytic domain* (residues 1 - 37 and 157 - 454) with the mononuclear iron (Fig. 3).

The Rieske domain

The *Rieske domain* presents essentially the same quaternary structure as other RHDs, with three α-helices (ara 1 to 3) and eleven β-strands (arb 1 to 11). The overall $<B>$ factor for this domain is 22 $Å^2$ except for two flexible and solvent exposed regions for which the *B* factor is above 35 $Å^2$. The first region, located on a β−turn between residues 69 to 71 is totally exposed to the solvent and does not interact with other subunits. The second region located between residues 116 to 134 forms a long coil, LCr, that shields the [2Fe-2S] cluster from the solvent, and interacts with the catalytic domain from the adjacent α subunit (Fig. 3).

The [2Fe-2S] cluster is located at the edge of the Rieske domain between two β-turns which form a gripper-like structure that, with LCr, places the cluster within 12 Å from the catalytic center of the neighboring α subunit (Fig. 2). The cluster presents a distorted lozenge geometry, with planarity ranging from 2.5 to 8.8^o for the three centers.

As for other RHDs the cluster is coordinated by the highly conserved Rieske iron-sulfur motif; Fe1 is coordinated by His_{α} 82 and His_{α} 103, located at the tip of the gripper structure, while Fe2 is coordinated by Cys_{α} 80, located on the β-turn between arb 4 and arb 5, and Cys_{α} 100 in the β−strand, arb7. A far reaching hydrogen network between highly conserved residues surrounds the Rieske cluster and its ligands promoting close interactions with the mononuclear Fe in the catalytic domain of the adjacent α subunit.

The catalytic domain

The catalytic domain is composed of sixteen α-helices and eleven β-strands (Fig. 1). The core region is dominated by a nine-stranded anti-parallel β-sheet in the center of the domain with the active site of the enzyme on one side and the Rieske center on the other side of the sheet (Fig. 3). Covering one side of the sheet are two consecutive helices, aca 10 and 11 (residues 336 to 350 and 356 to 373), which are highly conserved among RHD structures. Strategically located in the vicinity of the catalytic Fe, aca 11 contains residues 356 to 360 and carries the totally conserved amino acids Gly_{α} 354, Glu_{α} 357, Asp_{α} 359 and Asn_{α} 363 which are part of a far reaching hydrogen network surrounding the catalytic center, as well as Asp_{α} 360, one of the three ligands of the catalytic Fe atom.

Fully exposed to solvent, the C terminal region of the catalytic domain, residues 426 to 452, containing α-helices, aca13 and aca14, cover the cap of the catalytic domain (Fig. 3). Compared to other RHDs, the C terminus is shown to be different in length and amino acid sequence (Fig. 1). In fact the C terminal region is quite different from RHDs of know crystallographic structure and therefore is not expected to present any function other than structural.

A large depression, about 20 Å wide, on the surface of the catalytic domain receives the Rieske domain from the adjacent α subunit placing the [2Fe-2S] center in the right conformation with respect to the catalytic Fe. Helix ara 2 and the long coil, LCr, anchor the Rieske domain to the adjacent catalytic domain between loops acb9 and acb10, acb11 and aca13, and to loop LI (residues 221 to 228).

A 35 Å long cavity extending from the solvent to the anti-parallel β-sheet contains the substrate binding pocket. With its 12 x 8 x 6 $Å^3$ the PhnI catalytic pocket is

approximately 2 Å longer and the largest reported so far for a RHD. Mostly formed by hydrophobic amino acids the pocket is surrounded by two loops exposed to the solvent, LI (residues 221 to 238) and LII (258 to 265), α-helix, aca6, residues 206 to 220, containing two of the mononuclear Fe ligands (His_{α} 207 and His_{α} 212) and helices, aca 10 and aca11, which include Asp_{α} 360, the third iron ligand. Providing access to the catalytic pocket loops LI and LII are not completely represented in the final model. As shown in the Fig. 5, loop LII, assumes three different conformations, one for each of the three α subunits. LI on the other hand could only be partially modeled for one of the three α subunits, the high flexibility of the loop precluded modeling for the two other chains.

Interdomain interactions

The $\alpha_3\beta_3$ hexamer is maintained by multiple interdomain interactions found in αα, ββ and αβ interfaces. Within the same αβ heterodimer, strong interactions give rise to a complex and extended hydrogen network between residues located at the base of the β subunit and the Rieske and catalytic domains of the α subunit. In the heterohexamer, the Rieske domain interacts with the base of the adjacent β subunit and the catalytic domain of the adjacent α subunit. Most of the αβ interactions are conserved at least in the dioxygenases from the naphthalene family. For instance, the ionic interaction between Asp_{α} 91 and Arg_{β} 163 within one αβ subunit is highly conserved. Another example Trp_{β} 91 at the base of the β subunit (helix ba4) interacts with Trp_{α} 210 (helix aca6) from the α subunit catalytic domain and with Asn_{α} 101, located on the gripper structure from the adjacent α subunit Rieske domain. These and additional numerous interactions contribute to the cohesion of the $\alpha_3\beta_3$ hexamer and ultimately favour the catalytic reaction by maintaining the two redox centers at an appropriate distance from each other. If multiple α and β interactions are found in PhnI, the function of the β subunits seem to be purely serve a structural role.

The Mononuclear Fe

The mononuclear Fe is coordinated by a highly conserved 2-His-1-carboxylate motif [10], His_{α} 207, His_{α} 212 and bidentaly by Asp_{α} 360. The Fe coordination geometry can be described as that of a distorted octahedron with the oxygen atom of Asn_{α} 200, at 4 Å, from the mononuclear Fe, occupying the position of a missing ligand. As observed for other dioxygenases [16], while the carboxyl-oxygen OD1 from Asp_{α} 360 is located at 2 Å from the mononuclear iron, the 3 Å coordination distance observed for the Asp_{α} 360 OD2, seems rather large compared to the typical 1.9 Å average distance.

For several dioxygenases the catalytic iron is reported to be coordinated by one or two water molecules. In the refined PhnI structure, the three catalytic Fe atoms were found to be coordinated by at least one water molecule. The crystallographic refinement, showed a large positive difference in the |fo|-Fc| electron density map in two of the three subunits suggesting the existence of an external ligand. The position of this density is similar to that found for the NDO-O_{98164} crystallographic structure [13] and resembles that of an indole molecule. In the third subunit, chain E, the refined distance between the two oxygen atoms, 1.5 Å, suggests the presence of a dioxygen molecule at the catalytic iron site.

The substrate binding pocket

The PhnI catalytic pocket, the largest reported so far for RHDs, is at least 2 Å longer, wider and higher at the entrance when compared to related dioxygenases [32]. The amino acids lining the PhnI pocket are represented in Fig. 6 superposed to the NDO-O_{98146} catalytic pocket. Only small differences can be observed between the two structures in the proximal region, close to the mononuclear Fe atom. In the central region most significant are residues Phe_{α} 350, Phe_{α} 404, Leu_{α} 356, in PhnI. While Phe_{α} 404 is replaced by the smaller residue Ala 407 in NDO-O_{98146}, Leu_{α} 356 is replaced by a bulky aromatic residue (Trp or Phe) in naphthalene dioxygenases. Together these residues and the specific conformations of residues Gly_{α} 205, Val_{α} 208, Thr_{α} 308 contribute to enlarge the PhnI catalytic pocket giving its rather uniform shape without kinks or torsions as found for other dioxygenases.

Probably the distinctive broad substrate specificity presented by the dioxygenase from strain CHY-1 toward PAHs [9] can be mostly ascribed to differences observed in the distal region. Most significant in this region are residues Leu_{α} 223 and Leu_{α} 226 in loop LI, and Ile_{α} 253 and Ile_{α} 260 in loop LII, which most likely control the access to the catalytic pocket.

To further explore the broad specificity of PhnI towards high molecular weight PAHs a benz[a]antracene molecule was overlaid to the PhnI substrate binding pocket. The three most favorable orientations, each of which corresponded to one of the three dihydrodiol isomers obtained by enzymatic conversion of this PAH [9], are shown in Fig. 7. This and several PAHs, known from enzymatic assays to be dihydroxylated by PhnI, could be modeled into the PhnI catalytic pocket minimizing Van der Waals contacts. These results indicate that PhnI can bind large substrates made of 4 or 5 rings with minimal or no rearrangement of side chains [32]. These simulations indicate that amino acids belonging to loops LI and LII, at the entrance of the substrate binding pocket, determine the pocket length, and therefore might play a key role in the substrate selectivity of the enzyme. Similarly these simulations showed that Phe_{α} 350 in the central region of the PhnI catalytic pocket prevents some specific substrate orientations and therefore is thought to participate in the regio-specificity of the enzyme. Site specific mutagenesis of Phe 352 in NDO-O_{98164} was shown to significantly alter the regioselectivity of the enzyme [31].

The Asp 204 electron transfer bridge

Totally conserved amongst RHDs, Asp_{α} 204 is buried in a large depression at the junction of the Rieske domain and the catalytic domain of neighboring α subunit. In this key position, Asp_{α} 204 provides a bridge between the Rieske cluster and the mononuclear iron center (Fig. 8). In PhnI, Asp_{α} 204 side chain is located between His_{α} 207, ligand to the catalytic Fe, and His_{α} 103, ligand to the Rieske center in the adjacent α subunit. Asp_{α} 204 OD2 is 2.7 Å away from His_{α} 103 ND2, and OD1 is 3.3 Å from His_{α} 207 ND1 thus providing a plausible path for intramolecular electron transfer. As part of an extended hydrogen network (Fig. 8) that holds the two redox centers at 12 Å from each other, Asp_{α}

204 OD2 is 3.3 Å away from Tyr_{α} 102 OH (in the adjacent α subunit) and is H-bonded to Tyr_{α} 410 OH (2.8 Å). Asp_{α} 204 OD1 is 3.3 Å from His_{α} 207 ND1, and is H bonded to His_{α} 207 main chain N atom (2.7 Å). Asp_{α} 204 main chains atoms O and N interact with His_{α} 207 ND1 (2.9 Å) and Asn_{α} 200 O (3 Å) atoms respectively. Specific to this network are not only highly conserved amino acid side and main chain interactions but also interactions with a few structural waters. The replacement of this aspartic acid by a Ala, Glu, Gln or Asn in NDO-O_{98164} resulted in a totally inactive enzyme suggesting that it is essential either directly in electron transfer or in positioning the two adjacent α subunits to allow effective for electron transfer [33].

Occurrence of a water channel

An 11 Å long channel filled with eight water molecules extends from the base of the β subunit up to the catalytic site (Fig. 9). The water molecule closest to the catalytic site is at hydrogen bond distance from Glu_{α} 357 and at 4.2 Å from the mononuclear Fe atom. This channel is also found in other RHDs although residues lining the channel are not fully conserved. Only one of the residues at the entrance of the channel is conserved throughout the naphthalene dioxygenase family, Gly_{α} 354. The function of this channel is not well understood. Assuming that water molecules serve as proton source for the catalytic reaction, the channel might be a pathway to convey protons to the active site.

Possible role of Asn_{α} 200

Located in the vicinity of the mononuclear Fe but further buried in the catalytic pocket, Asn_{α} 200 is one of the closest residues to the catalytic Fe (4.0 Å) but not close enough to be a ligand. As Asp_{α} 204, Asn_{α} 200 participates in the extended hydrogen network at the junction of two neighboring α subunits (Fig. 8). Through Tyr_{α} 102, in the adjacent α subunit Rieske domain, Asn_{α} 200 provides a bridge to Cys_{α} 100 one of the Rieske ligands; Asn_{α} 200 ND2 atom is 2.8 Å away from Tyr_{α} 102 OH group while Cys_{α} 100, is hydrogen bonded through main chains to Trp_{α} 105, Gly_{α} 104 and Tyr_{α} 102.

A theoretical analysis predicts that Asn 201 in NDO-O_{98164} would be at hydrogen bond distance from the hydroxyl of the enzyme reaction product during a transition state

[34]. In PhnI the ND2 side chain atom of Asn_{α} 200 is approximately 3 Å away from one of the water molecules bound to the active site. In the catalytic site of BPDO-O_{RHA1} [16], although the asparagine is replaced by a glutamine, a hydrogen bond has also been observed between the side chain atom NE2 and the water molecule present at the active site. Asn (Gln) may assist in the stereo-specific reaction as it may constrain the O in place through hydrogen bonds. The role of Asn 201 in NDO-O_{98164} was tested by substitution of this residue by Gln, Ser or Ala [35]. The enzyme activity was significantly reduced but not totally abolished. It was therefore concluded that Asn 201 is not essential for catalysis, but may be important for maintaining protein-protein interactions between α subunits through its H bond with Tyr 103 (Tyr_{α} 102) in the adjacent subunit.

In conclusion, the PhnI oxygenase is similar in structure to the catalytic component of other RHDs, especially naphthalene dioxygenases. The exceptionally broad substrate specificity of this enzyme, and in particular, its ability oxidize large PAH molecules, may be explained by the large size of its substrate-binding pocket and the flexibility of residues located at the entrance. While residues Phe_{α} 350, Phe_{α} 404 and Leu_{α} 356, shape the pocket and likely influence the regiospecificity of the enzyme, the access to the catalytic site is most probably controlled by residues in loop LI, especially Leu_{α} 223 and Leu_{α} 226. The present structure represents a valuable frame to investigate the role of certain residues on the substrate specificity and/or catalytic activity of the enzyme through site-directed mutagensis.

Experimental Procedures

Purification and crystallization of PhnI

The over expression of recombinant His-tagged PhnI (ht-PhnI) in *P. putida* KT2442 and the purification of the protein were carried out as described by Jouanneau et al. [9]. The oxygenase was further purified by two chromatographic steps under argon. The ht-PhnI preparation was treated with 0.25 U thrombin/mg (Sigma-Aldrich) for 16 h at 20°C in 25

mM Tris-HCl, pH 8.0, containing 0.15 M NaCl, 2.0 mM $CaCl_2$, 0.1 mM $Fe(NH_4)_2(SO_4)_2$ and 5% glycerol, then applied to a small column of TALON affinity chromatography (BD Biosciences, Ozyme, France). The unbound protein fraction was concentrated on a small DEAE-cellulose column, then applied to a 2.6×110 cm column of gel filtration (AcA34, Biosepra) eluted at a flow rate of 50 ml/h with 25 mM Tris-HCl, pH 7.5, containing 0.1 M NaCl, and 5% glycerol. The purified protein was concentrated to about 31 mg/ml, and frozen as pellets in liquid nitrogen.

Search for preliminary crystallization conditions were carried out using the vapor diffusion method in the hanging drop configuration. EasyXtal Cryos Suite (Nextal Biotechnologies, Montreal, Quebec, CA) solution number 67 produced small poorly diffracting crystals within 12 h at 20°C. Upon refining the crystallization conditions, 250 μm long crystals were obtained in less then 8 h in a sitting-drop configuration, by mixing 1 μL of purified PhnI, with 1 μL of mother liquor (11 % PEG8000, 5% Ethanol, 100 mM HEPES pH 7.0, 15 % glycerol, 400 mM $(CH3COO)_2Ca$ and 150 mM NaCl). To improve the diffraction quality, the nucleation and crystal growth process were slowed down by covering each well with 300 μL of mineral oil [21].

Data collection and processing

Diffraction data were recorded at the X6A beam line at the National Synchrotron Light Source (NSLS), Upton, NY [22]. Native crystals directly recovered from the sitting drop, were cooled at 100K in a cold stream of liquid nitrogen. A total of 750 frames (oscillation width 0.2^o) were collected on native crystals. Diffraction data were inspected, indexed, integrated and scaled with the HKL2000 program suite [23]. Data collection and processing statistics are summarized in Table 1.

Structure solution and refinement

The structure of PhnI was solved by molecular replacement (MR) using MOLREP [24] after the failure of several experimental phasing techniques. Based on sequence homology and structural similarity, the search model for the α subunit consisted of the naphthalene dioxygenase $NDO\text{-}O_{9816\text{-}4}$ (PDB access code 1NDO) α subunit while for the

β subunit, the cumene dioxygenase CDO-O_{IP01} (PDB access code 1WQL) β subunit was chosen. For both subunits only main chain atoms were kept, regions presenting high flexibility and high root mean square (RMS) deviations were not considered in the model. Density modification (DM) with non-crystallographic three-fold symmetry (NCS) averaging [25] was applied according to the solvent content determined from Matthews Coefficient probability [26]. The αβ heterodimer presenting the best electron density was completed automatically with ARPwARP [27] and manually with COOT [28]; the two other heterodimers were generated using NCS operators. Restrained refinement was carried out with REFMAC [29]. During the final refinement steps, the Fe and the [2Fe-2S] were refined with no restrains on the geometry and coordination. The final model was analyzed with Procheck [30].

Protein Data Bank accession number

Coordinates and structure factors have been deposited for PhnI in the Protein Data Bank under accession code **2CKF**.

Acknowledgements

We would like to thank the staff of the National Synchrotron Light Source, Brookhaven National Laboratory for their continuous support. This work was supported by grants from the National Institute of Health, NIGMS # GM-0080, U.S. Department of Energy, BES, #DE-AC02-98CH10886, Centre National de la Recherche Scientifique, Commisariat à l'Energie Atomique and Université Joseph Fourier to UMR5092.

References

[1] A. L. Juhasz and R. Naidu, Bioremediation of high molecular weight polycyclic aromatic hydrocarbons: a review of the microbial degradation of benzo[a]pyrene, Int. Biodet.&Biodegr. 45 (2000) 57-88.

[2] R. Kanaly, S. Harayama, Biodegradation of High-Molecular-Weight Polycyclic Aromatic Hydrocarbons by Bacteria, J. Bacteriol. 182 (2000) 2059-2067.

[3] D. T. Gibson, V. Mahadevan, D. M. Jerina, H. Yogi, H. J. Yeh, Oxidation of the carcinogens benzo [a]pyrene and benzo[a]anthracene to dihydrodiols by a bacterium, Science 189 (1975) 295-297.

[4] D. Dean-Ross,C. E. Cerniglia, Degradation of pyrene by *Mycobacterium flavescens*. Appl Microbiol. Biotechnol. 46 (1996) 307-312.

[5] D. R. Boyd, N. D. Sharma, R. Agarwal, S. M. Resnick, M. J. Schocken, D. T. Gibson, J. M. Sayer, H. Yagi, D. M. Jerina, Bacterial dioxygenase-catalysed dihydroxylation and chemical resolution routes to enantiopure cis-dihydrodiols of chrysene. J. Chem. Soc. Perkin Trans. 1 (1997) 1715-1723.

[6] S. Krivobok, S. Kuony, C. Meyer, M. Louwagie, J. C. Willison, Y. Jouanneau, Identification of Pyrene-Induced Proteins in *Mycobacterium* sp. Strain 6PY1: Evidence for Two Ring-Hydroxylating Dioxygenase, J. Bacteriol. 185 (2003) 3828-3841.

[7] J. C. Willison, Isolation and characterization of a novel sphingomonad capable of growth with chrysene as sole carbon and energy source, FEMS Microbiol Lett. 241 (2004) 143-150.

[8] S. Demaneche, C. Meyer, J. Micoud, M. Louwagie, J. C. Willison, Y. Jouanneau, Identification and functional analysis of two aromatic-ring-hydroxylating dioxygenases from a Spingomonas strain that degrades variuous polycyclic aromatic hydrocarbons, App. Environment. Microbiol. 70 (2004) 6714-6725.

[9] Y. Jouanneau, C. Meyer, J. Jakoncic, V. Stojanoff, J. Gaillard, Characterization of a ring-hydroxylating dioxygenase able to oxidize a wide range of polycyclic aromatic hydrocarbons including four and five ring carcinogens, Biochemistry 45 (2006) 12380-12391.

[10] C. Werlen, H. P. E. Kohler, J. R. van der Meer, The Broad Substrate Chlorobenzene Dioxygenase and cis-Chlorobenzene Dihydrodiol Dehydrogenase of *Pseudomonas* sp. Strain P51 Are Linked Evolutionarily to the Enzymes for Benzene and Toluene Degradation, J. Biol. Chem. 271 (1996) 4009-4016.

[11] B. Kauppi, K. Lee, E. Carredano, R. E. Parales, D. T. Gibson, H. Eklund, S.

Ramaswamy, Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase, Structure 6 (1998) 571–586.

[12] E. Carredano, A. Karlsson, B. Kauppi, D. Choudhury, R. E. Parales, J. V. Parales, K. Lee, D. T. Gibson, H. Eklund, S. Ramaswamy, Substrate binding site of naphthalene 1,2-dioxygenase: functional implications of indole binding, J. Mol. Biol. 296 (2000) 701–712.

[13] A. Karlsson, J. V. Parales, R. E. Parales, D. T. Gibson, H. Eklund, S. Ramaswamy, Crystal Structure of Naphthalene Dioxygenase: Side-on binding of Dioxygen to Iron, Science 299 (2003) 1039-1042.

[14] L. Gakhar, Z. A. Malik, C. C. Allen, D. A. Lipscomb, M. J. Larkin, S. Ramaswamy, Structure and Increased Thermostability of Rhodococcus sp. Naphthalene 1,2-Dioxygenase, J.Bacteriol. 187 (2005) 7222-7231.

[15] R. Friemann, M.M. Ivkovic-Jensen, D.J. Lessner, C.L. Yu, D.T. Gibson, R.E. Parales, H. Eklund, S. Ramaswamy, Structural insight into the dioxygenation of nitroarene compounds: the crystal structure of nitrobenzene dioxygenase, J. Mol. Biol. 348 (2005) 1139–1151.

[16] Y. Furusawa, V. Nagarajan, M. Tanokura, E. Masai, M. Fukuda, T. Senda, Crystal structure of the terminal oxygenase component of biphenyl dioxygenase derived from *Rhodococcus* sp. strain RHA1, J. Mol. Biol. 342 (2004) 1041–1052.

[17] X. Dong, S. Fushinobu, E. Fukuda, T. Terada, S. Nakamura, K. Shimizu, H. Nojiri, T. Omori, H. Shoun, T. Wakagi, Crystal structure of the terminal oxygenase component of cumene dioxygenase from Pseudomonas fluorescens IP01, J. Bacteriol. 187 (2005) 2483–2490.

[18] B. M. Martins, T. Svetlitchnaia, H. Dobbek, 2-Oxoquinoline 8-monooxygenase oxygenase component: active site modulation by Rieske-[2Fe–2S] center oxidation/reduction, Structure 13 (2005) 817–824.

[19] H. Nojiri, Y. Ashikawa, H. Noguchi, J. W. Nam, M. Urata, Z. Fujimoto, H. Uchimura, T. Terada, S. Nakamura, K. Shimizu, T. Yoshida, H. Habe, T. Omori, Structure of the terminal oxygenase component of angular dioxygenase, carbazole 1,9a-dioxygenase, J. Mol. Biol. 351 (2005) 355-370.

[20] J. L. Que, R. Y. N. Ho, Dioxygen activation by enzymes with mononuclear non-heme iron active sites, Chem. Rev. 96 (1996) 2607-2624.

[21] N. E. Chayen, Comparative Studies of Protein Crystallization by Vapour-Diffusion and Microbatch Techniques, Acta Cryst. D54 (1998) 8-15.

[22] M. Allaire, M. Aslantas, A. Berntson, L. Berman, S. Cheung, B. Clay, R. Greene, J. Jakoncic, E. Johnson, C. C. Kao, A. Lenhard, S. Pjerov, D. P. Siddons, W. Stober, V. Venkatagiriyappa, Z. Yin, V. Stojanoff, A Modular Approach to Beam Line Automation: The NIGMS Facility at the NSLS, Synchr. Rad. News 16 (2003) 20-25.

[23] Z. Otwinowski, W. Minor, Processing of X-ray Diffraction Data Collected in Oscillation Mode, Methods in Enzymology 276 (1997) 307-326.

[24] A. Vagin, A. Teplyakov, MOLREP: an automated program for molecular replacement, J. Appl. Cryst. 30 (1997) 1022-1025.

[25] K. Cowtan, 'dm': An automated procedure for phase improvement by density modification, Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 31 (1994) 34-38.

[26] B. W. Mathews, Solvent Content of Protein Crystals, J. Mol. Biol. 33 (1968) 491-497.

[27] A. Perrakis, R. M. Morris, V.S. Lamzin, Automated protein model building combined with iterative structure refinement, Nature Struct. Biol. 6 (1999) 458-463

[28] P. Emsley, K. Cowtan, Coot: Model-Building Tools for Molecular Graphics, Acta. Cryst. D60 (2004) 2126-2132.

[29] G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of Macromolecular Structures by the Maximum-Likelihood Method, Acta Cryst. D53 (1997) 240-255.

[30] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. App. Cryst. 26 (1993) 283-291.

[31] R. E. Parales. The role of active site residues in naphthalene dioxygenase. J. Ind. Microbiol. Biotechnol. 30 (2003) 271-278.

[32] J. Jakoncic, Y. Jouanneau, C. Meyer, V.Stojanoff, The catalytic pocket of the ring-hydroxylating dioxygenase from *Sphingomonas* CHY-1. Biochem. Biophys. Res. Commun. (2007) 352, 861-868

[33] R. E. Parales, J. V. Parales, D. T. Gibson, Aspartate 205 in the catalytic domain of naphthalene dioxygenase is essential for activity, J. Bacteriol. 181 (1999) 1831-1837.

[34] A. Bassan, M. R. Blomberg, P. E. Siegbahn, A theoretical study of the cis-dihydroxylation mechanism in naphthalene 1,2-dioxygenase, J. Biol. Inorg. Chem. 9 (2004) 439-52.

[35] R. E. Parales, L. Lee, S. M. Resnick, H. Jiang, D. J. Lessner, D. T. Gibson, Substrate Specificity of Naphthalene Dioxygenase: Effect of Specific Amino Acids at the Active Site of the Enzyme, J. Bacteriol. 182 (2000) 1641-1649.

[36] D. Higgins, J. Thompson, T. Gibson, J. D. Thompson, D. G. Higgins, T. J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673-4680.

[37] P. J. Kraulis, MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures, J. Appl. Cryst. 24 (1991) 946-950.

[38] E. A. Merritt, M. E. Murphy, Raster3D Version 2.0: A Program for Photorealistic Molecular Graphics, Acta Cryst. D50 (1994) 869-873.

[39] W. L. DeLano, DeLano Scientific, San Carlos, CA, USA, 2002.

Figure Legends

Fig. 1. Sequence alignment of selected ring hydroxylating dioxygenases. (A) α subunit and (B) β subunit from PhnI (phn1), NDO-$O_{9816\text{-}4}$ (ndo), CDO-O_{IP01} (cudo), BPDO-O_{RHA1} (bpdo) and NBDO-O_{JS765} (nbdo). The PhnI α subunit was found to be 40, 31, 34 and 40 % identical to ndo, cudo, bpdo and nbdo, while for the β subunit the identity was found to be lower, 24, 35, 32 and 31 % respectively. Highly conserved residues are boxed and shown against a red background; boxed residues shown against a yellow background are not totally conserved. The numbering given above the sequence refers to PhnI. Secondary structural elements are indicated above the alignment. The figure was generated with CLUSTALW [36].

Fig. 2. Crystal structure of PhnI. Ribbon representation of the PhnI $\alpha_3\beta_3$ hexamer along the three fold symmetry axis (A) and perpendicular to this axis (B). The three αβ units are colored in red, green and blue; the β subunits are represented in lighter tones. Iron atoms are shown in yellow and sulfur atoms in green. The figures were drawn using the programs MOLSCRIPT [37] and Raster 3D [38].

Fig. 3. The PhnI αβ heterodimer. Ribbon representation of the three superposed heterodimers in red, green and blue. The *α subunit,* contains two domains the *Rieske domain* with the [2Fe-2S] cluster (residues 38 to 156) and the *catalytic domain* (residues 1 - 37 and 157 - 454) with the mononuclear iron. Relevant interactions between domains and subunits are shown. The figure was prepared using the programs MOLSCRIPT [37] and Raster 3D [38].

Fig. 4. Superposition of the PhnI αβ heterodimer (chains A and B, grey), with NDO-$O_{9816\text{-}4}$ (blue), CDO-O_{IP01}(red), BPDO-O_{RHA1}(green) and NBDO-O_{JS765} (yellow). (A) αβ heterodimers and (B) catalytic domains. The two solvent exposed loops LI and LII are shown at the entrance of the catalytic pocket, as well as, the highly conserved helices, aca

10 and aca11. The figure was drawn using the programs MOLSCRIPT [37] and Raster 3D [38].

Fig. 5. Surface envelope of the PhnI catalytic pocket. Shown are the three conformations adopted by loop LII at the entrance of the catalytic pocket. Loop LI is shown only for chain A as no density was observed in this region for the two other chains, C and E. Even for chain A, LI is not fully represented, as no density was observed for residues 233 to 236. The figure was made using the program PyMol [39].

Fig. 6. The superposition of the PhnI and NDO-$O_{9816\text{-}4}$ catalytic pocket. The mononuclear Fe ligands are shown in red, PhnI residues in grey and NDO-$O_{9816\text{-}4}$ residues in blue. Residues with similar conformation in both structures are shown in orange. The largest conformational differences are observed for those residues at the entrance of the pocket, Leu_{α} 223, Leu_{α} 226 and Ile_{α} 253. These residues are believed to control the access and the length of the catalytic pocket while residues in the central region, Phe_{α} 350, Leu_{α} 356 and Phe_{α} 404 seem to participate in the regio specificity of the enzyme.

Fig. 7. Superposition of a four ring PAH and the PhnI catalytic pocket. The molecular surface of a benz[a]antracene molecule, represented by a mesh, is overlaid on the substrate binding pocket of PhnI. The three most favorable orientations (A, B and C) shown requiring minimal rearrangement of residues in the catalytic pocket correspond to the three dihydrodiol isomers obtained by enzymatic conversion of this PAH [9].

Fig. 8. Rieske domain and catalytic domain of neigboring α subunits. Ligands to the reaction centers, and residues Asn_{α} 200 and Asp_{α} 204 believed to be involved in the electron transfer to the catalytic site are shown in red. Also shown in red are relevant water molecules in the hydrogen network. In the background the catalytic surface envelope of the PhnI pocket showing the available internal space.

Fig. 9. The PhnI water channel. The channel surface is shown in blue in the foreground and the surface of the catalytic pocket in orange in the back. Structural water molecules are shown in red at the entrance and inside the channel. At the end of the channel a green mesh represents molecule of benzo[a]pyrene a five ring PAH superposed into the catalytic pocket. Partial ribbon diagram of the β subunit, chain B, and α subunit, chain A, are shown in orange and green, respectively. The figure was made using PyMol [39].