# Etude structurale de métalloprotéines à centres [2Fe-2S].Cas d'une ferrédoxine et d'une dioxygénase impliquée dans la biodégradation deshydrocarbures aromatiques.Cristallographie des protéines à très haute énergie.Méthodes de phasage d'une protéine modele à 55 keV.

Jean Jakoncic

# UNIVERSITE JOSEPH FOURIER – GRENOBLE I

# SCIENCES ET GEOGRAPHIE

## THESE

présentée et soutenue publiquement par

**Jean JAKONCIC**

le 22 janvier 2007

Discipline : **Chimie Physique Moléculaire et Structurale**

# Etude structurale de métalloprotéines à centres [2Fe-2S].

Cas d'une ferrédoxine et d'une dioxygénase impliquée dans la biodégradation des hydrocarbures aromatiques.

# Cristallographie des protéines à très haute énergie.

Méthodes de phasage d'une protéine modele à 55 keV.

Composition du jury

| | | |
|---|---|---|
| Président : | M. | S. PEREZ |
| Rapporteurs : | Mme | V. BIOU |
| | M. | B. GUIGLIARELLI |
| Directeurs de thèse : | M. | Y. JOUANNEAU |
| | Mme | V. STOJANOFF |

Thèse preparée au sein du National Synchrotron Light Source, Brookhaven National Laboratory, d'Upton, NY, US et du laboratoire de Biochimie et Biophysique des Systèmes Intégrés CNRS UMR5092, DRDC – Commissariat à l'Energie Atomique de Grenoble, France.

# UNIVERSITE JOSEPH FOURIER – GRENOBLE I

# SCIENCES ET GEOGRAPHIE

## THESE

présentée et soutenue publiquement par

**Jean JAKONCIC**

le 22 janvier 2007

Discipline : **Chimie Physique Moléculaire et Structurale**

# Etude structurale de métalloprotéines à centres [2Fe-2S].

Cas d'une ferrédoxine et d'une dioxygénase impliquée dans la biodégradation des hydrocarbures aromatiques.

# Cristallographie des protéines à très haute énergie.

Méthodes de phasage d'une protéine modele à 55 keV.

Composition du jury

| | | |
|---|---|---|
| Président : | M. | S. PEREZ |
| Rapporteurs : | Mme | V. BIOU |
| | M. | B. GUIGLIARELLI |
| Directeurs de thèse : | M. | Y. JOUANNEAU |
| | Mme | V. STOJANOFF |

Thèse preparée au sein du National Synchrotron Light Source, Brookhaven National Laboratory, d'Upton, NY, US et du laboratoire de Biochimie et Biophysique des Systèmes Intégrés CNRS UMR5092, DRDC – Commissariat à l'Energie Atomique de Grenoble, France.

# Sommaire

# Abréviations

| | |
|---|---|
| 3D | Tridimensionnelle |
| ATP | Adénosine triphosphate |
| BaA | Benzo[a]anthracène |
| BaP | Benzo[a]pyrène |
| BPDO | Biphényle dioxygénase |
| CARDO | Carbazole dioxygénase |
| CUDO | Cumène dioxygénase |
| CYP450 | cytochromes P450 |
| DFT | Théorie de la fonction de densité |
| DO | Dioxygénase |
| Fd | Ferrédoxine |
| Fe-S | Centre à fer- et soufre |
| HAP | Hydrocarbure aromatique polycyclique |
| HEWL | Lysozyme d'œuf de poule (Hen egg white lysozyme) |
| ISC | Centre fer -soufre (Iron Sulfur Cluster) |
| MAD | Diffraction anomale à plusieurs longueurs d'onde |
| MX | Cristallographie des macromolécules |
| NADH | Nicotinamide adénine dinucléotide |
| NADPH | Nicotinamide adénine dinucléotide phosphate |
| NDO | Naphtalène dioxygénase |
| Nif | Fixation de l'azote (nitrogen fixation) |
| OMO | 2-oxoquinoline monooxygénase |
| PCB | Polychlorobiphényle |
| RPE | Resonance paramagnétique éelectronique |
| R-X | Rayons X |
| SAD | Diffraction anomale à une longueur d'onde |
| SIRAS | Remplacement isomorphe avec diffusion anomale |
| SUF | Assimilation du soufre (SUlFur assimilation) |

# Avant propos

Le travail présenté dans cette thèse est principalement consacré à l'étude cristallographique de métalloprotéines contenant du fer, soit sous forme mononucléaire, soit sous forme de centres fer-soufre. Il a été réalisé pour l'essentiel dans le laboratoire de cristallographie Brookhaven, près de la source nationale de rayonnement synchrotron (NSLS) à Upton (USA) et résulte d'une collaboration entre l'équipe de Yves Jouanneau au CEA-Grenoble et celle de Vivian Stojanoff au NSLS.

J'ai tout d'abord étudié la structure d'une ferrédoxine à centre [2Fe-2S] isolée de la bactérie *Rhodobacter capsulatus*. L'enjeu du projet était de mettre en évidence les changements structuraux entre les états oxydé et réduit de la protéine. J'ai obtenu des cristaux de la protéine réduite en conditions anoxiques, puis j'ai établi la structure cristallographique de la ferrédoxine avec une résolution de 2.0 Å. La présentation de la structure de la ferrédoxine FdVI de *R. capsulatus,* dans ses deux états redox a fait l'objet d'un article publié dans la revue *J. Biol. Inorg. Chem.* et présenté dans le chapitre I de cette thèse.

Mon travail a ensuite porté sur une métalloenzyme à 3 composantes de type dioxygénase isolée de la souche bactérienne *Sphingomonas* CHY-I. Cette enzyme catalyse la dioxygénation d'hydrocarbures aromatiques polycycliques (HAP), première étape dans la biodégradation de ces composés organiques toxiques. L'enzyme a été tout d'abord purifiée puis caractérisée au plan biochimique, catalytique et structural. Ce travail est présenté sous la forme de trois articles dans le chapitre II. Le premier article publié dans la revue *Biochemistry* décrit les propriétés des trois composantes de l'enzyme purifiée, et montre les résultats de cinétique enzymatique obtenus avec 9 HAP comportant entre 2 et 5 noyaux aromatiques. Le deuxième article, spublié dans la revue *FEBS Journal* décrit la structure cristallographique de la composante catalytique de l'enzyme obtenue à une résolution de 1.85 Å. Le troisième article publié dans la revue *Biochem. Biophys. Res. Commun.* décrit en detail la structure de la poche catalytique de la dioxygénase de Sphingomonas CHY-1 et présente une modélisation du benzo[a]pyrène au site actif de l'enzyme.

Durant mon séjour au NSLS, j'ai eu la chance de participer à une expérience pilote de cristallographie de protéine à l'aide d'un faisceau de rayons X de très haute énergie, 55 keV; l'objectif était de développer une méthode cristallographique qui réduise l'effet destructeur du rayonnement synchrotron sur les cristaux de protéine. En poursuivant ces travaux, j'ai mis au point une méthode expérimentale permettant de résoudre la structure d'une protéine en faisant

appel à la diffusion anomale à 55 keV. Les résultats obtenus à 55 keV sont comparés aux données collectées sur une ligne de lumière classique de 12 keV. Ces travaux ont fait l'objet de deux articles, le premier qui traite du phasage est publié dans la revue *J. Appl. Cryst.* le second aborde les effets du rayonnement synchrotron à deux énergies. Ces deux articles sont présentés dans le chapitre III.

J'ai aussi été amené à participer au développement de l'installation X6A ainsi qu'à plusieurs projets de recherche qui ont abouti à la résolution de structures de protéines dont trois sont présentées dans la dernière partie de cette thèse. Dans une annexe à cet ouvrage, je présente les résumés des articles auxquels j'ai contribué.

Cette thèse est donc divisée en trois chapitres et comprend une annexe.

# Résumé

## Version française

Les métalloprotéines contenant des centres Fe-S jouent un rôle important dans la nature car elles sont impliquées dans des fonctions physiologiques essentielles telles que la photosynthèse, la respiration et la fixation de l'azote.

Dans cette thèse, une ferrédoxine impliquée dans la biogénèse des centres Fe-S, et une dioxygénase bactérienne jouant un rôle crucial dans la biodégradation des hydrocarbures aromatiques ont fait l'objet d'analyses structurales par cristallographie aux rayons X. La structure d'une ferrédoxine de la bactérie photosynthétique *Rhodobacter capsulatus*, a été résolue dans les états oxydé et réduit. De petits changements structuraux ont été observés lors de la réduction, notamment au voisinage du centre [2Fe-2S]. Ces changements sont comparés à ceux décrits pour des ferrédoxines de structure similaire mais de fonction différente.

Une métalloprotéine plus complexe, appartenant à une grande famille de dioxygénases bactériennes, a été étudiée pour son activité d'oxydation des hydrocarbures aromatiques polycycliques (HAP). Cette enzyme à trois composantes, isolée d'une souche de *Sphingomonas* dégradant les HAP comprend une oxydoréductase à NAD(P)H, une ferrédoxine à centre [2Fe-2S], et composante oxygénase de six sous-unités assemblées en un hexamère de type $\alpha_3\beta_3$. La composante oxygénase, appelée PhnI, contient une centre [2Fe-2S] de type Rieske et un ion ferreux par sous-unité $\alpha$, qui ont été identifiés par leur signature RPE. L'enzyme est douée d'une spécificité du substrat extrêmement large, puisqu'elle est capable d'hydroxyler toute une gamme de HAP fait de 2 à 5 cycles aromatiques, y compris des cancérogènes comme le benz[a]anthracène et le benzo[a]pyrène. Avec le naphtalène comme substrat, des mesures de cinétique ont montré que cette enzyme a un *K*m bas (0.92 $\mu$M) et une constante de spécificité de 2.0 $\mu M^{-1}. s^{-1}$. La protéine Phn1 a été cristallisée, et sa structure 3D a été résolue avec une résolution de 1.85 Å. En dépit d'une modeste similitude de séquence avec des dioxygénases homologues, le repliement polypeptidique est très semblable. Des différences ont toutefois été observées au niveau de la poche catalytique.

Les protéines sous forme cristallisée, notamment les protéines Fe-S, peuvent subir des dommages dus au rayonnement X synchrotron, causant des artéfacts lors de la détermination de la structure. Pour essayer de palier cet inconvénient, des rayons X de très haute énergie (55 keV; 0.22 Å), qui sont peu absorbés par les protéines, ont été utilisés pour résoudre la

structure d'une protéine modèle, le lysozyme. Une structure a été établie pour la première fois par cette approche, en utilisant les phases expérimentales obtenues par différentes méthodes. Les applications potentielles en biologie structurale sont discutées.

**Mots clés :** métalloprotéines, ferrédoxine, centres fer-soufre, dioxygénase, biodégradation des hydrocarbures, rayonnement synchrotron, cristallographie de rayons X

## Version anglaise

Metalloproteins containing Fe-S clusters play an important role in nature as they are involved in essential physiological functions including photosynthesis, respiration, and nitrogen fixation. In this thesis, a [2Fe-2S] ferredoxin involved in Fe-S cluster biogenesis, and a bacterial dioxygenase playing a critical role in aromatic hydrocarbon biodegradation were subjected to structural analysis by synchrotron X-ray crystallography. The structure of a ferredoxin from the photosynthetic bacterium *Rhodobacter capsulatus* was solved in both its oxidized and reduced states. Subtle structural changes were observed upon reduction, especially in the vicinity of the [2Fe-2S] cluster. These changes are discussed in comparison with those described for ferredoxins with similar structures but different functions.

A more complex metalloprotein, belonging to a large family of bacterial dioxygenases, was studied for its ability to oxidize polycyclic aromatic hydrocarbons (PAHs). This multi-component enzyme, isolated from a PAH-degrading *Sphingomonas* strain, consists of a NAD(P)H-oxidoreductase, a [2Fe-2S] ferredoxin, and a terminal oxygenase. The terminal oxygenase component, called PhnI, consists of six subunits assembled into an $\alpha_3\beta_3$ hexamer, and contains one Rieske-type [2Fe-2S] cluster and one Fe(II) ion per $\alpha$ subunit, which were identified by their characteristic EPR signature. The enzyme showed an exceptionally broad substrate specificity, as it could hydroxylate a wide range of PAHs made of two to five fused rings, including the carcinogens, benz[a]anthracene and benzo[a]pyrene. With naphthalene as substrate, steady-state kinetics showed that the enzyme had a low apparent *K*m (0.92 μM) and a specificity constant of 2.0 $\mu M^{-1}$. $s^{-1}$. The Phn1 protein was crystallized and its three-dimensional structure was determined at 1.85 Å resolution. In spite of moderate sequence similarity with homologous dioxygenases, the 3D polypeptide fold was found to be very similar, most of the differences being observed near the substrate binding pocket.

Many protein crystals, especially those of Fe-S proteins, have been shown to undergo X-ray radiation damage, leading to artifacts in protein structure determinations. As an attempt to solve the problem, ultra-high energy X-rays (55 keV; 0.22 Å), which are only slightly absorbed by proteins, were used for the first time to determine the 3D structure of a model protein, lysozyme. Beamline specificities as well as optimum energy were determined. Potential applications for structural biology are discussed.

**Key words**: metalloproteins, ferredoxin, iron-sulfur cluster, dioxygenase, hydrocarbon biodegradation, synchrotron radiation, X-ray crystallography.

# CHAPITRE 1

## La structure de FdVI, la ferrédoxine à centre [2Fe-2S] de la bactérie photosynthétique Rhodobacter capsulatus probablement impliquée dans la synthèse des centres Fe-S

# Partie 1 : Introduction

## 1. Les ferrédoxines

Les ferrédoxines sont des métalloprotéines contenant des atomes de fer et de soufre inorganique organisés sous la forme d'un ou plusieurs groupes prosthétiques. Dans ce cas, le groupe prosthétique est appelé centre à fer-soufre (Fe-S) et est lié à la chaine polypeptidique par des liaisons covalentes. Les ferrédoxines sont de petites protéines solubles ubiquitaires qui, grâce à leur propriété de transfert d'électrons, participent à une grande variété de réactions redox biologiques ; elles interviennent dans des processus essentiels à la vie cellulaire, tels que la fixation de l'azote et la photosynthèse (Matsubara *et al.,* 1980).

Les ferrédoxines peuvent être classées en fonction de la nature du centre Fe-S qu'elles contiennent ; on trouve principalement trois types de centre Fe-S, les centres [2Fe-2S], [3Fe-4F] et [4Fe-4S]. Dans la majorité des cas, chaque atome de fer est coordonné à deux des atomes de soufre inorganique et à deux atomes de soufre provenant de la chaine latérale de cysteines (Fig. 1). Dans le cas particulier des ferrédoxines à centre [2Fe-2S], un des atomes de fer peut être lié par l'atome d'azote de deux histidines, le centre Fe-S étant alors appelé centre [2Fe-2S] de type Rieske ($[2Fe\text{-}2S]^R$). Le centre Rieske est couramment retrouvé dans les dioxygénases microbiennes impliquées dans la dégradation des composés aromatiques, ainsi que dans les chaînes de transport d'électrons associées à la photosynthèse dans les chloroplastes et à la respiration dans les mitochondries.

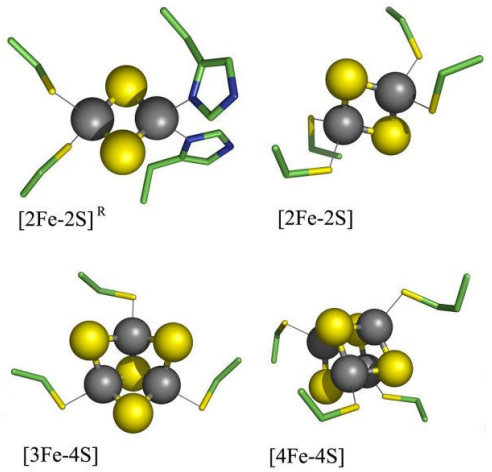



**Figure 1**. Principaux centres Fe-S rencontrés dans les ferédoxines. Les atomes de fer et de soufre sont représentés en gris et en jaune, respectivement.

## 2. Les ferrédoxines de type [2Fe-2S]

Les ferrédoxines (Fd) contenant un centre [2Fe-2S] lié à la chaîne polypeptidique par l'intermédiaire de quatre cystéines appartiennent majoritairement à deux familles, les ferrédoxines de type plante et les ferrédoxines de type vertébré (comme par exemple l'adrénodoxine). Cette classification est basée sur des critères structuraux et des propriétés spectroscopiques distinctives. Le centre Fe-S des Fds de type plante donne un signal rhombique en RPE, alors que celui des Fds de type vertébré est plutôt de symétrie axiale. Les deux familles de Fds se distinguent également par l'écartement des cystéines ligands du centre Fe-S dans la séquence. On retrouve un consensus de type ($CX_4CX_2CX_nC$) chez les Fd de plantes et ($CX_5CX_2CX_nC$) chez les Fds de vertébrés (Müller *et al.,* 1999).

### 2.1. Propriétés structurales

Les ferrédoxines sont des protéines de petit poids moléculaire, environ 12 kDa (une centaine d'acides aminés) constituées de deux domaines : un domaine central où se situe le centre Fe-S et un domaine d'interaction. Bien que les ferrédoxines de type plante et de type vertébré aient relativement peu de similitude de séquence (20 % d'identités tout au plus), le repliement de la chaîne polypeptidique des deux classes de Fds est très semblable. La structure de l'adrénodoxine bovine (code d'accès PDB 1CJE) et celle de *Spirulana platensis* (PDB code 4FXC) sont comparées dans la figure 2 (Pikuleva *et al*., 2000, Fukuyama *et al*., 1995).

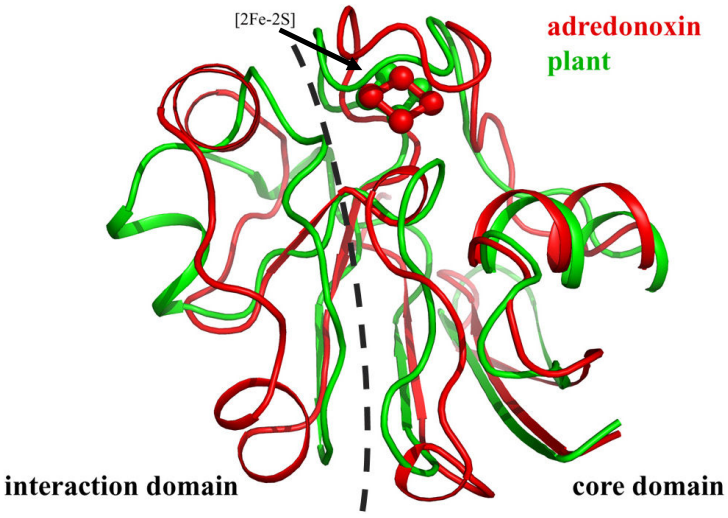



**Figure 2**. Superposition des structures cristallographiques de l'adrénodoxine bovine (en rouge) et de la ferrédoxine de type plante de *Spirulina platensis* (en vert). Les centres [2Fe-2S] sont représentés en mode bâton.

Chez toutes les ferrédoxines à centre [2Fe-2S], un motif structural absolument conservé a été observé, composé d'un feuillet β à 4 brins flanqué d'une hélice α (Müller *et al.,* 1999). Ce motif constitue l'essentiel du domaine central dans lequel se trouve le centre Fe-S, protégé du solvant par une boucle qui le recouvre. L'accessibilité du centre [2Fe-2S] au solvant diffère selon le type de Fd, car chez les ferrédoxines de type plante, il existe un canal contenant des molécules d'eau qui connecte le centre Fe-S au solvant (Morales *et al.*, 1999). Un tel canal n'existe pas dans l'adrénodoxine (Müller *et al*., 1999). Les ferrédoxines ont une charge globalement négative, du fait de la présence d'acides aminés acides en excès dans la séquence. Ces charges négatives ne sont pas uniformément réparties, de sorte que les fds ont souvent un caractère dipolaire marqué. Cette propriété, ainsi que la présence de charges négatives à la surface de la molécule, notamment au niveau du domaine d'interaction, contribuent probablement à la reconnaissance et à l'interaction des Fds avec leurs partenaires physiologiques. Il a été proposé que ces charges permettent d'orienter les Fds avant qu'elles ne s'associent à leur partenaire en un complexe où le transfert des électrons est optimisé (Müller *et al*., 1999).

### 2.2. Fonction

Le centre [2Fe-2S] des ferrédoxines oscille entre deux états d'oxydation dans lesquels les atomes de fer sont soit tous les deux ferriques (état oxydé), soit l'un ferrique, l'autre ferreux (état réduit). Le potentiel de demi-réduction des Fds de type plante est relativement bas (compris entre – 0.3 et – 0.5 V vs NHE). Pour la plupart, ces ferrédoxines servent d'échangeurs d'électrons entre le photosystème I et des oxydoréductases impliquées dans le métabolisme de la plante, comme la Fd-NADP réductase. Dans les cyanobactéries du genre *Anabaena*, des ferrédoxines à centre [2Fe-2S] participent à l'assimilation de l'azote atmosphérique dans des cellules différenciées appelées hétérocystes. La ferrédoxine des hétérocystes sert de donneur d'électrons à la nitrogénase, l'enzyme qui réduit l'azote en ammoniac (Jacobson *et al.,* 1993).

Les Fds de type vertébré ont un potentiel redox voisin de - 0.3 V et fonctionnent souvent comme transporteurs d'électrons entre une NAD(P)H-oxydoréductase et une monooxygénase de type cytochrome P450 (Müller *et al*., 1998). Toutefois, certaines ferrédoxines de cette catégorie seraient impliquées dans la biosynthèse des centres Fe-S, comme par exemple la ferrédoxine Fdx de *Escherichia coli* (Kakuta *et al.*, 2001). La biosynthèse des centres Fe-S, dont on ne connaissait presque rien il y a 15 ans, a connu un fort

regain d'intérêt ces dernières années. Un bref état des connaissances sur cette fonction essentielle, limité au cas des bactéries, sera présenté un peu plus loin dans ce chapitre.

Nous illustrons ci-dessous le rôle physiologique des ferrédoxines à centre [2Fe-2S], par deux exemples représentatifs des deux catégories de protéines considérées ici.

### 2.2.1. Rôle de l'adrénodoxine chez les mammifères

Chez les mammifères, l'adrénodoxine est localisée dans les mitochondries du cortex des glandes surrénales et participe à la synthèse d'hormones stéroïdes (Grinberg *et al.*, 2000). Elle transfère les électrons d'une NADPH-réductase à une enzyme de la famille des cytochromes P450 (CYP450). Les CYP450 sont des enzymes contenant un groupe héminique, qui catalysent des réactions de monohydroxylation très variées. Les CYP450 associés à l'adrénodoxine oxydent des substrats polycycliques, comme le cholestérol en prégnénolone, et servent donc à la biosynthèse des hormones stéroïdes. Le fonctionnement du CYP450 requiert une source d'électrons alimentée par deux protéines spécifiques, une NADPH oxydoréductase et l'adrénodoxine (Fig. 3). Les électrons provenant de l'oxydation du NADPH, et transférés au CYP450 par l'intermédiaire de l'adrénodoxine, permettent l'activation de l'oxygène moléculaire nécessaire à la catalyse. Ce système enzymatique a fait l'objet de nombreuses études et une revue détaillée a fait le point des connaissances sur les aspects moléculaires et structuraux de son fonctionnement (Grinberg *et al.*, 2000).

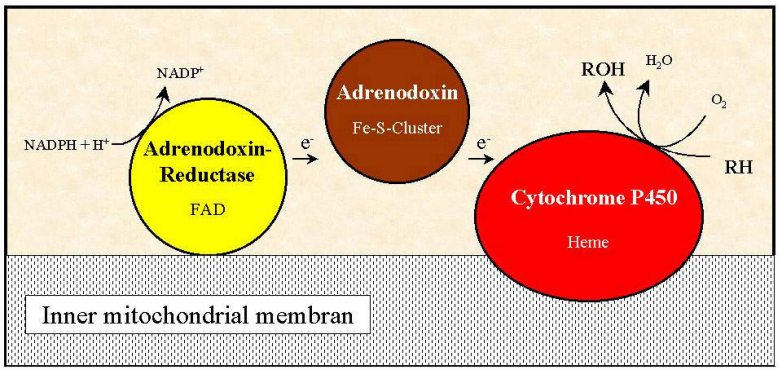



**Figure 3**. Rôle de l'adrénodoxine dans le transfert d'électrons associé au CYP450 dans les mitochondries

Chez les bactéries, il existe des Fds associées à des CYP450, mais dans ce cas, les oxygénases font partie de voies cataboliques permettant la décomposition oxydative de certains substrats carbonés. Chez *Pseudomonas putida*, par exemple, le CYP450cam est une enzyme soluble qui oxyde le camphre et permet aux bactéries d'utiliser ce composé comme substrat de croissance. Le CYP450cam est associé à une réductase et à une ferrédoxine à

centre [2Fe-2S], appelée putidarédoxine dont la structure est voisine de celle de l'adrénodoxine (Sevrioukova *et al*., 2003). Le principe de fonctionnement de ce système enzymatique est comparable à celui décrit ci-dessus pour l'adrénodoxine.

### 2.2.2. Rôle des ferrédoxines dans la photosynthèse

La photosynthèse permet aux plantes et aux cyanobactéries de convertir l'énergie lumineuse en énergie chimique utilisable pour synthétiser le glucose à partir de l'eau et du gaz carbonique. Au cours de ce processus, les ferrédoxines à centre [2Fe-2S] jouent un rôle essentiel en transférant les électrons à bas potentiel produits par le photosystème I (PSI) à un certain nombre d'oxydoréductases, comme la ferrédoxine réductase (FNR) qui régénère le NADPH (Fig.4). La réaction a lieu dans les thylacoïdes des chloroplastes.

Chez les plantes, les Fds participent aussi à la réduction des nitrates, ainsi qu'à la synthèse d'acides aminés (Neuhaus and Emes, 2000).

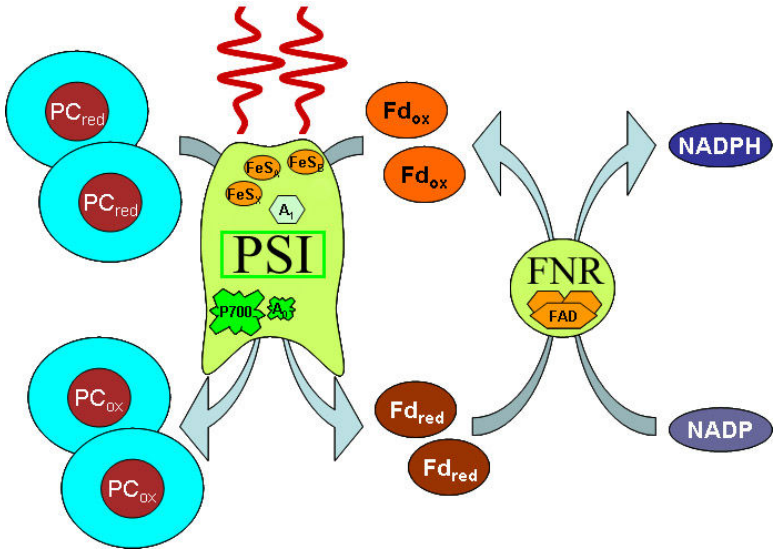



**Figure 4**. Une ferrédoxine à centre [2Fe-2S] assure le transfert d'électrons du photosystème I à la FNR chez les plantes et les cyanobactéries.

## 3. Biosynthèse des centres Fe-S chez les bactéries

Nous l'avons vu, les ferrédoxines à centre [2Fe-2S] sont des protéines essentielles au fonctionnement normal des cellules vivantes. Il en est de même pour de nombreuses enzymes contenant des centres Fe-S, comme par exemple celles qui font partie de la chaîne respiratoire

des mitochondries. *In vitro*, le centre [2Fe-2S] se forme spontanément par incubation de fer ferrique et de soufre inorganique en conditions réductrices avec une apoferrédoxine (Malkin and Rabinowitz, 1966). *In vivo*, la biosynthèse des centres Fe-S dans les protéines fait appel à une machinerie complexe, faisant intervenir plusieurs protéines spécifiques. Compte tenu de la toxicité des ions $Fe^{2+}$ et $S^{2-}$, la synthèse des centres Fe-S in vivo requiert des transporteurs et des protéines spécialisées dans l'assemblage et le transfert du centre dans la chaîne polypeptidique des apoprotéines.

C'est lors de l'étude de la Nitrogénase de la bactérie fixatrice d'azote *Azotobacter vinelandii,* que les protéines impliquées dans la biosynthèse des centres Fe-S ont été mises en évidence pour la première fois (Jacobson *et al*., 1989). Depuis, des protéines homologues ont été découvertes dans d'autres bactéries et dans les cellules eucaryotes. Dans ce qui suit, la biosynthèse des centres Fe-S sera abordée seulement dans le contexte bactérien. Des revues récentes ont fait le point sur ce sujet pour ce qui concerne les plantes (Balk and Lobréaux, 2005) et les autres encaryotes (Lill and Mühlenhoff, 2005).

### 3.1. Le système de biosynthèse des centres Fe-S NIF-spécifique de *A. vinelandii*

La Nitrogénase est une métalloenzyme multimérique qui catalyse la réduction de l'azote moléculaire ($N_2$) en ammoniac ($NH_3$). Elle comprend deux composantes, une réductase spécifique homodimérique, dite protéine à Fer, contenant un centre [4Fe-4S] à l'interface des deux sous-unités, ainsi qu'un site de fixation pour l'ATP, et une protéine à fer-molybdène (protéine MoFe) de type hétérotetramérique ($\alpha_2\beta_2$), contenant le site catalytique. Celui-ci se compose d'un cofacteur de structure unique, dit cofacteur à fer-molybdène ou FeMoco, fait de 7 atomes de fer et d'un atome de Mo associés à une molécule d'homocitrate. La réduction d'une molécule d'azote s'accompagne de l'hydrolyse de 16 molécules d'ATP. La structure du complexe enzymatique a été obtenu et contient deux dimères de la protéine à Fer associés à un hétérotetramère de la proteine MoFe (Schindelin *et al.,* 1997). Une vingtaine de gènes, appelés gènes *nif*, sont responsables de la fixation de l'azote chez *A. vinelandii*. Deux d'entre eux, *nifU* et *nifS,* sont nécessaires à la maturation de la Nitrogénase (Jacobson *et al.*, 1989, Dos Santos *et al.,* 2004). NifS est une protéine homodimérique possédant une activité cysteine désulfurase qui convertit la L-cysteine en L-alanine avec production de soufre sous forme de sulfure (Zheng *et al*., 1993). NifS contient un cofacteur 5'-phosphate pyridoxal (PLP) lié de façon covalente par une lysine. Une liaison persulfure (-S-SH) est crée entre le S du substrat et une cysteine du site catalytique à proximité du PLP. Apres réduction

du persulfure, le sulfure ($S^{2-}$) est libéré et l'alanine du complexe ala-PLP est relâchée.

La protéine NifU servirait de matrice d'assemblage des centres Fe-S. NifU est capable de former un complexe avec NifS. Deux autres gènes du locus *nif* (Fig. 5) pourraient être impliqués dans l'assemblage des centres Fe-S, *cysE1* et *isca*$^{Nif}$. CysE1 est une protéine qui intervient dans la synthèse de la cysteine, le substrat de NifS (Jacobson *et al.*, 1989). IscA$^{Nif}$ est capable d'assembler *in vitro* des centres Fe-S et jouerait un rôle voisin de celui de NifU (Krebs *et al.*, 2001).

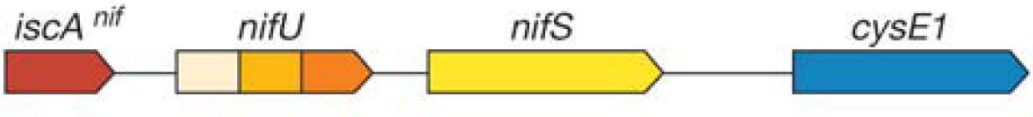



**Figure 5**. Organisation des gènes du locus *nif* impliqués dans la synthèse des centres Fe-S chez *A. vinelandii* (Johnson *et al.*, 2005).

L'inactivation des gènes *nifU* et *nifS,* dédiés à la fixation de l'azote conduit à une diminution de l'activité de la Nitrogénase (Zheng *et al.*, 1998). Des gènes homologues appelés *iscS, iscU* et *iscA* (ISC pour iron sulfur cluster) sont responsables de la maturation des autre protéines à centre Fe-S chez *A. vinelandii* (Jacobson *et al.*, 1989, Zheng *et al*., 1998). Les gènes *isc* ont été mis en évidence dans la plupart des genres bactériens et leur fonction a été particulièrement étudiée chez *E. coli*.

### 3.2. Le système ISC de biosynthèse des centres Fe-S de *E. coli*

Chez *E. coli,* au moins trois cystéine désulfurases ont été identifiées, et appelées IscS, SufS et CSD selon le nom du gène codant (Flint, 1996, Outten *et al*., 2003, Mihara *et al*., 2000). L'inactivation du gène *iscS* conduit à une diminution importante de l'activité spécifique des protéines à centre Fe-S de *E. coli* indiquant que IscS joue un rôle majeur dans la biosynthèse des centres Fe-S (Takahashi and Nakamura, 1999). Le gène *iscS* fait partie de l'opéron *isc* contenant entre autres les gènes *iscU* et *iscA* (Fig. 6).

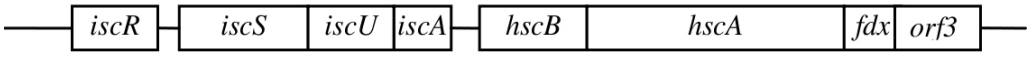



**Figure 6**. Opéron *isc* responsable de biosynthèse des centres Fe-S chez *E. coli* (Takahashi and Nakamura, 1999).

IscS est une cysteine désulfurase très proche de NifS au plan structural et fonctionnel. Le soufre qu'elle génère à partir de la cystéine est transféré à la protéine IscU, qui sert de matrice (scaffold) à l'assemblage des centres Fe-S. La protéine IscU est un homodimère qui s'associe à IscS pour former un hétérotetramère. L'interaction est stabilisée par des ponts disulfures faisant intervenir Cys63 et Cys328 de IscU et IscS, respectivement (Kurihara *et al*., 2003). La Cys328, qui représente le site catalytique de IscS, est localisé dans une longue boucle flexible facilitant les changements de conformation (Cupp-Vickery *et al.,* 2003). Le mécanisme d'assemblage du centre Fe-S auquel participe IscU reste mal compris. Il dépend sans doute de l'état d'oligomérisation de IscU ainsi que de la séquence d'addition du fer et du soufre. Il est clair que c'est la protéine IscU qui permet l'assemblage puis le transfert d'un précurseur de centre Fe-S vers l'apoprotéine cible. IscU possède un motif à trois cysteines conservées autorisant une ligation transitoire du centre Fe-S ou de son précurseur (Zheng *et al*., 1998).

Le rôle de la proteine IscA reste un sujet de discussion. En effet, d'après Ding et Clark, IscA fixerait le fer et servirait de réservoir transitoire de ce métal (Ding et Clark, 2004), alors qu'un autre groupe pense que IscA, qui a trois cysteines conservées, servirait à assembler un centre Fe-S (soit [2Fe-2S] soit [4Fe-4S]) dans une apoprotéine, rôle similaire à celui de IscU (Frazzon and Dean, 2003). La structure cristallographique de IscA sous forme apoprotéine indique que cette protéine est tétramérique, et possède en son centre une poche occupée par une cystéine conservée (Cys35). Les deux autres cysteines conservées (Cys99 et Cys101) ne sont pas visibles dans ce modèle car elles se trouvent dans une région flexible de la structure (Bilder *et al*., 2004). La conformation de cette poche permettrait soit à un ou des atomes de fer, soit à un centre Fe-S, d'être stocké transitoirement avant transfert vers la protéine cible. Récemment, la structure de l'holoprotéine IscA de *Thermosynechoccus elongatus* a apporté la preuve que cette protéine est capable d'assembler un centre [2Fe-2S] (Morimoto *et al*., 2006). Dans la protéine tetramérique, les sous-unités sont organisées deux à deux de manière asymétrique, chacune des paires de protomères $\alpha\beta$ liant un centre [2Fe-2S] par 4 ligands Cys dont 3 sont portés par le protomère $\alpha$. Le centre Fe-S est très exposé au solvant dans une structure flexible qui suggère un mécanisme plausible de transfert à une apoprotéine. Cependant, une autre étude indique que la protéine IscU est la matrice principale d'assemblage des centres Fe-S, et que IscA incorpore le fer libre (en condition aérobie) et le transfère à IscU en présence de L-cystéine (Yang *et al*., 2006).

Les protéines codées par les gènes *hscB* et *hscA* sont des chaperonnes, capables de s'associer au complexe apoIscU/apoIscA afin de stabiliser une conformation favorisant

l'assemblage des centres Fe-S.

Une ferrédoxine à centre [2Fe-2S] codée par le gène *fdx*; pourrait fournir les équivalents réduits nécessaires à différents stades de la biosynthèse des centres Fe-S. Par exemple, elle pourrait apporter les électrons requis pour libérer le sulfure lors de la réaction catalysée par IscS. Elle pourrait aussi participer à la formation du centre Fe-S en fournissant des électrons aux protéines matrices IscA ou IscU ou en facilitant le transfert du centre. Il a été montré que la protéine IscA s'associe de manière spécifique avec la ferrédoxine de *E. coli* en un complexe de stœchiométrie 1:1 (Ollagnier *et al.*, 2001). L'inactivation du gène *fdx* n'a pas d'effet délétère chez *E.coli*. Cependant, il a été démontré que la ferrédoxine est indispensable à la synthèse des centres Fe-S chez la cellule de levure (Lange *et al*., 2000). Il se peut que, dans la cellule bactérienne, la carence en Fd soit compensée par une autre protéine transporteur d'électrons, une flavodoxine par exemple, alors que dans les cellules eucaryotes, du fait de l'adressage spécifique de la Fd dans les mitochondries, ce remplacement fonctionnel n'ait pas lieu.

La protéine Orf3 contenant un grand nombre d'acides aspartiques et glutamiques fournirait le fer nécessaire à la formation des Fe-S (Shimomura *et al.,* 2005).

La protéine IscR, contient un centre [2Fe-2S], et a pour fonction la régulation de l'expression des gènes *iscSUA*. IscR, serait un détecteur/régulateur en fonction des besoins de la cellule (Frazzon and Dean, 2001).

La présence d'homologues isofonctionnels de certains des gènes de l'opéron *iscSUAhscBAfdx* des bactéries dans la plupart des cellules vivantes, y compris les cellules humaines, indique que la machinerie d'assemblage des centres Fe-S, a été conservée au cours de l'évolution.

Une troisième voie d'assemblage des centres Fe-S a été mise en évidence chez *E. coli*, et appelée SUF (sulfur mobilization). Elle est présente dans certaines bactéries mais aussi dans les plantes.

### 3.3. Le système SUF de *E. coli*

L'opéron *suf* de *E. coli* comporte 6 gènes consécutifs appelés *sufABCDSE* (Patzer and Hanke, 1999). Le gène *sufS* code une cystéine désulfurase comparable à celles déjà décrites (Mihara *et al.,* 1997). La proteine SufE est un homodimère qui ressemble à NifS mais qui n'a pas d'activité désulfurase (Loiseau *et al.*, 2003). Cependant, il a été montré que SufE peut s'associer à SufS pour former un complexe actif, dans lequel le soufre est transféré du site actif de la protéine SufS (Cys 364) au site actif de SufE (Cys 51) (Outten *et al.*, 2003). La

structure cristallographique de SufS (Mihara *et al*., 2002) indique qu'au site catalytique, la Cys364 est distante de la molécule de PLP d'au moins 8 Å, une distance plus grande que celle observée dans les autres désulfurases qui expliquerait les performances moindres de SufS (Mihara and Esaki, 2002). La structure cristallographique de SufE de *E. coli* indique que la Cys51 ne se trouve pas en position idéale, ce qui ne favorise pas le transfert rapide et efficace du soufre de SufS à SufE (Goldsmith-Fischman *et al*., 2004). Cependant, on pense que des changements de conformation doivent se produire lors de l'interaction de ces deux protéines, ce qui pourrait favoriser le transfert du soufre de l'une à l'autre.

Aucune fonction n'est connue pour les gènes *sufBD*. La proteine SufC quant à elle appartient à la classe des transporteurs ABC, et de ce fait, est localisée au niveau de la membrane cytoplasmique ; elle permettrait d'importer et/ou d'exporter des substrats. SufB, SufC et SufD forment un complexe qui améliore l'activité désulfurase du complexe SufS-SufE (Outten *et al.*, 2003). SufA, tout comme IscA avec laquelle elle partage 40 % d'identité de séquence, permettrait la formation transitoire d'un centre Fe-S et son transfert vers l'apoprotéine cible (Loiseau, 2004). Le motif CXnCGC trouvé dans IscA, est aussi présent dans SufA.

Le mécanisme d'assemblage des centres Fe-S de la machinerie SUF est schématisé dans la figure 7 (Fontecave *et al*., 2005). La machinerie SUF de *E. coli* apparaît donc remplir des fonctions semblables à celles du système ISC, et est indispensable dans des conditions de stress oxydant ou de carence en fer (Outten *et al.*, 2004).

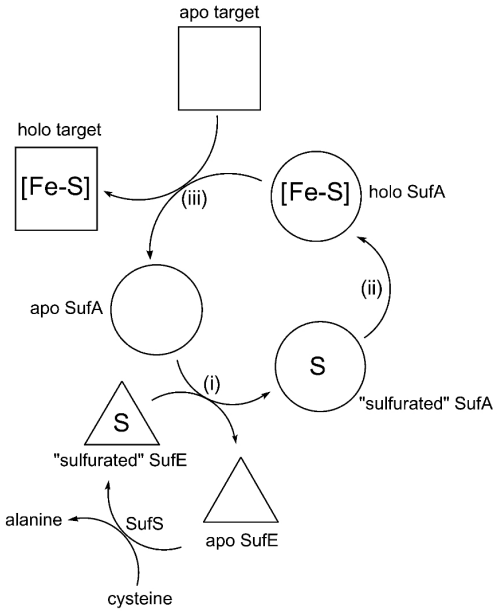



**Figure 7.** Schéma hypothétique de fonctionnement du système SUF. Le mécanisme proposé comporterait trois etapes : (i) transfert du S de SufSE à SufA, (ii) formation du Fe-S dans SufA, (iii) transfert du centre Fe-S dans la proteine cible.

### 3.4. Quelques conclusions sur la biosynthèse des centres Fe-S chez les bactéries

La biosynthèse bactérienne des Fe-S s'effectue au minimum en trois étapes :
(i) production d'un atome de soufre inorganique par une désulfurase à partir d'une cystéine, (ii) incorporation des atomes de fer et de soufre et assemblage du centre Fe-S par un ensemble de protéines servant de matrices (scaffold), (iii) finalement transfert du centre à une apoprotéine.
Le donneur de fer ainsi que son état d'oxydation ne sont toujours pas identifiés.
Une troisième désulfurase identifiée chez *E. coli*, la protéine CSD, est associée au produit du gène *csdE* (ou *ygdK*) qui serait un activateur de la désulfurase (Loiseau *et al*., 2005).

La plupart des bactéries, possèdent au moins deux systèmes spécifiques, NIF et ISC ou SUF et ISC. Un gène codant pour une ferrédoxine (*fdx*) est uniquement associé au système ISC.

Il n'existe pas de machinerie réellement spécifique à tel ou tel centre Fe-S mais des systèmes capables d'assembler plusieurs types de centre à une variété d'apoprotéines (Fontecave *et al.,* 2005).

Les systèmes présents chez *E. coli* sont représentés dans la figure 8.

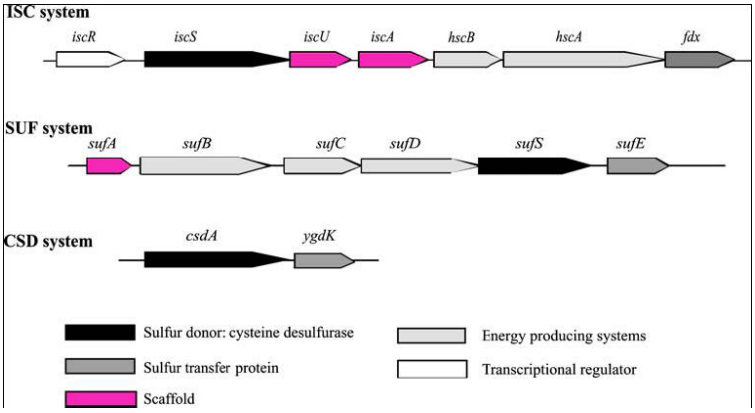



**Figure 8.** Les opérons codant pour les machineries de biosynthèse des Fe-S ISC, SUF et CSD de *E. coli* (Fontecave *et al*., 2005).

### *Etude structurale de la ferrédoxine VI (FdVI) de* Rhodobacter capsulatus

La bactérie photosynthétique *Rhodobacter capsulatus* synthétise six ferrédoxines parmi lesquelles trois contiennent un centre [2Fe-2S]. Deux d'entre elles sont de type plante et participent à la fixation de l'azote (Grabau *et al*., 1991 ; Armengaud *et al.*, 1994), la troisième, appelée FdVI, est de type vertébré et n'a pas de fonction connue (Naud *et al*., 1994). L'inactivation du gène *fdxE* codant Fd6 est létale, suggérant que FdVI participe à une fonction physiologique essentielle à la survie de cette bactérie (Armengaud *et al.*, 1997). Il est plausible que FdVI soit impliquée dans la biosynthèse des centres Fe-S de la même façon que la ferrédoxine Fdx chez *E. coli*.

Nous décrivons ici la structure cristallographique des formes oxydée et réduite de FdVI obtenues toutes deux à une résolution de 2.0 Å.

## Références bibliographiques


Armengaud, J., Meyer, C. and Jouanneau, Y. (1994). Recombinant expression of the *fdxD* gene of *Rhodobacter capsulatus* and characterization of its product, a [2Fe-2S] ferredoxin. *Biochem. J.* **300**:413-418.

Armengaud, J., Meyer, C. and Jouanneau, Y. (1997). A [2Fe-2S] ferredoxin (FdVI) is essential for growth of the photosynthetic bacterium *Rhodobacter capsulatus*. *J. Bacteriol.* **179**:3304-3309.

Balk, J. and Lobreaux, S. (2005). Biogenesis of iron-sulfur proteins in plants. *Trends Plant. Sci.* **10:**324-331.

Bilder, P.W., Ding, H. and Newcomer, M.E. (2004). Crystal Structure of the Ancient, Fe-S Scaffold IscA Reveals a Novel Protein Fold. *Biochemistry* **43**:133-139.

Cupp-Vickery, J.R., Urbina, H. and Vickery, L.E. (2003). The SufE Protein and the SufBCD Complex Enhance SufS Cysteine Desulfurase Activity as Part of a Sulfur Transfer Pathway for Fe-S Cluster Assembly in *Escherichia coli*. *J. Mol. Biol*. **330**:1049-1059.

Ding, H. and Clark, R.J. (2004). *Biochem. J.* **379**:433-440. Dos Santos, P.C., Dean, D.R., Hu, Y. and Ribbe, M.W. (2004). IscA Mediates Iron Delivery for Assembly of Iron-Sulfur Clusters in IscU under the Limited Accessible Free Iron Conditions. *Chem. Rev.* **104**:1159–1173.

Flint, D.H. (1996). Escherichia coli Contains a Protein That Is Homologous in Function and N-terminal Sequence to the Protein Encoded by the nifS Gene of *Azotobacter vinelandii* and That Can Participate in the Synthesis of the Fe-S Cluster of Dihydroxy-acid Dehydratase. *J. Biol. Chem.* **271**:16068-16074.

Fontecave, M., Ollagnier de Choudens, S., Py, B. and Barras, F. (2005). Mechanisms of iron-sulfur cluster assembly: the SUF machinery. *J. Biol. Inorg. Chem*. **10**:713-721.

Frazzon, J. and Dean, D.R. (2001). Feedback regulation of iron-sulfur cluster biosynthesis. *Proc. Natl. Acad. Sci*. **98**:14751-14753.

Frazzon, J. and Dean, D.R. (2003). Formation of iron-sulfur clusters in bacteria: an emerging field in bioinorganic chemistry. *Curr. Opin. Chem. Biol*. **7**:166-173.

Fukuyama, K., Ueki, N., Nakamura, H., Tsukihara, T. and Matsubara, H. (1995). Tertiary structure of [2Fe–2S] ferredoxin from *Spirulina platensis* refined at 2.5 Å resolution: structural comparisons of plant-type ferredoxins and an electrostatic potential analysis. *J.Biochem.(Tokyo)* **117**:1017-1023.

Goldsmith-Fischman, S., Kuzin, A., Edstrom, W.C., Benach, J., Shastry, R., Xiao, R., Acton, T.B., Honig, B., Montelione, G.T. and Hunt, J.F. (2004). The SufE sulfur-acceptor protein contains a conserved core structure that mediates interdomain interactions in a variety of redox protein complexes. *J.Mol.Biol*. **344**:549-565.

Grabau, C., Schatt, E., Jouanneau, Y., and Vignais, P. M. (1991) A new [2Fe-2S] ferredoxin from *Rhodobacter capsulatus*. Coexpression with a 2[4Fe-4S] ferredoxin in Escherichia coli. *J. Biol. Chem. 266*, 3294-3299.

Grinberg, A.V., Hannemann, F., Schiffler, B., Muller, J., Heinemann, U. and Bernhardt, R. (2000). Adrenodoxin: Structure, stability, and electron transfer properties. *Proteins-Structure Function and Genetics* **40:**590-612.

Jacobson, M.R., Brigle, K.E., Bennett, L.T., Setterquist, R.A., Wilson, M.S., Cash, V.L., Beynon, J., Newton, W.E. and Dean, D.R. (1989). Physical and genetic map of the major nif gene cluster from *Azotobacter vinelandii*. *J Bacteriol*. 171:1017-27.

Jacobson, B.L., Chae, Y.K., Markley, J.L., Rayment, I. and Holden, H.M. (1993). Molecular structure of the oxidized, recombinant, heterocyst [2Fe-2S] ferredoxin from Anabaena 7120 determined to 1.7-A resolution. *Biochemistry* **32**:6788-6793.

Johnson, D.C., Dean, D.R, Smith, A.D. and Johnson, M.K. (2005). Structure, Function, and Formation of Biological iron-sulfur clusters. *Annu. Rev. Biochem.* **74**:247–281.

Kurihara, T., Mihara, H., Kato, S., Yoshimura, T. and Esaki, N. (2003). Assembly of iron-sulfur clusters mediated by cysteine desulfurases, IscS, CsdB and CSD, from *Escherichia coli***.** *Biochim Biophys Acta*. **1647**:303-319.

Kakuta, Y., Horio, T., Takahashi, Y. and Fukuyama, K. (2001). Crystal structure of *Escherichia coli* Fdx, an adrenodoxin-type ferredoxin involved in the assembly of iron-sulfur clusters. *Biochemistry* **40**:11007-11012.

Krebs, C., Agar, J.N., Smith, A.D., Frazzon, J., Dean, D.R., Huynh, B.H. and Johnson, M.K. (2001). IscA, an Alternate Scaffold for Fe-S Cluster Biosynthesis. *Biochemistry* **40**:14069-14080.

Lange, H., Kaut, A., Kispal, G. and Lill, R. (2000). A mitochondrial ferredoxin is essential for biogenesis of cellular iron- sulfur proteins. *Proc. Nat. Aca. Sci.* **97**:1050-1055.

Lill, R. and Muhlenhoff, U. (2005). Iron-sulfur-protein biogenesis in eukaryotes. *Trends Biochem. Sci.* **30***:* 133-141.

Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M. and Barras, F. (2003). Biogenesis of Fe-S Cluster by the Bacterial Suf System SufS and SufE form a new type of cysteine désulfurase. *J. Biol. Chem*. **278**:38352–38359.

Loiseau, L., Ollagnier-de-Choudens, S., Lascoux, D., Forest, E., Fontecave, M. and Barras, F. (2005). Analysis of the Heteromeric CsdA-CsdE Cysteine Desulfurase, Assisting Fe-S Cluster Biogenesis in *Escherichia coli J. Biol. Chem*. **280**:26760–26769.

Malkin, R. and Rabinowitz, J.C. (1966). The reconstitution of clostridial ferredoxin. *Biochem Biophys Res Commun.* **23**:822–827.

Matsubara, H., Hase, T., Wakabayashi, S. and Wada, K. (1980). Structure and evolution of chloroplast- and bacterial-type ferredoxins, in *The Evolution of Protein Structure and Function* (Sigman, D. S., and Brazier, M. A. S., Eds.) pp 245-266, Academic Press, New-York.

Mihara, H., Kurihara, T., Yoshimura T., Soda, K. and Esaki, N. (1997). Cysteine sulfinate desulfinase, a NIFS-like protein of *Escherichia coli* with selenocysteine lyase and cysteine desulfurase activities. Gene cloning, purification, and characterization of a novel pyridoxal enzyme *J. Biol. Chem.* **272**:22417–22424.

Mihara, H., Kurihara, T., Yoshimura, T. and Esaki, N. (2000). Kinetic and Mutational Studies of Three NifS homologs from *Escherichia coli*: Mechanistic Difference between L-Cysteine Desulfurase and L-Selenocysteine Lyase Reaction. *J Biochem. (Tokyo)* **127**:559-567.

Mihara, H. and Esaki, N. (2002). Bacterial Cysteine Desulfurases: Their Function and Mechanisms. *Appl. Microbiol. Biotechno*l. **60**:12–23.

Mihara, H., Fujii, T., Kato, S., Kurihara, T., Hata, Y. and Esaki, N. (2002). Structure of External Aldimine of *Escherichia coli* CsdB, an IscS/NifS Homolog Implications for Its Specificity toward Selenocysteine. *J. Biochem.(Tokyo)* **131**:679-685.

Morales, R., Charon, M.H., Hudry-Clergeon, G., Petillot, Y., Norager, S., Medina, M. and Frey, M. (1999). Refined X-ray structures of the oxidized, at 1.3 Å, and reduced, at 1.17Å, [2Fe–2S] ferredoxin from the cyanobacterium Anabaena PCC7119 show redox-linked conformational changes. *Biochemistry* **38**:15764-15773.

Morimoto, K., Yamashita, E., Kondou, Y., Lee, S.J., Arisaka, F., Tsukihara, T. and Nakai, M. (2006). The asymmetric IscA homodimer with an exposed [2Fe-2S] cluster suggests the structural basis of the Fe-S cluster biosynthetic scaffold. *J. Mol. Biol.* **360**:117-132.

Müller, A., Müller, J.J., Muller, Y.A., Uhlmann, H., Bernhardt, R. and Heinemann, U. (1998). New aspects of electron transfer revealed by the crystal structure of a truncated bovine adrenodoxin, Adx(4-108). *Structure* **6**:269-280.

Müller, J.J., Müller, A., Rottmann, M., Bernhardt, R. and Heinemann, U. (1999). Vertebrate-type and plant-type ferredoxins: crystal structure comparison and electron transfer pathway modelling. *J. Mol. Biol*. **294**:501-513.

Müller, J.J., Lapko, A., Bourenkovi, G., Ruckpaul, K. and Heinemann, U. (2001). Adrenodoxin reductase-adrenodoxin complex structure suggests electron transfer path in steroid biosynthesis. *J. Biol. Chem.* **276**:2786–2789.

Naud, I., Vincon, M., Garin, J., Gaillard, J., Forest, E., and Jouanneau, Y. (1994) Purification of a sixth ferredoxin from *Rhodobacter capsulatus*. Primary structure and biochemical properties. *Eur. J. Biochem.* **222**:933-939.

Neuhaus, H.E. and Emes, M.J. (2000). Nonphotosynthetic Metabolism in Plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. **51**:111–140.

Ollagnier-de-Choudens, S., Mattioli, T., Takahashi, Y. and Fontecave, M. (2001). Iron-sulfur cluster assembly: characterization of IscA and evidence for a specific and functional complex with ferredoxin. *J. Biol. Chem.* **276**:22604-22607.

Outten, F.W., Wood, M.J., Munoz, F.M. and Storz, G. (2003). The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in *E. coli*. *J. Biol. Chem*. **278**:45713-45719.

Outten, F.W., Djaman, O. and Storz, G. (2004). A s*uf* operon equirement for Fe–S cluster assembly during iron starvation in *Escherichia coli. Mol. Microbiol*. **52**:861–872.

Patzer, S.I. and Hantke, K. (1999). SufS Is a NifS-Like Protein, and SufD Is Necessary for Stability of the [2Fe-2S] FhuF Protein in *Escherichia coli*. *J Bacteriol*. **181**:3307-3309.

Pikuleva, I.A., Tesh, K., Waterman, M.R. and Kim, Y. (2000). The tertiary structure of full-length bovine adrenodoxin suggests functional dimers. *Arch.Biochem.Biophys.* **373**:44-55.

Schatt, E., Jouanneau, Y. and Vignais, P. M. (1989). Molecular cloning and sequence analysis of the structural gene of ferredoxin I from the photosynthetic bacterium *Rhodobacter capsulatus*. *J. Bacteriol* .**171**:6218-6126.

Schindelin, H., Kisker, C., Schlessman, J.L., Howard, J.B. and Rees, D.C. (1997). Structure of ADP.$AlF_4^-$-stabilized nitrogenase complex and its implications for signal transduction. *Nature* **387**:

370-376.

Sevrioukova, I.F., Garcia, C., Li, H., Bhaskar, B. and Poulos, T.L. (2003). Crystal structure of putidaredoxin, the [2Fe-2S] component of the P450cam monooxygenase system from *Pseudomonas putida*. *J. Mol. Biol.* **333**:377-392.

Shimomura, Y., Takahashi, Y. Kakuta, Y. and Fukuyama, K. (2005). Crystal structure of *Escherichia coli* YfhJ protein, a member of the ISC machinery involved in assembly of iron-sulfur clusters. *Proteins: Structure, Function, and Bioinformatics*. **60**:566-569.

Takahashi, Y. and Nakamura, M. (1999). Functional Assignment of the ORF2-iscS-iscU-iscA-hscB-hscA-fdx-0RF3 Gene Cluster Involved in the Assembly of Fe-S Clusters in *Escherichia coli*. *J. Biochem. (Tokyo)* **126**:917-26.

Valentine, R.C. (1964). Bacterial ferredoxin. *Bacteriol Rev.* **28**: 497–517.

Yang, J., Bitoun, J.P. and Ding, H. (2006). Interplay of IscA and IscU in Biogenesis of Iron-Sulfur Clusters *J. Biol. Chem*. **281**:27956-27963.

Zheng, L., White, R.H., Cash, V.L., Jack, R.F. and Dean, D.R. (1993). Cysteine Desulfurase Activity Indicates a Role for NIFS in Metallocluster Biosynthesis. *PNAS.* **90**:2754-2758.

Zheng, L., Cash, V.L., Flint, D.H. and Dean, D.R. (1998). Assembly of Iron-Sulfur Clusters Identification of an iscSUA-hscBA-fdx gene cluster from *Azotobacter vinelandii*. *J. Biol. Chem*. **273**:13264-13272.

# Partie 2 : Article 1

## Structure of a [2Fe–2S] ferredoxin from *Rhodobacter capsulatus* likely involved in Fe–S cluster biogenesis and conformational changes observed upon reduction

Cet article a été publié dans la revue *Journal of Biological Inorganic Chemistry* (volume 11, pages 235-246).

# Structure of a [2Fe-2S] ferredoxin from *Rhodobacter capsulatus* likely involved in Fe-S cluster biogenesis and conformational changes observed upon reduction

**Germaine Sainz[1†] Jean Jakoncic[1,2†], Larry C. Sieker[3], Vivian Stojanoff[1,2], Nukri Sanishvili[4], Marcel Asso[5], Patrick Bertrand[5], Jean Armengaud[6,7] and Yves Jouanneau[6]✉**

European Synchrotron Radiation Facility, BP 220, 38054 Grenoble Cedex 9, France, and National Synchrotron Light Source, Upton, NY 11973, USA, and Structural Biology Center/Midwest Center for Structural Genomics, Argonne National Laboratory, IL 60439, USA and Laboratoire de Bioénergétique et Ingénierie des Protéines, UPR 9036 CNRS, 13402 Marseille Cedex 20, France, and Laboratoire de Biochimie et Biophysique des Systèmes Intégrés, CNRS UMR 5092, Département Réponse et Dynamique Cellulaires, CEA-Grenoble, F-38054 Grenoble Cedex 9, France, and CEA-Valrho, DSV-DIEP-SBTN, BP 171, F-30207 Bagnols-sur-Cèze cedex, France.

[1]European Synchrotron Radiation Facility, Grenoble.
[2]Current address: Brookhaven National Laboratory, National Synchrotron Light Source, NY.
[3]Current address: Department of Biological Structure, University of Washington, Seattle, WA.
[4]Structural Biology Center, Argonne National Laboratory.
[5]Laboratoire de Bioénergétique et Ingénierie des Protéines, Marseille.
[6]Laboratoire de Biochimie et Biophysique des Systèmes Intégrés, Grenoble.
[7]CEA-Valrho, DSV-DIEP-SBTN, BP 171, F-30207 Bagnols-sur-Cèze cedex, France.
[†] these authors contributed equally to this work

**Abbreviations**

Adx, adrenodoxin;

AnFd, *Anabaena* PCC7119 ferredoxin;

Fd, ferredoxin;

Fdx, ferredoxin from *E. coli*;

MAD, multiple anomalous dispersion;

mr, molecular replacement;

Pdx, putidaredoxin;

rms , root means square

## Abstract

FdVI from *Rhodobacter capsulatus* is structurally related to a group of [2Fe-2S] ferredoxins involved in iron-sulfur cluster biosynthesis. Comparative genomics suggested that FdVI and orthologs found in α–proteobacteria are involved in this process. Here, the crystal structure of FdVI has been determined on both the oxidized and the reduced protein. The [2Fe-2S] cluster lies 6 Å below the protein surface in an hydrophobic pocket without access to the solvent. This particular cluster environment might explain why the FdVI midpoint redox potential (–306 mV at pH 8.0) did not show temperature or ionic strength dependence. Besides the four cysteines that bind the cluster, FdVI features an extra cysteine which is located close to the S1 atom of the cluster and is oriented in a position such that its thiol group points towards the solvent. Upon reduction, the general fold of the polypeptide chain was almost unchanged. The [2Fe-2S] cluster underwent a conformational change from a planar to a distorted lozenge. In the vicinity of the cluster, the side chain of Met24 was rotated by 180° bringing its S atom within H-bonding distance of the S2 atom of the cluster. The reduced molecule also featured a higher content of bound water molecules, and more extensive hydrogen bonding networks compared to the oxidized molecule. The unique conformational changes observed in FdVI upon reduction are discussed in the light of structural studies performed on related ferredoxins.

**Key words :** ferredoxin; crystal structure; iron sulfur cluster; redox potential; conformational changes;

# 1. Introduction

Ferredoxins are small electron carrier proteins that participate in various redox reactions, and are widely distributed among all forms of living organisms. Soluble ferredoxins containing one [2Fe-2S] cluster comprise two major groups, the plant-type and the vertebrate-type ferredoxins, which exhibit distinctive biochemical and structural properties [1,2]. Vertebrate-type ferredoxins are exemplified by adrenodoxin which is involved in steroid hormone biosynthesis in mammals [3]. Adrenodoxin serves as an electron donor to two types of cytochrome P450 enzymes present in mitochondria of the adrenal cortex. Vertebrate ferredoxins have homologous counterparts in bacteria, some of which are functionally associated with P450 enzymes [4,5].

Over the past five years, information available from the sequencing of complete genomes showed that vertebrate-type ferredoxins are present in almost all living cells except in Archaea, suggesting that they might fulfill a more general function. Studies on prokaryotes belonging to the γ-*Proteobacteria*, including *Azotobacter vinelandii*, *Escherichia coli* and *Haemophilus influenzae*, provided evidence that certain vertebrate-type ferredoxins are responsible together with the *iscS, iscU, iscA, hscB* and *hscA* gene products for the formation and insertion of Fe-S clusters in proteins [6,7]. Likewise, the mitochondrial ferredoxin Yah1 from *Saccharomyces cerevisiae* was shown to be essential for the biogenesis of iron-sulfur proteins in yeast [8]. Although vertebrate-type ferredoxins fulfill very diverse functions depending on the original host cells, sequence alignment highlighted several conserved features including the cluster-binding sequence, Cys-$X_5$-Cys-X-Thr-Cys-$X_{36\text{-}38}$-Cys [9].

The crystal structure of a truncated form of a typical vertebrate ferredoxin, bovine adrenodoxin, has been described [10] and a detailed structure-function analysis of adrenodoxin has been published by Grinberg and co-workers [9]. More recently, the X-ray structures of putidaredoxin and the related bacterial ferredoxin from *E. coli* (Fdx) have been reported [11,12]. Although amino acid sequence comparisons showed only limited identities (e.g., 33% between adrenodoxin and putidaredoxin ), the structures of these proteins are very similar, especially in the core domain which contains the cluster [10,12].

In this report, we describe the X-ray structure of FdVI, a ferredoxin from the photosynthetic bacterium *Rhodobacter capsulatus* (α-*Proteobacteria*). *R. capsulatus* synthesizes six ferredoxins, three of which were found to contain a [2Fe-2S] cluster. Two ferredoxins, designated FdIV and FdV, belong to the plant-type ferredoxin subgroup and are

involved in nitrogen fixation [13,14] while FdVI is related to the vertebrate category [15]. Genetic studies indicated that FdVI must play an essential role since disruption of its structural gene *fdxE* is lethal [16]. In a previous study, FdVI was overproduced in *E. coli*, allowing the recombinant ferredoxin to be purified and crystallized [17]. Here, the crystal structure of FdVI was determined from X-ray diffraction data collected in the oxidized and partially reduced states, allowing a description at 2.0 Å resolution of structural changes occurring upon reduction. The significance of the redox-induced conformational changes observed in FdVI are discussed in the light of previous studies describing redox changes in related ferredoxins.

## 2. Materials and Methods

### 2.1. Purification and crystallization of FdVI.

The purification and crystallization conditions of FdVI have been described previously [17]. Diffraction quality crystals were obtained in two steps by vapor equilibration. In the first step seed crystals were grown by the hanging drop method. Crystallization was performed with a 10 mg/ml protein solution buffered with 100 mM imidazole (pH 7.6) plus sodium formate to make the drop initially 5.4 M. The drop was then equilibrated against a solution of 7.0 M sodium formate containing 100 mM imidazole (pH 7.6) at 15 °C. In a second step, the seed crystals were washed briefly in imidazole buffered 3.5-4.2 M formate and grown in a sitting drop as long red-brown prismatic shaped needles, 0.4 mm x 0.1mm x 0.05 mm in size, within 2-3 days. These crystals were sensitive to changes in temperature and were lost when exposed to temperatures above 20°C. The crystals were cryoprotected by a quick soak in 13% glycerol and flash-frozen at 100K in a cold $N_2$ stream generated by an Oxford Cryosystems apparatus.

### 2.2. Reduction of FdVI crystals.

The following experiments were carried out in a glove box under anoxic conditions (< 2ppm $O_2$). Two different ways of producing crystals of reduced FdVI have been implemented: either crystallization of the reduced protein, or soaking crystals of the oxidized protein in a reducing solution. The former was performed in formate-imidazole solution as described above except that 2 mM sodium dithionite (Fluka) was added to both the protein sample and

the crystallization solution. The formate concentration was varied between 5 and 7 M at pH values of 7.6 and 7.8. Alternatively, oxidized crystals were soaked in a reducing solution consisting of the mother solution containing a slight excess of precipitant (7.3 M formate) and a variable concentration of dithionite. Reduced crystals were obtained after two successive soaks in 20 and 30 mM dithionite for 20 min each. The extent of reduction of the protein in the crystals was assessed by micro-spectrophotometry as described below. Reduced crystals were soaked in 13% glycerol and flash-frozen in liquid propane inside the glove box.

### 2.3. Analytical and spectroscopic methods.

The concentration of *R. capsulatus* FdVI was determined from absorbance measurements at 416 nm using an absorption coefficient of 10 $mM^{-1}$ $cm^{-1}$. Ultraviolet-visible absorption, EPR and circular dichroïsm spectroscopy were performed according to published procedures [13]. Micro-spectrophotometric analysis of FdVI crystals was carried out with a Cryobench apparatus allowing measurements with a thin light beam which was delivered through optical fibers[18] .

### 2.4. Determination of midpoint redox potentials.

Redox titrations of *R. capsulatus* FdVI and *Spirulina maxima* ferredoxin were carried out at room temperature under argon in Tris-HCl, pH 8.0 buffer containing 42 μM protein. The purification of the sample of S. *maxima* ferredoxin used in this study has been described previously [19]. The reduction of the protein was monitored by recording absorbance changes at 456 and 430 nm for FdVI and *S. maxima* ferredoxin, respectively. Redox potentials were adjusted by stepwise additions of 10 mM sodium dithionite and were measured with a combined Pt-Ag/AgCl/KCl(3M) microelectrode. Equilibrium with the electrode was achieved by adding the following mediators: phenosafranin (-255 mV), benzyl viologen (-350 mV) and methyl viologen (-440 mV), each at 2 μM concentration. The absorption of the mediators at 456 nm was negligible. Redox titrations were performed at two different ionic strengths: 0.1 M and 1 M NaCl. A non-isothermal potentiometric device was used in temperature-dependent experiments. An optical cell containing the protein solution, a platinum Metrohm microelectrode and a calibrated thermocouple was placed in a variable temperature holder and the reference electrode was kept at 23°C. Under these conditions, the temperature coefficient $dE^{\circ\prime} / dT$ of the midpoint potential $E^{\circ\prime}$ is directly proportional to the entropy variation $\Delta S^{\circ\prime}_{rc}$=

$S^{\circ'}_{red} - S^{\circ'}_{ox}$ of the studied redox couple [20]. All potential values are given with respect to the standard hydrogen electrode.

### 2.5. X-ray data collection and processing.

Diffraction data were collected on both oxidized and reduced crystals. Initially, a MAD experiment at the iron absorption edge was carried out on two crystals of oxidized FdVI. From the first crystal, a fluorescence scan was recorded. The f' and f" values were calculated and plotted as a function of energy with the program CHOOCH [21]. From this plot, the inflection point and the peak of absorption were determined to be at 7.1198 KeV and 7.1315 KeV in energy, equivalent to 1.7414 Å and 1.7386 Å in wavelength, respectively. The high energy remote was chosen at 8 KeV (wavelength $\lambda$=1.5498 Å). X-ray diffraction data were recorded using a 3 x 3 mosaic CCD detector on the undulator beam line ID19 at the Advanced Photon Source (Argone National Laboratory, IL) [22]. A 180° data set was collected at each wavelength with 2° oscillation images and 15 s exposure each. The crystal was mounted in a random orientation at a crystal-to-detector distance of 100 mm. As the crystal did not undergo detectable radiation damage, additional data were recorded to increase redundancy. The crystal belongs to the space group $P2_12_12_1$ with unit cell dimensions of a= 45.34 Å, b= 49.03 Å and c= 54.91 Å. All diffraction data were reduced and scaled with HKL2000 package [23,24].

The diffraction data for the reduced crystal were collected on the undulator beam line ID14 EH4 at the European Synchrotron Radiation Facility. A $100^o$ data set was recorded at a wavelength of 0.98 Å with $1^O$ oscillation steps. Data were processed as above [24]. Data statistics are summarized in Table 1.

The program CNS [25] was used to carry out local scaling of the MAD data, perform automatic Patterson interpretations, locate the two iron scatterers present in the asymmetric unit, refine heavy-atom parameters and perform phase calculations. In the experimental map, most of the atomic model was built by the auto-tracing option of the Warp program [26], whereas 11 residues (Pro33 to Asp36, Cys45 to His49, Trp56 to Asp58, Asp69 to Ile71 and Ile106) mainly in loops were built manually with the Baton-build option from the graphics program O [27]. Features of positive density greater than 3$\sigma$ in the Fo-Fc difference Fourier maps were modeled as the 2 sulfur and 2 iron atoms of the [2Fe-2S] cluster. replacement (reduced crystal). Calculations were carried out with the CNS program.

| **Data** | **Oxidized crystal** | | | **Reduced crystal** |
|---|---|---|---|---|
| **Space group** | $P2_12_12_1$ | | | $P2_12_12_1$ |
| **Unit-cell parameters (Å)** | 45.34, 49.03, 54.91 | | | 45.54, 50.36, 55.35 |
| | Peak | Edge | Remote | |
| **Wavelength (Å)** | 1.7386 | 1.7429 | 1.5498 | 0.98 |
| **Resolution Range (Å)** | 2.19-19.27 | 2.19-19.27 | 1.96-19.27 | 2.00-33.0 |
| No. of observations | 43096 | 43264 | 60252 | 29989 |
| **Unique reflections** | 6551 (610) | 6502 (597) | 9128 (836) | 9101 (426) |
| **Completeness (%)** | 99.0 (94.1) | 98.9 (93.6) | 99.0 (94.4) | 98.6(96.6) |
| **$\langle I \rangle / \sigma I_{overall}$** | 24.2 (4.5) | 24.9 (8.6) | 25.2 (4.1) | 14.3(5.6) |
| **[a]$R_{sym}$** | 5.3 (24.8) | 5.0 (13.3) | 5.1 (25.6) | 8.2(27.3) |

[a]$R_{sym} = \sum | I_i - \langle I \rangle | / \sum \langle I_i \rangle$, where $I_i$ is the intensity of the ith observation and $\langle I_i \rangle$, is the mean intensity of the reflection. Numbers in parentheses gave statistics in the highest resolution range : 2.28 Å -2.20 Å for the peak and inflection data from the oxidized crystal 2.03 Å -1.96 Å for the reduced crystal

**Table 1**. Crystallographic data used for the phasing (oxidized crystal) and for the molecular

The final crystallographic model was refined at 2.07 Å resolution to an R factor of 19.6 % and an $R_{free}$ of 21.1 %, with the data from the high energy remote wavelength. There was one molecule per asymmetric unit, and the solvent content was estimated to be about 50 %. All 106 residues of the ferredoxin sequence are included in this final model, which consists of 797 protein atoms, 2 Fe and 2 inorganic S atoms as well as 62 water molecules. The atomic coordinates have been deposited in the Protein Data Bank with entry code 1E9M.

The structure of reduced FdVI was determined by molecular replacement with the AMoRE package[28], using as starting model the structure of the oxidized molecule, devoid of the cluster and the water molecules. The initial crystallographic R-factor was 39.7%. The crystallographic model was refined to 2 Å resolution using program CNS and program O as described above. Strong electron density was observed above the sulfur atoms of the cluster early in the refinement. Simulated annealing omit map suggested a non-planar conformation of the cluster. The cluster conformation was further refined according to the omit map by releasing all restrains on the cluster geometry. The positive and negative electron density found around residues 44 and 24 were interpreted as second conformations of these residues

After a final annealing cycle with the CNS program, the R and $R_{free}$ factors were 23.0% and 25.7%, respectively. Ramachandran plot indicated that all residues were in the most favored or additional allowed regions with the exception of Ala44. The coordinates have been deposited in the Protein Data Bank as entry **1UWM**. The refinement statistics of the models for the oxidized and reduced proteins are summarized in Table 2.

### 2.6. Structure comparisons of [2Fe-2S] containing proteins.

The rms deviations between structures were calculated using the MOE program (available at www.chemcomp.com). All structure superpositions were done using the lsqkab program from the CCP4 suite [29].

## 3. Results and discussion

### 3.1. Comparative genomics indicate that FdVI and its orthologs in α- and β-proteobacteria are involved in Fe-S cluster biosynthesis.

According to the current annotation of the *R. capsulatus* genome (http://ergo.integratedgenomics.com/Genomes/R.capsulatus/proteins), there exist six ferredoxin-encoding genes in this bacterium. All these ferredoxins have been previously characterized, and it clearly appears that FdVI is the only one to show properties similar to ferredoxins known to be involved in the assembly of iron-sulfur clusters [15]. As exemplified by the *E. coli* Fdx, these ferredoxins share a Cys-$X_3$-Cys-$X_1$-Cys-$X_2$-Cys cluster binding motif, where the second cysteine, which is not a ligand of the [2Fe-S] cluster, has been put forward as a possible ligand for an iron or a sulfur atom [11]. PSI-BLAST searches for bacterial homologues of FdVI in the data bases revealed a systematic occurrence of similar sequences in most *Proteobacteria* (purple bacteria and relatives). In some organisms such as *Rhodopseudomonas palustris*, paralogous [2Fe-2S] ferredoxins were found, sharing the Cys-$X_5$-Cys-$X_2$-Cys motif common to vertebrate type-ferredoxins. Using the STRING web-interfaced software (http://www.bork.embl-heidelberg.de/STRING/) for predicting functional associations between gene products in whole genomes [30], we analysed, in the available genomes of proteobacteria, the occurrence and neighborhood associations between the eight genes of the *iscSUA-hscBA-fdx-ORF3* operon found in *E. coli* and *Azotobacter vinelandii*. The

analysis showed that *fdx* genes from most *γ–Proteobacteria* (*Pseudomonas*, *Haemophilus*, *Escherichia*) and *β-Proteobacteria* (*Ralstonia, Chromobacterium*) are clearly associated with *iscSUA* and/or *hscBA* genes involved in Fe-S cluster biosynthesis (Fig. S1 in supplementary material). A few neighborhood associations are also observed in some *α-Proteobacteria*, but in most cases, the genes belonging to the *iscSUA-hscBA-fdx* operon are scattered on the chromosome. Remarkably, most *α-Proteobacteria* genomes show only one *isc*-related *fdx* *hscBA* genes involved in Fe-S cluster biosynthesis (Fig. S1 in supplementary material). A few neighborhood associations are also observed in some *α-Proteobacteria*, but in most cases, the genes belonging to the *iscSUA-hscBA-fdx* operon are scattered on the chromosome. Remarkably, most *α-Proteobacteria* genomes show only one *isc*-related *fdx* gene, as well as single copies of the *iscSUA* and *hscBA* related genes. In addition, a phylogenetic comparison of 37 vertebrate-type ferredoxin sequences showed a striking correlation between the sequence relatedness and the taxonomical relationship between microorganisms (data not shown). In this respect, FdVI appeared more closely related to ferredoxins from other *α-Proteobacteria* than to ferredoxins from *β* and *γ-Proteobacteria* (Fig. 1). These findings, together with the sequence alignments with *bona fide isc*-linked Fds (Fig.1), support the idea that FdVI and its orthologs in *α-* and *β-Proteobacteria* are involved in Fe-S cluster assembly, similar to their counterparts in *γ–Proteobacteria.*

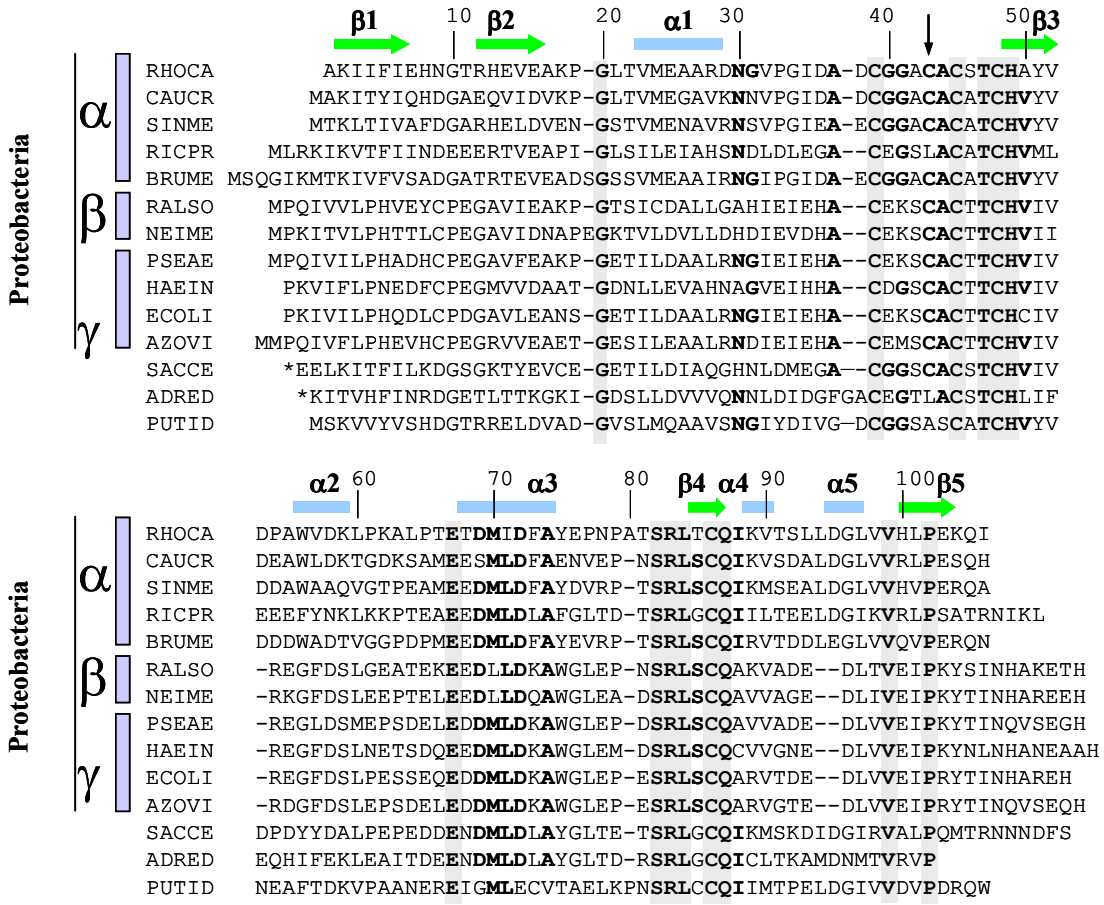


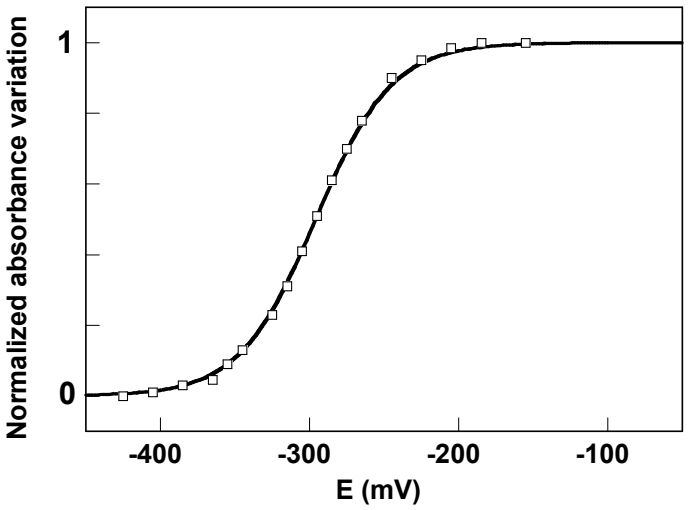

**Figure 1**. Sequence alignment of representative vertebrate-type [2Fe-2S] ferredoxins. The alignment was performed using VectorNTI software. The numbering refers to the FdVI sequence. Secondary structures are shown as arrows (β-strands) and boxes (α-helices) above the sequence. Highly conserved residues are shown in bold letters and invariant residues are shaded. The vertical arrow points at a non-ligand cysteine conserved in *isc*-linked Fds. The abbreviations used are : RHOCA for *R. capsulatus* B10 FdVI (Fer6_RHOCA), CAUCR for *Caulobacter crescentus* CB15 Fd2 (CC3524), SINME for *Sinorhizobium meliloti* Fd (SMCO0192), RICPR for *Rickettsia prowazekii* Fd (RP199), BRUME for *Brucella melitensis* 16M Fd (BMEI0959), RALSO for *Ralstonia solanacearum* GMI1000 Fd (RSC1025), NEIME for *Neisseria meningitidis* MC58 Fd (NMA1344), PSEAE for *Pseudomonas aeruginosa* PA01 Fd (PA3809), HAEIN for *Haemophilus influenzae* Fd (HI0372), ECOLI for *Escherichia coli* K12 Fd (B2525), AZOVI for *Azotobacter vinelandii* Fd (T44286), SACCE for *Saccharomyces cerevisiae* Yah1p (YPL252c) (residues 59-172), ADRED for bovine adrenodoxin (BAA00362) (residues 4-108), and PUTID for putidaredoxin from *Pseudomonas putida* (P00259).

### 3.2. Circular dichroïsm and redox properties of FdVI

In a previous study, we demonstrated that the recombinant ferredoxin considered herein was indistinguishable from the native *R. capsulatus* FdVI, based on UV-visible and EPR spectroscopy, as well as mass spectrometric measurements [17]. The protein was further characterized by circular dichroïsm spectroscopy. The spectrum of FdVI was very similar to those previously found for adrenodoxin, putidaredoxin, *E. coli* Fdx and *A. vinelandii* FdIV [31,32] (data not shown). It featured a shoulder at 450 nm and separate absorption bands in the 250-350 nm range which are typical of vertebrate-type Fds, and are not observed in the case of a representative plant-type ferredoxin [33]. This result further confirms that FdVI belongs to the former class of [2Fe-2S] ferredoxins.

Redox titrations of FdVI were first carried out at 23°C and at two different ionic strengths. The data were well fitted by a Nernst curve, yielding very similar midpoint potentials at 0.1M NaCl (E°'= -306 mV ± 5 mV) and at 1 M NaCl (E°'= -296 ± 5 mV) (Fig. 2). The midpoint potential of FdVI was also found to be essentially temperature-independent in the 13 to 35°C range (Fig. 2b) resulting in an entropy variation $\Delta S^{\circ'}_{rc}$ equal to 0 ± 5 J $mol^{-1}$ $K^{-1}$. In contrast, strong ionic strength dependence and a markedly negative $\Delta S^{\circ'}_{rc}$ value were observed for *S. maxima* ferredoxin (Fig.2b) and other plant-type ferredoxins [34]. A large entropy variation $\Delta S^{\circ'}_{rc}$ = -207 J $mol^{-1}$ $K^{-1}$ has been previously reported for bovine adrenodoxin [35]. However, recent non-isothermal potentiometric titrations of this protein performed at 0.1 M and 1M NaCl, yielded a much less negative value ($\Delta S^{\circ'}_{rc}$ = -20 ± 10 J $mol^{-1}$ $K^{-1}$; Bernardt, R., Asso, M. and Bertrand, P., unpublished results). The very low entropy variation found for FdVI and adrenodoxin compared to plant-type ferredoxins, may reflect differences in the solvent accessibility of the cluster. This hypothesis is supported by comparisons of the molecular structures of relevant ferredoxins as will be discussed below.

a




b

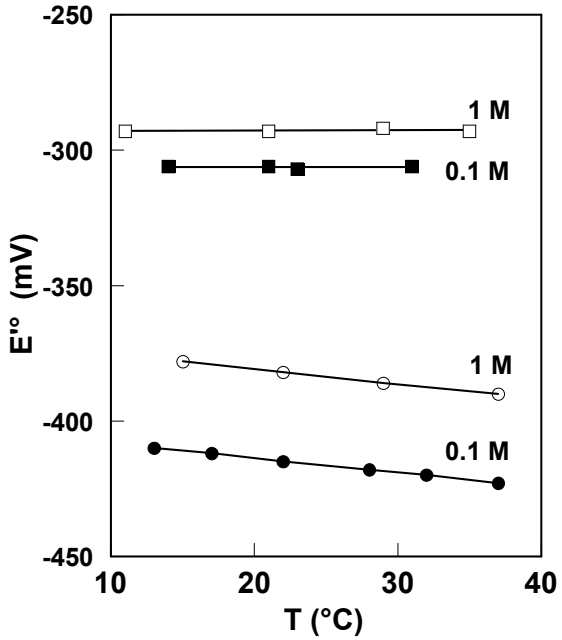



**Figure 2.** Redox properties of FdVI.
a: Reductive titration of FdVI with dithionite. The FdVI concentration was 42 µM in 1.0 M NaCl, Tris-HCl buffer, pH 8.0. Stepwise dithionite reduction of the ferredoxin was monitored by optical absorbance measurements at 456 nm, at 23 °C. The normalized variations correspond to the ratio $(A_E - A_{red})/(A_{ox} - A_{red})$. The solid line shows the best fit with a Nernst curve centered at -296 mV.
b: Temperature dependence of the midpoint potential of FdVI. The temperature dependence of the midpoint potential E°' of FdVI from *R. capsulatus* for two NaCl concentrations (□ = 1 M , ■ = 0.1 M) was compared with that of *S. maxima* ferredoxin (○ = 1 M, ● = 0.1 M). The solid lines correspond to $\Delta S^{\circ'}_{rc}$ values equal to 0 and -52 $J.mol^{-1}.K^{-1}$ for FdVI and *S. maxima* ferredoxins, respectively.

### 3.3. Overall fold of FdVI

The structure of FdVI was determined from high quality MAD data sets (Table 1) obtained from oxidized crystals. The crystallographic model was refined at 2.07 Å resolution to an *R* factor of 19.6 % and an $R_{free}$ of 21.1 %. As depicted in figure 3, the FdVI polypeptide displayed the typical $\alpha + \beta$ fold that was previously observed in other [2Fe-2S] Fds [2], featuring a twisted sheet that consists of the four major β-strands, and one major α-helix adjacent to the sheet.

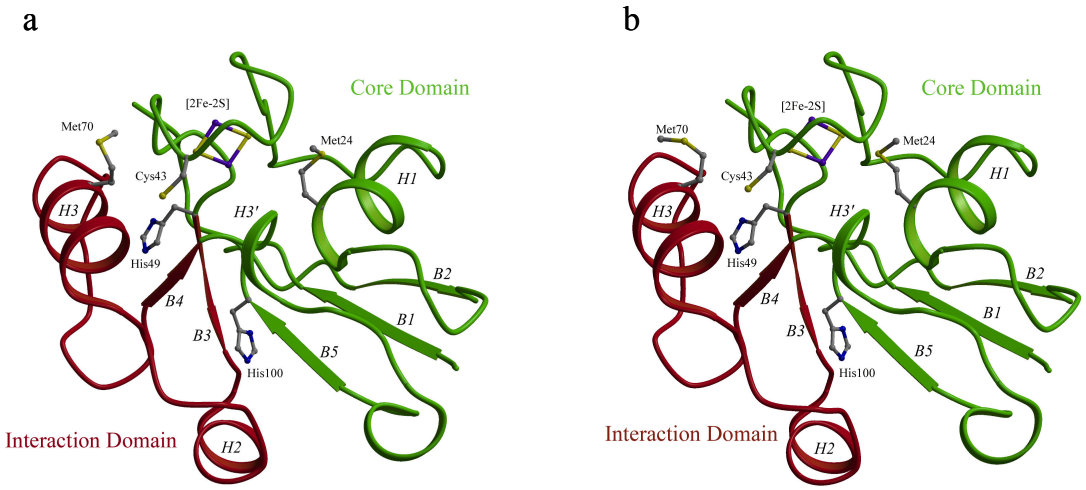



**Figure 3.** Ribbon representation of the FdVI protein structure in the oxidized (a) and reduced (b) states. The core and interaction domains are drawn in green and red, respectively. The [2Fe-2S] cluster is represented in a stick-and-ball mode (Fe and S atoms are shown in purple and yellow, respectively). The side chains of residues Met24, Met70, Cys43, His49 and His100 are shown. In the case of the reduced crystal, only the conformation which is thought to correspond to the fully reduced molecule is displayed. The figure was drawn using BOBSCRIPT [45].

The polypeptide chain of FdVI is organized into a core domain which contains the cluster, and a large hairpin comprising residues between His49 and Leu84, which includes strand B3, and helices H2 and H3 (Fig. 3). The large hairpin, which is later referred to as interaction domain, was found to be more flexible than the core domain. Examination of the B factor distribution along the Cα main chain indicated that the interaction domain had a higher average value (32 $Å^2$) compared to the rest of the protein (28.1 $Å^2$) and to the overall structure (29.2 $Å^2$).

Two extended hydrogen bond networks between the core and the interaction domains significantly reduce the mobility of the latter. One of these networks is centered around His49

and involves Ser82 Oγ - His49 $N\varepsilon_2$ and Arg83 O - Ala50 N which stabilize the B3 strand. The second network is organized around His100 and includes the following H-bonds: Val98 O - Asp53 N, His100 N - Tyr51 O, His100 O - Tyr51 N and His100 $N\varepsilon_2$ - Asp53 $O\delta_1$. Residues His49 and His100 form a hinge-like device linking the core and interaction domains, which would make the molecule quite flexible, not taking into account the contribution of the hydrogen bond networks. Furthermore, a hydrogen bond network bridged by water molecules and involving residues 51 and 86 on the one hand, and residues 87 and 62, 64, 66, and 67 on the other, further stabilizes the loop preceding helix H3.

### 3.4. Comparison with related [2Fe-2S] ferredoxins

Overall rms deviations of 1.0, 2.0 and 2.4 Å were determined between the main chain Cα atoms of FdVI and those of Pdx, Adx and Fdx, respectively. The largest differences were observed between the interaction domains, C-termini and loop regions. From the sequence alignment shown in figure 1, there are fourteen amino acid residues that are strictly invariant, four of which are the ligand cysteines 39, 45, 48 and 86 that link the [2Fe-2S] cluster to the protein (Fig. 3). The remaining invariant residues, Gly20, Thr47, His49, Glu67, Ile71, the Ser82 - Gln87 segment and Pro102 distinguish FdVI and homologous Fds from the plant-type Fds. Thr47 and His49 form with the ligand residue Cys48 a highly conserved pattern (Thr-Cys-His) among vertebrate-type ferredoxins. Thr47 $O\gamma_1$ interacts through a OH-Sγ bond with the Cys45 cluster ligand which might be relevant to protein stability as suggested from studies on variants of Adx obtained by site-directed mutagenesis [36]. In addition, the hydroxyl group of Thr47 is within H-bonding distance of the amide group of residues Ala37 and Asp38, a feature that is only observed in FdVI.

Like in other vertebrate-type Fds, residue Glu67 has an important role in FdVI as it forms a salt-bridge with Arg83. Disruption of this salt-bridge in human ferredoxin through site-directed mutagenesis led to misfolding of the protein [37], suggesting that such a salt-bridge is essential for the correct folding of the Fds [9]. Furthermore, Glu67 O is bonded to Met70 N and Ile71 N, the latter residue being in turn linked to Ala74 through a Ile71 O-Ala74 N bond.

The Ser82-Gln87 segment around the fourth ligand cysteine, is highly although not fully conserved among vertebrate-type Fds [9](Fig. 1). The hydrogen bond between Gln87 and Leu84 creates a hairpin loop that traps Cys86 in position. Also, the H bond between Gln87 $N\varepsilon_2$ and Cys86 Sγ is a feature shared by all vertebrate ferredoxins.

Pro102 is also conserved throughout this class of ferredoxins. Like Thr47 already mentioned above, Pro102 plays a key role in maintaining the 3-D fold. It is involved in H-bonding networks with neighbor residues Leu101 and Glu103. In the FdVI molecule, Pro102 is connected through a water bridge to residues Glu7 and His8, thus providing a link between the C and N termini. Furthermore, Pro102 is in contact with the aromatic polar residues His49, Tyr51 and Tyr75 providing additional links between the core and the interaction domains.

A further analysis of the interactions that might be important for the 3D folding highlights the role of several water molecules. Such buried water molecules may not only help stabilizing the C-terminus and the interaction domain but also contribute to the redox properties of the [2Fe-2S] cluster via the H-bonding network that extends all the way from the cluster to the surface of the molecule.

### 3.5. The [2Fe-2S] cluster environment

The [2Fe-2S] cluster is located at the edge of the molecule and shows a general configuration remarkably similar to that found in Adx, Pdx and Fdx, featuring a unique amide-to-sulfur NH-S bond Gln87 NE2 – Cys86 SG as well as the hydroxyl-to-sulfur OH-S bond between Thr47 OG – Cys45 SG. The plant-type Fds do not contain these two interactions.

The cluster lies 5 to 6 Å below the surface in a pocket delimited by residues forming the cluster binding loop (Cys39 to Cys48), as well as by residues Met24, Met70 and Gln87. It is buried in an hydrophobic environment in which none of the cysteine ligands is in contact with the solvent. An internal channel lined by residues Gly40, Arg28, Glu25 and Asp29 would give S2 access to the solvent, if the passage were not obstructed by the Met24 side chain. The low accessibility of the cluster to the solvent, which was also observed in Adx, contrasts with the situation found in plant-type ferredoxins, as previously reported by Müller et al.[2] Indeed, in *Anabaena* Fd as well as in related Fds, a funnel leads from the surface of the molecule to the vicinity of the cluster, resulting in the exposure of the Cys49 cluster ligand to the solvent. This difference in solvent accessibility established on the basis of structural grounds might explain why the mid-point redox potential of plant-type ferredoxins is sensitive to temperature and changes of the solvent ionic strength, whereas that of vertebrate-type Fds is much less sensitive to such changes.

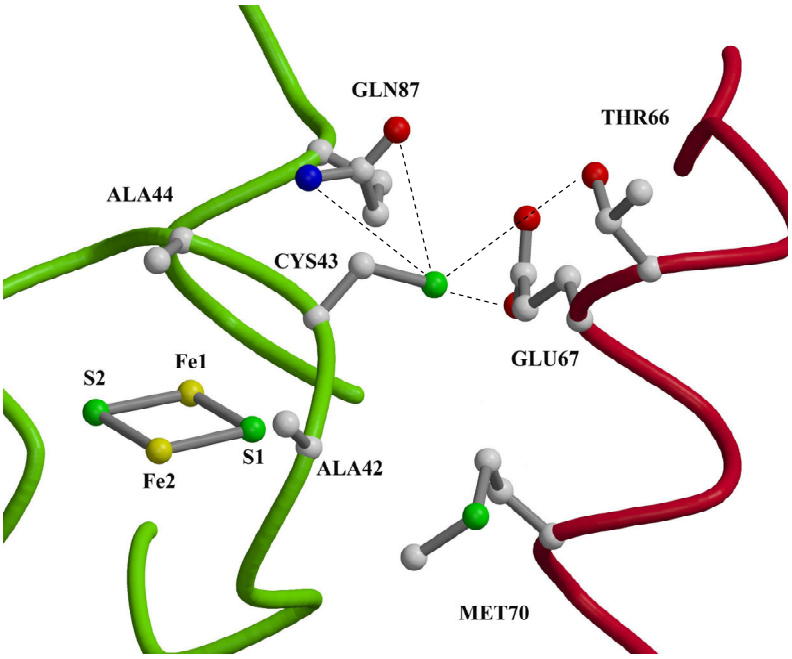



**Figure 4.** Closeup view of the polypeptide region surrounding Cys43 in FdVI. The drawing shows the [2Fe-2S} cluster, the polypeptide backbone of the core (green) and the interaction (red) domains, and the side chains of residues in close proximity to Cys43. Distances between Cys43 S and nearest atoms are indicated as broken lines: Cys43 S – Gln87 Nε, 4.1 Å; Cys43 S – Gln87 Oε, 4.14 Å; Cys43 S – Thr66 Oγ, 4.41 Å; Cys43 S – Thr66 O, 3.82 Å; Cys43 S – Glu67 Oε1, 6.16 Å; Cys43 S – Glu67 Oε2, 6.91 Å. The figure was drawn with a modified version of MOLSCRIPT [45] and rendered with Raster 3D [46]

FdVI features one non-ligand cysteine, Cys43, located in the cluster-binding loop preceding strand B3. An extra cysteine is present in the same position only in ferredoxins involved in the biogenesis of Fe-S clusters, such as *E. coli* Fdx, while it is missing in Adx and Pdx (Fig. 1). In FdVI, Cys43 makes bonds only with its neighbor residues Ala42 and Ala44. Examination of the Cys43 environment indicates that its side chain is in van der Waals contact with the side chain of Glu67 and the S1 atom of the [2Fe-2S] cluster. Interestingly, the side chain of Cys43 is oriented towards the protein surface so that its reactive thiol group is exposed to the solvent and surrounded by hydrophilic residues (Thr 66, Glu 67, Gln 87) (Fig. 4). A comparison of FdVI Cys43 with its counterpart in Fdx (Cys46) shows that they are located on the same side of the cluster, and that their thiol group is similarly pointed towards the solvent. In *E. coli* Fdx, Cys46 has been proposed to provide, through its thiol group, a potential ligand for a sulfur atom or a $Fe^{3+}$ ion, during the process of Fe-S cluster assembly [11].

### 3.6. Charge distribution and Interaction domain

Overall, the electrostatic potential of the surface of FdVI is almost neutral (Fig.5) which results into a remarkably small dipole moment compared to most ferredoxins. The asymmetric distribution of charges at the surface of [2Fe-2S] ferredoxins, with more negative charges near the active site, is generally thought to favor the correct orientation of Fds while docking with their redox partner. Four acidic residues of Adx (Asp79, Asp76, Glu73 and Asp72) were shown to be involved in the interaction with cytochrome P450[37]. Two of these residues, Asp76 and Asp79, are also required for the recognition of the reductase. In FdVI, only two of these negatively charged residues are conserved (Asp69/Asp72 in place of Asp76/Asp79 in Adx), whereas Pro65 and Thr66 replace Asp72 and Glu73 in Adx. A pair of aspartate residues in position equivalent to Asp69 and Asp72 in FdVI is also encountered in most ferredoxin sequences (Fig. 1), including Fds thought to participate in iron-sulfur cluster biosynthesis in γ-*Proteobacteria* and yeast. A representative of this subclass of Fds, *E. coli* Fdx, displays on its surface, numerous acidic residues which form two markedly negative patches on one side of the molecule[11]. In FdVI, most of the corresponding residues are replaced by hydrophobic or basic residues (Fig. 1). In this respect, FdVI is similar to Pdx which also showed an atypical distribution of charges resulting in a neutral surface charge[12]. Pdx lacks the pair of Asp residues supposedly involved in the interaction with the reductase in other Fds. Instead, it was found that Glu72 and Cys73 were important for the interaction of Pdx with its cognate reductase, whereas Asp38 and Trp106 are required for binding to the cytochrome P450cam[38]. Kinetics studies showed that the association of Pdx with the reductase was driven by non-electrostatic interactions [39].

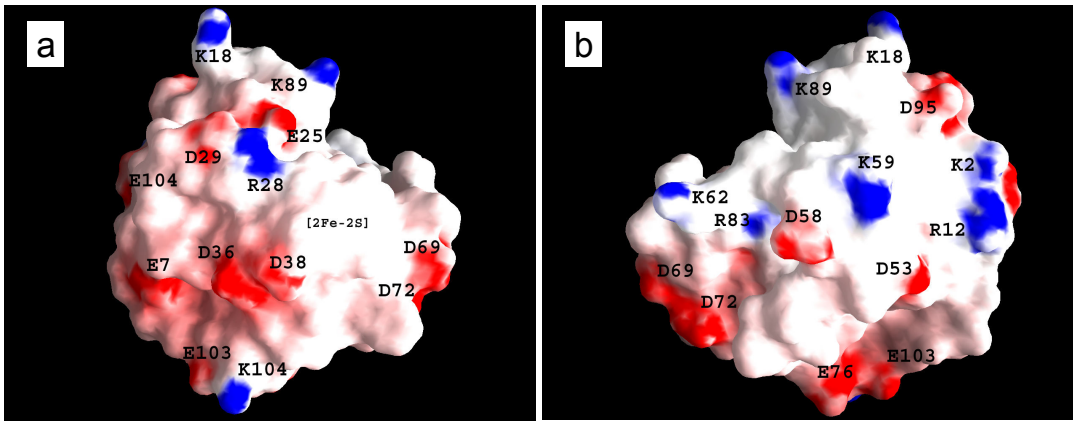



**Figure 5.** Distribution of electrostatic potentials on the surface of oxidized FdVI. Basic and acidic residues are shown in blue and red, respectively. The views shown in (a) and (b) were obtained by rotating the molecule by 180 degrees around a vertical axis. The figure was drawn using GRASP [47]

### 3.7. Structural changes induced upon reduction

EPR and spectrophotometric measurements showed that FdVI was readily and fully reduced by 2 mM dithionite when in solution in the crystallization buffer at a protein concentration around 0.5 mM (Fig. S2 in supplementary material). However, the reduced protein failed to crystallize when incubated under conditions similar to those promoting the crystallization of the oxidized form. As another option, oxidized crystals were soaked in a defined reducing solution and the extent of FdVI reduction was checked by micro-spectrophotometry. As illustrated in Fig. 6, the spectrum of the reduced crystal exhibited a lower absorption in the range 380-650 nm and a prominent absorption band around 544 nm, similar to the reduced protein in solution (Fig. S2). The $A_{414}/A_{544}$ ratio, taken as an indicator of the extent of reduction, was calculated to be 2.80 and 1.30 for the oxidized and the reduced crystal, respectively. The corresponding values for the protein in solution were 2.65 for the oxidized, and 1.60 for the fully reduced protein. While these data suggested that the ferredoxin in the crystal underwent extensive reduction, accurate determination of the extent of reduction was precluded by the fact that the spectrum of the reduced crystal exhibited distinct features compared to that of the protein in solution, including the occurrence of an additional absorption band near 440 nm.

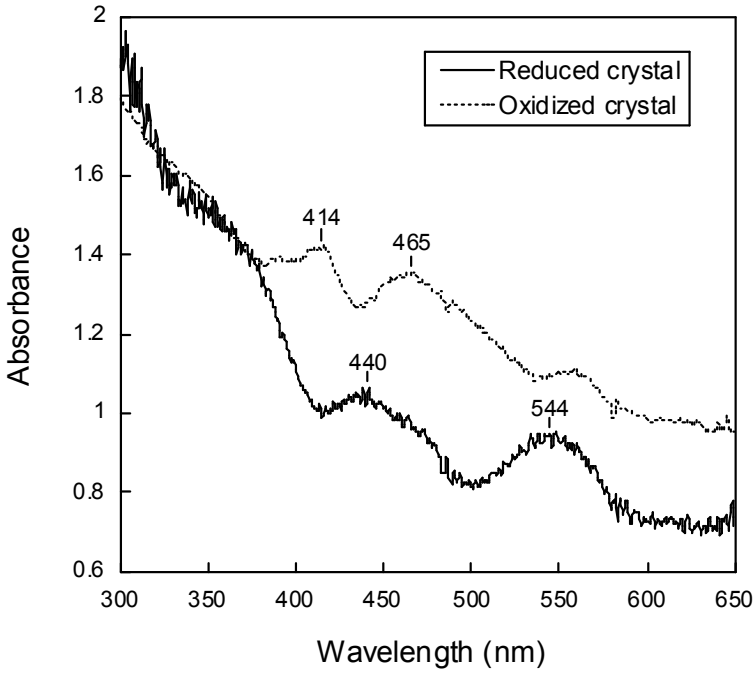



**Figure 6.** Optical absorption spectra of FdVI crystals. Spectra were recorded at 100 K on either an oxidized crystal (broken line) or a crystal reduced with dithionite (continuous line). The same reduced crystal was used for X-ray structural determination. The two spectra are presented with a vertical offset for the sake of clarity. Relevant absorption maxima are indicated.

The structure of the reduced crystal was solved at 2.0 Å resolution. The fold of FdVI in the reduced crystal was found to be almost identical to that of the oxidized molecule, with an overall rms deviation of 0.9 Å. As shown in Fig. 4, largest differences between the reduced and oxidized molecules are observed in the interaction domain and in the vicinity of the cluster. Analysis of the density maps of the reduced crystal strongly suggested that the cluster underwent distortion. Density greater than $3\sigma$ in the difference Fourier maps was detected near the cluster after one cycle of refinement, indicating that the cluster geometry changed from a planar to a distorted lozenge (Fig.7).

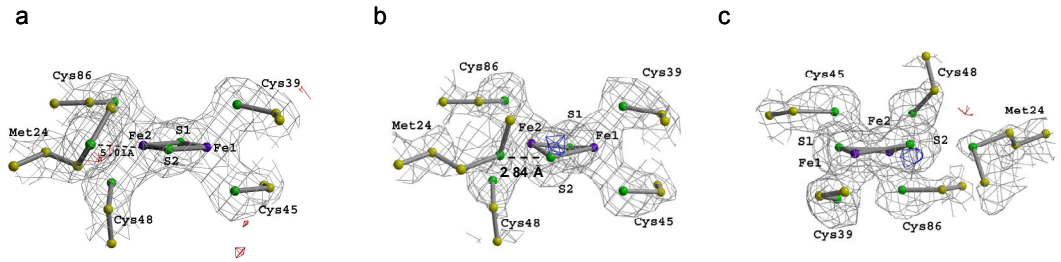



**Figure 7.** Close-up view of the [2Fe-2S] environment in the oxidized and reduced crystals. The drawings show the cluster, the four cysteine ligands and the Met24 side chain with the 2Fo-Fc and Fo-Fc densities contoured at 1σ (grey), +3σ (blue) and -3σ (red). The conformation observed in the oxidized crystal (a) is compared to the average conformation observed in the reduced crystal (b). Panel c is another view of the reduced cluster in a different orientation, showing positive difference electron density in the vicinity of the S2 atom. The distance between Met24 Sδ and S2 is indicated.

Calculation of an omit map at this stage of the refinement confirmed the distortion of the cluster (Fig. S3). In the fully refined model, positive density was also observed above the S2 atom of the cluster, suggesting some dynamical disorder or an alternate conformation of the cluster. Based on the electron density maps, the S2 atom would occupy two positions 0.5 Å apart, which could result from a translation of this atom in a plane perpendicular to the cluster plane. In figure 7b and 7c, the S2 atom is shown in an average position which corresponds to a minimum of density in the difference Fourier map. Positive density was also detected in the vicinity of the S1 atom, suggesting a possible movement of this atom which could not be accurately determined at 2 Å of resolution. Despite this conformational change, the Fe-S distances within the cluster underwent no significant variation (Table 3).

Examination of difference Fo-Fc maps also indicated the presence of alternate conformations for two residues, Met24 and Ala44. One of the two conformations was virtually identical to that found in the oxidized crystal. In the other conformation, the side chain of Met24 underwent a 180$^{o}$ rotation, approaching the S2 atom of the cluster at 2.84 Å (Fig. 7b). At the same time, the H-bond between Met24 O and Arg28 N which was observed in the oxidized molecule was broken. This second conformation of Met24 accounted for about 40% of the molecules in the crystal.

| Parameter | Oxidized | Reduced |
|---|---|---|
| **Distances** | | |
| **S1 S2** | 3.6 | 3.6 |
| **Fe1 Fe2** | 2.8 | 2.8 |
| **S1 Fe1** | 2.2 | 2.3 |
| **S1 Fe2** | 2.3 | 2.2 |
| **S2 Fe2** | 2.3 | 2.3 |
| **S2 Fe1** | 2.3 | 2.4 |
| **Fe1 S39** | 2.3 | 2.2 |
| **Fe1 S45** | 2.3 | 2.3 |
| **Fe2 S86** | 2.4 | 2.3 |
| **Fe2 S48** | 2.2 | 2.3 |
| **Angles** | | |
| **Fe1 S1 Fe2** | 75.7 | 75.7 |
| **Fe1 S2 Fe2** | 74.8 | 72.6 |
| **S1 Fe2 S2** | 105.1 | 105.9 |
| **S1 Fe1 S2** | 104.4 | 101.9 |
| **Planarity** | 1.6 | 15 |

**Table 3.** Geometry of the FdVI [2Fe-2S] cluster in the oxidized and reduced crystals. Interatomic distances in the [2Fe-2S] cluster are given in Å. Angles are in degrees.

In the oxidized ferredoxin, the Ala44-CO pointed away from the cluster ("CO out"), whereas In the reduced crystal, Ala44 showed an additional conformation where the carbonyl came closer to the cluster S1 atom, at a distance of 3.9 Å ("CO in"). The "CO out" and "CO in" conformations had calculated occupation rates of about 60% and 40%, respectively, in the reduced crystal. In the "CO in" conformation, the Ala44 carbonyl turned toward Met70, thus reducing the distance from Ala44-O to Met70-S from 4.7 to 3.0 Å. This conformational flip of the Ala44 carbonyl group was associated with higher B-factors for neighboring residues. The alternate conformations of Met24 and Ala44 observed in the reduced crystal likely represent the main structural changes occurring in the vicinity of the cluster upon reduction.

When compared to oxidized FdVI, the reduced ferredoxin was characterized by a higher level of H-bonding, especially at the interface between the core and the interaction domains. Moreover, the reduced crystal showed a higher content of water molecules (Table 2), many of which were involved in bridges linking distant residues. For example, in the reduced ferredoxin, the C-terminal Leu101 residue is linked to the N-terminal His8 through two water

molecules. An extension of the hydrogen bonding network was also noticed around residues Arg28 and His49. In the reduced ferredoxin, four water molecules were involved in links between Arg28 and residues Thr22, Glu25, Ala37 and Gly40, thereby stabilizing the helix H1 and the loop preceding β-sheet B3. In the vicinity of His49, two new bonds linked Tyr75 O and Tyr75 N to His49 Nδ1 through a water molecule, and another water molecule connected Gln105 to His49 and Tyr75. Taken together, the changes observed in the FdVI crystal upon reduction all contributed to decrease the flexibility of the protein and to tighten the molecular interaction between the core domain, the interaction domain and the C-terminus. This is reflected by a lower average B factor found for the reduced crystal compared to the oxidized crystal. Similar differences in B factors were observed when comparing the data collected on three oxidized crystals with those obtained for another reduced crystal.

### 3.8. Comparison of redox-linked structural changes in FdVI, Pdx, Adx and AnFd

So far, high resolution crystal structures of reduced [2Fe-2S] ferredoxins have been described in only two cases, the plant-type *Anabaena* PCC7119 Fd (AnFd) [40] and Pdx [41]. As for FdVI, the redox-mediated changes were deduced from a comparison of the X-ray structures of the relevant proteins obtained from oxidized and reduced crystals. In the plant-type AnFd, reduction gave rise to relatively limited conformational changes around residues Cys46, Ser47 and Phe65 [40]. In the two vertebrate Fds considered herein, structural changes included a movement of the cluster, conformational shifts of residues in the vicinity of the cluster, and an extension of the H-bonding network linking the core and the interaction domains. The FdVI cluster switched from a planar to distorted lozenge geometry, a conformational change that has not been observed in the Fds described so far. As the significance of this change is still unclear, it is worth noting that some [2Fe-2S] Fds, including AnFd, have been shown to harbor a non-planar cluster[40]. In FdVI, two redox-induced conformational shifts occurred in the vicinity of the cluster, involving residues Ala44 and Met24. The movement of the Met24 side chain resulted in the S□ atom approaching the S2 atom of the cluster at less than 2.9 Å. The proximity of this electron rich sulfur atom might stabilize the reduced state of the cluster or tune its redox potential by changing the electrostatic environment. The potential role of the Met24 residue in controlling the redox potential of FdVI can be investigated by amino acid substitution through site-directed mutagenesis.

The flip of the Ala44 carbonyl which points towards the cluster in the reduced state (COin), is reminiscent of a similar carbonyl flip previously observed in Pdx and AnFd, with two important differences: in the latter cases, the CO flip concerns the cysteine residue adjacent to Ala44, and the CO orientation is opposite, pointing outwards in the reduced state. In AnFd, the Cys46-Ser47 link lies in close contact with the Phe65 aromatic ring which plays a crucial role in electron transfer to the cognate reductase FNR. From this observation, it has been proposed that the carbonyl flip might trigger the dissociation of the AnFd-FNR complex after electron transfer [42]. In reduced Pdx, the conformational shift of the Cys45 carbonyl allows formation of a H bond between Ala46N and the S1 atom of the cluster, and initiates a readjustment leading to a tighter interaction between the cluster and the peptide shell[41] . In reduced FdVI, the corresponding carbonyl stayed in the COin orientation while the carbonyl of the preceding peptide bond turned to face Met70. Interestingly, Met70 occupies a position similar to Phe65 in AnFd, and given that Phe65 plays a critical role in the electron transfer to FNR, it is inferred that the CO switch of Ala44 might be involved in the interaction of FdVI with its natural partner.

FdVI reduction was also accompanied by an extension of the H bond network which further rigidified the structure of the molecule and prevented movement between the core and the interaction domains. A similar increase of the intramolecular H bonding was observed in Pdx upon reduction, which not only stabilized the peptide binding loop around the cluster but also modified the surface area supposed to interact with cytP450 [41]. In Adx, an NMR study showed that reduction induced conformational changes in the C-terminal region of the molecule, causing the dissociation of the dimeric oxidized form[43]. In another study, it was found that oxidized Adx was more dynamic than the reduced protein, and that His56, which lies at the interface between the core and the interaction domains, was required for the redox dependent changes observed upon reduction[44]. Likewise, the His49 residue of FdVI which occupies a position similar to His56 in Adx, was found to play a pivotal role in linking the core and the interaction domains and in stabilizing the reduced protein.

Despite obvious structural similarities between FdVI, Pdx and Adx, and to a lesser extent AnFd, the work described here and in previous studies[40,41,43] indicates that the three proteins exhibited clearly distinct behaviour upon reduction. The observed differences likely reflect functional adaptation between each ferredoxin and its cognate protein partners.

## Acknowledgement

We thank Christine Meyer for helping in the purification of FdVI. We are indebted to Emile Duée and Eric Fanchon for helpful discussion. Thanks to D. Bourgeois, X. Vernede, R. Morales and J. Fontecilla for helpful advices and for giving us access to the facilities which allowed to obtain and analyze reduced crystals of FdVI. We wish to thank Valerie Biou, Janet Smith and Andrew Thompson, for valuable suggestions and support. Funding for this project was provided by the Centre National de la Recherche Scientifique, the Commissariat à l'Energie Atomique, the European Synchrotron Radiation Facility, and the NIGMS under agreement Y1 GM-0080.

## References


[1] Bruschi, M. and Guerlesquin, F. (1988) *FEMS Microbiol. Rev*. **54**, 155-176.

[2] Muller, J.J., Muller, A., Rottmann, M., Bernhardt, R. and Heinemann, U. (1999) *J. Mol. Biol.* **294**, 501-513.

[3] Bernhardt, R. (1996) *Rev Physiol Biochem Pharmacol* **127**, 137-221.

[4] Baldwin, J.E., Morris, G.M. and Richards, W.G. (1991) *Proc R Soc Lond B Biol Sci* **245,** 43-51.

[5] Berg, A., Gustafsson, J.A. and Ingelman-Sundberg, M. (1976) *J Biol Chem* **251**, 2831-2838.

[6] Zheng, L., Cash, V.L., Flint, D.H. and Dean, D.R. (1998) *J Biol Chem* **273**, 13264-13272.

[7] Takahashi, Y. and Nakamura, M. (1999) *J. Biochem. (Tokyo)* **126**, 917-926.

[8] Lange, H., Kaut, A., Kispal, G. and Lill, R. (2000) *Proc. Natl. Acad. Sci. U.S.A*. **97**, 1050-1055.

[9] Grinberg, A.V., Hannemann, F., Schiffler, B., Muller, J., Heinemann, U. and Bernhardt, R. (2000) *Proteins:Struct. Funct. Genet*. **40**, 590-612.

[10] Muller, A., Muller, J.J., Muller, Y.A., Uhlmann, H., Bernhardt, R. and Heinemann, U. (1998) *Structure* **6**, 269-280.

[11] Kakuta, Y., Horio, T., Takahashi, Y. and Fukuyama, K. (2001) *Biochemistry* **40**, 11007-11012.

[12] Sevrioukova, I.F., Garcia, C., Li, H.Y., Bhaskar, B. and Poulos, T.L. (2003) *J. Mol. Biol.* **333**, 377-392.

[13] Armengaud, J., Meyer, C. and Jouanneau, Y. (1994) *Biochem. J.* **300**, 413-418.

[14] Grabau, C., Schatt, E., Jouanneau, Y. and Vignais, P.M. (1991) *J. Biol. Chem*. **266**, 3294-3299.

[15] Naud, I., Vinçon, M., Garin, J., Gaillard, J., Forest, E. and Jouanneau, Y. (1994) *Eur. J. Biochem*. **222**, 933-939.

[16] Armengaud, J., Meyer, C. and Jouanneau, Y. (1997) J. Bacteriol. **179**, 3304-3309.

[17] Armengaud, J., Sainz, G., Jouanneau, Y. and Sieker, L.C. (2001) *Acta Crystallogr.* **D57**, 301-303.

[18] Bourgeois, D., Vernede, X., Adam, V., Fioravanti, E. and Ursby, T. (2002) *J. Appl. Cryst.* **35**, 319-326.

[19] Bertrand, P. and Gayda, J.P. (1979) *Biochim. Biophys. Acta* **579**, 107-121.

[20] Tanigushi, V.T., Sailasuta-Scott, N., Anson, F.C. and Gray, H.B. (1980) *Pure Appl. Chem*. **52**, 2275-2281.

[21] Evans, G. and Pettifer, R.F. (2001) *J. Appl. Cryst.* **34**, 82-86.

[22] Westbrook, E.M. and Naday, I. (1997) *Methods Enzymol.* **276**, 244-268.

[23] Otwinowski, Z. (1993) in: *Data Collection and Processing, Proceedings of the CCP4*, pp. 56-62 (Bailey, S., Ed.) SERC Laboratory, Daresbury, Warrington, UK.

[24] Otwinowski, Z. and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326.

[25] Brünger, A.T. et al. (1998) *Acta Crystallogr.* ***D*****54**, 905-921.

[26] Perrakis, A., Morris, R. and Lamzin, V.S. (1999) *Nat Struct Biol* **6**, 458-463.

[27] Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) *Acta Crystallogr* **A47**, 110-119.

[28] Navaza, J. (1994) *Acta Crystallogr* **A50**, 157-163.

[29] CCP4 (1994) *Acta Crystallogr*. **D50**, 760-763.

[30] von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. and Snel, B. (2003) *Nucleic Acids Research* **31**, 258-261.

[31] Ta, D.T. and Vickery, L.E. (1992) *J. Biol. Chem*. **267**, 11120-11125.

[32] Jung, Y.S., Gao-Sheridan, H.S., Christiansen, J., Dean, D.R. and Burgess, B.K. (1999) *J. Biol. Chem*. **274**, 32402-32410.

[33] Hase, T., Mizutani, S. and Mukohata, Y. (1991) *Plant Physiol*. **97**, 1395-1401.

[34] Battistuzzi, G., D'Onofrio, M., Borsari, M., Sola, M., Macedo, A.L., Moura, J.J. and Rodrigues, P. (2000) *J Biol Inorg Chem* **5**, 748-760.

[35] Huang, Y.Y. and Kimura, T. (1983) *Anal. Biochem* **133**, 385-393.

[36] Burova, T.V., Beckert, V., Uhlmann, H., Ristau, O., Bernhardt, R. and Pfeil, W. (1996) *Protein Sci* **5**, 1890-7.

[37] Coghlan, V.M. and Vickery, L.E. (1991) *J Biol Chem* **266**, 18606-12.

[38] Holden, M., Mayhew, M., Bunk, D., Roitberg, A. and Vilker, V. (1997) *J. Biol. Chem*. **272**, 21720-21725.

[39] Sevrioukova, I.F., Hazzard, J.T., Tollin, G. and Poulos, T.L. (2001) Biochemistry **40**, 10592-600.

[40] Morales, R., Charon, M.H., Hudry-Clergeon, G., Petillot, Y., Norager, S., Medina, M. and Frey, M. (1999) *Biochemistry* **38**, 15764-15773.

[41] Sevrioukova, I.F. (2005) *J Mol Biol* **347**, 607-21.

[42] Morales, R., Charon, M.H., Kachalova, G., Serre, L., Medina, M., Gomez-Moreno, C. and Frey, M. (2000) *EMBO rep*. **1**, 271-276.

[43] Beilke, D., Weiss, R., Lohr, F., Pristovsek, P., Hannemann, F., Bernhardt, R. and Ruterjans, H. (2002) *Biochemistry* **41**, 7969-78.

[44] Kostic, M., Bernhardt, R. and Pochapsky, T.C. (2003) *Biochemistry* **42**, 8171-82.

[45] Esnouf, R.M. (1997) *J. Mol. Graphics* **15**, 133-138.

[46] Merritt, E.A. and Murphy, M.E. (1994) *Acta Crystallog*r. **D50**, 869-873.

[47] Nicholls, A., Sharp, K.A. and Honig, B. (1991) *Proteins: Struct.Funct.Genet.* **11**, 281-296

## Supplementary material

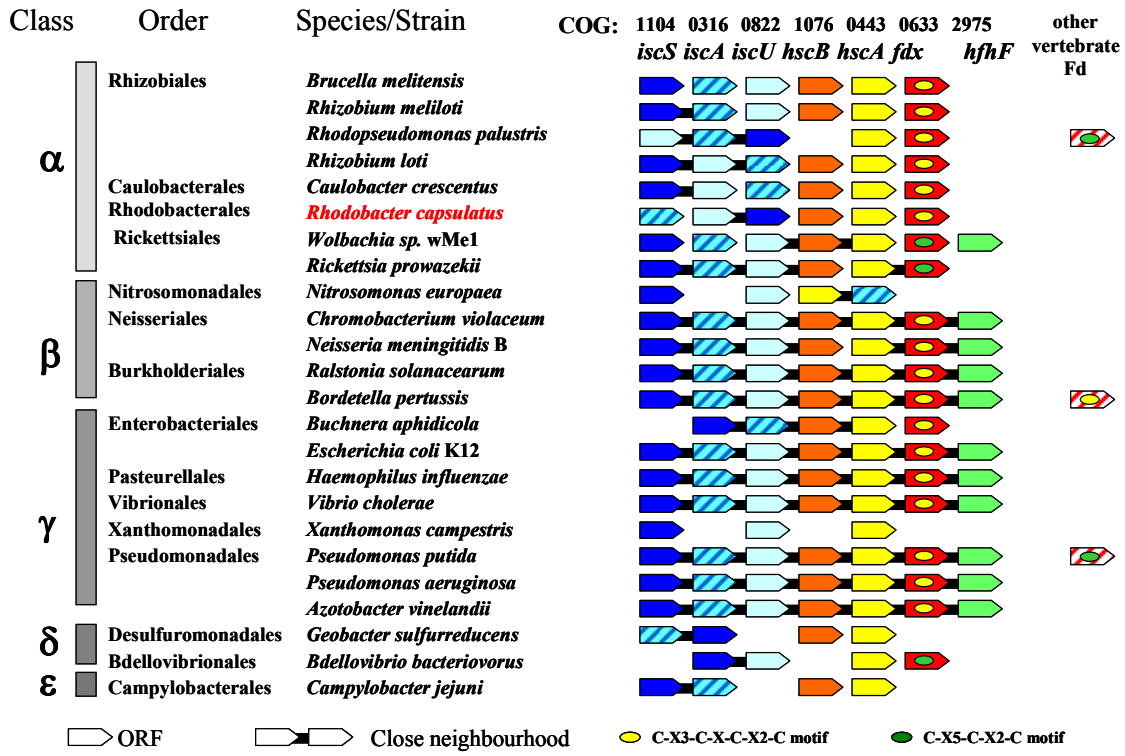



**Figure S1.** Occurrence of *isc*-genes and *isc*-linked *fdx* genes in the genomes of α-*Proteobacteria* and their neighbourhood relationships. Search for *isc*-genes was carried out on 65 completely sequenced genomes of proteobacteria using the STRING database (http://string.embl.de/) and confirmed with PSI-BLAST reciprocal searches. Neighbourhood relationships were checked by the STRING neighbourhood module and manually checked for *R. capsulatus SB1003* (http://www.integratedgenomics.com/genomereleases.html#list1). The result is depicted for 24 representative genomes (out of 65). Additional vertebrate-type ferredoxins identified by PSI-BLAST searches in the genomic sequences are indicated. These include the plasmid encoded putidaredoxin from *Pseudomonas putida*. A yellow symbol indicates the cluster-binding motif encountered in most *isc*-linked Fds (Cys-X3-Cys-X-Cys-X2-Cys), whereas a green symbol designates the general motif found in vertebrate Fds (Cys-X5-Cys-X2-Cys).

# Partie 3 : Conclusions et perspectives

La bactérie photosynthétique *Rhodobacter capsulatu*s synthétise six ferrédoxines, qui ont toutes été purifiées et caractérisées dans le laboratoire de Grenoble (Armengaud *et al.*, 1997 et références citées dans cet article). Quatre ferrédoxines, parmi lesquelles deux ferrédoxines à centre [2Fe-2S] de type plante, participent à la fixation de l'azote, car leur gène structural est situé dans un opéron de gènes *nif* (Grabau *et al.*, 1991 ; Armengaud *et al.*, 1994). La fonction physiologique des deux dernières ferrédoxines, appelées FdII et FdVI, est incertaine. FdII possède un centre [3Fe-4S] et un centre [4Fe-4S] (Jouanneau *et al*., 1990), et est analogue de la FdI de *Azotobacter vinelandii*, qui a été caractérisée en détail aux plans structural et fonctionnel (Chen *et al.*, 2000). Le gène *fdxA* codant FdII est exprimé de manière constitutive, et son inactivation par insertion s'est révélée impossible, suggérant qu'il s'agit d'un gène essentiel (Duport *et al*., 1992). Chez *A. vinelandii*, l'inactivation du gène codant FdI a pour conséquence la surexpression d'une protéine de type NADH-Fd réductase impliquée dans la réponse au stress oxydant (Regnstrom *et al*., 1999). Curieusement, la délétion de FdI déclenche aussi un accroissement de la synthèse d'une autre Fd à centre [2Fe-2S], FdIV, qui est proche de FdVI de *R. capsulatus*. Cette protéine est codée par un gène qui fait partie de l'opéron *isc* chez *A. vinelandii*, et est donc impliqué dans la biosynthèse de centres Fe-S (Jung *et al.*, 1999).

Le rôle de FdVI chez *R. capsulatus* est toujours incertain car, contrairement au cas de *A. vinelandii* et de *E. coli*, le gène *fdxE* n'est pas associé aux gènes *isc*. L'analyse du génome de cette bactérie (presque complètement séquencé) indique qu'il n'y pas d'autres gènes de ferrédoxine que ceux correspondant aux six Fds déjà identifiées. D'autre part, les gènes *isc* ne sont pas regroupés en opéron comme c'est la cas chez les gamma Protéobactéries. Notre analyse des génomes bactériens connus a révélé que cette situation prévaut chez les bactéries de groupe des alpha Protéobactéries auquel appartient *R. capsulatus*. D'autre part, comme l'inactivation du gène *fdxE* affecte une fonction vitale de la bactérie, l'hypothèse d'une participation de FdVI à la biogenèse des centres Fe-S devenait très plausible.

L'étude structurale présentée ici conforte ce jugement car FdVI possède certaines caractéristiques propres aux ferrédoxines associées au système ISC. Notamment, FdVI contient une cinquième cysteine, en position 43 accessible au solvant et à proximité du centre Fe-S (5 Å du S1 du Fe-S). La cystéine équivalente dans la séquence de Fdx de *E. coli* est positionnée de la même manière dans la structure 3D avec le groupement thiol pointé vers

l'extérieur. Le rôle de cet acide aminé ainsi que la fonction précise des ferrédoxines de type ISC dans la biosynthèse des centres Fe-S reste cependant une énigme. Le fait que certaines bactéries soient dépourvues de gènes *isc* et possèdent un autre système d'assemblage des centres Fe-S (système suf) sans ferrédoxine spécifiquement associée, suggère que la biosynthèse de ces centres n'exige pas de transporteur d'électrons particulier.

Cette étude a également porté sur les changements structuraux liés au passage de l'état oxydé à l'état réduit dans une petite protéine à Fe-S. Lorsque nous avons entrepris de résoudre la structure de FdVI sous forme réduite, seule une Fd à centre [2Fe-2S] avait été étudiée par cristallographie dans les deux états redox, la Fd de type plante de *Anabaena* (Morales *et al*., 1999). Apres avoir réduit un cristal de FdVI et contrôlé l'état de réduction de la protéine par une méthode originale de spectroscopie appliquée à un microcristal, la structure de la forme réduite a été obtenue à une résolution de 2 Å. Les principaux changements structuraux observés lors de la réduction portent sur (i) la conformation du centre Fe-S, (ii) la rotation de la chaîne latérale d'une méthionine dont le soufre s'approche à 2.8 Å de l'atome S2 du centre Fe-S, et (iii) l'augmentation du nombre de liaisons H qui rigidifient la structure. Ces observations ont relativement peu de points communs avec celles faites dans le cas de la Fd de *Anabaena*, et dans le cas de la putidaredoxin (Sevrioukova, 2005). Dans la Fd de *Anabaena*, la principal changement structural est une rotation du carbonyl autour de la liaison peptidique Cys46-Ser47 proche du cluster qui se traduit par un éloignement de l'atome d'oxygène du cluster ($CO_{in}$-$CO_{out}$ ; Morales *et al*., 1999). Il a été proposé que cette modification agisse comme un switch moléculaire qui déclenche la dissociation du complexe entre la Fd et son partenaire naturel, la FNR (Morales *et al*., 2000). Un changement structural redox-dépendant très similaire a été observé dans la putidaredoxine, au niveau du CO de la liaison Cys45-Ala46 (Sevrioukova, 2005), mais on ne sait pas si, dans ce cas, un tel changement a une incidence sur l'interaction avec le partenaire redox. En tous cas, nous n'avons pas trouvé de déplacement équivalent dans FdVI. Il ne s'agit donc pas d'un mécanisme général de réponse moléculaire au changement d'état redox, qui serait propre aux Fds à centre [2Fe-2S]. Dans FdVI, la déformation du cluster Fe-S sous l'effet de sa réduction, si faible soit elle, n'avait jamais encore été observée dans une protéine Fe-S. De même, le rapprochement du soufre de la Met24 de l'atome S2 du cluster n'a pas d'équivalent. Ces observations structurales suscitent plus de questions qu'elles n'apportent de réponses, mais on peut donner quelques pistes de recherche pour poursuivre notre travail.

La contribution de la methionine 24 peut être étudiée par des expériences de

mutagenèse dirigée visant à remplacer cet acide aminé par une alanine, une leucine ou une isoleucine. Ces deux derniers acides aminés sont fréquemment rencontrés dans cette position dans les Fds de la famille de FdVI. Des substitutions par des acides aminés polaires (Ser, Cys, Asn) ou chargés (Asp) peuvent aussi être réalisées afin d'examiner l'effet de la proximité de la chaîne latérale de ces résidus sur les propriétés du cluster. La stabilité ainsi que les propriétés redox et spectroscopiques de ces variants pourraient être étudiées.
Le rôle éventuel de la cysteine 43 peut être déterminé par des études de spectroscopie RPE, EXAFS et /ou des analyses cristallographiques afin de préciser si FdVI peut fixer du fer et/ou du soufre par le biais du groupement thiol. Nous pouvons aussi déterminer si l'apoprotéine peut spontanément incorporer un centre [2Fe-2S] et/ou lier des atomes de fer et/ou de soufre.

L'étude de FdVI et notamment les aspects structure-fonction, dépendent de l'identification des partenaires redox de cette Fd. Le génome de *R. capsulatus* étant presque entièrement séquencé, on peut rechercher des gènes de la machinerie ISC dont les produits sont susceptibles d'interagir avec FdVI. Chez *E. coli,* les protéines Fdx et IscA interagissent de façon spécifique, et le centre [2Fe-2S] est efficacement transféré de IscA vers l'apoferrédoxine (Ollagnier-de-Choudens *et al.*, 2001). On pourra rechercher une telle interaction entre FdVI et les protéines de type iscA, iscS ou iscU de *R. capsulatus*.

Armengaud, J., Meyer, C. and Jouanneau, Y. (1994). Recombinant expression of the fdxD gene of Rhodobacter capsulatus and characterization of its product, a [2Fe-2S] ferredoxin. *Biochem. J.* **300**:413-418.

Armengaud, J., Meyer, C. and Jouanneau, Y. (1997) A [2Fe-2S] ferredoxin (FdVI) is essential for growth of the photosynthetic bacterium Rhodobacter capsulatus. *J. Bacteriol*. **179**:3304-3309.

Chen, K., Hirst, J., Camba, R., Bonagura, C.A., Stout, C.D., Burgess, B.K. and Armstrong, F.A. (2000). Atomically defined mechanism for proton transfer to a buried redox centre in a protein. *Nature* **405**:814-817.

Duport, C., Jouanneau, Y. and Vignais, P. M. (1992). Transcriptional analysis and promoter mapping of the fdxA gene which encodes the 7Fe ferredoxin (FdII) of Rhodobacter capsulatus. Mol. Gen. Genet. 231:323-328.

Grabau, C., Schatt, E., Jouanneau, Y. and Vignais, P.M. (1991). A new [2Fe-2S] ferredoxin from Rhodobacter capsulatus. Coexpression with a 2[4Fe-4S] ferredoxin in Escherichia coli. *J. Biol. Chem.* **266**:3294-3299.

Jouanneau, Y., Meyer, C., Gaillard, J. and Vignais, P. M. (1990). Purification and characterization of a 7Fe-ferredoxin from Rhodobacter capsulatus. *Biochem. Biophys. Res. Commun.* **171**:273-279.

Jung, Y.S., Gao-Sheridan, H.S., Christiansen, J., Dean, D.R. and Burgess, B.K. (1999). Purification and biophysical characterization of a new [2Fe-2S] ferredoxin from Azotobacter vinelandii, a putative [Fe-S] cluster assembly/repair protein. *J. Biol. Chem*. **274**:32402-32410.

Regnstrom, K., Sauge-Merle, S., Chen, K. and Burgess, B.K. (1999). In Azotobacter vinelandii, the E1 subunit of the pyruvate dehydrogenase complex binds fpr promoter region DNA and ferredoxin I. *Proc. Natl. Acad. Sci. U. S. A*. **96**:12389-12393.

Morales, R., Charon, M.H., Hudry-Clergeon, G., Petillot, Y., Norager, S., Medina, M. and Frey, M. (1999). Refined X-ray structures of the oxidized, at 1.3 Å, and reduced, at 1.17 Å, [2Fe-2S] ferredoxin from the cyanobacterium Anabaena PCC7119 show redox-linked conformational changes. *Biochemistry* **38**:15764-73.

Sevrioukova, I. F. (2005). Redox-dependent structural reorganization in putidaredoxin, a vertebrate-type [2Fe-2S] ferredoxin from Pseudomonas putida. *J. Mol. Biol.* **347**:607-621.

# CHAPITRE 2

## Caractérisation biochimique et structure cristallographique de la dioxygénase de *Sphingomonas* CHY-1 (PhnI), catalyseur de l'attaque initiale des hydrocarbures aromatiques polycycliques

# Partie 1 : Introduction

## 1. Structure des hydrocarbures aromatiques polycycliques

Les hydrocarbures aromatiques polycycliques (HAP) sont des composés organiques ubiquitaires et rémanent dans l'environnement. Ils sont constitués de noyaux aromatiques (benzéniques) condensés ne contenant pas d'hétéroatome. On trouve à peu près une centaine de HAP dont la taille varie de deux à sept noyaux. Les structures et nomenclatures de 18 HAP sont illustrées dans la figure 1

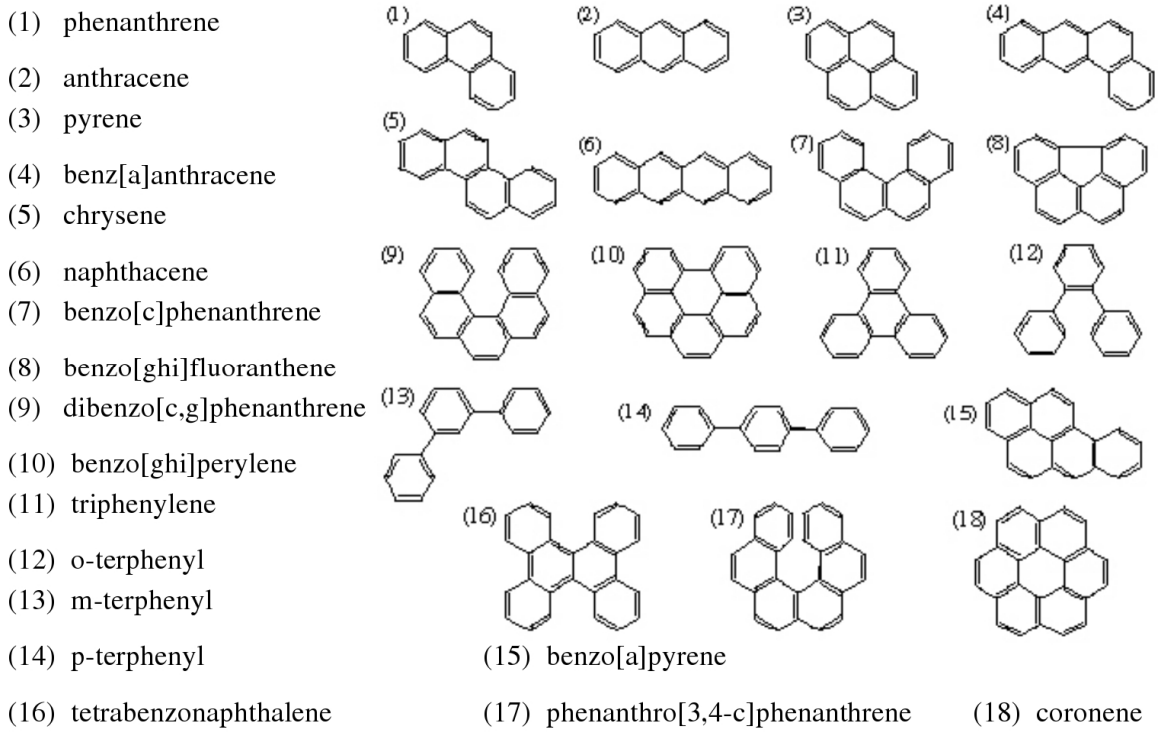



**Figure 1**. Structure des HAP contenant 3 noyaux et plus. La structure du naphtalène n'est pas présentée étant constituée de 2 noyaux seulement. Provenance de la figure : chrom.tutms.tut.ac.jp.

Les HAP sont très stables, peu volatils et hydrophobes, donc très peu solubles dans l'eau. La solubilité dépend en partie de la masse moléculaire, plus le HAP contient de noyaux moins il est soluble**.** Des solubilités dans l'eau de 32, 0.07, 0.16 et 0.004 mg/L ont été déterminées pour le naphtalène, l'anthracène, le pyrène et le benzo[a]pyrène (BaP),

respectivement (Bouchez *et al.*, 1995).

La plupart des HAP sont toxiques pour les organismes vivants, et certains d'entre eux, comme le benzo[a]pyrène, sont catalogués composés génotoxiques, mutagènes et cancérigènes (Randerath *et al*., 1999)

.

## 2. Origine et exposition aux HAP

Les HAP sont présents dans l'air, l'eau et les sols; ils sont produits lors de la combustion incomplète de matériaux organiques tel que le bois, les combustibles fossiles (charbon, pétrole), les ordures ménagères, le tabac. La gazéification de la houille et des charbons en a produit beaucoup jusque dans les années 60, ce qui est à l'origine de centaines de friches industrielles contaminées. Ils sont aussi formés naturellement (feux de foret ou émission de volcans actifs), mais la source principale de production est l'industrialisation (raffineries de pétrole, centrales thermoélectriques au charbon et au pétrole, sites de production de goudron, et usines de traitement du bois). L'air est aussi contaminé par les HAP, les principales sources étant la fumée de cigarette (Baek et Jenkins, 2004), les incinérateurs d'ordures ménagères, les gaz d'échappement ainsi que le chauffage domestique. La concentration moyenne des HAP d'un sol semi-urbain a été multipliée par 6 en l'espace de 100 ans et par 20 pour le BaP (Jones *et al*., 1989). Certains HAP, libérés dans l'atmosphère sont adsorbés sur des particules et se retrouvent dans les sols ou rivières transportées par les pluies. D'autres, principalement les HAP à 3 cycles sont présents dans l'atmosphère sous forme gazeuse (INERIS). Toutefois, les raffineries de pétrole ainsi que les usines de production précédemment citées sont des sites où la plupart des HAP sont trouvés à des concentrations relativement élevées.

Les voies principales d'exposition humaine aux HAP sont l'inhalation d'air pollué, le contact cutané, ainsi que l'ingestion de nourriture contaminée; le caractère hydrophobe des HAP entraîne leur concentration dans la chaîne alimentaire.

## 3. La toxicité des HAP

Parmi la centaine de HAP, l'agence américaine de protection de l'environnement (EPA) a identifié 16 d'entre eux comme composés prioritaires, sur la base des risques

sanitaires que ces composés peuvent induire, et en particulier le cancer chez les animaux et l'homme. Ces 16 HAP sont listés dans la base de données du système d'enregistrement des substances (SRS) et sont présentés dans la figure 2.

Les HAP sont souvent présents sous forme de mélanges sur les sites contaminés. Les 16 HAP prioritaires servent de référence pour les analyses toxicologiques, et leur quantification est suffisante pour juger du taux de pollution des sites.

Le 11$^{ème}$ rapport sur les cancérigènes a listé 15 HAP en tant que "potentiellement cancérigènes chez l'homme" ; leur cancérogénicité a été étudiée in-vivo sur animaux. Il y a suffisamment de preuves pour classer le benzo[a]pyrène (BaP) parmi les molécules cancérigènes chez l'homme.

Le centre international de recherche sur le cancer (IARC) édite une liste des agents et composés qui sont classés cancérigènes (liste 1), ou probablement cancérigènes (liste 2A) chez l'homme. Le BaP appartient au groupe 2A, et est classé en catégorie 2 par l'Union Européenne, indiquant son danger pour la reproduction. Le benzène, les goudrons et la fumée de tabac appartiennent à la liste 1, tandis que le benz[a]anthracène, le dibenz[a,h]anthracène et les gaz d'échappement de moteur diesels font partie de la liste 2A.

C'est en 1775 que l'exposition professionnelle des ramoneurs à la suie, riche en HAP, a été avancée pour la première fois comme étant la cause du cancer du scrotum. Par la suite, on a remarqué que l'exposition aux goudrons provoquait des cancers de la peau. Le poumon est désormais la principale localisation des cancers causés par exposition aux HAP; les cancers cutanés étant devenus rares grâce aux avancées de l'hygiène. L'application d'anthracène, de fluorène et de phénanthrène provoque des réactions cutanées. Une étude chez des ouvriers de l'industrie du traitement de bois exposés à la créosote a montré une augmentation de l'incidence du cancer cutané et de la lèvre (rapport USDHHS, 2002).

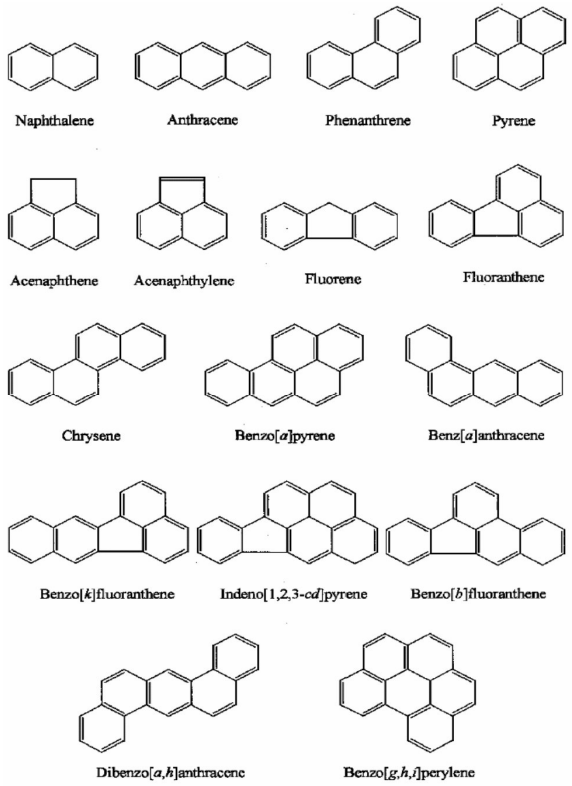



**Figure 2**. Structure des 16 HAP jugés prioritaires par l'EPA.

Chez l'homme et les mammifères, les HAP deviennent toxiques après oxydation enzymatique, provoquant la formation de métabolites électrophiles solubles qui finalement amènent à un dérèglement de la division cellulaire et à la formation de tumeurs (Szeglia et Dipple, 1998). Le mécanisme implique une monooxygénase de type cytochrome P450, et transforme les HAP en *trans*-dihydrodiols (Jerina *et al.*, 1971). Le cytochrome P450 hydroxyle les HAP en dérivés monohydroxylés. Dans le cas du naphtalène, le dihydrodiol est formé en deux étapes ; l'étape initiale, énantio-spécifique est catalysée par la monooxygénase, qui produit le naphtalène-1,2-oxyde, transformé à son tour par une époxyde hydrolase en *trans*-dihydrodiol. Le dihydrodiol est ensuite converti par une dihydrodiol déshydrogénase en 1,2-naphthaquinone. Ce sont les formes époxyde et quinone qui sont les métabolites réactifs responsables de la formation de complexes covalents avec les protéines et l'ADN (Zheng *et al.*, 1997). Le benzo[a]pyrène est aussi activé par un cytochrome de type P450 (P4501A1) en une forme 7,8-oxyde ensuite métabolisé par une époxyde hydrolase en dihydrodiol. Le dihydrodiol peut être oxydé de nouveau par le cytochrome P450 en une forme époxyde. C'est cette forme qui est cancérigène par son action sur l'ADN (Brookes, 1977).

Le BaP est certainement le plus connu et étudié des HAP quant à sa cancérogénicité. Certaines études indiquent une plus grande mutagénicité des mélanges de HAP en présence de BaP (Randerath *et al*., 1999).

## 4. La biodégradation des HAP par les bactéries

L'utilisation des bactéries pour éliminer les HAP des sites contaminés par le biais des méthodes biologiques (ou biorémédiation) est certainement très prometteuse. Cependant, l'efficacité des méthodes biologiques dépend de la biodisponibilité des polluants, elle-même limitée par leur faible solubilité dans l'eau (Johnsen *et al*., 2005; Dur *et al*., 2000), par la spécificité des micro-organismes utilisés, par l'apport d'oxygène (Quantin *et al*. 2005). Elle dépend aussi de la taille des HAP à traiter, les HAP constitués de 2-3 noyaux étant plus faciles à traiter que les HAP à 4 cycles et plus. La température et le pH peuvent aussi être des facteurs limitant (Hogan, 1997). Afin d'augmenter la biodégradation des HAP, notamment ceux à 4 et 5 noyaux, une combinaison de micro-organismes a parfois été préconisée (Boonchan *et al*., 2000). Deux études ont montré une augmentation du taux de dégradation des HAP lourds en présence de surfactants (Boonchan *et al*., 1998) et de détergents (Jimenez et Bartha, 1996).

Certaines bactéries sont capables d'utiliser les HAP comme seules sources de carbone et d'énergie. L'étude de ces souches a permis d'établir les voies métaboliques de la dégradation des HAP. La plupart des souches dégradant les HAP appartiennent aux genres *Pseudomonas, Mycobacterium, Gordona, Sphingomonas, Rhodococcus* (Cerniglia, 1992). Du point de vue de la résistance à la biodégradation, on distingue deux sous-familles de HAP, ceux de faible poids moléculaire (lHAP), contenant jusqu'à trois noyaux benzéniques accolés, et les HAP de haut poids moléculaire (hHAP) contenant quatre noyaux aromatiques et plus. Les bactéries isolées capables de dégrader les HAP légers sont souvent incapables d'attaquer les HAP lourds (Cerniglia, 1992). Les bactéries capables de dégrader le naphtalène sont également capables de dégrader le phénanthrène, et dans ce cas les mêmes enzymes catalysent l'oxydation des deux substrats (Parales *et al*., 2000). Si la majorité des souches connues peuvent métaboliser des HAP à deux et trois noyaux, peu en revanche sont capables d'utiliser des HAP à quatre ou cinq noyaux. Néanmoins, certaines souches, appartenant au genre *Sphingomonas* et *Mycobacterium* peuvent co-métaboliser des HAP à 5 cycles tels que le BaP (Gibson *et al*., 1975, Schneider *et al*., 1996).

La dégradation du naphtalène est la plus étudiée. La première étape de la voie de dégradation est une di-hydroxylation catalysée par la naphtalène 1,2-dioxygénase. Le produit de cette réaction est un *cis*-dihydrodiol (Fig. 3). La voie de dégradation du naphtalène en catéchol, indiquant les produits d'oxydation intermédiaires, et les enzymes responsables de chaque étape réactionnelle est présentée dans la figure 3 (Yen et Serdar, 1988).

Les voies cataboliques d'autres HAP ont été élucidées (Ellis *et al*., 2006). La première étape fait intervenir régulièrement une enzyme de type dioxygénase, qui catalyse la dihydroxylation des HAP de façon spécifique en *cis*-dihydrodiol.

D'autres voies cataboliques font intervenir une monooxygénase lors de l'attaque initiale du substrat, comme c'est le cas de la dégradation du phénanthrène par la cyanobactérie marine *Agmenellum quadruplicatum* PR-6 (Narro *et al*., 1992). Lors de cette réaction, le phénanthrène est converti en phénanthrène-1,2- ou 9,10-oxyde. Cette réaction est voisine de celle catalysée par les cytochromes P450 dans les cellules mammifères et certains champignons.

## 5. Dégradation bactérienne des HAP à quatre cycles et plus

La plupart des bactéries à même de dégrader les HAP à 4 noyaux sont des actinomycètes du genre *Mycobacterium (*Dean-Ross et Cerniglia, 1996 ; Krivobok *et al*., 2003), *Rhodococcus* (Zhou *et al.*, 2006) ou *Gordona* (Kastner *et al*., 1994) ainsi que des espèces Gram$^-$ telles que *Pseudomonas fluorescens* (Caldini *et al.,* 1995)*, Sphingomonas paucimobilis* (Mueller *et al*., 1990) ou *Burkholderia cepacia (*Juhasz *et al*., 1997).

Un grand nombre de souches bactériennes ont été isolées pour leur capacité à dégrader le pyrène, un HAP à quatre cycles, et des études métaboliques ont conduit à proposer une voie de dégradation (Kanaly e*t al*., 2000, Krivibok *et al*., 2003). En revanche, très peu de microorganismes capables de métaboliser le chrysène, un autre HAP à quatre cycles, ont été étudiés. Une souche mutante de *Sphingobium yanoikuyae* B1 (Boyd *et al*., 1997) s'est montré capable d'hydroxyler le chrysène en position *cis*-3,4, ce qui représente la première étape dans la voie de dégradation du chrysène. La souche *Rhodococcus* sp. UW1 peut se développer sur des HAP à 4 cycles, comme le pyrène, le fluoranthéne et le chrysène, mais le naphtalène et le fluorène ne sont que co-métabolisés (Walter *et al*., 1991).

La minéralisation complète du BaP a pu être obtenue en employant des "populations mixtes" de bactéries impliquant des souches de *Mycobacterium* et de *Sphingomonas* (Kanaly *et al.*, 2000b). Cependant, aucune souche bactérienne capable de croître sur BaP n'a pu être

isolée.

Quatre voies de dégradation du benz[a]anthracène (BaA), un HAP à quatre cycles ont été proposées chez *S. yanoikuyae* B1 (Mahaffey *et al*., 1988) et *Mycobacterium* sp. souche RJGII-135 (Schneider *et al*., 1996) ; elles diffèrent par la nature des isomères produits lors de la réaction initiale catalysée par une enzyme spécifique de type dioxygénase. Les dihydrodiols détectés sont les isomères *cis* en positions -5,6, -8,9, -10,11 pour la souche RJGII-135, les isomères -1,2, -8,9 et -10,11 pour la souche B1. Le catabolisme du benzo[a]pyrène (BaP), un HAP à 5 cycles a été étudié chez *Mycobacterium* RJGII-135 (Schneider *et al*., 1996), ainsi que chez *S. yanoikuyae* (Gibson, 1999). Selon l'espèce bactérienne considérée, l'oxydation du BaP donne un isomère hydroxylé en position *cis*-7,8 (souche RJGII-135), ou en position *cis*-9,10 (B1). Dans les deux cas, l'étape suivante, fait appel à une déshydrogénase. Les étapes ultérieures de la dégradation sont mal connues et font intervenir des enzymes qui, pour la plupart, n'ont pas été identifiées (Fig. 3).

Dans la majorité des cas la réaction initiale d'oxydation des HAP est catalysée de façon régio- énantio- et stéréo-spécifique par des dioxygénases (DO). Plusieurs DO peuvent coexister dans une souche bactérienne, comme par exemple chez *Mycobacterium* 6PY1, où deux DO ont été identifiées dont l'une oxyde préférentiellement le pyrène et l'autre le phénanthrène (Krivobok *et al*., 2003). Si les DO clivent en général un nombre restreint de HAP, certaines sont capables d'en oxyder une gamme plus large comportant de 2 à 5 cycles (Gibson, 1999, Demanèche *et al*., 2004).

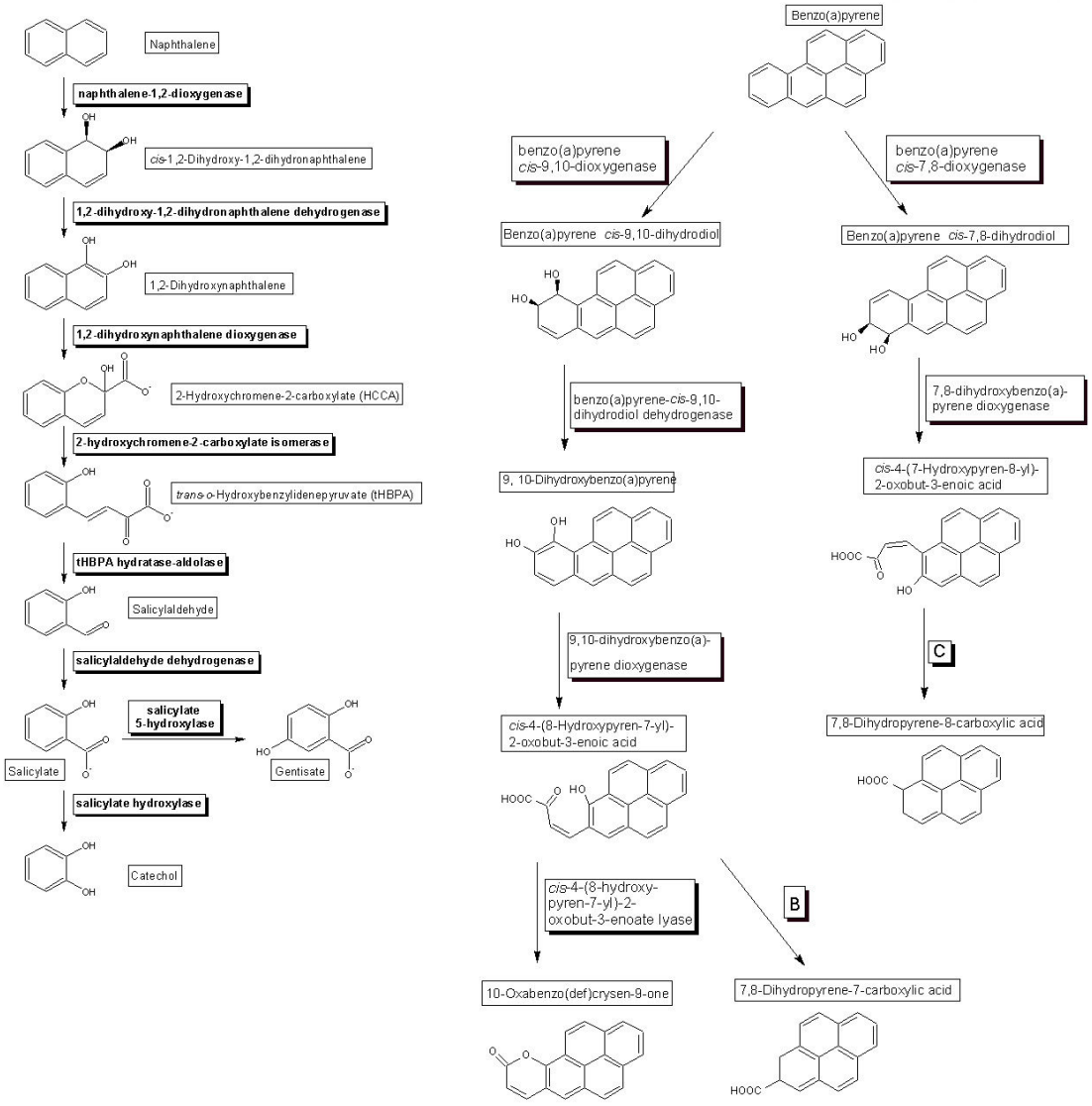



**Figure 3**. Les voies du catabolisme du naphtalène chez *Pseudomonas* sp. NCIB 9816-4 (voie de gauche) et du Benzo[a]pyrène chez S. *yanokuyae* (voie centrale) et *Mycobacterium* sp. RJGII (voies de droite). Les schémas ont été obtenus à partir du serveur http://umbbd.msi.umn.edu.

# 6. Les dioxygénases bactériennes

## 6.1. Propriétés générales des dioxygénases

La première étape de la dégradation bactérienne des HAP est catalysée par un système enzymatique de type dioxygénase. Les dioxygénases constituent une grande famille d'enzymes d'origines diverses, et dont la spécificité du substrat est très variée. Elles interviennent dans la dégradation de composés monoaromatiques (benzoate, toluène), biaromatiques comme le biphényle et les dérivés poly-chlorés du biphényle, les PCB. Du fait

de la toxicité des PCB, les biphényle dioxygénases qui initient la dégradation des PCB, ont fait l'objet d'études détaillées (Furukawa *et al*., 2004).

Les dioxygénases sont des métalloenzymes solubles à deux ou trois composantes comprenant une ferrédoxine réductase, une ferrédoxine et la composante terminale, une dioxygénase. Dans les complexes à deux composantes la composante ferrédoxine est absente (Karlsson *et al*., 2002).

La réductase est une flavoprotéine qui oxyde le NAD(P)H et transmet les électrons un par un à la ferrédoxine, laquelle les transfère à la dioxygénase qui finalement catalyse la réaction de dihydroxylation du substrat. Au cours de la réaction, l'addition de deux atomes d'oxygène moléculaire adjacents transforme les hydrocarbures très hydrophobes en dérivés hydrosolubles et assimilables par la bactérie (Fig. 4).

La réaction d'hydroxylation proprement dite est catalysée par la composante terminale, une métalloprotéine contenant un centre de type Rieske et un fer non hémique au site catalytique.

Du fait des réactions régio- et stéréospécifiques qu'elles catalysent sur une grande diversité de substrats, aromatiques ou non, les dioxygénases bactériennes suscitent beaucoup d'intérêt et ont donné lieu à des applications industrielles autres que la biorémédiation de composés toxiques, comme par exemple la synthèse d'indigo par la NDO (Ensley *et al*., 1983). Certains *cis*-dihydrodiols peuvent être synthétisés à partir de substrats spécifiques afin d'être utilisés comme précurseurs de la synthèse de médicaments (Buckland *et al*., 1999).

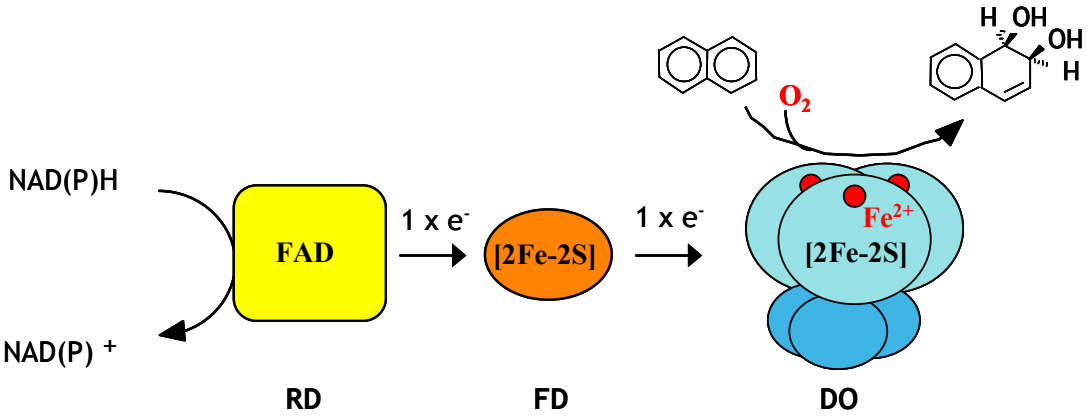



**Figure 4.** Schéma de fonctionnement des dioxygénases. La réductase (RD, FAD), la ferrédoxine (FD, [2Fe-2S]) et la composante oxygénase (DO, Fe2+, [2Fe-2S]) sont indiquées avec le cofacteur ou le centre métallique qu'elles contiennent (indiqué entre parenthèses).

### 6.2. Classification des dioxygénases

Les dioxygénases bactériennes ont d'abord été classées sur la base de critères biochimiques globaux, tels que la structure quaternaire de la composante terminale, le nombre des protéines transporteur d'électrons associées et le type de cofacteurs qu'elles contiennent (Batie *et al.*, 1992). Cette classification a été remise en cause car elle ne correspondait pas à celle qui consiste à comparer les séquences protéiques. D'autres critères, tels que la spécificité du substrat de l'enzyme et les liens phylogénétiques semblaient plus appropriés (Werlen *et al*., 1996).

Plus récemment, une analyse approfondie des séquences des sous-unités α a conduit à classer les dioxygénases en 4 groupes (Nam *et al*., 2001). Selon cette étude, la naphtalène dioxygénase ainsi que les dioxygénases oxydant les HAP appartiennent au groupe III. Les biphényle dioxygénases appartiennent quant à elles au groupe IV.

Le groupe I regroupe les dioxygénases constitués uniquement de sous-unités α, comme la carbazole dioxygénase homotrimérique et la phthalate dioxygénase. Le groupe II rassemble les anthranilate, benzoate, et toluate dioxygénases (Nam *et al*., 2001).

Les dioxygénases "communes" sont regroupées selon ces classifications résumées et présentées dans la table 1.

| Dioxygénase | Batie et coll. | Nam et coll. | Composition Oxygénase |
|---|---|---|---|
| **Phthalate** | IA | I | $\alpha_3$ |
| **2-oxoquinoline** | IB | I | $\alpha_3$ |
| **Carbazole** | III | I | $\alpha_3$ |
| **Benzoate** | IB | II | $\alpha_3\beta_3$ |
| **Nitrotoluène** | III | III | $\alpha_3\beta_3$ |
| **Naphtalène** | III | III | $\alpha_3\beta_3$ |
| **HAP** | III | III | $\alpha_3\beta_3$ |
| **Cumène** | IIB | IV | $\alpha_3\beta_3$ |
| **Biphényle** | IIB | IV | $\alpha_3\beta_3$ |
| **(chloro-)Benzène** | IIB | IV | $\alpha_3\beta_3$ |
| **Toluène** | IIB | IV | $\alpha_3\beta_3$ |

**Tableau 1**. Classification des dioxygénases basée sur les propriétés des transporteurs d'électrons (Batie *et al*., 1992) et sur les spécificités de substrat associée à une analyse de séquence des sous-unités α (Nam *et al*., 2001).

### 6.3. Données structurales sur les transporteurs d'électrons associés

Les ferrédoxine réductases de la biphényle dioxygénase de *Pseudomonas* sp. souche KKS102 (Senda *et al*., 2000) et celle de la benzoate dioxygénase de *Acinetobacter* souche ADP1 (Karlsson *et al*., 2002) ont été cristallisées et leur structure déterminée. Elles appartiennent à deux familles différentes (Batie *et al*., 1992). La première appartient à la famille des glutathion réductases, la seconde de la famille FMN réductases de type plante. Elles possèdent deux domaines très semblables, les domaines de liaison au FAD ou au FMN et celui de fixation du NADH. Elles différent par le domaine C-terminal, qui lie un centre [2Fe-2S] de type plante pour la seconde catégorie d'enzymes. Les réductases de la famille glutathion réductase ne contiennent pas de centre Fe-S (Ferraro *et al*., 2005).

Quelques structures de composantes ferrédoxine sont publiées, notamment celle de la biphényle dioxygénase de *Burkholderia cepacia* (Colbert *et al.,* 2000) et celle de la carbazole dioxygénase de *Pseudomonas resinovorans* (Nam *et al*., 2005). Toutes deux contiennent un centre [2Fe-2S] de type Rieske, responsable du transfert d'électrons vers la composante terminale.

## 7. La structure de la naphtalène dioxygénase (NDO-P)

La première structure tridimensionnelle obtenue d'une dioxygénase bactérienne est celle de la naphtalène 1,2-dioxygénase (NDO-P) de *Pseudomonas putida* souche NCIB 9816-4 (Kauppi *et al.*, 1998). Les phases expérimentales ont été obtenues en combinant les résultats de "multiple wavelength anomalous dispersion" (MAD) et de "multiple isomorphous replacement" (MIR), ce en dépit de la présence de Fer dans le cristal. La structure de NDO-P, enzyme de référence, a ensuite été déterminée sous forme complexée avec l'indole, le naphtalène et le naphtalène dihydrodiol (Karlsson *et al*., 2003).

NDO-P est une hétérohexamère de type $\alpha_3\beta_3$ dont la forme évoque celle d'un champignon où les trois sous-unités α formeraient la coiffe et les trois sous-unités β le pied (Fig. 5). La sous-unité β d'environ 20 kDa n'aurait qu'un rôle structural (Zielinski *et al*., 2002). Toutefois, des études réalisées avec d'autres dioxygénases suggèrent l'implication de la sous-unité β dans la spécificité du substrat (Ge et Eltis, 2003, Kimura *et al*., 1997). D'autre part, la carbazole dioxygénase est une enzyme naturellement dépourvue de sous-unité β (Nijori *et al*., 2005).

**Figure 5**. Structure de l'hétérohexamère $\alpha_3\beta_3$ de la naphtalène dioxygénase. Les 3 sous-unités α et β sont respectivement représentées en orange et en vert. Chaque hétérodimère αβ est relié à l' hétérodimère adjacent par une symétrie d'ordre 3.

La sous-unité α, d'environ 50 kDa, est composée de deux domaines, le domaine Rieske contenant un centre [2Fe-2S] et le domaine catalytique qui contient l'atome de fer mononucléaire où a lieu la réaction enzymatique. La distance qui sépare les deux centres métalliques d'une même sous-unité (> 40 Å) est trop grande pour permettre un transfert rapide des électrons. Les électrons sont transférés du centre [2Fe-2S] d'une sous-unité au site catalytique de la sous-unité α voisine, sans doute par l'intermédiaire de l'acide aminé Asp205 (Parales *et al.*, 1999). La structure quaternaire de l'enzyme $\alpha_3\beta_3$ est donc essentielle à son fonctionnement puisque deux sous-unités α adjacentes sont impliquées; l'une capture les électrons de la ferrédoxine et l'autre catalyse la réaction. Les électrons permettent l'activation de l'oxygène nécessaire à l'hydroxylation du substrat.

La réaction prend place dans une poche catalytique contenant le site actif. La poche est majoritairement hydrophobe et le substrat atteint le site actif en empruntant une voie débutant par deux boucles flexibles exposées au solvant (Carredano *et al.*, 2000). Lors de la réaction catalytique, les deux atomes du dioxygène réagissent simultanément avec deux atomes de carbone adjacents du substrat. Les 2 protons nécessaires à l'hydroxylation viendraient de molécules d'eau se trouvant à proximité. Cette hypothèse a été confirmée expérimentalement par l'observation de la molécule de dioxygène dans le cristal de NDO-P, où la molécule d'$O_2$ "approche" le Fer en configuration latérale et non distale (Karlsson *et al.*, 2003).

## 8. Mécanisme réactionnel des dioxygénases

Le mécanisme de fonctionnement des dioxygénases a surtout été étudié avec la NDO-P comme modèle par des approches structurales et biochimiques. Le site actif de la NDO-P (des DO de la forme $\alpha_3\beta_3$ et $\alpha_3$) se situe près de l'interface entre deux sous-unités α adjacentes (Fig. 6). La disposition spatiale des centres métalliques suggère que le cheminement des électrons jusqu'au site actif passe par un pont entre deux sous-unités adjacentes. En effet, la distance entre le site catalytique d'une sous-unité et le cluster de la sous-unité voisine n'est que de 12 Å. Entre ces deux sites se trouve l'acide aminé Asp205, qui est lié par liaison hydrogène, d'une part au résidu His208 ligand de l'ion ferreux, d'autre part au résidu His104 ligand du cluster. La fonction de l'aspartate en position 205 a été étudiée par des expériences de mutagénèse dirigée dans lesquelles le remplacement de ce résidu s'est soldé par une perte d'activité des enzymes mutantes (Parales *et al.*, 1999). Ce résultat suggérait que le résidu Asp205 participait sans doute au transit des électrons intramoléculaire. Cependant, cette hypothèse a été remise en question par les travaux de Beharry et al. (2003) sur l'anthranilate dioxygénase. Le remplacement de l'aspartate équivalent dans cette enzyme (Asp218) a pour effet de faire varier le potentiel de ½ réduction du cluster de – 100mV, cet effet est provoqué par l'incapacité des enzymes mutantes de protoner l'une des histidines ligand du cluster, fonction assurée par Asp218.

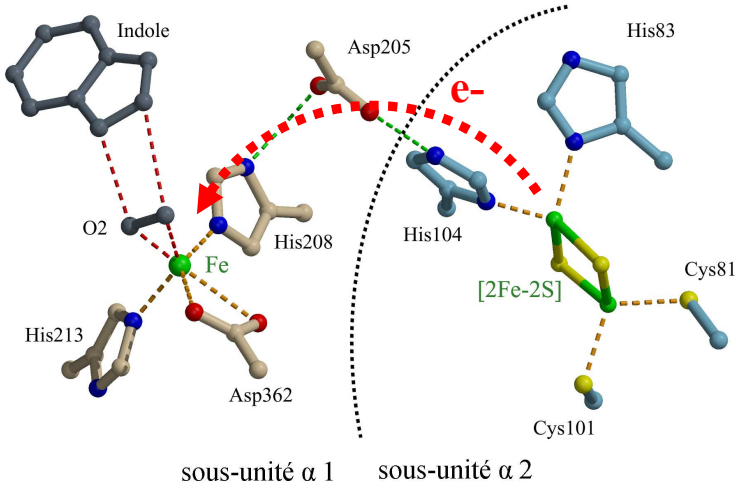



**Figure 6**. Structure du site actif de la naphtalène dioxygénase (NDO-P) (Kauppi *et al.*, 1998). Le dioxygène et l'indole sont représentés en gris, respectant orientations et positions initiales de la structure ayant pour code d'accès 1O7N. Les liaisons et interactions sont indiquées en pointillées : les liaisons protéine-cofacteur (orange), les interactions Fer-$O_2$-Indole (rouge). Un cheminement plausible des électrons via Asp205 est également indiqué. La distance minimale entre le Fer mononucléaire d'une sous-unité α et le cluster [2Fe-2S] de la sous unité α voisine est d'environ 12 Å.

Le mécanisme catalytique de la NDO a été étudié en détail par Wolfe et coll., par des études de cinétique rapide et des analyses par spectroscopie RPE. Si l'action simultanée des trois composantes de l'enzyme est nécessaire à l'activité catalytique, des expériences à un seul turnover ont montré que la composante oxygénase seule était capable de catalyser l'hydroxylation du naphtalène dès lors que les centres métalliques de la protéine sont réduits au préalable (Wolfe *et al*., 2001).

La réactivité du fer mononucléaire au site catalytique vis-à-vis de l'oxygène est régulée par l'état d'oxydation du centre FeS et la présence du substrat au site actif (Wolfe *et al.*, 2001). En d'autres termes, le cluster FeS doit être réduit et le substrat doit être présent au site actif pour que le dioxygène ait accès au site actif et que la réaction prenne place. Sans substrat, l'enzyme réduite réagit très lentement avec l'oxygène, ce qui empêche la formation de peroxyde qui pourrait inactiver l'enzyme. Le mécanisme réactionnel de la NDO est représenté dans la figure 7, indiquant l'état d'oxydation du FeS et du Fer à chacune des étapes du cycle catalytique. Le cycle catalytique se termine par la libération du *cis*-dihydrodiol, l'enzyme ayant alors ces deux centres métalliques oxydés.

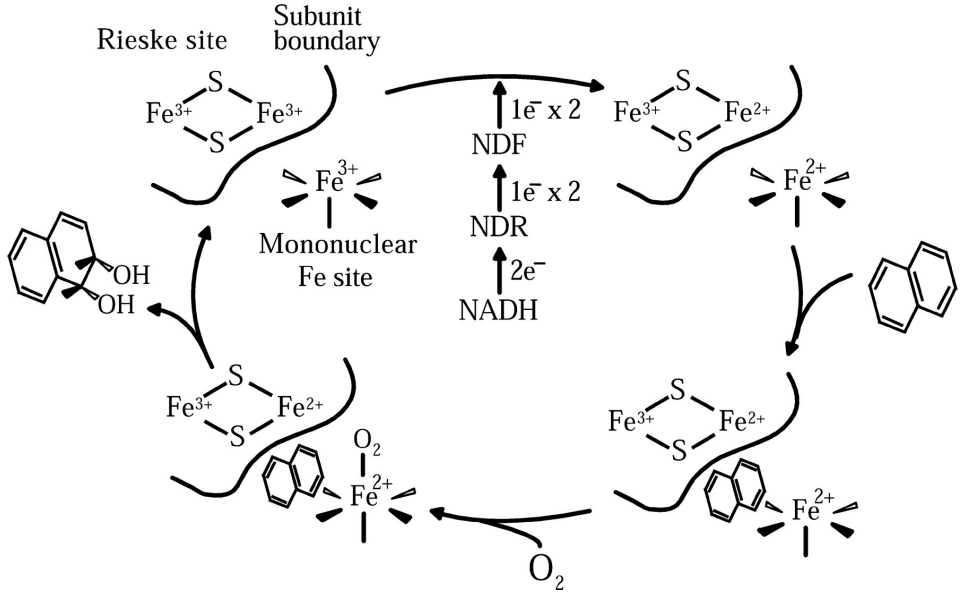



**Figure 7**. Représentation du mécanisme réactionnel de la NDO au cours du cycle catalytique (Wolfe *et al*., 2001). Les états d'oxydation du Fer du site actif et du cluster sont indiqués.

Le mécanisme fait intervenir l'espèce $Fe^{III}$-hydro-peroxo en tant qu'intermédiaire réactionnel, qui peut alors suivre 3 voies possibles comme présenté dans la figure 10 (Wolfe *et al*., 2001). Une étude menée par Lee (1999) utilisant le benzène comme substrat de la NDO

a montré que l'enzyme génère très peu de dihydrodiol mais produit beaucoup de peroxyde d'hydrogène. Le $H_20_2$ relâché au cours de la réaction découplée inactive la composante terminale, sans doute parce que le peroxyde réagit avec le fer ferreux du site actif pour former de radicaux hydroxyles ˙OH selon la réaction de Fenton.

$$\mathbf{Fe^{2+} + H_2O_2 ----> Fe^{3+} + {}^{\cdot}OH + OH^-}$$

Une autre étude a montré qu'en présence d'$H_2O_2$, la composante oxygénase de la NDO peut seule hydroxyler le naphtalène en l'absence d'$O_2$. $H_2O_2$ est donc capable de réagir directement avec l'ion ferreux du site actif et de court-circuiter ainsi les étapes de réduction du cluster FeS et de réaction avec $O_2$. Cependant, dans ces conditions, le dihydrodiol reste piégé au site actif, suggérant que la réduction du Fer déclenche la libération du produit de la réaction ainsi que le début du cycle catalytique suivant (Wolfe et Lipscomb, 2003).

Parmi les différentes voies proposées d'activation d'$O_2$ par NDO (Fig. 8, voies A, B et C), la voie C semble la moins probable car, dans cette hypothèse, le fer du site actif resterait ferreux en fin de cycle catalytique, en contradiction avec les observations RPE indiquant que le Fer est alors ferrique De plus, la voie C requière une réduction additionnelle non observée in vitro.

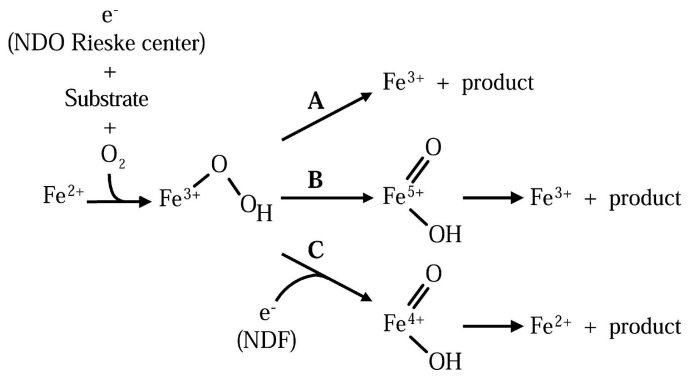



**Figure 8**. Hypothèse sur les intermediares reactionnels formés par NDO (Wolfe *et al*., 2001).

En se basant sur la structure de la NDO en présence de dioxygène lié latéralement, la réaction de *cis*-dihydroxylation du naphtalène a été modélisée en utilisant la théorie des fonctions de densité (DFT) (Bassan *et al*., 2004)**.** Les résultats suggèrent une espèce intermédiaire de type fer-oxo (degré d'oxydation IV ou V), et que la voie C serait plus favorable avec une énergie d'activation inferieure à celle correspondant à la voie B.

Parmi toutes les voies testées, la plus probable (demandant un minimum d'énergie d'activation) fait intervenir un intermédiaire époxyde, suivi d'un arène cationique, précédant la formation du *cis*-dihydrodiol (Fig. 9). L'orientation de l'intermédiaire cationique doit être contrôlée par l'environnement du site actif afin de générer un dihydrodiol ayant la stéréospécificité correcte, c'est-à-dire le *cis*-dihydrodiol (Bassan *et al*., 2004).

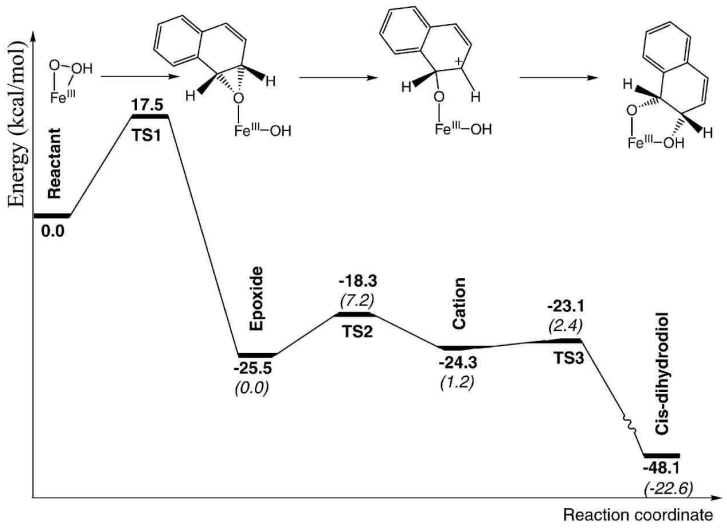



**Figure 9**. Chemin réactionnel le plus probable de la *cis*-dihydroxylation du naphtalène catalysée par NDO. Les intermédiaires réactionnels ainsi que le diagramme d'énergie sont représentés (Bassan *et al*., 2004).

## 9. Les structures d'autres composantes terminales de dioxygénases

La naphtalène dioxygénase (NDO-P) de *Pseudomonas* NCIB 9816-4 (Kauppi *et al*., 1998 ; 1NDO) a initialement servi de modèle à la compréhension du métabolisme des HAP. L'obtention récente d'autres structures de DO telles que celles de la biphényle dioxygénase (BPDO-R) de *Rhodococcus* sp. RHA1 (Furasawa *et al*., 2004 ; 1ULI), la carbazole dioxygénase (CARDO) de *Pseudomonas resinovorans* (Nojiri *et al*., 2005 ; 1WW9), la nitrobenzène dioxygénase (NBDO) de *Comamonas* sp. JS765 (Friemann *et al*., 2005 ; 2BMO), la naphtalène dioxygénase (NDO-R) de *Rhodococcus* sp. NCIMB12038 (Gakhar *et al*., 2005 ; 2B1X), la 2-oxoquinoline monooxygénase (OMO) de *Pseudomonas putida* strain 86 (Martins *et al*., 2005 ; 1ZO1) et la cumène dioxygénase (CUDO) de *Pseudomonas fluorescens* IP01 (Dong *et al*., 2005 ; 1WQL) se traduit par une meilleure connaissance des

relations entre structure et fonction enzymatique. La plupart de ces structures appartient à deux sous-familles celle des naphtalène dioxygénases et celle des biphényle dioxygénases.

Malgré des similitudes de séquences relativement modestes entre ces enzymes, toutes les structures tertiaires et quaternaires des DO sont très voisines (Fig. 10).

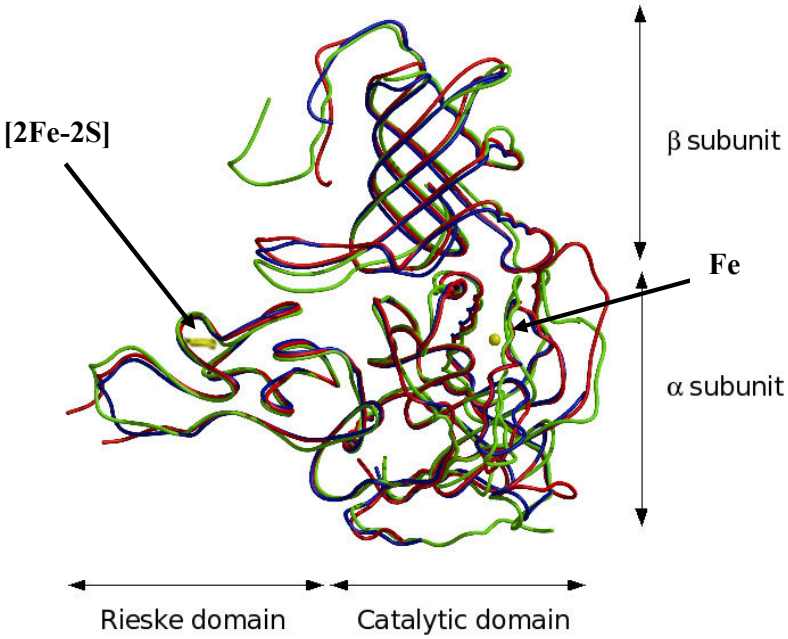



**Figure 10.** Superposition des hétérodimères αβ de NDO-P, BPDO-R et CUDO représentés en vert, bleu et rouge, respectivement. Les deux sous-unités α et β ainsi que les deux domaines de la sous-unité α sont représentés. L'atome de fer et le centre [2Fe-2S] sont représentés en jaune. La zone contenant les régions LI et LII formant un couvercle contrôlant l'accès au site actif de la NDO-P (lid) est aussi indiquée.

Les structures représentées ci-dessus concernent des dioxygénases qui appartiennent aux groupes III et IV définis par Nam et al (2001). Deux structures connues de composantes terminales appartenant au groupe I, et constituées uniquement de sous unités α (forme trimèrique $\alpha_{3)}$ ne sont pas représentées. Ce sont les structures des composantes terminales de la carbazole dioxygénase (Nojiri *et al*., 2005) et de la 2-oxoquinoline monooxygénase (Martins *et al*., 2005) .

Les dioxygénases catalysent des réactions régio- et stéréospécifiques, imposées par la structure de l'enzyme, qui conditionne l'orientation du substrat dans la poche catalytique. Les résidus de la poche catalytique ne sont pas tous conservés au sein des dioxygénases considérées, ce qui est sans doute à l'origine des différences de spécificités, vis-à-vis du substrat. L'alignement des séquences peptidiques des sous-unités α de cinq DO de structures connues est présenté dans la figure 11.

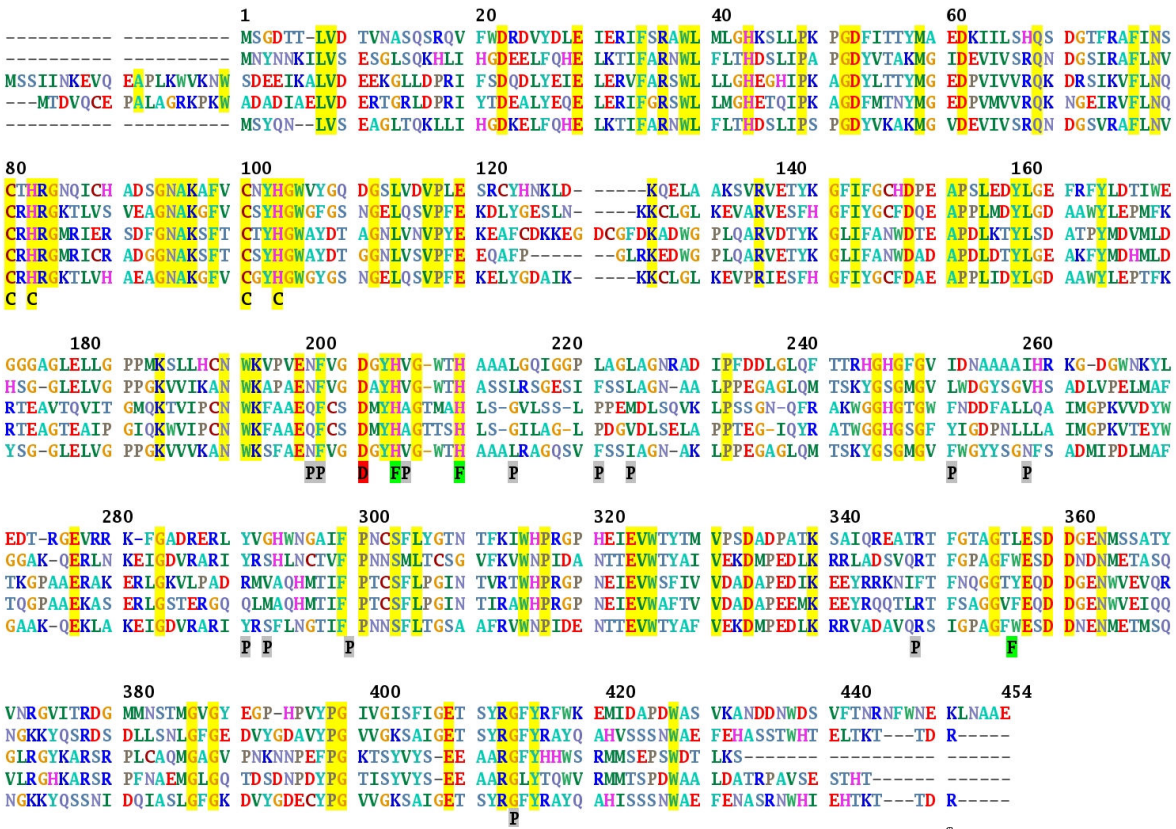


**Figure 11**. Alignement des séquences peptidiques des sous-unités α de PHN1[$], NDO-P, CUDO, BPDO-R et NBDO. Les résidus totalement conservés sont indiqués en jaune. [$]séquence de la chaine polypeptidique de la protéine étudiée. Les résidus conservés sont ceux qui lient le centre [2Fe-2S] (annotés C) et l'atome de fer (annotés F), et un aspartate équivalent de Asp 205 de NDO-P (D). Les résidus annotés P sont ceux qui définissent la poche catalytique.

## 10. Spécificité des dioxygénases

Des études de mutagenèse dirigée entreprises sur la naphtalène dioxygénase de *Pseudomonas* sp. souche NCIB 9816-4 (NDO-P), indiquent que la plupart des acide-aminés situés au niveau de la poche catalytique n'ont que peu, voire aucune influence, sur la régio et stéréospécificité du produit de la réaction. Cependant, des substitutions de la phénylalanine en position 352 par des acides aminés dont la chaine latérale est moins volumineuse (Leu, Val) provoquent une altération de la stéréochimie du produit (Parales *et al.*, 2000). Le résidu Phe 352 est situé à environ 5 Å du carbone 1 du naphtalène dans la structure de NDO-P complexée au substrat (Karlsson *et al.*, 2003).

L'élargissement de la poche catalytique tout près du site d'hydroxylation du substrat

diminuerait donc les contraintes stériques de positionnement du substrat, ce qui expliquerait les altérations de la spécificité. D'autre part, le remplacement de l'acide aspartique en position 362, l'un des ligands du fer au site actif en alanine s'avère fatal pour l'activité de l'enzyme.

Les remplacement d'autres résidus de la poche catalytique par mutagenèse dirigée affectent principalement la régio- et la stéréo-sélectivité de la NDO-P (Ferraro *et al*., 2005).

Les acides aminés qui forment la poche catalytique de cinq DO sélectionnées appartenant aux groupes III et IV (Nam *et al*., 2001) sont indiqués dans le tableau 2 (Ferraro *et al*., 2005).

| NDO-P | NBDO | NDO-R | CUDO | BPDO-R |
|---|---|---|---|---|
| Asn201 | Asn199 | Asn209 | Gln227 | Gln217 |
| Phe202 | Phe200 | Phe210 | Phe228 | Phe218 |
| Val203 | Val201 | Val211 | Cys229 | Cys219 |
| Gly204 | Gly202 | Gly212 | Ser230 | Ser220 |
| Ala206 | Gly204 | Ala214 | Met232 | Met222 |
| Val209 | Val207 | Thr217 | Ala235 | Ala225 |
| Leu217 | Leu215 | Val225 | Val244 | Ile234 |
| Phe224 | Phe222 | Phe293 | Leu284 | Leu274 |
| Leu227 | Ile225 | Phe236 | Leu259 | Pro250 |
| Gly251 | Gly249 | Gly252 | Gly276 | Gly266 |
| Leu253 | Phe251 | Ile254 | Phe278 | Tyr268 |
| Val260 | Asn258 | Met309 | Ile288 | Ile278 |
| His295 | Phe293 | Phe293 | Ala321 | Ala311 |
| Asn297 | Asn295 | His295 | His323 | His313 |
| Leu307 | Leu305 | Phe307 | Leu333 | Leu323 |
| Ser310 | Ser308 | Phe320 | Ile336 | Ile326 |
| Phe352 | Ile350 | Phe362 | Phe378 | Phe368 |
| Trp358 | Trp356 | Phe368 | Tyr384 | Phe374 |

**Tableau 2**. Acides aminés structurellement analogues à ceux formant la poche catalytique de NDO-P. Les acides aminés sur fond gris clair sont identiques à ceux observés dans NDO-P, ceux sur fond gris foncé sont identiques à ceux de BPDO-R. Les ligands du Fer catalytique ne sont pas indiqués, étant donné qu'ils sont totalement conservés (Ferraro *et al.*, 2005).

Selon cette comparaison, on peut distinguer deux groupes ; le groupe A comprenant les deux naphtalène dioxygénase et la nitrobenzène dioxygénase, et le groupe B contenant la cumène et la biphényle dioxygénase.

Les composés auxquels s'attaquent les dioxygénases présentées dans le tableau 1 sont représentés dans la figure 12. Ces composés sont de taille similaire, cependant, le nitrobenzène possède un groupe hydrophile qui donne lieu à des interactions particulières lors de la réaction catalytique. Naphtalène et biphényle étant structurellement proches, les DO dégradant l'un de ces composés sont souvent capables de dégrader l'autre (Parales *et al*., 1999).

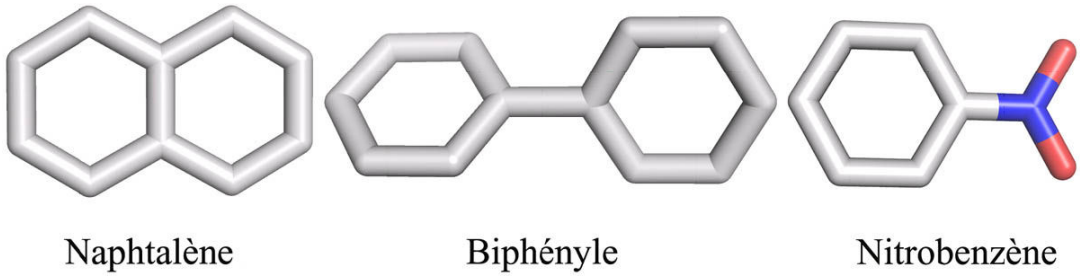



**Figure 12**. Structure du naphtalène, du biphényle et du nitrobenzène, substrats des dioxygénases NDO-P, BPDO et NBDO. Les atomes de carbone, azote et oxygène sont représentés en blanc, bleu et rouge. Les atomes d'hydrogène ne sont pas montrés.

Pour chacune des trois enzymes décrites ci-dessus, au moins une structure de complexe enzyme-substrat a été déterminée et permet l'interprétation du rôle de certains résidus à proximité du substrat. Par exemple, Asn258 dans la poche catalytique de la NBDO est indispensable à l'activité catalytique envers les nitroarènes, car le groupement NH2 de Asn 258 interagit avec le groupement nitro du substrat, ce qui permet un bon positionnement du substrat lors de la catalyse (Friemann *et al.*, 2005). Le remplacement de ce résidu par une valine, l'acide aminé correspondant de NDO-P, change la spécificité de l'enzyme de sorte que le produit d'oxydation du nitrobenzène est l'alcool nitrobenzylique au lieu du catéchol, (Ferraro *et al.*, 2005). Dans le cas d'OMO, l'orientation du substrat, avec l'atome de carbone 8 de la 2-oxoquinoline le plus proche du Fer catalytique, est contrôlée par l'oxygène de la glycine 216 (2.65 Å de l'atome N2 du substrat) (Fig. 13). Dans ces deux cas, un seul résidu peut contrôler l'orientation le substrat au site actif de l'enzyme et autorise alors la réaction régio-sélective.

La poche catalytique de NDO-P avec le substrat présent est représentée dans la figure 14. Le naphtalène est orienté de telle manière que les carbones 1 et 2 sont proches du Fer catalytique. La structure d'autres DO complexées avec un substrat, celles de CARDO, BPDO-R et NDO-R a dévoilé une orientation du substrat telle que les carbones devant être hydroxylés sont dans une position semblable vis-à-vis du fer du site actif.

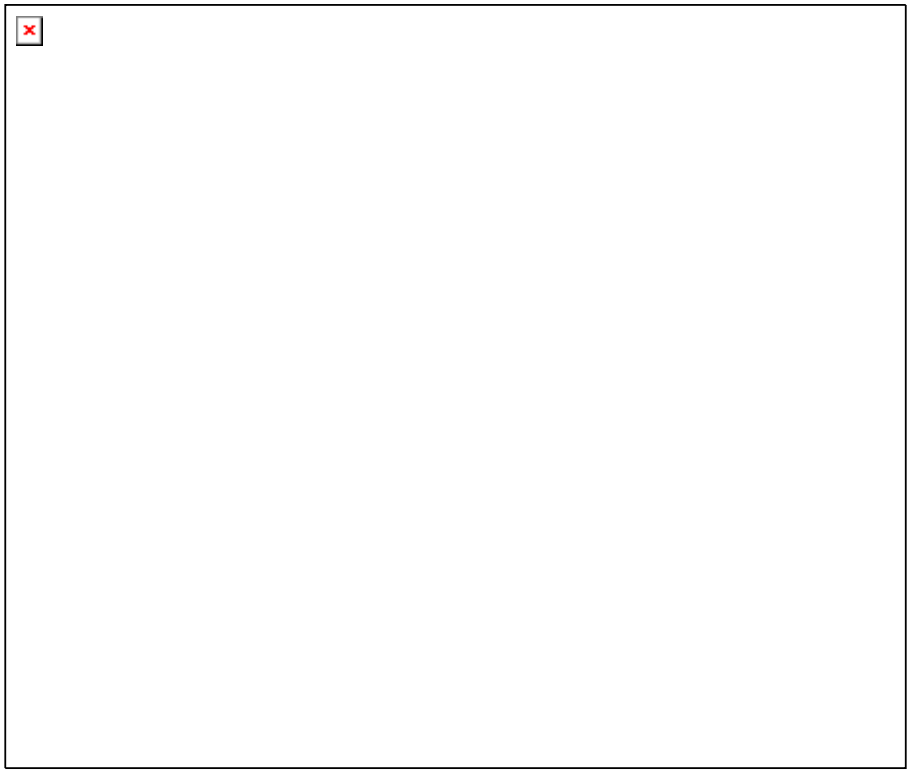


**Figure 13**. Poche catalytique de la composante oxygénase de la 2-oxoquinoline 8-monooxygénase. Le substrat est coloré en rosé, la molécule d'eau et le Fer catalytique sont indiqués, les résidus à proximité du substrat sont aussi indiqués.

L'un des paramètres important est la géométrie et les dimensions de la poche catalytique, permettant l'accès du substrat Une augmentation du volume de la poche catalytique est observée lorsque le substrat est présent au site actif dans BPDO-R et OMO (Furasawa *et al*., 2004, Martins *et al*., 2005). Les plus grandes variations sont observées à l'entrée de la poche, dans une zone plus flexible contenant deux boucles LI et LII qui contrôleraient l'accès de la poche.

Ces deux dernières années, l'étude comparative des dioxygénases, de leur structure et de leur spécificité a fait faire un bond spectaculaire à nos connaissances de ce type d'enzymes et de leur mode de fonctionnement. Cependant, avant les travaux présentés dans cette thèse, aucune dioxygénase capable de s'attaquer aux HAP de plus de 3 cycles n'avait été isolée, ce qui justifiait la caractérisation de l'enzyme identifiée précédemment chez *Sphingomonas* CHY-1 (Demanèche *et al*., 2004)

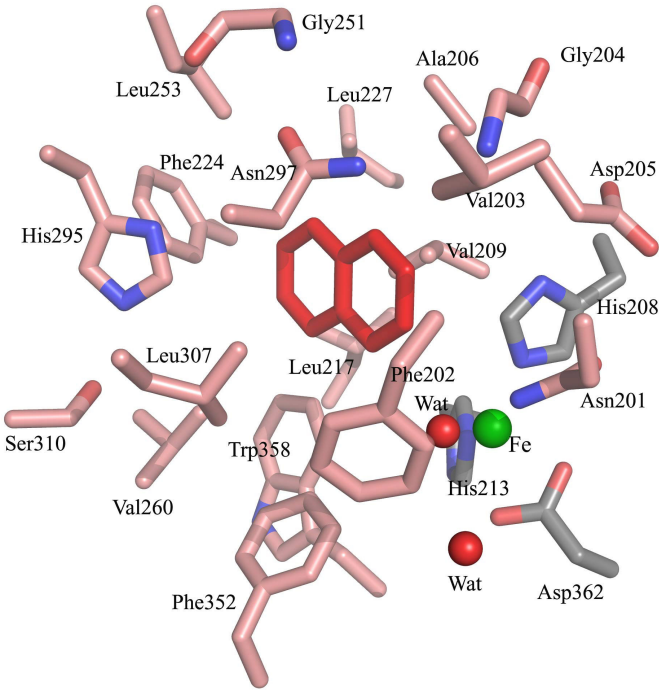



**Figure 14**. Poche catalytique de NDO-P associée au substrat. La chaine latérale des résidus formant la poche de fixation du substrat, principalement hydrophobes sont indiqués. Les ligands du fer sont représentés en gris.

## 11. *Sphingomonas* sp. CHY-1

Une souche bactérienne a été isolée à partir d'un sol extrait d'une ancienne usine à gaz localisée en région Rhône-Alpes (France). Cette souche, *Sphingomonas* sp. CHY-1 (CHY-1) possède la rare propriété de croître sur chrysène, HAP à quatre cycles, comme seule source de carbone et d'énergie (Willison, 2004).

Des gènes responsables de l'activité métabolique de CHY-1 vis-à-vis du chrysène ont été identifiés et ont fait l'objet d'études fonctionnelles dans l'équipe de Y. Jouanneau à Grenoble (Demanèche et al., 2004). Certains de ces gènes codent pour une enzyme de type naphtalène dioxygénase comprenant une réductase (*phnA4*), une ferrédoxine (*phnA3*) et une composante terminale, PhnI. D'autres gènes ont été identifiés dans la même région, notamment *phnA1bA2b*, *phnB*, *phnC* et *phnD* qui codent respectivement une autre oxygénase (PhnII), une alcool déshydrogénase, une extradiol dioxygénase et une isomérase. Ces enzymes interviennent à des étapes ultérieures du métabolisme des hydrocarbures. PhnII catalyse la décarboxylation oxydative de salicylate en catéchol. La réductase et la ferrédoxine nécessaires

à l'activité catalytique de PhnI et PhnII seraient partagées dans la bactérie car il n'y a pas d'autres copies de gènes codant pour des transporteurs d'électrons dans le cluster de gènes cataboliques.

L'inactivation du gène codant pour la sous-unité α de PhnI entraine la perte de la faculté de croître sur chrysène, et abolit de la même façon la capacité de *Sphingomonas* de métaboliser d'autres HAP, suggérant que PhnI est responsable de l'attaque initiale des HAP (Demaneche *et al*., 2004). De plus, PhnI surproduite dans *Escherichia coli* a démontré l'unique capacité de catalyser la dioxygénation d'un grand nombre de HAP parmi lesquels le naphtalène, l'anthracène, le phénanthrène, le benz[a]anthracene et le chrysène.
Biphényle, naphtalène, phénanthrène et chrysène sont hydroxylés en positions 2,3-, 1.2-, 3,4- et 3,4- tandis que l'anthracène est converti en cis-1,2-dihydrodiol. PhnI partage certaines propriétés catalytiques de la dioxygénase de *S. yanoikuyae* B1, enzyme qui n'a pas encore été isolée (Gibson, 1999).

Les réactions spécifiques vis-à-vis du chrysène, phénanthrène et anthracène catalysées par Phn1 sont illustrées par la figure 15.

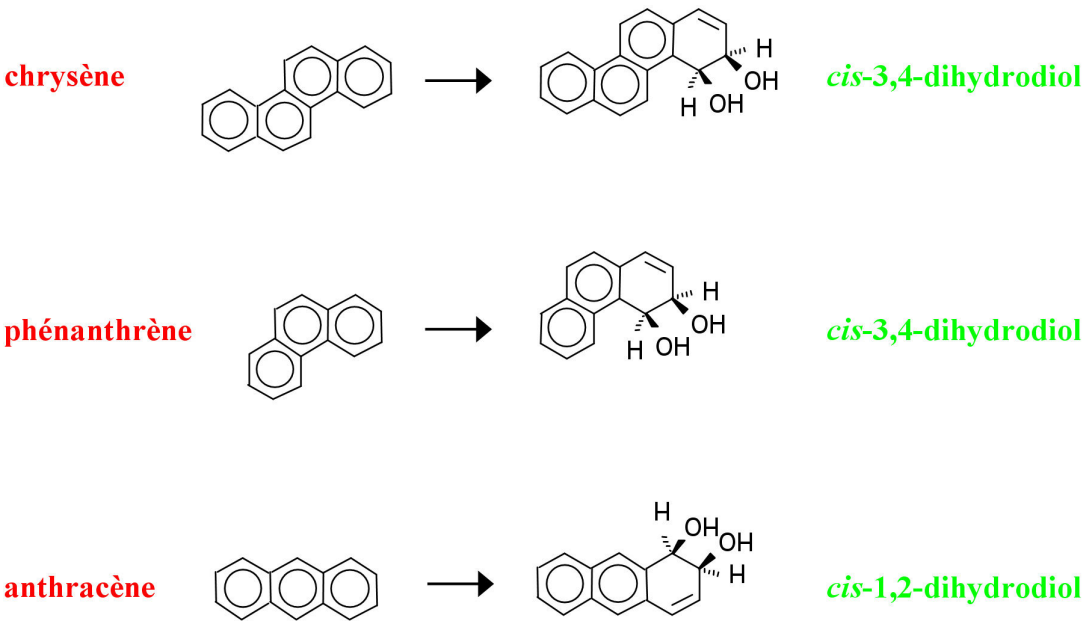



**Figure 15**. Réactions régio- et stéréospécifiques d'oxydation des HAP catalysées par PhnI.

Dans sa forme recombinante, PhnI montre une faible activité sans la présence de la ferrédoxine et de la réductase associées (Demaneche *et al*., 2004).

## *Présentation des travaux expérimentaux sur la dioxygénase de Sphingomonas CHY-1*

Au cours de cette étude, nous nous sommes intéressés à la dioxygénase de *Sphingomonas* CHY-1, identifiée et purifiée dans le laboratoire de Grenoble. Cette enzyme a fait preuve d'une sélectivité exceptionnellement large, puisque c'est la première enzyme isolée capable d'oxyder 9 des 16 HAP jugés prioritaires par l'EPA., parmi lesquels des HAP à quatre et cinq cycles. Nos travaux ont porté sur la détermination des caractéristiques biochimiques de cette enzyme, notamment ses propriétés catalytiques, et ont fait l'objet d'un article publié dans la revue Biochemistry. Par ailleurs, la composante oxygénase a été cristallisée, et sa structure tridimensionnelle a été résolue, ce qui a donné lieu à la présentation de deux articles soumis à publication. Ces trois articles sont présentés dans ce chapitre.

Dans un premier temps, les conditions de purification de chacune des composantes de l'enzyme sont décrites. Les caractéristiques biochimiques et spectroscopiques de ces protéines sont ensuite présentées, notamment les résultats d'expériences de RPE visant à déterminer la nature et la stœchiométrie des centres métalliques présents dans la ferrédoxine et la composante oxygénase. L'activité catalytique de l'enzyme vis-à-vis de HAP de 2 à 5 noyaux aromatiques est enfin examinée par l'analyse qualitative et quantitative de tous les produits de la réaction enzymatique.

Dans un deuxième temps, les conditions expérimentales de cristallisation de la composante terminale PhnI ont été établies, ce qui a nécessité le criblage de plusieurs centaines de conditions et des mois de tâtonnements. La structure tridimensionnelle a été obtenue par remplacement moléculaire, après plusieurs tentatives infructueuses de résolution de la structure par les méthodes MAD et SAD. Le modèle obtenu apporte des éléments d'information très précieux sur le site actif de l'enzyme, la poche hydrophobe de liaison du substrat, et fournit un début d'explication à la sélectivité remarquablement large de cette enzyme. Il constitue, d'autre part, une excellente base de travail afin d'identifier et de comprendre le rôle des acides aminés responsables de cette sélectivité.

## Références bibliographiques


Baek, S.O. and Jenkins, R.A. (2004). Characterization of trace organic compounds associated with aged and diluted sidestream tobacco smoke in a controlled atmosphere volatile organic compounds and polycyclic aromatic hydrocarbons. *Atmos. Envir*. **38**: 6583-6599.

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

Batie, C.J., Ballou, D.P. and Correll, C.C. (1992). Phthalate dioxygenase reductase and related flavin-iron-sulfur containing electron transferases. *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., Ed.) CRC Press, Boca Raton, Fla. 543-556.

Boonchan, S., Britz, M.L. and Stanley, G.A. (1998). Surfactant-Enhanced Biodegradation Of High Molecular Weight Polycyclic Aromatic Hydrocarbons. *Biotec. and Bioeng*. **59**: 482-494.

Boonchan, S., Britz, M.L. and Stanley, G.A. (2000). Degradation and mineralization of high-molecular-weight Polycyclic Aromatic Hydrocarbons by defined fungal-bacterial cocultures. *Appl. Environ. Microbiol*. **66**: 1007-1019.

Bouchez, M., Blanchet, D. and Vandecasteele, J.P. (1995). Degradation of polycyclic aromatic hydrocarbons by pure strains and by defined strain associations: inhibition phenomena and cometabolism. *Appl. Microbiol. Biotechnol*. **43**: 156-164.

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

Brookes, P. (1977). Mutagenicity of polycyclic aromatic hydrocarbons. *Mutat. Res*. **39**: 257-283.

Buckland, B.C., Drew, S.W., Connors, N.C., Chartrain, M.M., Lee, C., Salmon, P.M.,Gbewonyo, K., Zhou, W., Gailliot, P., Singhvi, R., Olewinski, R.C.Jr., Sun, W.J., Reddy, J., Zhang, J., Jackey, B.A., Taylor, C., Goklen, K.E., Junker, B. and Greasham, R.L. (1999). Microbial conversion of indene to indandiol: a key intermediate in the synthesis of CRIXI-VAN. *Metab. Eng*. **1**: 63–74.

Caldini, G., Cenci, G., Manenti, R. and Morozzi, G. (1995). The ability of an environmental isolate of *Pseudomonas fluorescens* to utilize chrysene and other four-ring polynuclear aromatic hydrocarbons. *Appl. Microbiol. Biotechnol*. **44**: 225-229.

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

Colbert, C.L., Couture, M.M., Eltis, L.D. and Bolin, J.T. (2000). A cluster exposed: structure of the Rieske ferredoxin from biphenyl dioxygenase and the redox properties of Rieske Fe-S proteins. *Structure. Fold. Des*. **8**: 1267-1278.

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

Demaneche, S., Meyer, C., Micoud, J., Louwagie, M., Willison, J.C. And Jouanneau, Y. (2004). Identification and functional analysis of two aromatic ring-hydroxylating dioxygenases from a

*Sphingomonas* strain degrading various polycyclic aromatic hydrocarbons. *Appl. Environ. Microbiol.* **70**: 6714-6725.

Dong, X., Fushinobu, S., Fukuda, E., Terada, T., Nakamura, S., Shimizu, K., Nojiri, H., Omori, T., Shoun, H. and Wakagi, T. (2005). Crystal Structure of the Terminal Oxygenase Component of Cumene Dioxygenase from *Pseudomonas fluorescens* IP01. *J. Bacteriol*. **187**: 2483-2490.

Dur, J.C., Rama, R., Mougin, C. and Chaplain, V. (2000). Use of surfactants to enhance the removal of PAHs from soil. *Polycyclic Aromatic Compounds*. **20**: 143-154.

Ellis, L.B.M., Roe, D. and Wackett, L.P. (2006). The University of Minnesota Biocatalysis / Biodegradation Database: The First Decade. *Nucleic Acids Research D* **34**: 517-521.

Ensley, B.D., Ratzkin, B.J., Osslund, T.D., Simon, M.J., Wackett, L.P. and Gibson, D.T. (1983). Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. *Science*. **222**: 167–169.

Ferraro, D.J., Gakhar, L. and Ramaswamy, S. (2005). Rieske business: Structure–function of Rieske non-heme oxygenases. *Biochemical and Biophysical Research Communications* **338**: 175–190.

Friemann, R., Ivkovic-Jensen, M.M., Lessner, D.J., Yu, C., Gibson, D.T., Parales, R.E., Eklund, H. and Ramaswamy, S. (2005). Structural Insight Into the Dioxygenation of Nitroarene Compounds: The Crystal Structure of Nitrobenzene Dioxygenase. *J. Mol. Biol*. **348**: 1139-1151.

Furukawa, K., Suenaga, H. and Goto, M. (2004). Biphenyl Dioxygenases: Functional Versatilities and Directed Evolution. *J. Bacteriol*. **186**: 5189-5196.

Furusawa, Y., Nagarajan, V., Tanokura, M., Masai, E., Fukuda, M. and Senda, T. (2004). Crystal Structure of the Terminal Oxygenase Component of Biphenyl Dioxygenase Derived from *Rhodococcus* sp. Strain RHA1. *J. Mol. Biol*. **342**: 1041-1052.

Gakhar, L., Malik, Z.A., Allen, C.C.R., Lipscomb, D.A., Larkin, M.J. and Ramaswamy, S. (2005). Structure and Increased Thermostability of *Rhodococcus sp*. Naphthalene 1,2-Dioxygenase. *J. Bacteriol*. **197**: 7222-7231.

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

Ge, Y. and Eltis, L.D. (2003). Characterization of hybrid toluate and benzoate dioxygenases. *J. Bacteriol.* **185**: 5333-5341.

Gibson, D.T. (1999). *Beijerinckia* sp strain B1: a strain by any other name. *J. Ind. Microbiol. Biotechnol*. **23**: 284-293.

Hogan, J.A. (1997). Composting for soil remediation. In Biological treatment of hazardous wastes. Edited by G.A. Lewandowski and L.J. De Filippi. John Wiley & Sons, New York, N.Y., pp. 373–395.

Jerina, D.M., Daly, J.W., Jeffrey, A.M. and Gibson, D.T. (1971). Cis-1,2-dihydroxy-1a a bacterial metabolite from naphthalene. *Arch. Biochem. Biophys*. **142**: 394-396.

Johnsen, A.R., Wick, L.Y. and Harms, H. (2005). Principles of microbial PAH-degradation in soil. *Envir. Pollution* **133**: 71-84.

Jimenez, I.Y. and Bartha, R. (1996). Solvent-Augmented Mineralization of Pyrene by a

*Mycobacterium* sp. *Appl. Environ. Microbiol*. **62**: 2311–2316.

Jones, K.C., Stratford, J.A., Waterhouse, K.S., Furlong, E.T., Giger, W., Hites, R.A., Schaffner, C. and Johnston, A.E. (1989). Increases in the polynuclear aromatic hydrocarbon content of an agricultural soil over the last century. *Environ. Sci. Technol*. **23**: 95-101.

Jouanneau, Y., Willison, J.C., Meyer, C., Krivobok, S., Chevron, N., Besombes, J.L. and Blake, G. (2005). Stimulation of Pyrene Mineralization in Freshwater Sediments by Bacterial and Plant Bioaugmentation. *Environ. Sci. Technol*. **39**: 5729-5735.

Juhasz, A.L., Britz, M.L. and Stanley, G.A. (1997). Degradation of fluoranthene, pyrene, benz[a]anthracene and dibenz[a,h]anthracene by *Burkholderia cepacia*. *J. Appl. Microbiol*. **83**: 189-198.

Kanaly, R.A. and Harayama, S. (2000). Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. *J. Bacteriol*. **182**: 2059-2067.

Kanaly, R.A., Bartha, R., Watanabe, K. and Harayama, S. (2000b). Rapid Mineralization of Benzo[a]pyrene by a Microbial Consortium Growing on Diesel Fuel. *Appl. Environ. Microbiol*. **66**: 4205-4211.

Karlsson, A., Beharry, Z.M., Matthew, E.D., Coulter, E.D., Neidle, E.L., Kurtz, D.M.Jr., Eklund, H. and Ramaswamy, S. (2002). X-ray crystal structure of benzoate 1,2-dioxygenase reductase from *Acinetobacter sp*. strain ADP1. *J. Mol. Biol*. **318**: 261-272.

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

Kastner, M., Breuer-Jammali, M. and Mahro, B. (1994). Enumeration and characterization of the soil microflora from hydrocarbon-contaminated soil sites able to mineralize polycyclic aromatic hydrocarbons (PAH). *Appl. Microbiol. Biotechnol.* **41**:267-273

Kauppi, B., Lee, K., Carredano, E., Parales, R.E., Gibson, D.T., Eklund, H. and Ramaswamy, S. (1998). Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase. *Structure* **6**: 571-586.

Kimura, N., Nishi, A., Goto, M. and Furukawa, K. (1997). Functional analyses of a variety of chimeric dioxygenases constructed from two biphenyl dioxygenases that are similar structurally but different functionally. *J. Bacteriol.* **179**: 3936-3943.

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

Lee, K. (1999). Benzene-Induced Uncoupling of Naphthalene Dioxygenase Activity and Enzyme Inactivation by Production of Hydrogen Peroxide. *J. Bacteriol*. **181**: 719-2725.

Nam, J.W., Nojiri, H., Yoshida, T., Habe, H., Yamane, H. and Omori, T. (2001). New Classification System for Oxygenase Components Involved in Ring-Hydroxylating Oxygenations. *Biosci. Biotechnol. Biochem.* **65**: 254-263.

Nam, J.W., Noguchi, H., Fujiomoto, Z., Mizuno, H., Ashikawa, Y., Abo, M., Fushinobu, S., Kobashi, N., Wakagi, T., Iwata, K., Yoshida, T., Habe, H., Yamane, H., Omori, T. and Nojiri, H. (2005). Crystal structure of the ferredoxin component of carbazole 1,9a-dioxygenase of *Pseudomonas resinovorans* strain CA10, a novel Rieske non-heme iron oxygenase system. *Proteins.* **58**: 779-789.

Narro, M.L., Cerniglia, C.E., Van Baalen, C. and Gibson, D.T. (1992). Metabolism of phenanthrene by the marine cyanobacterium *Agmenellum quadruplicatum* PR-6. *Appl. Environ. Microbiol*. **58**: 1351–1359.

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

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

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

Quantin, C., Joner, E.J., Portal, J.M. and Berthelin, J. (2005). PAH dissipation in a contaminated river sediment under oxic and anoxic conditions. *Environ. Pollution* **134**: 315-322.

Randerath, K., Randerath, E., Zhou, G.D., Supunpong, N., He, L.Y., mcDonald, T.J. and Donnelly, K.C. (1999). Genotoxicity of complex PAH mixtures recovered from contaminated lake sediments as assessed by three different methods. *Environ. Mol. Mutagen*. **33**: 303-312.

Schneider, J., Grosser, R., Jayasimhulu, K., Xue, W. and Warshawsky, D. (1996). Degradation of pyrene, benz[a]anthracene, and benzo[a]pyrene by *Mycobacterium* sp. strain RJGII-135, isolated from a former coal gasification site. *Appl. Environ. Microbiol*. **62**: 13-19.

Senda, T., Yamada, T., Sakurai, N., Kubota, M., Nishizaki, T., Masai, E., Fukuda, M. and Mitsuidagger, Y. (2000). Crystal structure of NADH-dependent ferredoxin reductase component in biphenyl dioxygenase. *J. Mol. Biol*. **304**: 397-410.

Szeliga, J. and Dipple, A. (1998). DNA adduct formation by polycyclic aromatic hydrocarbon dihydrodiol epoxides. *Chem. Res. Toxicol*. **11**: 1-11.

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

Walter, U., Beyer, M., Klein, J. and Rehm, H.(1991). Degradation of pyrene by *Rhodococcus* sp. UW1. *Appl. Microbiol. Biotechnol.* **34**: 671-676.

Werlen, C., Kohler, H.P.E. and van der Meer, J.R. (1996). 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**: 4009-4016.

Wolfe, M.D., Parales, J.V., Gibson, D.T. and Lipscomb, J.D. (2001). Single Turnover Chemistry and Regulation of $O_2$ Activation by the Oxygenase Component of Naphthalene 1,2-Dioxygenase**.** *J. Biol. Chem.* **276**: 1945-1953.

Wolfe, M.D. and Lipscomb, J.D. (2003). Hydrogen Peroxide-coupled *cis*-Diol Formation Catalyzed by Naphthalene 1,2-Dioxygenase. *J. Biol. Chem.* **278**: 829-835

Yen, K.M. and Serdar, C. M. (1988). Genetics of naphthalene catabolism in pseudomonads. *Crit. Rev. Microbiol*. **15**: 247-268.

Zhou, H.W., Guo, C.L., Wong, Y.S. and Tam, N.F.Y. (2006). Genetic diversity of dioxygenase genes in polycyclic aromatic hydrocarbon-degrading bacteria isolated from mangrove sediments. *FEMS Microbiology Letters* **0**:??-??.

Zielinski, M., Backhaus, S. and Hofer, B. (2002). The principal determinants for the structure of the substrate-binding pocket are located within a central core of a biphenyl dioxygenase alpha subunit. *Microbiology* **148**: 2439-2448.

### Travail universitaire

Kuony, S., *Caracterisation d'arene dioxygenases impliquees dans la biodegradation frd hydrocarbures aromatiques polycycliques chez Mycobacterium sp.* These de Doctorat, Université Joseph Fourier, Grenoble, 20 juin 2005, 153 p.

### Ouvrage imprimé

Jouanneau, Y., Willison, J., Rodarie, D. (1999). Dégradation microbiologique des hydrocarbures aromatiques polycycliques, ADEME Editions, 156 p.

### Sites web consultés

UMBBD, Base de données des voies de dégradation des HAP. http://umbbd.msi.umn.edu/

INERIS, Fiche de données toxicologiques et environnementales des substances chimiques. En particulier les HAP, http://www.ineris.fr.

NIH, 11ème rapport sur les substances cancérigènes listant 15 HAP.
http://ntp-server.niehs.nih.gov/ntp/roc/eleventh/profiles/s150pah.pdf

USDHHS, Rapport des données toxicologiques des goudrons et la créosote, 2002. http://www.atsdr.cdc.gov/toxprofiles/tp85.pdf

# Partie 2 : Article 1

## Characterization of a naphthalene dioxygenase endowed with an exceptionally broad substrate specificity towards polycyclic aromatic hydrocarbons

Cet article a été publié en 2006 dans la revue *Biochemistry* (volume 45, pages 12380-12391). Il traite de la caractérisation biochimique de l'enzyme métabolisant les HAP chez une souche bactérienne isolée pour sa capacité à pousser sur du chrysène comme seule source d'énergie et de carbone. La majorité des travaux ont été effectués au laboratoire de Grenoble et l'étude RPE a été menée an collaboration avec Jacques Gaillard (CEA, DRFMC).

# Characterization of a naphthalene dioxygenase endowed with an exceptionally broad substrate specificity towards polycyclic aromatic hydrocarbons

**Yves Jouanneau *[,‡], Christine Meyer[‡], Jean Jakoncic[§], Vivian Stojanoff[§], Jacques Gaillard[‖]**

[‡]CEA, DSV, DRDC, Lab. Biochim. Biophys. Syst. Intégrés; CNRS, UMR 5092, F-38054 Grenoble, France ; [§]Brookhaven National Laboratory, National Synchrotron Light Source, Upton, NY 11973, USA; [‖]CEA, DRFMC, SCIB, LRM, UMR UJF-CEA 3, F-38054 Grenoble , France.

**Abbreviations**

GC-MS, gas chromatography coupled to mass spectrometry;
HEPES, N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid!;
HPLC, high performance liquid chromatography;
Ht, His-tagged;
IMAC; immobilized metal affinity chromatography;
IPTG, isopropyl-β-D-thiogalactopyranoside;
PAH; polycyclic aromatic hydrocarbon;
$Red_{B356}$, reductase component of the biphenyl dioxygenase from *C. testosteroni* B356;
RHD, ring hydroxylating dioxygenase

# Abstract

In *Sphingomonas* CHY-1, a single ring-hydroxylating dioxygenase is responsible for the initial attack of a range of polycyclic aromatic hydrocarbons (PAHs) composed of up to five rings. The components of this enzyme were separately purified and characterized. The oxygenase component (ht-PhnI) was shown to contain one Rieske-type [2Fe-2S] cluster and one mononuclear Fe center per alpha subunit, based on EPR measurements and iron assay. Steady-state kinetic measurements revealed that the enzyme had a relatively low apparent Michaelis constant for naphthalene ($K_m$= 0.92 ± 0.15 μM), and an apparent specificity constant of 2.0 ± 0.3 $\mu M^{-1}$ $s^{-1}$. Naphthalene was converted to the corresponding 1,2-dihydrodiol with stoichiometric oxidation of NADH. On the other hand, the oxidation of eight other PAHs occurred at slower rates, and with coupling efficiencies that decreased with the enzyme reaction rate. Uncoupling was associated with hydrogen peroxide formation, which is potentially deleterious to cells and might inhibit PAH degradation. In single turnover reactions, ht-PhnI alone catalyzed PAH hydroxylation at a faster rate in the presence of organic solvent, suggesting that the transfer of substrate to the active site is a limiting factor. The four-ring PAHs chrysene and benz[a]anthracene were subjected to a double ring-dihydroxylation, giving rise to the formation of a significant proportion of *bis-cis*-dihydrodiols. In addition, the dihydroxylation of benz[a]anthracene yielded three dihydrodiols, the enzyme showing a preference for carbons in positions 1,2 and 10,11. This is the first characterization of a dioxygenase able to dihydroxylate PAHs made up of four and five rings.

# 1. Introduction

Ring-hydroxylating dioxygenases (RHDs) are widely spread bacterial enzymes that play a critical role in the biological degradation of a large array of aromatic compounds, including polycyclic aromatic hydrocarbons (PAHs)(*1, 2*). RHDs catalyze the initial oxidation step of such compounds, which consists in the hydroxylation of two adjacent carbon atoms of the aromatic ring, thus generating a *cis*-dihydrodiol. This reaction converts hydrophobic, often toxic, molecules, into more hydrophilic products, allowing for their subsequent metabolism by other bacterial enzymes. Some RHDs were found to attack highly recalcitrant environmental pollutants, including dibenzo *p*-dioxin (*3, 4*), polychlorobiphenyls (*5*), and PAHs (*6-8*), thus promoting studies on this type of enzymes with the ultimate goal of improving bioremediation processes (*2, 9*). RHDs are multi-component enzymes, generally composed of a NADH-oxidoreductase, a ferredoxin and an oxygenase component that contains the active site. Sometimes, the reductase and the ferredoxin are fused in a single polypeptide. The oxygenase component is a multimeric protein, with either an $\alpha_n\beta_n$ (n=2 or 3) or $\alpha_3$ structure, that contains one [2Fe-2S] Rieske cluster and one non-heme iron atom per $\alpha$ subunit (*1*). During a catalytic cycle, two electrons from the reduced pyridine nucleotide are transferred, via the reductase, the ferredoxin and the Rieske center, to the Fe(II) ion at the active site. The reducing equivalents allow the activation of molecular oxygen, which is a prerequisite to dihydroxylation of the substrate (*10*).

So far, only a few RHDs have been purified and extensively characterized, including phthalate dioxygenase (*11, 12*), naphthalene dioxygenase (*13, 14*) and biphenyl dioxygenase (*15*). None of these enzymes is able to oxidize substrates with more than three fused rings, and data on the mechanism, kinetics and efficiency of the oxidation of high molecular weight PAHs by bacterial dioxygenases are relatively scarce (*16*). However, the four-ring PAHs chrysene and benz[a]anthracene, and the five-ring benzo[a]pyrene are of particular concern because they are well-documented carcinogens (*17*). Recently, a Sphingomonad endowed with the remarkable ability to grow on chrysene as sole carbon and energy source was isolated in our laboratory (*18*). In this strain, called *Sphingomonas* sp. CHY-1, a single dioxygenase was shown to be responsible for the oxidation of polycyclic hydrocarbons made of 2 to 4 rings (*6*). In the present study, the three components of the dioxygenase were purified and characterized, and the catalytic properties of the enzyme with respect to the oxidation of nine PAHs were examined. Due to the broad specificity of this enzyme, the kinetics and coupling

efficiency of the dioxygenase-catalyzed reaction with 2 to 5-ring PAHs could be compared for the first time. Steady-state kinetic parameters were determined for representative 2-ring PAHs. In addition, the reactivity and regioselectivity of the enzyme towards benz[a]anthracene was further investigated by means of single turnover chemistry and EPR spectroscopy.

## 2. Materials and Methods

### 2.1. Bacterial strains and growth conditions

Strains of *Escherichia coli* and *Pseudomonas putida* carrying the relevant expression plasmids, as well as general culture conditions, have been previously described (*6*). Large-scale cultures required for the purification of the enzyme components were grown on rich medium, either Luria-Bertani or Terrific broth (*19*), in a 12-L fermentor (Discovery 100, SGI-Inceltech/New Brunswick Scientific, Paris, France). Cultures destined to the overproduction of the oxygenase or the ferredoxin component were supplemented with 50 μM ferrous ammonium sulfate. The medium was inoculated with 400 ml of an overnight culture, then incubated at 37°C under constant aeration and agitation (500 rpm), until the bacterial density ($OD_{600}$) reached about 1.0. The temperature was then lowered to 25°C, IPTG was added to 0.2 mM final concentration, and the culture was further incubated for 20 h before being harvested by centrifugation. The bacterial pellet was washed with 50 mM Tris-HCl buffer (pH 7.5), and kept frozen until use.

### 2.2. Protein purification

All purification procedures were carried out under argon, using buffers equilibrated for at least 24 h in a glove box maintained under anoxic conditions ($O_2$ <2 ppm, Jacomex , France). The temperature was kept at 0-4 °C except when otherwise indicated. Crude extracts were prepared by thawing the bacterial pellets in twice as much lysis buffer by volume, followed by lysozyme treatment (0.5 mg/ml) for 15 min at 30°C. The lysis buffer was either 50 mM Tris-HCl, pH 7.5 (oxygenase preparation), 50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 10% glycerol (reductase preparation) or 50 mM potassium phosphate, pH 7.5, 0.5 M NaCl, 10% glycerol, 2 mM β-mercaptoethanol (ferredoxin preparation). The suspension was then subjected to ultrasonication for a total time of 5 min at 80% of maximal intensity, using a Vibra Cell apparatus run in pulse mode at 5 s/pulse (Fisher Bioblock Scientific, Illkirch,

France). The lysate was centrifuged at 12,000 g for 30 min, and the resulting cell extract was used as the starting material for protein purification.

### 2.2.1. Purification of the oxygenase component PhnI

A cell extract was prepared as described above from *P. putida* KT2442 carrying plasmid pSD9 (*6*). The extract obtained from approx. 50 g of cells was diluted two-fold with TGE buffer (25 mM Tris-HCl, pH 7.5, containing 5% glycerol, 5% ethanol, and 2 mM β-mercaptoethanol), and applied to a 40-ml column of DEAE-cellulose (DE52, Whatman) equilibrated with TGE buffer. After washing the column with four bed volumes of the same buffer, the oxygenase was eluted as a brown band with buffered 0.3 M NaCl. The eluate was immediately applied to a small column (7 ml) of immobilized metal affinity chromatography (IMAC) resin loaded with $Co^{2+}$ (TALON, BD Biosciences Ozyme, France). The column was washed successively with 8 bed volumes of TGE buffer containing 0.5 M NaCl, and 5 bed volumes of the same buffer supplemented with 20 mM imidazole. A brown protein fraction was then eluted with TGE buffer containing 0.15 M imidazole. This fraction was diluted 6-fold with TGE buffer and applied to a small column of DEAE-cellulose (4 ml). The purified protein was eluted in a small volume of TGE buffer containing 0.3 M NaCl, and frozen as pellets in liquid nitrogen. This preparation, designated ht-PhnI, was judged to be at least 95% pure by SDS-PAGE.

### 2.2.2. Purification of the ferredoxin component PhnA3

PhnA3 was overproduced in *E. coli* BL21AI (Invitrogen) carrying plasmid pEBA3 (*15*). The cell extract prepared from 154 g packed cells was loaded onto two columns of IMAC-TALON, 13 ml each, equilibrated in PG buffer (50 mM potassium phosphate, pH 7.5, 10% glycerol, 2 mM β-mercaptoethanol), containing 0.5 M NaCl. Each column was washed with 100 ml of equilibration buffer, followed by 50 ml of buffered 20 mM imidazole. A brown protein fraction was eluted with PG buffer containing 0.15 M imidazole. This fraction was immediately diluted 5-fold with PG buffer, and loaded onto a 10-ml DEAE cellulose column. After washing with two bed volumes of PG buffer, the brown ferredoxin fraction was eluted with buffered 0.3 M NaCl in a volume of 8.6 ml. This preparation was designated ht-PhnA3.

Part of the purified His-tagged protein (4.7 μmoles) was cleaved by incubation with

thrombin (10 U/ µmole) for 16 h at 20°C, in buffer containing 0.15 M NaCl and 2 mM $CaCl_2$, pH 8.2. The digested protein was passed through a 2-ml IMAC-TALON column, diluted 5-fold with PG buffer, then loaded onto a 2-ml column of DEAE-cellulose. The ferredoxin was eluted in a small volume PG buffer containing 0.3 M NaCl, and frozen as pellets in liquid nitrogen. This preparation was referred to as rc-PhnA3.

#### 2.2.3. Purification of the reductase component PhnA4

PhnA4 was overproduced in *E. coli* BL21(DE3) carrying plasmid pEBA4 (*15*). The crude extract from 16 g of cells was applied to a 2-ml column of IMAC-TALON equilibrated in TG buffer (Tris-HCl, pH 8.0, 10% glycerol) containing 0.5 M NaCl. The column was washed with 10 bed volumes of equilibration buffer, and 3 bed volumes of TG buffer containing 0.5 M NaCl and 10 mM imidazole. A yellow protein fraction was then eluted with TG buffer containing 0.15 M imidazole. This fraction was dialyzed for 16 h against TG buffer, and further purified on a second column of IMAC-TALON (1 ml). The column was successively washed with TG buffer containing 0.5 M NaCl (10 ml), and the same buffer containing 10 mM (9 ml), and 20 mM (4 ml) imidazole. The reductase was eluted in a small volume of TG buffer containing 0.15 M imidazole, dialyzed as above, and concentrated to 0.8 ml by ultrafiltration using an Ultrafree centifugal device with 30-kDa cut-off (Millipore, Amilabo, France). The purified protein was stored as pellets in liquid nitrogen.

#### 2.2.4. Purification of ht-$Red_{B356}$

The reductase component of the biphenyl dioxygenase from *C. testosteroni* B-356, designated as ht-$Red_{B356}$, was purified from *E. coli* SG12009(pREP4)(pEQ34*::bphG*) (*20*), as previously described (*21*).

### 2.3. Enzyme assays

Dioxygenase activity was assayed either by following NADH oxidation at 340 nm or by measuring the rate of $O_2$ consumption using a Clark-type $O_2$ electrode (Digital model 10; Bioblock Scientific, Illkirch, France). Polarographic measurements (standard assay) were performed at 30°C in reaction mixtures (1 ml) containing 0.13 µM ht-PhnI, 1.57 µM PhnA3, 0.40 µM ht-$Red_{B356}$ , and 0.5 mM NADH in 50 mM potassium phosphate buffer, pH 7.0. The

PAH substrate was supplied at 0.1 mM from a concentrated solution in acetonitrile. The concentration of all three enzyme components was doubled for assays with 4- and 5-ring PAHs. The reaction was initiated by injecting, with a gas-tight syringe, a 100-fold concentrated mixture of the proteins kept under argon on ice in phosphate buffer containing 10% glycerol, 10 mM dithiothreitol and 0.05 mM ferrous ammonium sulfate. The enzyme activity was determined from the initial rate of $O_2$ consumption and expressed as µmol $O_2$ per min per mg ht-PhnI. Reaction rates were calculated from duplicate assays and corrected for the $O_2$ consumption measured in control assays carried out in the absence of PAH substrate.

The $O_2$ electrode was also used to determine the coupling between PAH oxidation and $O_2$ consumption as follows. After approx. 50 nmol $O_2$ had been consumed, 300 U of bovine liver catalase (Sigma) was added as a means to estimate the amount of $H_2O_2$ generated during the reaction. Then, 0.6 ml of reaction mixture was withdrawn and immediately mixed with an equal volume of ice-cold acetonitrile. The dihydrodiols present in these samples were directly quantified by HPLC as described below. When appropriate, the dihydrodiols were extracted with ethyl acetate, derivatized and analyzed by GC-MS (see below).

To determine the coupling efficiency between NADH and PAH oxidation, some reactions were carried out in Eppendorf tubes containing 0.1 mM of substrate and 0.2 mM of NADH in 0.6 ml of reaction mixture. In those assays, the concentrations of the enzyme components were 0.38 µM (ht-PhnI), 6 µM (PhnA3) and 2 µM (ht-$Red_{B356}$). After an incubation time of 2 to 10 min at 30°C, depending on the substrate, the reaction was stopped by addition of an equal volume of acetonitrile. Residual NADH, and the dihydrodiols formed during the enzymatic reaction, were separated and quantified by HPLC.

Steady-state kinetic parameters of the dioxygenase-catalyzed reaction were determined from sets of enzyme assays where the substrate concentration was varied over a 0.5-100 µM range. The component ratio was the same as in the standard assay, but the protein concentration in the assays was 1.67-fold higher. The initial NADH concentration was 0.2 mM. Reactions were carried out at 30°C in quartz cuvettes, and the absorption at 340 nm was recorded at 0.1-s intervals over 1 min with a HP8452 spectrophotometer (Agilent Technologies, Les Ulis, France). The enzyme activity was calculated from the initial linear portion of the time course, using an absorption coefficient of 6,220 $M^{-1}.cm^{-1}$ for NADH. When biphenyl was used as substrate, NADH oxidation was recorded at 360 nm ($\varepsilon_{360}$ = 4,320 $M^{-1}.cm^{-1}$), because biphenyl 2,3-dihydrodiol absorbed at 340 nm. All assays were performed in duplicate, and at least 12 concentrations were tested per substrate. Plots of the initial reaction rate versus substrate concentration were fitted to the Michaelis-Menten equation using the curve fit option of

Kaleidagraph (Synergy Software). Only curve fits showing correlation coefficients better than 0.98 were considered.

### 2.4. Single turnover reactions

Ht-PhnI was diluted to 57 μM in 20 mM HEPES, pH 7.0 containing 10% glycerol and 5 μM methyl viologen under argon, then reduced with a stoichiometric amount of dithionite. The reduction was checked by monitoring the protein absorbance in the 300-600 nm range. A portion of the reduced protein (50 μL) was diluted in 0.55 ml of air-saturated HEPES buffer containing 0.1 mM of PAH substrate. In some experiments, the buffer also contained a proportion of acetonitrile, as indicated. After incubation at 30°C for up to 10 min, the reaction was stopped by mixing with an equal volume acetonitrile containing 0.8% acetic acid. The mixture was heated for 2 min at 90°C, centrifuged and subjected to HPLC analysis as described below. Part of the solution was also extracted with ethyl acetate, and analyzed by GC-MS.

### 2.5. Identification and quantification of reaction products

Determination of dihydrodiols and the residual NADH concentration at the end of the dioxygenase-catalyzed reactions was performed by HPLC using a Kontron system equipped with F430 UV detector. Samples (0.2 ml) were injected onto a 4×150-mm C8 reverse-phase column (Zorbax, Agilent Technologies, France) run at 0.8 ml/min. The column was eluted with water for 2 min, then with a linear gradient to 80% acetonitrile for 8 min, and finally with 80% acetonitrile for 5 min. Detection was carried out at 340 nm (for residual NADH), and one of the following wavelengths, which was varied as a function of the absorbance maxima of the PAH dihydrodiols: 220 nm (naphthalene), 303 nm (biphenyl), 260 nm (phenanthrene), 244 nm (anthracene), 263 nm (benz[a]anthracene), 278 nm (chrysene), 280 nm (benzo[a]pyrene). Quantification was performed on the basis of peak area using calibration curves obtained by injecting known amounts of each dihydrodiol. Residual NADH was determined from the peak eluting at 2.6 min. The three benz[a]anthracene dihydrodiols formed by ht-PhnI were not resolved under the HPLC conditions used, and were estimated as a sum of their individual contribution, given that their absorbance coefficients at 263 nm were close to 31,000 $M^{-1}.cm^{-1}$. Purified 1,2-dihydroxy-1,2-dihydrobenz[a]anthracene was used for

HPLC calibration. The extent of oxidation of fluorene and fluoranthene by the dioxygenase was determined from HPLC measurements of the amount of residual substrate. Wavelengths used for their detection were 262 and 236 nm, respectively.

PAH oxidation products generated by PhnI were also analyzed by GC-MS. Ethyl acetate extracts of samples were dried on sodium sulfate, evaporated under $N_2$, and derivatized with bis(trimethysilyl)trifluoroacetamide :trimethylchlorosilane (99:1) from Supelco (Sigma-Aldrich), prior to GC-MS analysis using a HP6890/HP5973 apparatus (Agilent Technologies). Operating conditions were as previously described (*22*), and mass spectrum acquisitions were carried out either in the total ion current or the single ion monitoring mode.

### 2.6. Determination of the iron content of proteins

To extract iron from proteins, samples (150 µl) were treated with 2.5 N HCl for 30 min at 95°C, then diluted with 0.7 volume of water. Iron was reacted with bathophenanthroline disulfonate (Sigma-Aldrich), and the complex formed was assayed by absorbance measurements at 536 nm (*23*). Assays were performed in triplicates. A calibration curve was generated by assaying serial dilutions of a standard solution of ferric nitrate, containing 1g/L of iron (Merck).

### 2.7. Protein analyses

Routine protein determinations were performed using the Bradford assay (*24*), or the bicinchoninic acid reagent kit (Pierce) using bovine serum albumin as a standard. The protein concentration of purified preparations of ht-PhnI was determined by a modification of the biuret assay (*25*). The absorbance coefficient of ht-PhnI at 458 nm was calculated to be 12,500 $M^{-1}.cm^{-1}$, on the basis of the latter assay. The concentrations of ht-PhnA3 and rc-PhnA3 were estimated from absorbance measurement at 460 nm, using an absorbance coefficient of 5,000 $M^{-1}.cm^{-1}$. SDS-PAGE on mini-slab gels was performed as previously described (*26*). The molecular masses of purified ht-PhnI and rc-PhnA3 were determined by size-exclusion chromatography on HR 10/30 columns of Superdex SD200 and SD75, respectively (both from Amersham Biosciences). The columns were run at a flow rate of 0.2 ml/min and calibrated with the following protein markers: Ferritin (443 kDa), catalase (240 kDa), aldolase (150 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa) and myoglobin (17 kDa), aprotinin (6.5 kDa), all from Sigma-Aldrich, and ferredoxin VI from

*Rhodobacter capsulatus* (11.58 kDa; (*27*)).

### 2.8. EPR spectroscopy

Protein samples were adjusted to a concentration of 20-40µM (ht-PhnI) or 100-500 µM (PhnA3) in argon-saturated phosphate buffer, pH 7.0, containing 10% glycerol. The redox status of the protein sample was checked by recording the absorbance spectrum, and, when appropriate, the protein was fully oxidized by injecting stoichiometric amounts of air with a gas-tight syringe. Ht-PhnI-nitrosyl complexes were prepared in a glove-box under argon (Jacomex) by incubating 190 µl of protein sample with 10 µl of 20 mM diethylamine NO-NOate (Cayman Chemical, Interchim, France) for 15 min. Samples were then introduced into EPR tubes and frozen in liquid nitrogen. In some experiments, protein samples were preincubated for 10 min with a PAH (0.1 mM) or a dihydrodiol, prior to NO-NOate addition. For analysis of the Rieske clusters, protein samples were reduced with an excess of sodium dithionite (1 mM). Full reduction was checked by absorbance recording prior to transferring the samples in EPR tubes and freezing them in liquid nitrogen. Spectra were recorded at a temperature set between 4 and 20 K with an X-band EMX Bruker spectrometer equipped with an ESR900 liquid helium cryostat (Oxford Instruments). Spin quantification was performed by integrating the appropriate signal, and comparing the signal intensity to that of the [2Fe-2S] ferredoxin (FdVI) from *Rhodobacter capsulatus*, taken as a reference (0.1 mM; (*27*)). The iron content of this reference sample was checked by chemical assay. All EPR tubes were calibrated in diameter.

### 2.9. Chemicals

$NAD^+$, NADH, PAHs, and most other chemicals were purchased from Sigma-Aldrich (Saint-Quentin-Fallavier, France). The *cis*-dihydrodiols used in this study were prepared from cultures of *E. coli* recombinant strains overproducing the PhnI dioxygenase, and incubated with a PAH. The purification and characterization of the diol compounds will be described elsewhere. The dihydrodiol concentrations were calculated using the following absorption coefficients: $\varepsilon_{262}$ = 8,114 $M^{-1}.cm^{-1}$ for *cis*-1,2-dihydroxy 1,2-dihydronaphthalene (*28*); $\varepsilon_{252}$ = 38,300 $M^{-1}.cm^{-1}$ and $\varepsilon_{260}$ = 43,000 $M^{-1}.cm^{-1}$ for *cis*-3,4-dihydroxy 3,4-dihydrophenanthrene (*29*); $\varepsilon_{244}$ = 55,600 $M^{-1}.cm^{-1}$ and $\varepsilon_{287}$ = 17,000 $M^{-1}.cm^{-1}$ for *cis*-1,2-dihydroxy 1,2-

dihydroanthracene (*29*); $\varepsilon_{278}$ = 57,650 $M^{-1}.cm^{-1}$ for *cis*-3,4-dihydroxy 3,4-dihydrochrysene (*30*); $\varepsilon_{280}$ = 66,500 $M^{-1}.cm^{-1}$ for *cis*-9,10-dihydroxy-9,10-dihydrobenzo[a]pyrene (*31*); $\varepsilon_{263}$ = 31,000 $M^{-1}.cm^{-1}$ for *cis*-1,2- dihydroxy-1,2-dihydrobenz[a]anthracene and $\varepsilon_{275}$ = 37,000 $M^{-1}.cm^{-1}$ for *cis*-10,11-dihydroxy-10,11-dihydrobenz[a]anthracene (*32*).

# 3. Results

## 3.1. Purification and properties of the oxygenase component PhnI

The His-tagged oxygenase component of strain CHY-1 dioxygenase, hereafter referred to as ht-PhnI, was anaerobically purified from *P. putida* KT2442(pSD9) in three steps as described under Materials and Methods. The procedure yielded approx. 12 mg of purified protein per liter of culture. The oxygenase was also produced in *E. coli* BL21(DE3)(pSD9), but the purification resulted in a lower yield. In addition, strain BL21(DE3) always produced a variable amount of insoluble recombinant protein (inclusion bodies), which was not the case when using strain KT2442 as host (data not shown). The latter strain was therefore preferred for overproduction and subsequent purification of ht-PhnI. SDS-PAGE analysis revealed that the ht-PhnI preparation was at least 95% pure, and was composed of two subunits with apparent $M_r$ of 52.000 and 20.000 (Fig. 1), consistent with the molecular masses of the polypeptides deduced from relevant gene sequences (*6*). Purified ht-PhnI exhibited a molecular mass of approx. 200 kDa by gel filtration chromatography, indicating that it is an $\alpha_3\beta_3$ hexamer. The brown protein showed a UV-visible absorbance spectrum with maxima at 280, 458 nm and a shoulder near 570 nm (data not shown), which is typical of proteins containing Rieske-type [2Fe-2S] clusters. The absorbance coefficient at 458 nm was found to be 12,500 $M^{-1}.cm^{-1}$ on average, as calculated from the protein content of three independent preparations of ht-PhnI with a similar content of [2Fe-2S] cluster (see Table 1). In contrast to related oxygenases previously characterized, ht-PhnI did not show a well-defined absorption band near 325 nm, but instead a shoulder likely resulting from two poorly resolved absorption bands. The $A_{280}/A_{458}$ ratio was relatively high (26.9) compared to that of naphthalene dioxygenase (17.6, (*33*)), a feature which might be partly explained by the higher content of aromatic residues of PhnI (Trp and Tyr account for 2.70 and 4.46 % of the total number of residues in PhnI versus 2.18 and 3.73% in naphthalene dioxygenase). EPR analysis of the reduced protein gave a rhombic signal with apparent g values at 2.02, 1.92 and 1.71, which is characteristic of Rieske-type [2Fe-2S] clusters (data not shown).

The iron content of the oxygenase was found to vary between 1.73 and 2.55 Fe atoms per pair of αβ subunits depending on preparations (Table I). In order to estimate the proportion of iron in each metal center, the ht-PhnI preparations were subjected to two independent EPR measurements. Upon reaction with NO, ht-PhnI gave rise to the formation of an Fe(II)-nitrosyl complex which was detected as an heterogeneous S = 3/2 EPR signal near g=4 (see Fig. 3). Based on the integration of that signal, the occupation rate of Fe(II) at the active site of the enzyme was found to vary between 0.20 and 0.92 (Table I). On the other hand, the estimation of the ratio cluster/αβ, calculated from the integration of the S= 1/2 signal in fully reduced protein samples, yielded values ranging between 0.75 and 0.85. Remarkably, the iron content of the preparations calculated from the sum of the two EPR determinations was in fairly good agreement with the total iron found by chemical assay.

The specific activity of the dioxygenase increased as a function of its iron content, but no clear correlation was observed between activity and the occupation rate of the active site (Table 1). In addition, preincubation of ht-PhnI with ferrous ions under reducing conditions prior to enzyme assay resulted in a marginal increase of activity (data not shown).

### 3.2. Purification of the ferredoxin and reductase components

The ferredoxin component was anaerobically purified as a His-tagged recombinant protein by IMAC chromatography, and designated ht-PhnA3. The purification procedure described herein yielded about 20 mg of ferredoxin per liter of culture, when strain BL21AI was used as a host for expression. Lower yields were observed with strain BL21(DE3)(pEBA3). The preparation was >90% pure as judged from SDS-PAGE. Cleavage of the protein with thrombin, followed by two short purification steps, gave an essentially pure preparation containing a 12-kDa polypeptide (Fig.1). The molecular mass of this protein, referred to as rc-PhnA3, was 13.5 kDa by gel filtration, which was slightly higher than the theoretical mass of the polypeptide calculated from the *phnA3* gene sequence (11,225 Da, (*6*)), but indicated that the ferredoxin was monomeric. Both ht-PhnA3 and rc-PhnA3 exhibited absorbance spectra indicative of partial reduction upon isolation under anoxic conditions, but rapidly oxidized in air. In the oxidized state, the two preparations of ferredoxin had identical spectra, featuring absorbance maxima at 278, 325 and 460 nm. The iron content of the ht-PhnA3 and rc-PhnA3 preparations was estimated to be 1.5 and 1.7 mol/mol of ferredoxin, respectively. EPR analysis of the reduced ferredoxin gave a signal with g values at 2.02, 1.90

and 1.82, which integrated to 0.86 spin/molecule. Taken together, these data provide strong evidence that the ferredoxin component contains one Rieske-type [2Fe-2S] cluster.

The reductase component of the dioxygenase encoded by *phnA4* was overproduced as a 45 kDa polypeptide in *E. coli* BL21(DE3)(pEBA4). However, a large proportion of the recombinant protein accumulated in the cells as inclusion bodies, and this problem was not solved by changing the host strain, or by lowering the temperature during induction. Although a low level of the reductase was recovered from the soluble cell extract, the recombinant protein was purified as a His-tagged fusion (ht-PhnA4) by affinity chromatography (0.2 mg/L of culture). The isolated ht-PhnA4 protein was yellow in color, and showed an absorbance spectrum typical for a flavoprotein, with absorbance maxima at 375 and 450 nm. Attempts to overexpress the reductase in *P. putida*, either intact or as a His-tagged fusion, under conditions similar to those described for PhnI, were unsuccessful (data not shown). These observations suggested that PhnA4 was an unstable protein. For our studies on the catalytic activity of the dioxygenase, we replaced PhnA4 by the more stable component, $Red_{B356}$, of the biphenyl dioxygenase from *C. testosteroni (20)*. Enzyme assays performed under standard conditions showed that $Red_{B356}$ efficiently substituted for PhnA4, and titration experiments with increasing concentrations of the reductase indicated that the two isoforms had almost identical affinities for rc-PhnA3 (data not shown).

### 3.3. Catalytic properties of the dioxygenase complex : dependence of activity on electron carrier concentrations

Purified ht-PhnI catalyzed the oxidation of naphthalene to *cis*-1,2-dihydroxy-1,2-dihydronaphthalene, in a reaction that required the presence of the reductase ($Red_{B356}$) and ferredoxin (ht-PhnA3) components. When the reductase concentration was varied, at constant concentrations of the ferredoxin and the oxygenase, activity reached half-saturation for a reductase concentration of 0.05 μM (data not shown). When the ht-PhnA3 concentration was varied, half-saturation was obtained when ferredoxin was added to an approx. 14-fold molar excess over the oxygenase concentration (4.5 μM). rc-PhnA3 was found to be equally active, indicating that the His-tag did not alter the enzyme function. At a ferredoxin concentration close to saturation (20 μM; 60-fold molar excess), the specific activity of the enzyme complex was calculated to be 1.25 ± 0.04 U/mg ht-PhnI. For most of the assays performed in this study, the reductase and ferredoxin concentrations were set at concentrations 2.4-fold and 12-

fold higher than that of PhnI, respectively. A suboptimal level of ferredoxin was chosen to limit non-specific NADH oxidation by the protein mixture in the absence of PAH substrate.

### 3.4. Specific activity and coupling efficiency

To examine the ability of the dioxygenase to oxidize PAHs, the enzyme activity was first determined by measuring the initial rate of oxygen consumption in the presence of an excess of substrate. The dihydrodiol products formed in the reaction mixture were quantified by HPLC as described under Materials and Methods. When fluorene and fluoranthene were tested, substrate oxidation was rather estimated by measuring the amount of residual PAH at the end of the enzymatic reaction, because the oxidation products of these PAHs have not yet been fully characterized (see below and Table 3). In a second and independent set of experiments, the coupling efficiency of the PAH oxidation reactions catalyzed by the dioxygenase was determined by measuring the rates of NADH oxidation and dihydrodiol formation during catalysis. Table 2 compares the results obtained for nine PAHs in terms of specific activity, and reaction coupling between oxygen consumption, NADH oxidation and dihydrodiol formation.

Naphthalene appeared to be the only substrate yielding a stoichiometry close to 1, indicating a tight coupling between substrate and cofactor oxidation. It was also the best substrate as it gave the highest rates of $O_2$ consumption or NADH oxidation. Other substrates were utilized at rates that decreased with the number of fused rings, in reactions that gave rise to significant uncoupling between NADH oxidation and dihydrodiol formation. Chrysene appeared to be the worst substrate in terms of both oxidation rate and coupling efficiency. Although discrepancies were observed with some substrates when comparing enzyme activities assayed by $O_2$ consumption and NADH oxidation, the coupling efficiencies calculated as either dihydrodiol/$O_2$ or dihydrodiol/NADH ratios were similar within experimental error, except for phenanthrene and anthracene. It is unclear why different ratios were obtained in the two latter cases.

During steady-state catalysis, hydrogen peroxide was produced, with a $H_2O_2/O_2$ ratio that increased with the uncoupling of the reaction (Table 2). Depending on the substrate, the fraction of oxygen utilized for dihydrodiol formation varied between 8% (chrysene) and 100% (naphthalene), the balance of $O_2$ consumed being mainly allocated to $H_2O_2$ formation. However, some peroxide was produced even in the tightly coupled naphthalene hydroxylation reaction, the amount of which corresponded to the background $O_2$ consumption observed in

the absence of PAH. Since the enzyme was saturated with naphthalene, the involvement of ht-PhnI in $H_2O_2$ formation was unlikely, suggesting that the electron carriers were responsible for this side reaction. In a control experiment, we observed that electron carriers alone gave rise to an $O_2$ consumption of 3.9 nmol.min$^{-1}$ compared to 5.1 nmol.min$^{-1}$ for the complete enzyme system, and generated 0.42 $H_2O_2$ per $O_2$ consumed (versus 0.57 for the complete system). Hence, a large proportion of the peroxide produced during *in vitro* catalysis of PAH hydroxylation was contributed by the electron carriers alone, most likely through air-oxidation of the reduced PhnA3 ferredoxin component.

### 3.5. Steady-state kinetics

Using naphthalene and biphenyl as substrates, the steady state rate of the PhnI-catalyzed reaction was determined in the 0.5-100 μM concentration range. The reaction was monitored spectrophotometrically, by measuring the kinetics of NADH oxidation. This assay method was preferred to the polarographic method, since at low substrate concentrations, the response time of the oxygen electrode was too long to account for the rapid consumption of the substrate. The dioxygenase exhibited a Michaelis-type behavior with respect to substrate concentration, and results indicated that the enzyme had an apparent $K_m$ as low as 0.92 ± 0.15 μM for naphthalene. The apparent turnover number for this substrate was 1.82 ± 0.03 s$^{-1}$. The enzyme showed a similarly low $K_m$ for biphenyl (0.42 ± 0.20 μM), the latter value being only an estimate as enzyme kinetics were extremely short (<4 s) and difficult to calculate accurately at substrate concentrations below 1.0 μM. The turnover number, expressed in terms of rate of dihydrodiol formed, was smaller (1.01 ± 0.04 s$^{-1}$), taking into account a dihydrodiol/NADH ratio of 0.67 in the calculation (Table 2). The apparent specificity constant was calculated to be 2.0 ± 0.3 μM$^{-1}$ s$^{-1}$ for naphthalene, and 2.4 ± 1.0 μM$^{-1}$ s$^{-1}$ for biphenyl.

### 3.6. Dihydroxylations and monohydroxylations catalyzed by PhnI

GC-MS analysis of the PAH oxidation products revealed that a single dihydrodiol was generated by the dioxygenase in most cases, except when fluorene, fluoranthene, chrysene and benz[a]anthracene were used as substrates (Table 3 and Fig. 2). Biphenyl, naphthalene, phenanthrene were hydroxylated at positions 2,3-, 1,2- and 3,4-, respectively. as previously determined (*6*), whereas anthracene was most likely converted to the 1,2-dihydrodiol, as found for the dioxygenase present in *S. yanoikuyae* B1 (*29*). Fluorene oxidation gave rise to the formation of five detectable products, four of which had mass spectra corresponding to

monohydroxylated derivates (Table 3). While 9-fluorenol resulted from a monohydroxylation, the other products might have arisen from either a single hydroxylation or spontaneous dehydration of unstable dihydrodiols primarily produced by the enzyme, as proposed in a previous study on fluorene oxidation by naphthalene dioxygenase (*34*). Fluoranthene oxidation yielded only one detectable product with a mass spectrum characteristic of a monohydroxylated molecule (the prominent fragment at m/z=290 in Table 3 corresponds to the mass of the trimethylsilyl derivate of hydroxyfluoranthene), and a UV absorbance spectrum identical to that of 8-hydroxyfluoranthene (*35*). This result suggested that the dioxygenase catalyzed a monohydroxylation of fluoranthene on the C8 position. With chrysene, the major product detected was the *cis*-3,4-dihydrodiol, as determined by comparison of GC-MS and UV absorption data with those of the previously characterized diol (*30*). A more polar compound was also detected by HPLC, which accounted for less than 10% of the total products based on peak area. This compound, which gave a trimethysilyl derivate with a mass of 584 (Table 3), had the same chromatographic properties as the 3,4,9,10-*bis*-*cis*-chrysene dihydrodiol. We have independently identified this product based on proton and $^{13}C$ NMR (Jouanneau, Meyer, and Duraffourg, unpublished results). Finally, the dioxygenase-catalyzed oxidation of benzo[a]pyrene yielded a single product with a mass spectrum characteristic of a dihydrodiol derivate (Table 3). The UV spectrum of this product was identical to that of the *cis* -9,10-benzo[a]pyrene dihydrodiol (*31*).

### 3.7. Dihydroxylation of benz[a]anthracene

`Benz[a]anthracene was converted by CHY-1 dioxygenase to three *cis*-dihydrodiol isomers and one *bis*-*cis*-dihydrodiol (Table 3 and Fig. 2). The three dihydrodiols have been independently purified and identified as the 1,2-, 8,9- and 10,11-isomers, based on a good match of GC-MS and UV absorbance data with previously published data (*32*). Quantitative analysis of the diols by GC-MS in several experiments showed that the 1,2-isomer was most abundant (68 ± 7%), with the 8,9- and the 10,11-isomers representing 9 ± 3% and 23 ± 4 % of the diols formed, respectively (average of 6 determinations). The proportion of *bis*-*cis*-dihydrodiol increased during the course of the enzymatic reaction, suggesting that at least one of the dihydrodiol reacted a second time with the enzyme to form the *bis*-*cis*-dihydrodiol. To test this hypothesis, the 1,2- and 10,11-isomers were independently provided as substrates to the dioxygenase. Interestingly, the two dihydrodiols triggered a fast and uncoupled oxidation of NADH, with small amounts of *bis*-*cis*-dihydrodiol produced (Table 2). Nevertheless, since

the two isomers yielded the same product as judged from HPLC and GC-MS analysis, it is inferred that the *bis*-*cis*-dihydrodiol bore hydroxyls on carbons in positions 1, 2, 10 and 11 of the benz[a]anthracene molecule. The enzymatic reaction generated hydrogen peroxide at a rate much higher than that attributed to the electron carriers, indicating that, in this case, the formation of $H_2O_2$ was mainly due to futile cycling of the oxygenase.

### 3.8. Reactivity of ht-PhnI toward benz[a]anthracene as investigated by single turnover experiments

To further investigate the reactivity of the dioxygenase toward benz[a]anthracene, single turnover reactions were carried out under conditions similar to those previously described for naphthalene dioxygenase (*10*). In these experiments, the oxygenase component alone was allowed to react with the substrate in air-saturated buffer, and a rapid formation of dihydrodiol was expected at the enzyme active site. In a control experiment with naphthalene as substrate, the formation of dihydrodiol was observed on a time scale lower than 1 min. Surprisingly, the conversion of benz[a]anthracene was much slower and reached completion only after approx. 20 min. In addition, *bis*-*cis*-dihydrodiol was detected and its concentration increased linearly during the course of the reaction. These results suggested that the rate of the reaction was limited by the solubility of the substrate, which in turn reduced the accessibility of the substrate to the enzyme active site. On the other hand, the dihydrodiols produced which are soluble in water, might compete with the PAH for enzyme active sites, thus explaining the formation of *bis*-*cis*-dihydrodiol. This interpretation was tested in experiments where the solubility of the PAH was increased by adding an organic solvent to the reaction (Table 4). By carrying out the reaction in 20% acetonitrile, the solubility of benz[a]anthracene was increased 100-fold, and the reaction was completed in less than 1 min. In 30% acetonitrile, the reaction was also fast, but the product yield was lower, probably because of enzyme inactivation. GC-MS analysis of the diols formed showed that the 10,11-isomer was most abundant, and no *bis*-*cis*-dihydrodiol was detectable in reactions carried out in the presence of solvent. These results contrasted with those obtained in steady-state experiments, since in the latter case, the 1,2-isomer was the predominant product. However, the two sets of data could be reconciled by assuming that, under steady state conditions, the 10,11-isomer is converted to *bis*-*cis*-dihydrodiol faster than the 1,2-isomer (see below). The results of single turnover experiments demonstrate that the oxygenase preferentially hydroxylates benz[a]anthracene on carbons in positions 10,11. It is also shown that the reactivity of the enzyme towards water-

insoluble substrates, which is limited by substrate transfer to the active site, could be enhanced by carrying out the reaction in aqueous medium containing up to 20% organic solvent.

### 3.9. Interaction of ht-PhnI with benz[a]anthracene and dihydrodiols as probed by EPR spectroscopy

The high level of uncoupling observed when benz[a]anthracene dihydrodiols were incubated with the dioxygenase indicated that these compounds did not interact correctly with the enzyme active site. As a means to probe this interaction, we carried out EPR analysis of complexes between the active site Fe(II) and NO, in the presence or absence of substrate (Fig. 3). The spectrum of the substrate-free enzyme showed a complex signal centered at g=4.0, which might reflect the existence of more than one NO binding site. Alternatively, the signals might arise from different conformations due to different orientations of the Fe-NO bond. Interestingly, two pairs of resonance lines at 3.68/4.40 and 3.98/4.07, which were prominent in the spectrum of the free enzyme, underwent dramatic changes upon substrate binding (Fig. 3). These lines almost disappeared in the spectrum of the benz[a]anthracene-bound enzyme, and were undetectable in the case of the naphthalene-bound enzyme complex. The shape of the signals obtained for the dihydrodiol-bound enzyme complexes showed patterns intermediate between the substrate-free and the benz[a]anthracene-bound enzyme, the spectrum of the 10,11-isomer-bound enzyme being closer to the latter. These observations could be taken as indirect evidence that the dihydrodiols were not correctly oriented in the substrate-binding pocket to allow for a productive catalytic conversion into *bis-cis*-dihydrodiol. Based on these EPR data, the 10,11-isomer would bind the active site in a more favorable position than the 1,2-isomer. Alternatively, our data might indicate that the enzyme was not saturated by the dihydrodiols, although concentrations in molar excess over the enzyme catalytic sites were used. Accordingly, in experiments where the concentration of the dihydrodiols was doubled (0.2 mM), we observed that the relative intensity of the resonance lines at 3.68/4.40 and 3.98/4.07, attributed to the free enzyme (see above), was significantly reduced compared to that found in spectra b and c in Fig. 3. The spectra of the enzyme-dihydrodiol complexes were then only slightly different from those obtained with naphthalene or benz[a]anthracene as ligands (data not shown). We therefore conclude that the differences seen in spectra b and c, compared to spectrum d, are likely due, in great part, to the

contribution of substrate-free enzyme, thus reflecting a relatively low affinity of the enzyme for the two benz[a]anthracene dihydrodiols.

## 4. Discussion

The ring-hydroxylating dioxygenase described in this study exhibits one of the broadest substrate specificities toward PAHs ever reported. It is a rare example of an enzyme able to attack aromatic substrates composed of 2 to 5 rings, which gave us an opportunity to compare the kinetics of dioxygenation for a wide range of PAHs. Consistent with previous *in vivo* observations (*6*), the specific activity of strain CHY-1 dioxygenase was highest with naphthalene, and declined as a function of substrate size. Ironically, while strain CHY-1 was isolated for its ability to grow on chrysene as sole carbon source (*18*), chrysene appeared to be the worst substrate in terms of both oxidation rate and coupling efficiency (Table 2). This finding suggested that a dioxygenase other than PhnI might be responsible for the initial attack of chrysene in strain CHY-1. However, such a hypothesis can be ruled out on the basis of our previous work showing that a mutant strain lacking PhnI failed to grow on and oxidize chrysene, as well as any other PAH used as substrate by the parental strain CHY-1 (*6*). Hence, PhnI is essential for growth on chrysene, and its low activity towards this substrate is probably one of the main reason why strain CHY-1 shows slow growth and poor cell yields when provided with chrysene as sole carbon source (*18*). In comparison, benz[a]anthracene, another 4-ring PAH which was oxidized at a much higher rate, failed to support growth of strain CHY-1 (*18*). Possible reasons which might explain this paradox are discussed below. The present study also unveiled interesting new features of strain CHY-1 dioxygenase, including its ability to utilize fluoranthene and benzo[a]pyrene. Hence, this enzyme has the remarkable potential to initiate the degradation of at least half of the 16 EPA priority PAHs, including the carcinogenic benz[a]anthracene and benzo[a]pyrene. A dioxygenase activity with a similarly broad substrate specificity has only been found in *S. yanoikuyae* B1, but the corresponding enzyme has not yet been described (*36*).

The dioxygenase from strain CHY-1 is a three-component enzyme that shares many of the biochemical properties of counterparts found in other bacteria degrading aromatic hydrocarbons. Based on the properties of the associated electron carriers, it would belong to class IIB of the dioxygenase classification proposed by Batie et al. (*37*), together with benzene and biphenyl dioxygenases (*1*). However, amino acid sequence analysis of the PhnI $\alpha$ and $\beta$

subunits rather indicated that the enzyme was more closely related to naphthalene and phenanthrene dioxygenases (*6*), consistent with the substrate specificity of the dioxygenase determined herein. Purified ht-PhnI contained Rieske [2Fe-2S] and mononuclear Fe (II) centers as expected, but quantitative analysis based on EPR spectroscopy revealed that the two types of metal binding sites were not fully occupied. We cannot rule out metal loss during purification, despite employing conditions likely to minimize such losses. However, it is also possible that the biosynthesis of the oxygenase in *Pseudomonas* recombinant cells yielded a protein that did not have its full content of metal centers. Examples of purified oxygenases having a full complement of iron have occasionally been reported (*10, 38, 39*), but enzyme preparations partially lacking iron are more frequently obtained (*4, 40-42*). Hence, our EPR-based method to determine the occupancy of both iron binding sites might be of general interest for the characterization of such types of enzymes.

The *in vitro* activity of the dioxygenase was highly dependent on the component ratio ferredoxin over oxygenase, half saturation occurring for a 14-fold molar excess of the ferredoxin. Likewise, a high molar excess of ferredoxin was required to reach maximal activity in the case of naphthalene (*10*), and biphenyl dioxygenase (*39*). As a consequence, comparison of the apparent specific activities or $k_{cat}$ of enzymes from different sources should be regarded with caution. At a ferredoxin/oxygenase ratio of 14, the CHY-1 dioxygenase showed an apparent $k_{cat}$ of 1.82 ± 0.03 $s^{-1}$ and a specificity constant of 2.0 ± 0.3 $\mu M^{-1}$ $s^{-1}$ with naphthalene as substrate. A velocity constant 2.5-fold as high (4.48 $s^{-1}$) was found at a molar ratio of 60. In comparison, the biphenyl dioxygenase from *C. testosteroni* exhibited a $k_{cat}$ of 7.0 ± 0.2 $s^{-1}$ and a specificity constant of 1.2 ± 0.1 $\mu M^{-1}$ $s^{-1}$, at a ratio of 23 (*39*). An apparent $k_{cat}$ of 2.4 $s^{-1}$ and a specificity constant of 7.0 $\mu M^{-1}$ $min^{-1}$ (equivalent to 0.11 $\mu M^{-1}$ $s^{-1}$), was reported for 2-nitrotoluene dioxygenase from *Comamonas* JS765, at a ratio of 3.7 (*42*).

The present study revealed that the coupling between substrate oxidation and $O_2$ (or NADH) utilization varied widely depending on PAHs. While the conversion of naphthalene to dihydrodiol by the dioxygenase involved a stoichiometric amount of $O_2$, the oxidation of all other PAHs, except fluorene, gave rise to significant uncoupling. This uncoupling was associated with the release of $H_2O_2$, although a large proportion of the peroxide could be attributed to auto-oxidation of the electron carriers alone. Assuming that most of the peroxide detected in our *in vitro* assays is an artifact due to the great molar excess of electron carriers used in these assays, it is unclear whether peroxide would be produced in significant amounts *in vivo* as a consequence of the dioxygenase-catalyzed oxidation of PAHs, given that electron carriers are certainly in limiting amounts in natural host cells. Nevertheless, because of this

uncoupling, part of the energy recovered from the catabolism of PAHs as NAD(P)H, is probably lost as unproductive transfer of reducing equivalents to $O_2$. This energy burden might affect growth yield, and could explain, at least in part, the higher resistance of 4- and 5-ring PAHs to bacterial biodegradation, inasmuch as the coupling efficiency of the dioxygenase reaction was lowest with those PAHs. Other dioxygenase systems were found to give rise to partially uncoupled reactions when challenged with poor substrates, and this was associated to a release of $H_2O_2$. Uncoupling occurred when naphthalene dioxygenase was incubated with benzene, and the peroxide formed was found to irreversibly inactivate the enzyme, probably because of the damage done by the product of the Fenton reaction between peroxide and the active site Fe(II) (*33*). Biphenyl dioxygenase also catalyzed uncoupled reactions and $H_2O_2$ production when provided with certain dichlorobiphenyls, which might result in the inhibition of the dioxygenation of other chlorobiphenyls. This effect, combined with the deleterious action of peroxide on cells, was predicted to inhibit the microbial catabolism of polychlorobiphenyls (*39*). Interestingly, some PAH-degrading bacteria were shown to specifically induce a catalase-peroxidase when grown on PAHs, thereby providing a means to cope with the dioxygenase-mediated formation of peroxide (*43*).

The oxidation of benz[a]anthracene by the dioxygenase from strain CHY-1 is of particular interest because this PAH was converted into three dihydrodiols, two of which were subjected to a second dihydroxylation in a highly uncoupled reaction. The amount of *bis*-*cis* dihydrodiol formed in this secondary reaction within the time of an assay (around 5 min) was estimated to be between 20 and 40 % of the total amount of diols primarily produced by the enzyme. These observations have several implications as for the reactivity and the coupling efficiency of the enzyme with respect to benz[a]anthracene. The rapid accumulation of *bis*-dihydrodiol indicated that the 1,2- and 10,11-dihydrodiols competed with benz[a]anthracene for the enzyme active site, which they reached faster than the PAH because of their much higher water solubility. This interpretation is supported by single turnover experiments showing that organic solvent accelerated benz[a]anthracene oxidation and suppressed *bis*-dihydrodiol formation. Because the dihydrodiols are poor substrates for the dioxygenase, as confirmed by EPR probing of nitrosyl-enzyme complexes, the competition they exert on benz[a]anthracene oxidation was expected to alter the coupling efficiency. A ratio dihydrodiol/$O_2$ of 0.31 was calculated without taking into account the formation of *bis*-dihydrodiol (Table 2). A calculation of the $O_2$ consumed for *bis*-dihydrodiol formation, assuming that 2 $O_2$ molecules were required per each molecule formed, allowed to bracket the ratio (*bis*-dihydrodiol + dihydrodiols)/$O_2$ between 0.42 and 0.55. This is definitely higher than

the dihydrodiol/$O_2$ ratios found for chrysene and benzo[a]pyrene. Hence, based on coupling efficiency and oxidation rates, benz[a]anthracene appeared to be a better substrate than chrysene, and yet, it could not support growth of strain CHY-1. In *S. yanoikuyae* B1, a strain which cannot grow on benz[a]anthracene either, previous studies showed that the oxidation of this PAH led to the accumulation of three metabolites identified as 1-hydroxyanthranoic acid, 2-hydroxy 3-phenanthroic acid and 3-hydroxy 2-phenanthroic acid (*44*). These metabolites were predicted to arise from five similar degradation steps of benz[a]anthracene, involving an initial dihydroxylation on positions 1,2-, 8,9- and 10,11-, respectively. The data suggested that a subsequent step in the catabolic pathway of this PAH might be too slow to allow efficient processing of the metabolites, thereby preventing bacterial growth. Alternatively, the possibility was considered that the peroxide formed in the dioxygenase-catalyzed oxidation of benz[a]anthracene, or in the secondary oxidation of dihydrodiols, could inhibit growth. The secondary reaction generating *bis*-dihydrodiol is highly uncoupled, but it is unknown whether it would occur *in vivo* in strain CHY-1. *S. yanoikuyae* B1 did not produced any detectable *bis*-dihydrodiol when degrading benz[a]anthracene (*44*), indicating that dihydrodiols were rapidly metabolized. Accordingly, we have recently characterized a dihydrodiol dehydrogenase which can efficiently convert the three dihydrodiol isomers of benz[a]anthracene to corresponding catechols (*22*). Hence, in PAH-degrading Sphingomonads, the coupling between the first and the second enzymatic step of the catabolic pathway likely prevents the dioxygenase from catalyzing unproductive and potentially deleterious reactions.

In this work, we have purified and characterized a ring-hydroxylating dioxygenase with an exceptionally broad substrate specificity, which provides a good model for further structure-function studies on this class of enzymes. The oxygenase component has been recently crystallized and subjected to X-ray diffraction analysis. The structure of the protein has been solved to 1.85 Å resolution and will be described elsewhere (J. Jakoncic, Y. Jouanneau, C. Meyer, V. Stojanoff, unpublished data).

## Acknowledgements

We thank John Willison for helpful discussions and critical reading of the manuscript.

## References


(1) Butler, C. S., and Mason, J. R. (1997) Structure-function analysis of the bacterial aromatic ring- hydroxylating dioxygenases. *Adv. Microb. Physiol. 38*, 47-84.

(2) Gibson, D. T., and Parales, R. E. (2000) Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Curr. Opin. Biotechnol. 11*, 236-243.

(3) Armengaud, J., Happe, B., and Timmis, K. N. (1998) Genetic analysis of dioxin dioxygenase of *Sphingomonas* sp. Strain RW1: catabolic genes dispersed on the genome. *J. Bacteriol. 180*, 3954-3966.

(4) Bunz, P. V., and Cook, A. M. (1993) Dibenzofuran 4,4a-dioxygenase from *Sphingomonas* sp. strain RW1: angular dioxygenation by a three-component enzyme system. *J. Bacteriol. 175*, 6467-6475.

(5) Furukawa, K., Suenaga, H., and Goto, M. (2004) Biphenyl dioxygenases: functional versatilities and directed evolution. *J. Bacteriol. 186*, 5189-5196.

(6) Demaneche, S., Meyer, C., Micoud, J., Louwagie, M., Willison, J. C., and Jouanneau, Y. (2004) Identification and functional analysis of two aromatic ring-hydroxylating dioxygenases from a *Sphingomonas* strain degrading various polycyclic aromatic hydrocarbons. *Appl. Environ. Microbiol. 70*, 6714-6725.

(7) Krivobok, S., Kuony, S., Meyer, C., Louwagie, M., Willison, J. C., and Jouanneau, Y. (2003) Identification of pyrene-induced proteins in *Mycobacterium* sp. 6PY1 : Evidence for two ring-hydroxylating dioxygenases. *J. Bacteriol. 185*, 3828-3841.

(8) Saito, A., Iwabuchi, T., and Harayama, S. (2000) A novel phenanthrene dioxygenase from *Nocardioides* sp strain KP7: Expression in *Escherichia coli*. *J. Bacteriol. 182*, 2134-2141.

(9) Furukawa, K. (2000) Engineering dioxygenases for efficient degradation of environmental pollutants. *Curr. Opin. Biotechnol. 11*, 244-249.

(10) Wolfe, M. D., Parales, J. V., Gibson, D. T., and Lipscomb, J. D. (2001) Single turnover chemistry and regulation of O-2 activation by the oxygenase component of naphthalene 1,2-dioxygenase. *J. Biol. Chem. 276*, 1945-1953.

(11) Batie, C. J., LaHaie, E., and Ballou, D. P. (1987) Purification and characterization of phthalate oxygenase and phthalate oxygenase reductase from *Pseudomonas cepacia*. *J. Biol. Chem. 262*, 1510-1518.

(12) Correll, C. C., Batie, C. J., Ballou, D. P., and Ludwig, M. L. (1992) Phthalate dioxygenase reductase: a modular structure for electron transfer from pyridine nucleotides to [2Fe-2S]. *Science 258*, 1604-1610.

(13) Kauppi, B., Lee, K., Carredano, E., Parales, R. E., Gibson, D. T., Eklund, H., and Ramaswamy, S. (1998) Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2- dioxygenase. *Structure 6*, 571-586.

(14) Parales, R. E., Lee, K., Resnick, S. M., Jiang, H. Y., Lessner, D. J., and Gibson, D. T. (2000) Substrate specificity of naphthalene dioxygenase: Effect of specific amino acids at the active site of the enzyme. *J. Bacteriol. 182*, 1641-1649.

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

(16) Kanaly, R. A., and Harayama, S. (2000) Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. *J. Bacteriol. 182*, 2059-2067.

(17) IARC. (1983) Polynuclear Aromatic Compounds, Part 1: Chemical, Environmental and Experimental Data, in *Monographs on the evaluation of carcinogenic risks to humans*.

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

(19) Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning : a laboratory manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.

(20) Hurtubise, Y., Barriault, D., Powlowski, J., and Sylvestre, M. (1995) Purification and characterization of the *Comamonas testosteroni* B-356 biphenyl dioxygenase components. *J. Bacteriol. 177*, 6610-6618.

(21) Rodarie, D., and Jouanneau, Y. (2001) Genetic and biochemical characterization of the biphenyl dioxygenase from *Pseudomonas* sp. strain B4. *J. Microbiol. Biotechnol. 11*, 762-771.

(22) Jouanneau, Y., and Meyer, C. (2006) Purification and characterization of an arene *cis*-dihydrodiol dehydrogenase endowed with broad substrate specificity toward polycyclic aromatic hydrocarbon dihydrodiols. *Appl. Environ. Microbiol. 72*, 4726-4734.

(23) Blair, D., and Diehl, H. (1961) Bathophenantroline disulphonic acid and bathocuproine disulphonic acid, water soluble reagents for iron and copper. *Talanta 7*, 163-174.

(24) Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem. 72*, 248-254.

(25) Pelley, J. W., Garner, C. W., and Little, G. H. (1978) A simple rapid buiret method for the estimation of protein in samples containing thiols. *Anal. Biochem. 86*, 341-343.

(26) Jouanneau, Y., Meyer, C., Naud, I., and Klipp, W. (1995) Characterization of an *fdxN* mutant of *Rhodobacter capsulatus* indicates that ferredoxin I serves as electron donor to nitrogenase. *Biochim. Biophys. Acta 1232*, 33-42.

(27) Naud, I., Vincon, M., Garin, J., Gaillard, J., Forest, E., and Jouanneau, Y. (1994) Purification of a sixth ferredoxin from *Rhodobacter capsulatus*. Primary structure and biochemical properties. *Eur. J. Biochem. 222*, 933-939.

(28) Jeffrey, A. M., Yeh, H. J., Jerina, D. M., Patel, T. R., Davey, J. F., and Gibson, D. T. (1975) Initial reactions in the oxidation of naphthalene by *Pseudomonas putida*. *Biochemistry 14*, 575-584.

(29) Jerina, D. M., Selander, H., Yagi, H., Wells, M. C., Davey, J. F., Mahadevan, V., and Gibson, D. T. (1976) Dihydrodiols from anthracene and phenanthrene. *J. Am. Chem. Soc. 98*, 5988-5996.

(30) Boyd, D. R., Sharma, N. D., Agarwal, R., Resnick, S. M., Schocken, M. J., Gibson, D. T., Sayer, J. M., Yagi, H., and Jerina, D. M. (1997) Bacterial dioxygenase-catalysed

dihydroxylation and chemical resolution routes to enantiopure *cis*-dihydrodiols of chrysene. *J. Chem. Soc. Perkin Trans. 1*, 1715-1723.

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

(32) Jerina, D. M., Vanbladeren, P. J., Yagi, H., Gibson, D. T., Mahadevan, V., Neese, A. S., Koreeda, M., Sharma, N. D., and Boyd, D. R. (1984) Synthesis and absolute-configuration of the bacterial *cis*-1,2-dihydrodiol, *cis*-8,9-dihydrodiol, and *cis*-10,11-dihydrodiol metabolites of benz[a]anthracene formed by a strain of *Beijerinckia*. *J. Org. Chem. 49*, 3621-3628.

(33) Lee, K. (1999) Benzene-induced uncoupling of naphthalene dioxygenase activity and enzyme inactivation by production of hydrogen peroxide. *J. Bacteriol. 181*, 2719-2725.

(34) Selifonov, S. A., Grifoll, M., Eaton, R. W., and Chapman, P. J. (1996) Oxidation of naphthenoaromatic and methyl-substituted aromatic compounds by naphthalene 1,2-dioxygenase. *Appl. Environ. Microbiol. 62*, 507-514.

(35) Rehmann, K., Hertkorn, N., and Kettrup, A. A. (2001) Fluoranthene metabolism in *Mycobacterium* sp strain KR20: identity of pathway intermediates during degradation and growth. *Microbiology-Sgm 147*, 2783-2794.

(36) Gibson, D. T. (1999) *Beijerinckia* sp strain B1: a strain by any other name. *J.Ind. Microbiol. Biotechnol. 23*, 284-293.

(37) Batie, C. J., Ballou, D. P., and Correll, C. C. (1992) Phthalate dioxygenase reductase and related flavin-iron-sulfur containing electron transferases, in *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., Ed.) pp 543-556, CRC Press, Boca Raton, Fla.

(38) Eby, D. M., Beharry, Z. M., Coulter, E. D., Kurtz, D. M., Jr., and Neidle, E. L. (2001) Characterization and evolution of anthranilate 1,2-dioxygenase from *Acinetobacter* sp. strain ADP1. *J. Bacteriol. 183*, 109-118.

(39) Imbeault, N. Y., Powlowski, J. B., Colbert, C. L., Bolin, J. T., and Eltis, L. D. (2000) Steady-state kinetic characterization and crystallization of a polychlorinated biphenyl-transforming dioxygenase. *J. Biol. Chem. 275*, 12430-12437.

(40) Ensley, B. D., and Gibson, D. T. (1983) Naphthalene dioxygenase: purification and properties of a terminal oxygenase component. *J. Bacteriol. 155*, 505-11.

(41) Hurtubise, Y., Barriault, D., and Sylvestre, M. (1996) Characterization of active recombinant his-tagged oxygenase component of *Comamonas testosteroni* B-356 biphenyl dioxygenase. *J. Biol. Chem. 271*, 8152-8156.

(42) Parales, R. E., Huang, R., Yu, C. L., Parales, J. V., Lee, F. K., Lessner, D. J., Ivkovic-Jensen, M. M., Liu, W., Friemann, R., Ramaswamy, S., and Gibson, D. T. (2005) Purification, characterization, and crystallization of the components of the nitrobenzene and 2-nitrotoluene dioxygenase enzyme systems. *Appl. Environ. Microbiol. 71*, 3806-3814.

(43) Wang, R. F., Wennerstrom, D., Cao, W. W., Khan, A. A., and Cerniglia, C. E. (2000) Cloning, expression, and characterization of the *katG* gene, encoding catalase-peroxidase, from the polycyclic aromatic hydrocarbon-degrading bacterium *Mycobacterium* sp. strain PYR-1. *Appl. Environ. Microbiol. 66*, 4300-4304.

(44) Mahaffey, W. R., Gibson, D. T., and Cerniglia, C. E. (1988) Bacterial oxidation of chemical carcinogens: formation of polycyclic aromatic acids from benz[a]anthracene. *Appl. Environ. Microbiol. 54*, 2415-2423.

(45) Boyd, D. R., Sharma, N. D., Hempenstall, F., Kennedy, M. A., Malone, J. F., Allen, C. C. R., Resnick, S. M., and Gibson, D. T. (1999) *bis*-*cis*-Dyhydrodiols: a new class of metabolites resulting from biphenyl dioxygenase-catalyzed sequential asymetric *cis*-dihydroxylation of polycyclic arenes and heteroarenes. *J. Org. Chem. 64*, 4005-4011.

| Ht-PhnI preparation | 1 | 2 | 3 |
|---|---|---|---|
| Concentration[a] (μM) | 21.0 | 25.7 | 19.7 |
| Total Fe (atoms/αβ) | 1.73 | 2.54 | 2.55 |
| Fe(II)-NO (spin/αβ) | 0.20 | 0.78 | 0.92 |
| [2Fe-2S] (spin/αβ) | 0.75 | 0.85 | 0.75 |
| Specific activity[b] (U/mg) | 0.41 ± 0.035 | 0.86 ± 0.01 | 0.58 ± 0.01 |

[a] The protein concentration were estimated from the microbiuret assays, using a molecular mass of 215 kDa for ht-PhnI. Based on these determinations, the absorbance coefficient of ht-PhnI at 458 nm was calculated to be $\varepsilon_{458}$ = 12,500 $M^{-1}.cm^{-1}$

[b] as determined by the NADH oxidation assay with naphthalene as substrate. The molar ratio PnA3/PhnI was approx. 9.5 in these assays.

**Table 1.** Specific activity and iron content of different preparations of ht-PhnI. Iron was determined by chemical analysis and EPR spectroscopy. The standard error for each determination was less than 10%.

| Substrate | $O_2$ consumption | Dihydrodiol formed | Dihydrodiol/$O_2$ [b] | $H_2O_2$/ $O_2$ | NADH oxidation | Diol/NADH |
|---|---|---|---|---|---|---|
| | nmol.min$^{-1}$.mg$^{-1}$ | nmol.min$^{-1}$.mg$^{-1}$ | | | nmol.min$^{-1}$.mg$^{-1}$ | |
| Naphthalene | 460 (20) | 418 (20) | 1.03 (0.07) | 0.122 (0.010) | 510 (50) | 0.92 (0.06) |
| Biphenyl | 410 (40) | 260 (12) | 0.70 (0.01) | 0.28 (0.015) | 290 (35) | 0.67 (0.03) |
| Phenanthrene | 370 (30) | 300 (30) | 0.83 (0.03) | 0.128 (0.028) | 465 (10) | 0.46 (0.01) |
| Anthracene | 360 (30) | 130 (10) | 0.36 (0.02) | 0.53 (0.02) | 141 (6) | 0.62 (0.01) |
| Fluorene | 91.5 (7) | 91 (5) [c] | 0.99 (0.05) | - | - | - |
| Fluoranthene | 147 (24) | 76 (5) [c] | 0.51 (0.04) | - | - | - |
| Benz[a]anthracene | 167 (20) | 52 (5) [d] | 0.31 (0.04) | 0.48 (0.04) | 46 (2) | 0.40 (0.07) |
| Chrysene | 49 (2) | 2.5 (0.3) | 0.082 (0.011) | 0.59 (0.08) | 9.6 (0.6) | 0.098 (0.007) |
| Benzo[a]pyrene | 43 (8) | 7.8 (0.7) | 0.33 (0.02) | 0.50 (0. 005) | 10 (1) | 0.19 (0.01) |
| 1,2-Benz[a]anthracene diol | 2500 (400) | 11 (2) | 0.0044 (0.001) | 0.83 | - | - |
| 10,11-Benz[a]anthracene diol | 1500 (150) | 16.5 (0.5) | 0.011 (0.0005) | 0.66 | - | - |

[a] The indicated values represent means obtained from two to four determinations, with standard deviations given in parentheses. - means not determined. $O_2$ consumption and NADH oxidation represent initial rates corrected for the background activity observed in the absence of PAH substrate. The rates of dihydrodiol formation are average rates calculated over the duration of the assay which lasted between 4 min(naphthalene) and 11 min (benzo[a]pyrene).
[b] Values were calculated as ratios between the rates of dihydrodiol formation (column 3), and the average rates of $O_2$ consumption over the duration of the assay, which are lower than the initial rates of $O_2$ consumption given in column 2.
[c] These values represent rates of substrate oxidation, because the oxidation products from florene and fluoranthene could not be measured accurately (see text).
[d] Three diol isomers were produced, which were not separated by HPLC and quantified as a mixture (see Methods). The contribution of *bis-cis*-dihydrodiol, also produced in the reaction, was not taken into account in the calculations.

**Table 2.** Specific activity and coupling efficiencies of the dioxygenase as a function of PAH substrates[a]

| Substrate | Properties of trimethylsilyl derivates of products | | | Identification |
|---|---|---|---|---|
| | GC retention time (min) | Relative peak area (%) | m/z and relative abundance (%) of major fragments | |
| Fluorene | 15.18 | 6 | 254($M^+$, 41), 239(8), 166(15), 165(100) | 9-Fluorenol[a] |
| | 16.18 | 6 | 254($M^+$, 100), 239(39), 224(19), 223(89), 178(8), 165(13) | Hydroxyfluorene |
| | 16.49 | 14 | 254($M^+$, 100), 239(70), 223(3), 195(2.5), 178(11), 165(63) | Hydroxyfluorene |
| | 16.67 | 16 | 254($M^+$, 100), 239(67), 223(2), 195(2), 178(10), 165(25) | Hydroxyfluorene |
| | 17.70 | 58 | 342($M^+$, 100), 327(4), 253(62), 238(5), 178(5), 164(5), 163(5) | Dihydroxyfluorene |
| Fluoranthene | 20.74 | 100 | 290($M^+$, 100), 275(70), 219(11), 215(11), 201(14), 200(14), 189(15) | 8-Hydroxyfluoranthene |
| Chrysene | 23.08 | >90[b] | 406($M^+$, 28), 317(14), 316(18), 303(32), 244(14), 228(40), 226(33), 215(66), 191(100) | 3,4-dihydrodiol[c] |
| | 24.55 | <10[b] | 584($M^+$, 8), 393(8), 355(8), 282(13), 281(40), 207(100), 191(73) | 3,4,9,10-*bis*-dihydrodiol[c] |
| Benz[a]anthracene | 22.65 | 68 | 406($M^+$, 28), 316(26), 303(30), 281(14), 228(28), 226(25), 215(36), 191(50), 73(100) | 1,2-Dihydrodiol[c] |
| | 22.95 | 23 | 406($M^+$, 32), 316(29), 303(59), 228(21), 226(14), 191(46), 73(100) | 10,11-Dihydrodiol[c] |
| | 23.43 | 9 | 406($M^+$, 31), 316(27), 303(80), 281(16), 228(18), 226(14), 215(29), 191(68), 73(100) | 8,9-Dihydrodiol[c] |
| | 23.53 | 20-40[d] | 584($M^+$, 46), 481(21) 392(9), 355(16), 281(22), 207(48), 191(100) | 1,2,10,11-*bis*-Dihydrodiol[c] |
| Benzo[a]pyrene | 25.88 | 100 | 430($M^+$, 5.5), 415 (1.5), 341(5), 327(8), 281(20), 252(8), 207(100) | 9,10-Dihydrodiol[c] |

[a] Identification based on match of mass spectrum and GC retention time with those of an authentic sample.
[b] Percentages estimated from HPLC peak area.
[c] Identification based on comparisons of the GC-MS and UV absorbance data of the products with previously published data (see text).
[d] The indicated range represents an estimation of the *bis*-dihydrodiol/dihydrodiols ratio, as indicated in the text.

**Table 3.** GC-MS identification of PAH oxidation products formed by CHY-1 dioxygenase.

| Acetonitrile % in the reaction | Benz[a]anthracene solubilized (μM) | Incubation time (min) | Total diols/PhnI | Percentage of dihydrodiols produced as[a] | | |
|---|---|---|---|---|---|---|
| | | | | 1,2-diol | 10,11-diol | 8,9-diol |
| 0 | 0.025 | 1 | 0.44 | | | |
| | | 10 | 2.12[b] | 28 | 64 | 8 |
| 20 | 2.5 | 1 | 2.01 | 30 | 53 | 17 |
| | | 10 | 1.94 | 36 | 50 | 14 |
| 30 | 27.0 | 1 | 1.12 | 29 | 43 | 28 |
| | | 10 | 0.96 | 28.5 | 54 | 17.5 |

[a] As determined from the peak area of trimethylsilyl derivates analyzed by GC-MS
[b] Some *bis-cis*-dihydrodiol was also produced which accounted for about 1.5% of the total diols formed, based on HPLC determination.

**Table 4.** Solvent-facilitated formation of benz[a]anthracene *cis*-dihydrodiols catalyzed by PhnI in single turnover reactions.

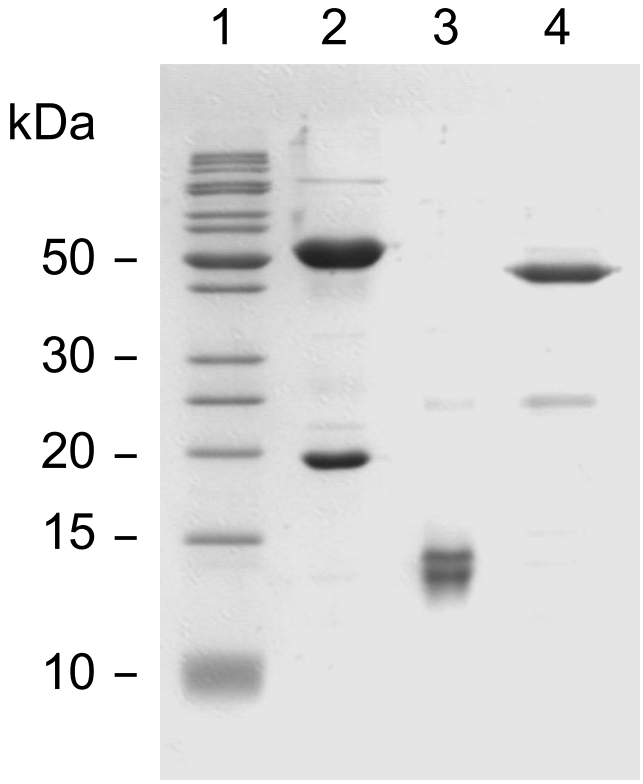



**Figure 1.** SDS-PAGE of the purified components of the ring-hydroxylating dioxygenase of *Sphingomonas* strain CHY-1. Polypeptides were electrophoresed on a 15% polyacrylamide slab gel. Lane 1: molecular mass markers. Lane 2: ht-PhnI, 3.2 µg. Lane 3: ht-PhnA3, 1.0 µg. Lane 4: ht-PhnA4, 1.8 µg.

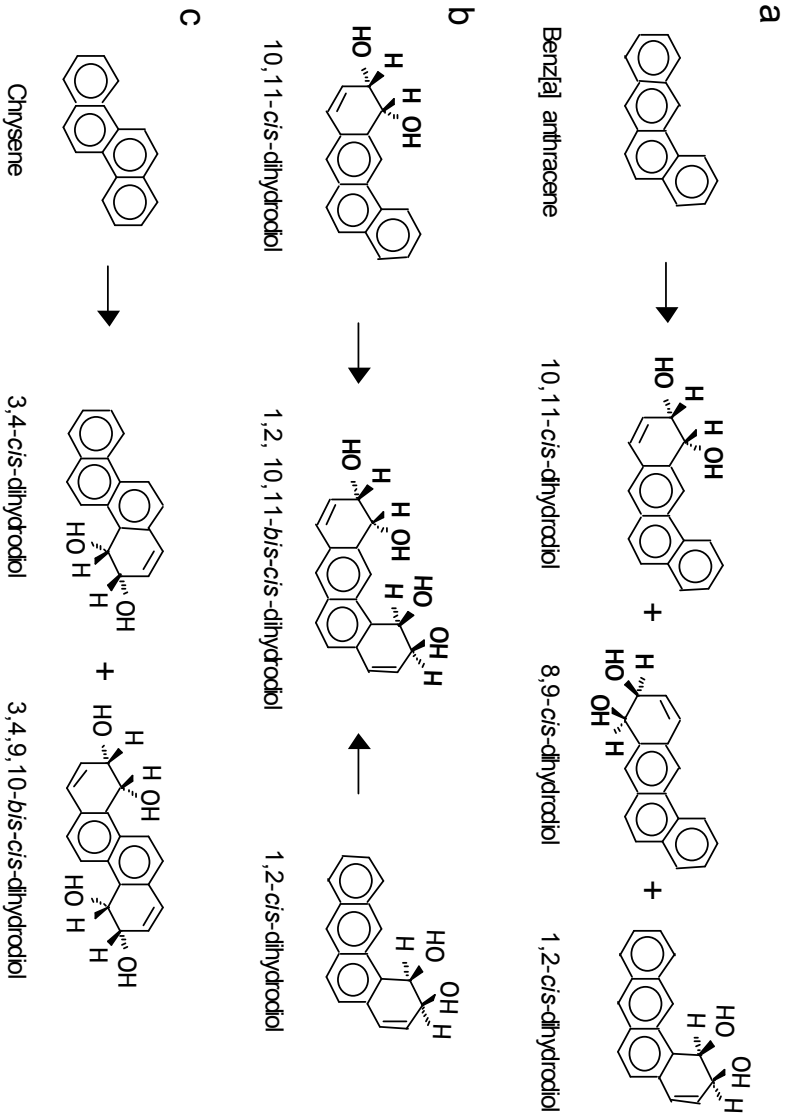



**Figure 2.** Dioxygenation reactions of four-ring PAHs catalyzed by ht-PhnI. Benz[a]anthracene was converted into three dihydrodiols isomers (a), in proportions which varied depending on experimental conditions (see text). The 1,2- and 10,11-isomers were subjected to a second dihydroxylation, yielding the same *bis*-*cis*-dihydrodiol (b). Chrysene was oxidized to a single dihydrodiol, which could subsequently react with the dioxygenase to yield the 3,4,9,10-*bis*-*cis*-dihydrodiol (c). Reaction products were identified as indicated in the text. Stereochemical configurations were assumed to be identical to those reported in previous studies on the *S. yanoikuyae* enzyme (*32, 45*).

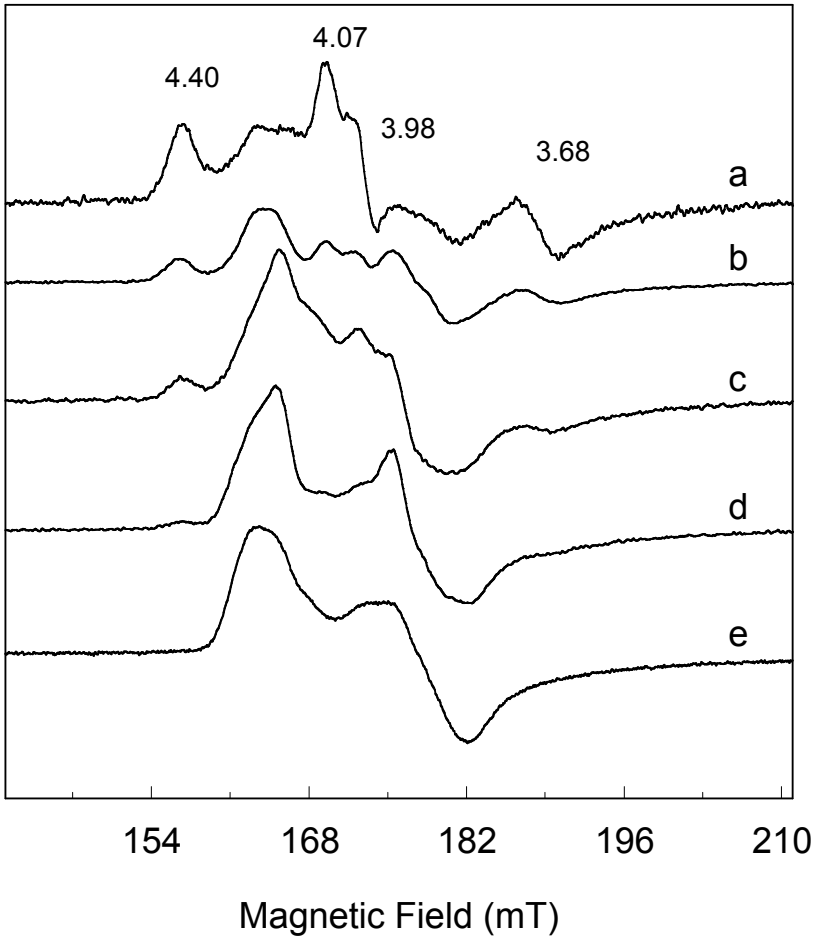



**Figure 3.** EPR spectra of nitrosyl complexes of ht-PhnI in the presence or absence substrates. Protein samples contained 22.0 µM ht-PhnI in 0.18 ml of 50 mM potassium phosphate, pH 7.5, and, either no substrate (spectrum a) or one of the following substrates (0.1 mM) added in 10 µl acetonitrile : 1,2-benz[a]anthracene dihydrodiol (spectrum b); 10,11-benz[a]anthracene dihydrodiols (spectrum c); benz[a]anthracene (spectrum d); naphthalene (spectrum d). After 15 min at room temperature under argon, nitrosyl complexes were prepared (see Materials and Methods), samples were transferred into EPR tubes and frozen. Acquisition conditions : Temperature; 4K; Microwave power: 250 µW; modulation frequency: 100 kHz; modulation amplitude: 1mT. Relevant *g* values are indicated.

# Partie 3 : Article 2

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

Cet article décrit la structure de la composante terminale, Phn1. Nous décrivons ici la cristallisation et l'obtention de la structure de Phn1. Les structures tertiaires et quaternaires sont ici comparées à celles d'autres dioxygenases. Les travaux ci présentés sont issus d'une collaboration entre l'équipe d'Yves Jouanneau et celle de Vivian Stojanoff. La majorité des travaux ont été effectués au NSLS. Cet article a été publié en 2007 dans le revue FEBS J. (volume 274, pages 2470-2481).

# 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.

## Abstract

Ring-hydroxylating dioxygenases (RHDs) play a critical role in the bacterial degradation of aromatic hydrocarbons. The RHD from *Sphingomonas* CHY-I is remarkable in that it is able to initiate the oxidation of a wide range of polycyclic aromatic hydrocarbons (PAHs), including 4 and 5-ring molecules, many of which are toxic pollutants. The crystal structure of the terminal component of the dioxygenase was determined by molecular replacement to 1.85 Å resolution by using the crystal structures of naphathalene dioxygenase from *Pseudomonas sp*. strain NCIB9816-4 and cumene dioxygenase from *P. fluorescens* strain IP01. Although the oxygenase from CHY-1 exhibits limited sequence similarity with well characterized RHDs, the tertiary and quaternary structures are found to be quite similar. The catalytic domain of the enzyme is characterized by a large substrate binding pocket, the largest ever reported for this type of enzyme. The present study provides structural evidence for the broad substrate specificity of the RHD from CHY-1including residues already shown to be important in substrate specificity and catalytic reaction.

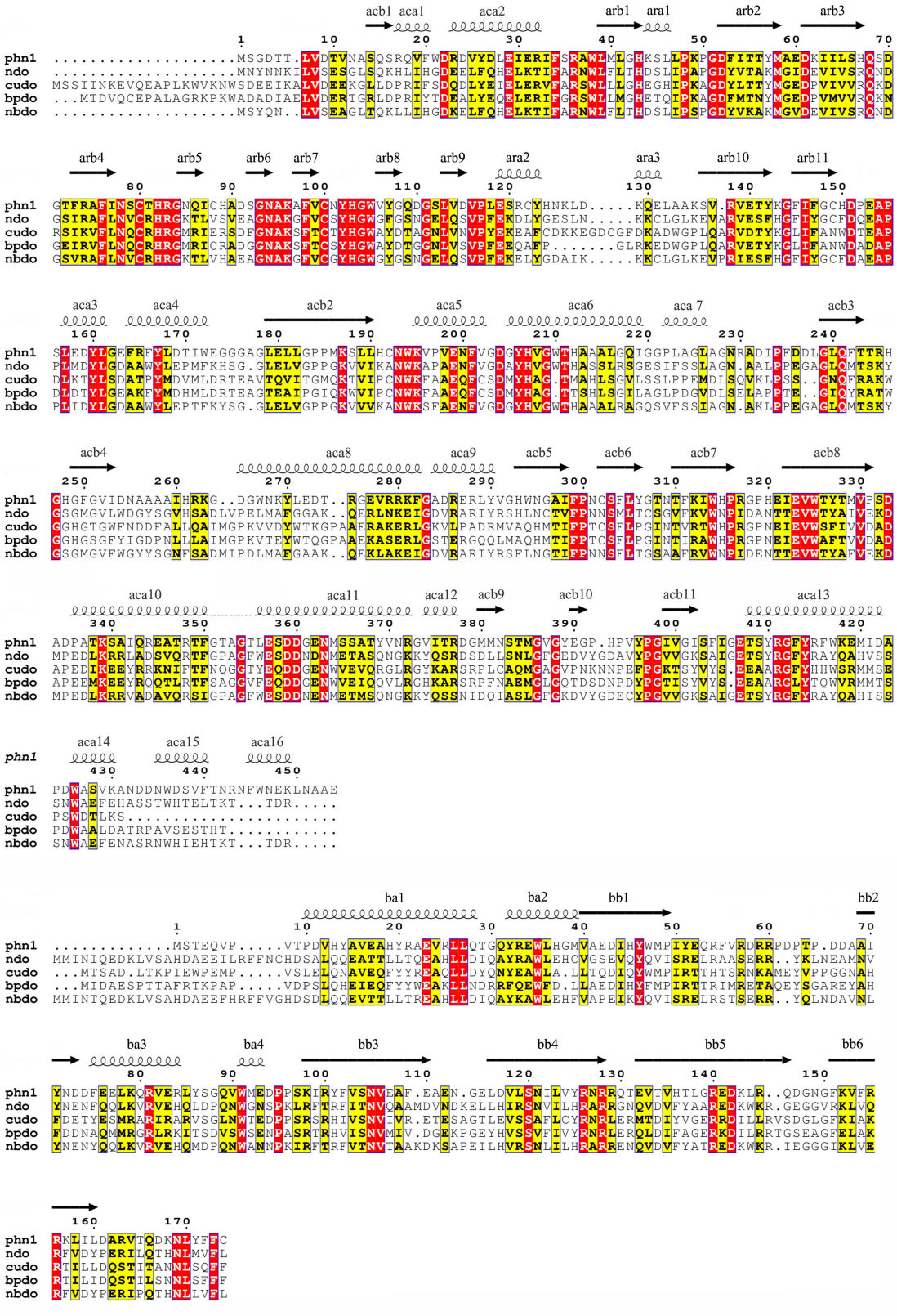
## 1. Introduction

Polycyclic Aromatic Hydrocarbons (PAHs) are considered major pollutants of the environment 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 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-oxoquilone monoxygenase 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 an 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 four and five ring PAHs.

**Figure 1**. Sequence alignment of the (a) α subunit and (b) β subunit from PhnI (phn1), NDO-$O_{9816-4}$ (ndo), CDO-$O_{IP01}$ (cudo), BPDO-$O_{RHA1}$ (bpdo) and NBDO-$O_{JS765}$ (nbdo). Highly conserved residues are boxed and shown against a red background; boxed residues shown against a yellow background are not totally conserved. The sequence numbering given above the sequence is related to PhnI. Secondary structure elements for PhnI are indicated above the sequence. The figure was generated with CLUSTALW [35]


acb1 aca1 aca2 arb1 ara1 arb2 arb3
arb4 arb5 arb6 arb7 arb8 arb9 ara2 ara3 arb10 arb11
aca3 aca4 acb2 aca5 aca6 aca 7 acb3
acb4 aca8 aca9 acb5 acb6 acb7 acb8
aca10 aca11 aca12 acb9 acb10 acb11 aca13
aca14 aca15 aca16
ba1 ba2 bb1 bb2
ba3 ba4 bb3 bb4 bb5 bb6
phn1
ndo
cudo
bpdo
nbdo

## 2. Material and Methods

### 2.1. 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. [8,9]. The oxygenase was further purified by two chromatographic steps under argon as follows. 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 Calcium Acetate and 150 mM NaCl). To improve the diffraction quality, the nucleation process was slowed down by covering each well with 300 μL of mineral oil [21].

### 2.2. 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.

### 2.3. 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-$O_{9816\text{-}4}$ (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].

### 2.4. Protein Data Bank accession number

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

## 3. Results and Discussion

### 3.1. 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 a crystallographic R-factor of 19.7 % and $R_{free}$-factor of 23.6 % (5 % of the reflections were used for the cross validation calculation), Table 1.

| **Crystal data and data processing** | |
|---|---|
| **Space Group** | $P2_12_12_1$ |
| **Unit Cell parameters**<br>**a, b, c (Å)**<br>**α= β= γ (°)** | <br>92.64, 112.73, 190.63<br>90.00 |
| **Resolution range (Å)** | 35.0-1.85 (1.88-1.85) |
| **Measured reflections** | 977916 |
| **Overall redundancy** | 5.8 |
| **Data completeness (%)** | 99.6 (99.0) |
| **$R_{sym}$** [#] | 0.07 (0.59) |
| **I/σI** | 22.1 (2.1) |
| **Molecules in asymmetric units** | 6 |
| | |
| **Refinement** | |
| **Resolution limits (Å)** | 35.0-1.85 |
| **R factor / R free (%)** | 19.7 / 23.6 |
| **No. of amino acids** | 1822 |
| **No. of protein atoms** | 14722 |
| **Number of ligand atoms** | 15 |
| **Number of water molecules** | 1096 |
| | |
| **Root mean square from ideal values** | |
| **Bond length (Å)** | 0.016 |
| **Bond angles (degrees)** | 1.6 |
| **Dihedral angles (degrees)** | 6.9 |
| **Temperature factor ($Å^2$)** | |
| **Protein atoms** | 27.5 |
| **Ligand atoms** | 24.2 |
| **Water molecules** | 30.4 |
| **Ramachandran plot (%)** | |
| **Most-favoured region** | 88.8 |
| **Additionally allowed** | 10.6 |
| **Disallowed region** | 0.6 |

Values in parentheses refer to the highest resolution shell.
$R_{sym}(I) = \Sigma_{hkl} \Sigma_i | I_{hkl,i} - \langle I_{hkl} \rangle | / \Sigma_{hkl} \Sigma_i | I_{hkl,i} |$, with $\langle I_{hkl} \rangle$ mean intensity of the multiple $I_{hkl,i}$ observations for symmetry-related reflections.

**Table 1**. Data processing and refinement statistics

Consistent with biochemical analysis [9], the PhnI crystal structure can be described by an $\alpha_3\beta_3$ type heterohexamer (Fig. 2) with a 452 amino acid long α subunit and a 174 amino acid long β subunit[1]. In addition to the six polypeptide 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 significantly 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 located in the vicinity of the catalytic site. Five water molecules were found to be in direct contact with the catalytic Fe atoms. Over 99.1% of the residues were found in the most favorable regions 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 $Leu_{\alpha}$ 238.

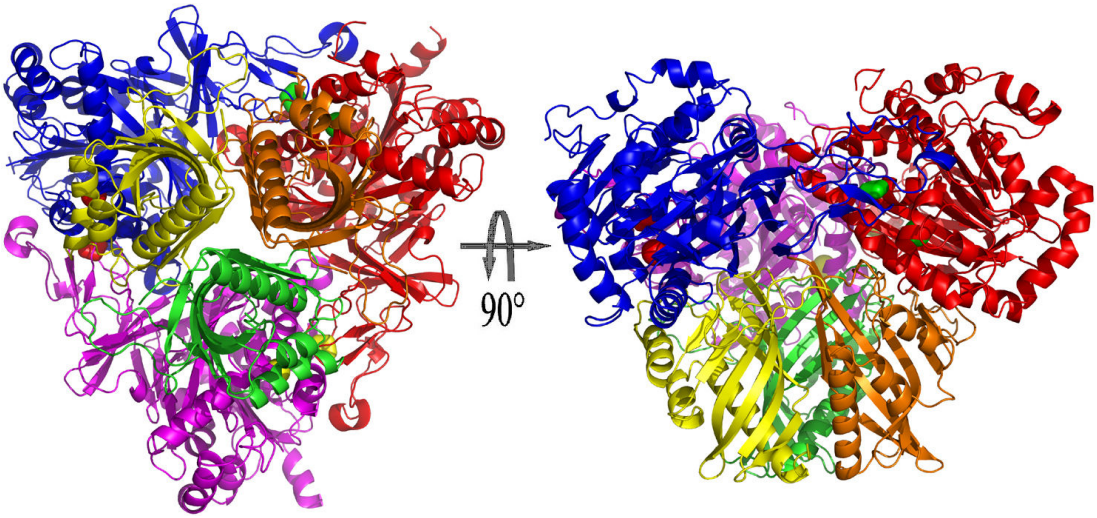



**Figure 2**. Crystal structure of PhnI. (a and b). Ribbon representation of the PhnI $\alpha_3\beta_3$ along and perpendicular to the three fold symmetry axis. The three αβ units are colored in red, magenta and blue; the β subunits are represented in yellow, green and orange. Mononuclear Iron and [2Fe-2S] centers are shown in green red and yellow. The figures were made using the program PYMOL [38].

[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.

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 and E than for chain A, indicating a higher dynamical disorder, and about the same for the three β subunits. 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.

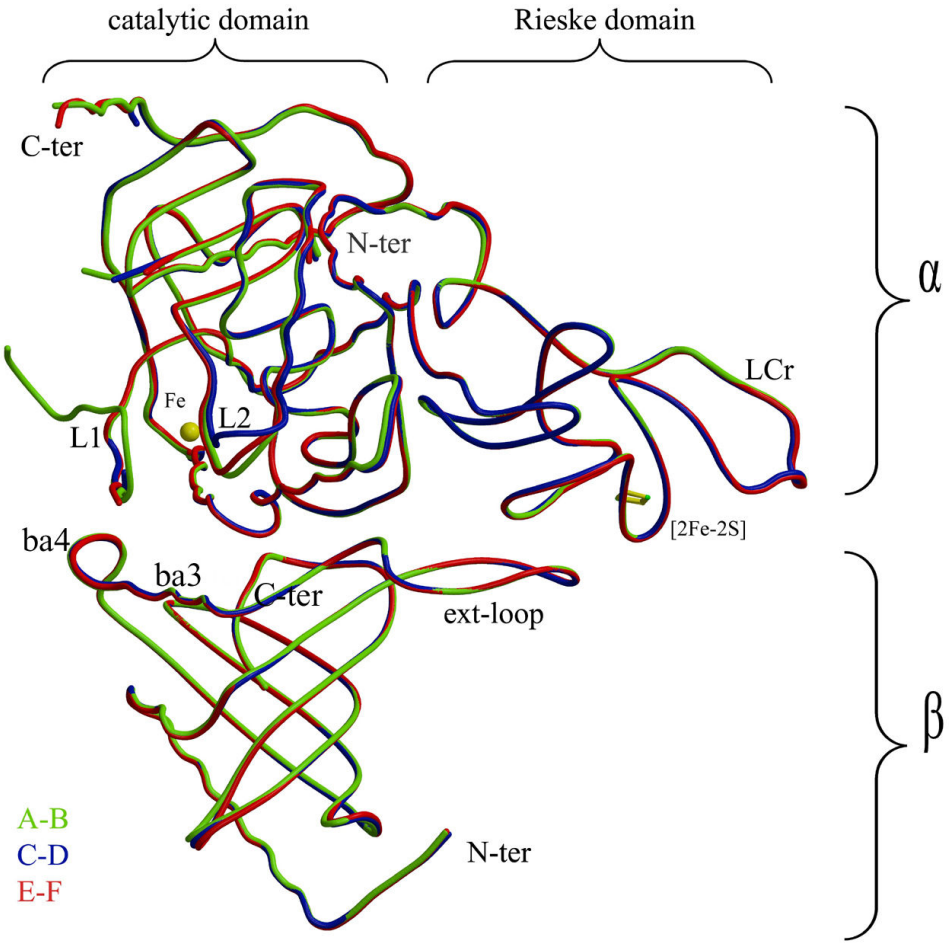



**Figure 3.** The αβ heterodimer. Ribbon representation of the three superposed heterodimers in red, green and blue. Shown are the relevant crystallographic structures in the interaction between domains and subunits. The figure was generated with MOLSCRIPT [36] and Raster 3D [37].

### 3.2. 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 $Pro_{\beta}$ 49 to $Ala_{\beta}$ 69). The C-terminal coil and the third and fourth α-helices (ba3, ba4)[2] form the 20 Å wide 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 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.

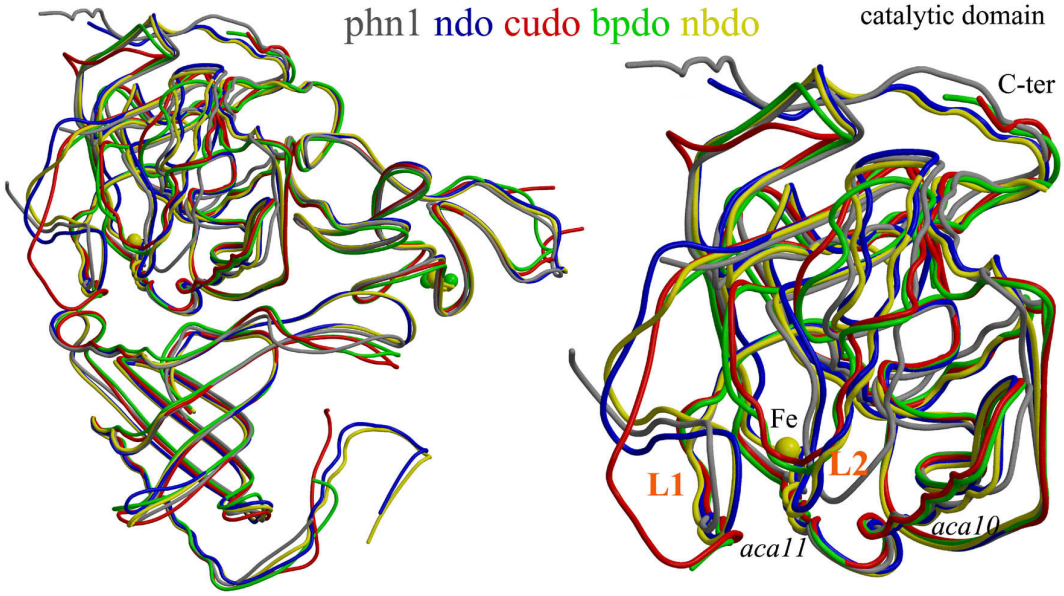



**Figure 4.** Superposition of the PhnI αβ heterodimer, chain A and B, with the RHDs in Figure 1. (a) αβ heterodimer and (b) catalytic domain. Shown are the two solvent exposed loops, LI and LII, at the entrance of the catalytic pocket, as well as, the highly conserved helices, aca 10 and aca11 Sequence identities between the PhnI α subunit and NDO-$O_{9816\text{-}4}$, CDO-$O_{IP01}$, BPDO-$O_{RHA1}$ and NBDO-$O_{JS765}$ are respectively 40, 31, 34 and 40; and for the β subunit 24, 35, 32 and 31. The figure was made using the program MOLSCRIPT [36] and Raster 3D [37].

[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

In spite of low amino acid sequence identity between the β subunits of related RHDs, the PhnI β subunit shares the global pattern with 24 to 35 % identical residues and main chain $C_\alpha$ rmsd ranging between 1.0 and 1.12 Å (Fig. 4). Although it has been demonstrated for other RHDs [31] that the β subunit affects substrate regio-selectivity and specificity the function of the PhnI β subunit is believed to be only structural in nature; as it seems only to maintain the three α subunits together.

### 3.3. 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). In spite of a low sequence identity between the α subunits of related RHDs (Fig. 1), Rieske domains share the same overall folding while most of the differences occur in the catalytic domains (Fig. 4).

#### 3.3.1. The Rieske domain

The *Rieske domain* presents essentially the same structure as that in 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 highly 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 designated below as LCr, which 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 holds 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 α subunits.

Like in other RDHs, the cluster is coordinated by the highly conserved Rieske iron-sulfur motif; Fe1 is coordinated by $His_\alpha$ 82 and $His_\alpha$ 103, located at the tip of the gripper structure, while $Cys_\alpha$ 80 and $Cys_\alpha$ 100 coordinate Fe2. The side chains of the two

histidine ligands are hydrogen bonded to the carboxylates of residues in the adjacent α subunit, $His_{\alpha}$ 82 NE being linked to $Glu_{\alpha}$ 407 side chain atom OE2, and $His_{\alpha}$ 103 NE to $Asp_{\alpha}$ 204 side chains OD1 and OD2. The $Cys_{\alpha}$ 80 sulfur atom is hydrogen bonded to the $Tyr_{\alpha}$ 107 hydroxyl, and the $Cys_{\alpha}$ 80 main chain atoms N and O are bonded to $Asn_{\alpha}$ 85 main chain O and $Gly_{\alpha}$ 84 main chain N, respectively. The $Cys_{\alpha}$ 80 O atom also interacts with the NH2 group of $Arg_{\alpha}$ 17 from the adjacent α subunit catalytic domain.

Besides, the main chain atoms N and O of $Cys_{\alpha}$ 100 are hydrogen bonded to main chains atoms from $Trp_{\alpha}$ 105, $Gly_{\alpha}$ 104 and $Tyr_{\alpha}$ 102. Hence, the Cys and His cluster ligands are part of a hydrogen bond network with the highly conserved residues, $Arg_{\alpha}$ 83, $Asn_{\alpha}$ 85, $Tyr_{\alpha}$ 102, $Trp_{\alpha}$ 105 and $Pro_{\alpha}$ 117 surrounding the Rieske cluster, thus promoting easy interactions with the mononuclear Fe in the catalytic domain of the adjacent α subunit.

### 3.3.2. 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, which separates the active site of the enzyme on one side of the sheet from the Rieske domain on the other side. 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 $Gly_{\alpha}$ 354, $Glu_{\alpha}$ 357, $Asp_{\alpha}$ 359 and $Asn_{\alpha}$ 363 which are part of a far reaching hydrogen network surrounding the catalytic center, as well as $Asp_{\alpha}$ 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, covers the cap of the catalytic domain (Fig. 3). Compared to other RHDs, the C terminus is shown to be different both in length and in amino acid sequence (Fig. 1).

A large depression, about 20 Å wide, on the surface of the catalytic domain receives the Rieske domain from the adjacent α subunit, allowing the [2Fe-2S] center to be held at the right position 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).

### 3.3.3. The substrate binding pocket

In PhnI and related RHDs, substrate binding takes place in a cavity lined up with mostly hydrophobic amino acids. Compared to other RHDs, the PhnI pocket (12 x 8 x 6 $Å^3$) is approximately 2 Å longer, and the largest reported so far. The hydrophobic pocket is surrounded by two loops called LI (residues 221 to 238) and LII (258 to 265), the α-helix aca6 containing two of the Fe ligands ($His_α$ 207 and $His_α$ 212) and the helices aca 10 and aca11, which include $Asp_α$ 360, the third iron ligand. The two solvent exposed loops LI and LII likely control the access of substrates to the catalytic pocket. As shown in Fig. 5, the conformation of loop LII was found to be different in the three α subunits of the protein crystal. Loop LI conformation could be defined, at least partially, in only one of the three α subunits, but the high flexibility of the loop precluded modeling in the two other cases.

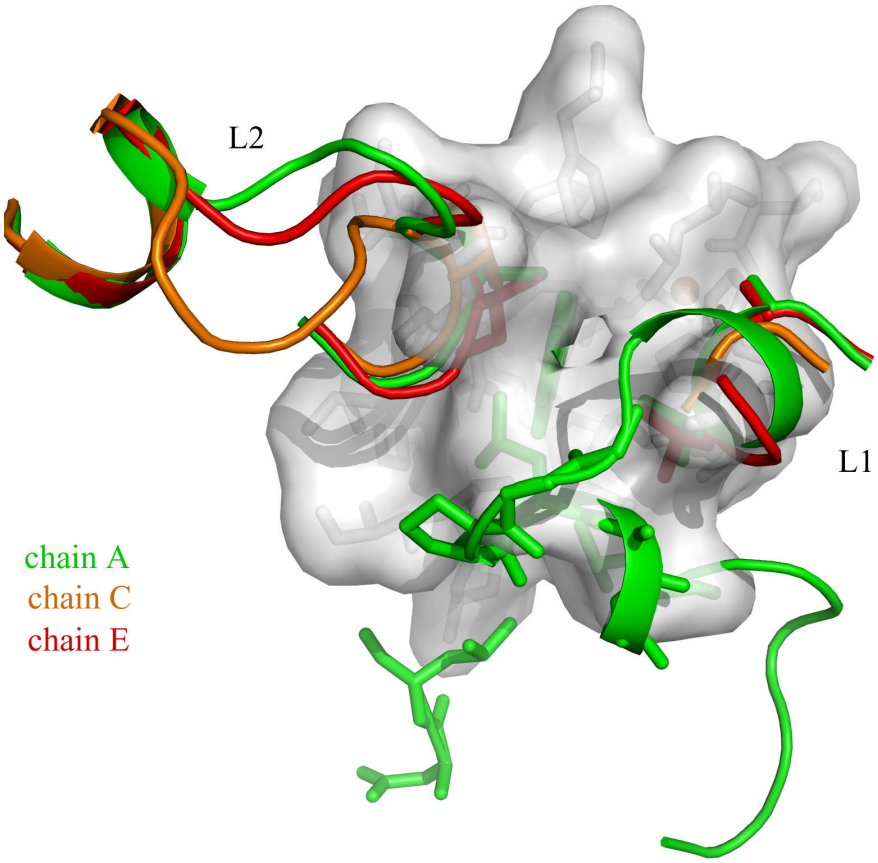



**Figure 5**. Surface envelope of the PhnI catalytic pocket. Shown are the three conformations adopted by loop LII at the entrance of the catalytic pocket. Loop LI is shown only for chain A as no density was observed in this region for the two other chains, C and E. Even for chain A, LI can not be represented completely, as no density was observed for residues 233 to 236. The figure was generated with the program PyMol [38].

Considering the broad substrate specificity of the dioxygenase from strain CHY-1 toward PAHs [9], it was of interest to study the binding of PAHs to the enzyme binding site on the basis of the present crystal structure. Using a molecular modeling approaches, the binding of several PAHs to the catalytic pocket was simulated. The results of these investigations, which will be published in detail elsewhere (Jakoncic, Jouanneau, Meyer, and Stojanoff, unpublished work), indicated that the catalytic pocket would bind large substrates made of 4 or 5 rings with minimal or no rearrangement of side chains. The four-ring benz[a]anthracene could be modeled in three favorable orientations, each of which corresponded to one of the three dihydrodiol isomers obtained by enzymatic conversion of this PAH [9]. Benzo[a]pyrene, a five-ring PAH, would bind the catalytic pocket in a single orientation with carbons 9 and 10 proximal to the mononuclear Fe, a prediction consistent with the identification of one product hydroxylated on the 9 and 10 positions upon in vitro assays [9]. Simulations of substrate binding to the catalytic cavity also indicated that amino acids at the entrance of the cavity and belonging to loops LI or LII determine the pocket length, and therefore might play a key role in the substrate selectivity of the enzyme. Modeling studies also indicated that the side chains of residues lining the central region of the pocket contribute to stabilize the substrate in the correct orientation prior to catalysis. In this respect, Phe350 has been predicted to be crucial for substrate orientation, and therefore for the regio-specificity of the enzyme. A comparison of RHDs with known structures indicates that amino acids lining the pocket side proximal to the solvent are variable suggesting that they contribute to the difference in substrate specificity observed among RHDs. In contrast, amino acids buried deep inside the pocket, close to the mononuclear iron atom, are totally conserved (Fig. 1).

### 3.3.4. The Mononuclear Fe

The mononuclear Fe is coordinated by a conserved 2-His-1-carboxylate triad motif [10], $His_{\alpha}$ 207, $His_{\alpha}$ 212 and bidentaly by $Asp_{\alpha}$ 360. The Fe coordination geometry can be described as that of a distorted octahedra with one unoccupied ligand position. The oxygen atom of $Asn_{\alpha}$ 200 (OD1) is located close to the position of the missing ligand (4 Å from the Fe atom). In several crystal structure of RHDs, the catalytic iron is also 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. During the course of the

crystallographic refinement, positive fo-fc map showed a large residual density in two of the three subunits suggesting the existence of an external ligand in the substrate binding pocket. The position and shape of this density is similar to that found in the NDO-$O_{98164}$ crystallographic structure [13] and resembles that of an indole molecule approaching the active site iron in a tilted position. In the third subunit which has such a ligand molecule, the refined distance between the two oxygen atoms, 1.5 Å, suggests the presence of a dioxygen molecule at the catalytic iron site.

### 3.3.5. Intramolecular Electron Transfer

$Asp_{\alpha}$ 204 is a totally conserved amino acid located between aca5 and aca6, and buried in a large depression at the junction of a Rieske domain and the catalytic domain of neighboring α subunit. In this key position, $Asp_{\alpha}$ 204 provides a bridge between the cluster and the mononuclear iron center. In PhnI, $Asp_{\alpha}$ 204 side chain is located between $His_{\alpha}$ 207, a ligand to the catalytic Fe, and $His_{\alpha}$ 103, a ligand to the Rieske center in the adjacent subunit. $Asp_{\alpha}$ 204 OD2 is 2.7 Å away from $His_{\alpha}$ 103 ND2, and OD1 is at 3.3 Å of $His_{\alpha}$ 207 ND1, thus providing a plausible path for intramolecular electron transfer. Replacement of this residue by a glutamine in NDO-O98164 resulted in a totally inactive enzyme [32].

### 3.3.6. alpha subunit interactions

Besides, $Asp_{\alpha}$ 204 participates in the hydrogen bond network holding the two redox centers at a distance of about 12 Å. $Asp_{\alpha}$ 204 OD2 is 3.3 Å away from $Tyr_{\alpha}$ 102 OH (in the adjacent α subunit) and is H-bonded to $Tyr_{\alpha}$ 410 OH (2.8 Å). $Asp_{\alpha}$ 204 OD1 is 3.3 Å from $His_{\alpha}$ 103 ND2 (adjacent subunit) and is H bonded to $His_{\alpha}$ 207 main chain N atom (2.7 Å). $Asp_{\alpha}$ 204 main chains atoms O and N interact with $His_{\alpha}$ 207 ND1 (2.9 Å) and $Asn_{\alpha}$ 200 O (3 Å) atoms respectively. This hydrogen bond network, depicted in figure 6, not only involves amino acids side and main chains interactions but also a few structural water molecules. Most of the residues involved in this hydrogen bond network are conserved and involve mainly Asp (204), Glu (407), Arg (83), Asn (200, 363) and mostly Tyr (102, 206, 410) (Fig.1 ). Specific to Phn1 are the interactions between $Arg_{\alpha}$ 17 and $Ser_{\alpha}$79 and $Cys_{\alpha}$ 80.

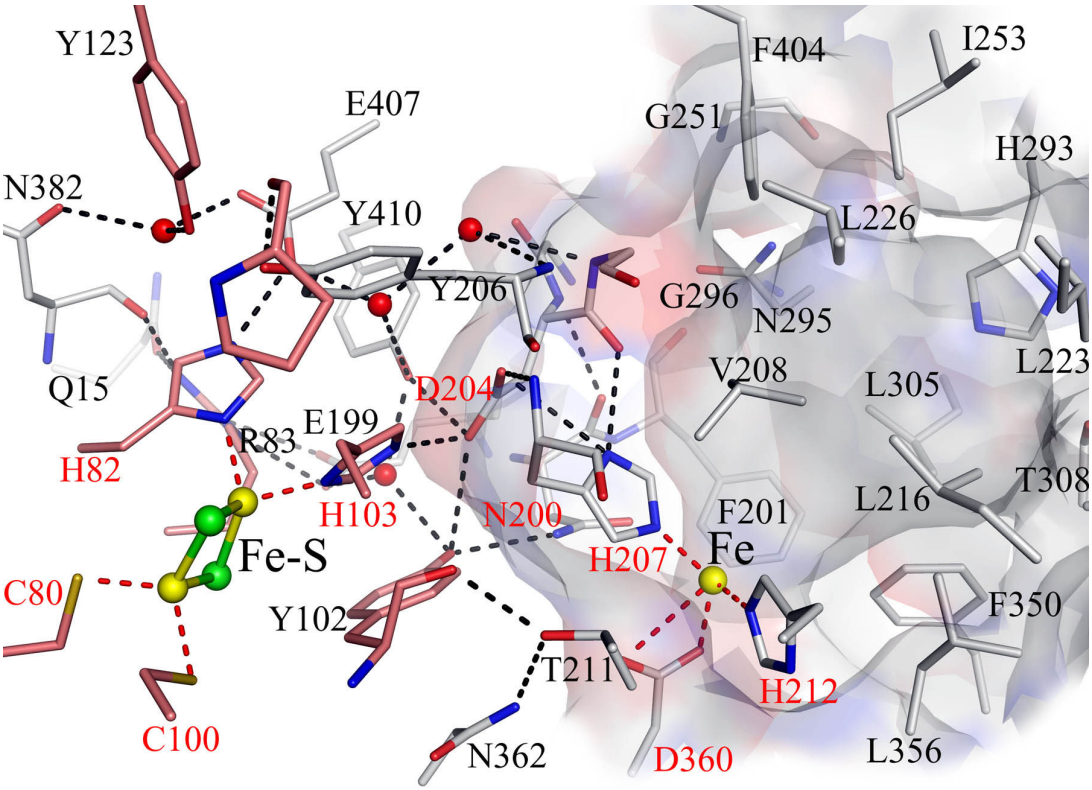



**Figure 6**. Interdomain interaction. Rieske domain and catalytic domain of neigboring α subunits. In red ligands to the reaction centers and residues Asn 200 and Asp 204 believed to be involved in the electron transfer process to the catalytic site. Also shown in red are relevant water molecules to the hydrogen network. In the background the catalytic surface envelope of the PhnI catalytic pocket showing the available internal space.

### 3.3.7. Interdomain interactions

Besides the αα interactions, the $\alpha_3\beta_3$ form is maintained by multiple interdomain interactions found in 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. For instance $Asp_{\alpha}91$ interacts with $Arg_{\beta}$ 163 and this interaction is conserved in naphthalene dioxygenases. In the heterohexamer, the Rieske domain interacts with the base of the adjacent β subunit and the adjacent α subunit catalytic domain. Most of the αβ interactions are conserved at least in the dioxygenases from the naphthalene family. For example $Trp_{\alpha}$ 91 at the base of the β subunit (helix ba4) interacts with $Trp_{\alpha}$ 210 in helix aca6 from the α subunit catalytic domain and with $Asn_{\alpha}$ 101, located on the gripper structure from the adjacent α subunit Rieske domain. These additional and numerous

interactions participate to keep the $\alpha_3\beta_3$ form and ultimately support the catalytic reaction to take place maintaining the two redox centers in the best conformation. If multiple αβ and ββ interactions are found in PhnI, the β subunits seem to play solely a structural function.

### 3.3.8. Occurrence of a water channel

A 11 Å long channel filled with eight ordered water molecules extends from the base of the β subunit up to the catalytic site (Fig. 7). The water molecule closest to the catalytic site is at hydrogen bond distance from $Glu_\alpha$ 357 and at 4.2 Å from the Fe atom. This channel is also found in other RHDs although residues lining the channel are not fully conserved. 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.

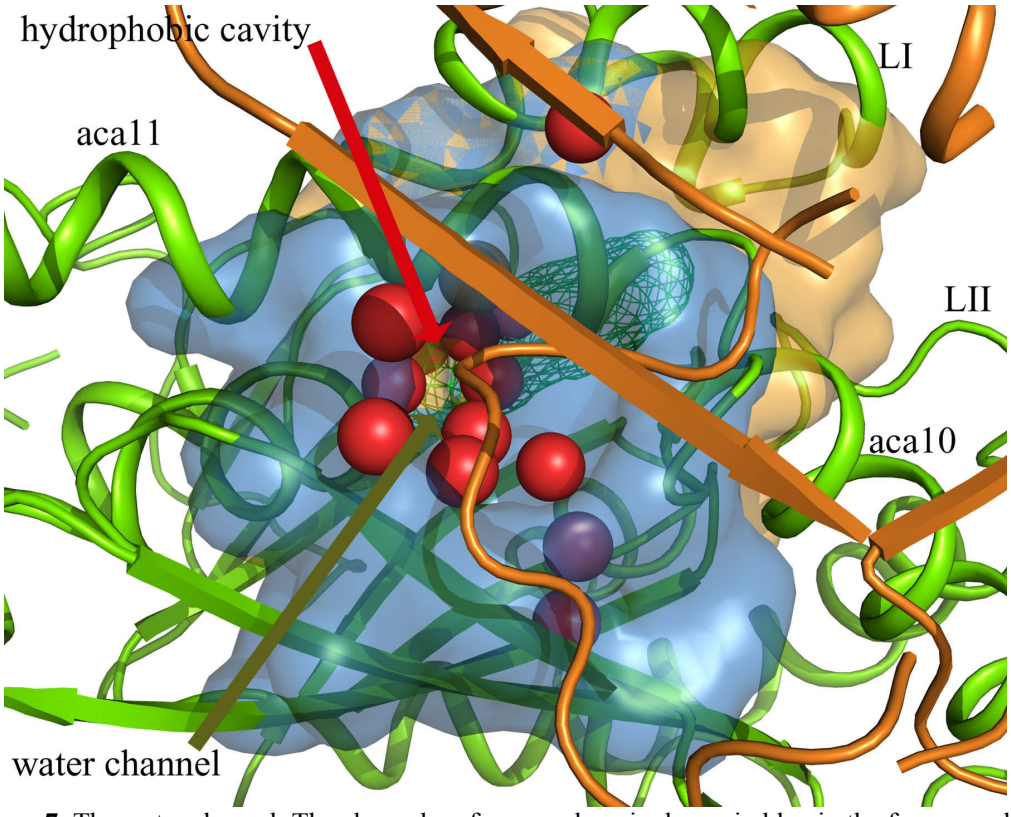



**Figure 7.** The water channel. The channel surface envelope is shown in blue in the foreground and the surface of the catalytic pocket in orange in the back. Structural water molecules are shown in red at the entrance and inside the channel. At the end of the channel a green mesh represents a five ring PAH modeled into the catalytic pocket. Partial ribbon diagram of the β subunit, chain B, in orange, and α subunit, chain A, in green. The Figure was made using PyMol [38]

### 3.3.9. Possible role of Asn 200

The hybrid density functional theory (DFT) was used to study the *cis*-dihydroxylation mechanism in naphthalene dioxygenase (NDO-$O_{98164}$) [33]. Starting from a hydroperoxo-iron(III) moiety as the intermediate reactive species, Bassan *et al*. [33] showed that the energetically most favorable pathway would generate an epoxide as the first intermediate, which would give rise to an arene cation that would eventually convert to the *cis*-dihydrodiol. This theoretical analysis predicts Asn 201 (NDO-$O_{98164}$) to be at hydrogen bond distance from the hydroxide during a transition state. In fact, the ND2 side chain atom of the corresponding Asn residue ($Asn_{\alpha}$ 200 in PhnI) is approximately 3 Å away from one of the water molecules. 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 Asn201 in NDO- $O_{98164}$ was tested by substitution of this residue by Gln, Ser or Ala [34]. 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 in the adjacent subunit (Fig. 6).

## 4. Conclusions

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 Pheα 350, Pheα 404 and Leuα 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 Leuα 223 and Leuα 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.

## Acknowledgements

We would like to thank the staff of the National Synchrotron Light Source, Brookhaven National Laboratory for their continuous support. The NSLS is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-98CH10886. The NIGMS East Coast Structural Biology Research Facility, the X6A beam line, is funded under contract # GM-0080.

## 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] S. Harayama, M. Kok, E. L. Neidle Functional and evolutionary relationships among diverse oxygenases, Annu. Rev. Microbiol. 46 (1992) 565-601.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

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

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

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

# Partie 4 : Article 3

## The catalytic pocket of the ring-hydroxylating dioxygenase from *Sphingomonas* CHY-1

Cet article a été publié en 2007 dans la revue *BBRC* (volume 352, pages 861-866) et décrit en détail la structure de la poche catalytique de la dioxygénase de Sphingomonas CHY-1. Il présente en outre une modélisation du benzo[a]pyrène et d'autres HAP au site actif de l'enzyme.

# The catalytic pocket 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.

## Abstract

Ring-hydroxylating dioxygenases are multicomponent bacterial enzymes that catalyze the first step in the oxidative degradation of aromatic hydrocarbons. The dioxygenase from *Sphingomonas* CHY-1 is unique in that it can oxidize a wide range of polycyclic aromatic hydrocarbons (PAHs). With a crystal structure similar to that of the seven other known dioxygenases, its catalytic domain features the largest hydrophobic substrate binding cavity characterized so far. Molecular modeling studies indicated that the catalytic cavity is large enough to accommodate a five-ring benzo[a]pyrene molecule. The predicted positions of this and other PAHs in the substrate binding pocket are consistent with the product regio- and stereo-selectivity of the enzyme.

**Keywords:** dioxygenase; catalytic domain; mononuclear iron; bioremediation; high molecular weight polycyclic aromatic hydrocarbons

## 1. Introduction

The first step in the biodegradation of aromatic hydrocarbons by aerobic bacteria often involves a dihydroxylation on two adjacent carbon atoms of the aromatic ring, catalyzed by a ring-hydroxylating dioxygenase (RHD). RHDs form a large family of enzymes, very diverse in terms of substrate specificity and protein sequence [1]. Their role is crucial in the degradation of many organic pollutants, including polycyclic aromatic hydrocarbons (PAHs), which are notorious for their resistance to biodegradation. Several bacteria were found to degrade PAHs but only a few have been reported to attack four and five ring PAHs [2,3]. In *Sphingomonas* strain CHY-1 a single RHD has been shown to be responsible for the oxidation of a wide range of PAHs [4]. This dioxygenase consists of three components, a NADH-dependent reductase (PhnA4), a ferredoxin containing a Rieske type [2Fe-2S] cluster (PhnA3), and a terminal oxygenase, PhnI, containing both a mononuclear iron [$Fe^{2+}$] and a [2Fe-2S] Rieske cluster [5]. Recent biochemical studies showed that the dioxygenase from strain CHY-1 was able to oxidize at least eight PAHs made of 2 to 5 aromatic rings, in contrast to most other dioxygenases, whose selectivity is limited to 2 and 3 ring PAHs [6,7,8]. A positive electron density observed in the PhnI refined three-dimensional structure served as probe in modeling different substrates in the catalytic pocket. This study presents evidence that the broad substrate specificity of the enzyme is primarily due to the large volume and particular shape of its catalytic pocket.

## 2. Material and methods

The purification of recombinant His-tagged PhnI was performed as described previously [5]. The protein was further purified by gel filtration chromatography following His-tag removal by thrombin cleavage. Crystals were grown at room temperature using the sitting drop method and a crystallization solution derived from Cryoscreen solution 67 (Nextal Biotechnologies, Montreal, Quebec, CA). Crystals that diffracted up to 1.85 Å resolution (Jakoncic, unpublished work) were obtained by further covering each well with mineral oil [9]. Diffraction data were recorded on beam line X6A at the National Synchrotron Light Source [10]. The 3D structure was determined by molecular replacement, using the α-subunit from naphthalene dioxygenase from *Pseudomonas* sp. Strain NCIB 9816-4, NDO-$O_{9816\text{-}4}$ [6] and the β-subunit from cumene dioxygenase from *P. fluorescens* strain IP01, CDO-$O_{IP01}$ [11] as templates. The model was refined using

REFMAC [12] and COOT [13] to R and $R_{free}$ factors of 19.7 and 23.6 %, respectively (Fig. 1). Most of the amino acids are present in the final model (PDB accession code **2CKF**), with the exception of residues in the C- and N- termini regions and a highly flexible region at the entrance of the catalytic pocket LI (residues 220 to 237) and LII (residues 258 to 265) described in the following section. Docking experiments were performed manually in the program COOT [13] using the residual positive density and the relative position of substrates in other oxygenase as they are found in binary complexe forms. The most probable substrate orientations are the one with minimum or no steric constrains with amino acids lining the hydrophobic cavity.

## 3. Results and Discussion

### 3.1. The Phn1 catalytic domain

The PhnI quaternary and tertiary structures were found to be similar to those of other RHDs, featuring a $\alpha_3\beta_3$ arrangement with each $\alpha\beta$ heterodimer (Fig. 1) related to the other by a non-crystallographic three-fold axis [6]. Each α-subunit can be divided into a Rieske-cluster containing domain and a catalytic domain hosting one mononuclear iron. The core region of the catalytic domain is dominated by a nine-stranded anti-parallel β-sheet that divides the domain into two parts: the active site of the enzyme on one side of the β-sheet and the Rieske domain on the other. Strategically located in the vicinity of the catalytic iron atom, and covering one side of the sheet, lies a highly conserved secondary structure amongst RHDs, a α-helix that extends from residues 336 to 373 in PhnI. The mononuclear Fe atom is located in the center of a 35 Å long cavity extending from the solvent to the anti-parallel β-sheet. On one side of the cavity, a hydrophobic pocket extends from the solvent to the mononuclear iron (Fig. 1). Two solvent exposed loops, LI and LII, form the entrance of the pocket. The overall temperature factor of the Cα main chain in this region is relatively higher compared to the rest of the molecule; the B factor being highest for LI. While LI was relatively well defined in one of the α monomers with only five missing amino acids (Fig.1), it was barely detectable in the other two. LII on the other hand was fully defined in the electron density maps, but presented a different conformation for each of the monomers. Highly flexible structures are also observed in the corresponding regions for other RHDs, suggesting that LI might control the access to the

catalytic pocket.

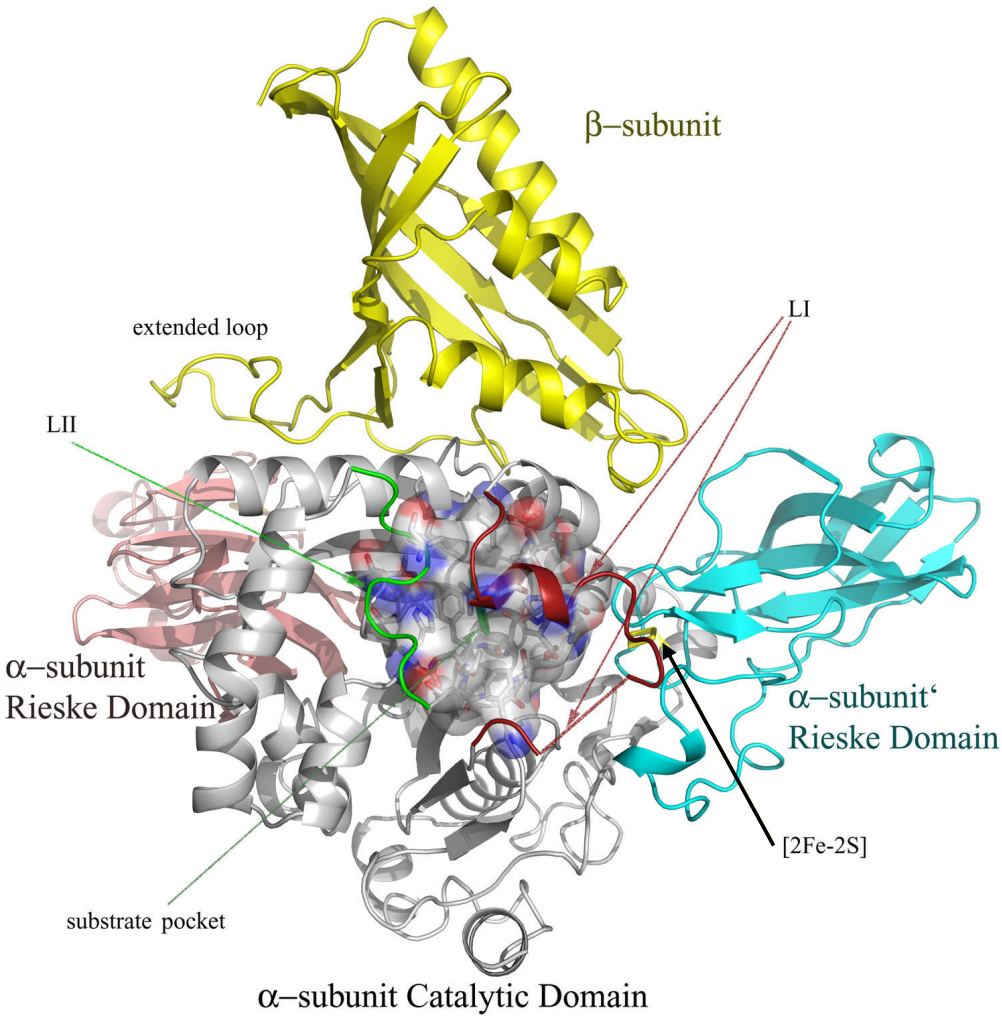



**Figure 1**. Structure of the αβ heterodimer of PhnI: the α-subunit is shown in grey, catalytic domain, and in pink, Rieske domain, the β-subunit is shown in yellow. The surface of the substrate binding pocket is shown in the foreground; loops LI and LII are indicated. The Rieske domain of the adjacent α-subunit is shown in cyan. This view emphasizes important features of the enzyme, especially the proximity of the 2Fe-2S cluster and the mononuclear Fe at the interface between two adjacent α-subunits. The figure was produced with Pymol [22].

In the PhnI crystal structure, an unidentified ligand could be observed in two of the three catalytic pockets. As shown in Fig. 2, the size and shape of the density recalls that of an indole molecule with its C3 carbon located 4.5 Å from the catalytic iron atom. The electron density maps in Fig. 2 clearly show the presence of a residual positive density between the catalytic iron and the modeled indole molecule. Therefore the final PhnI crystal model only includes two water molecules in this region.

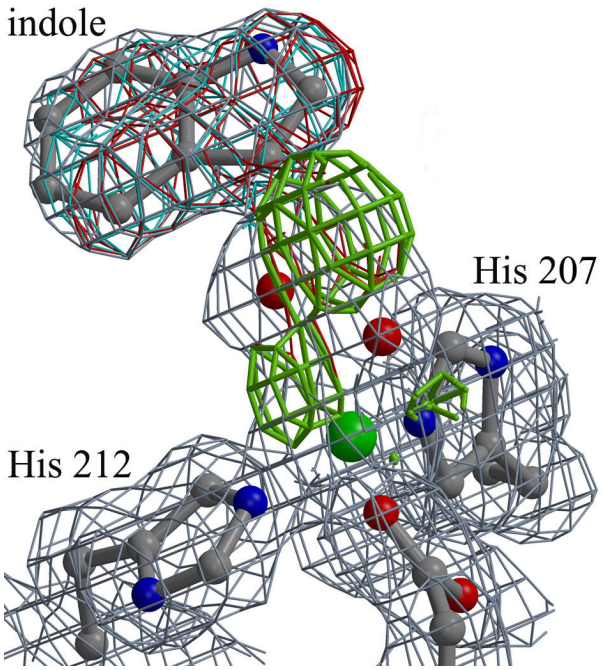



**Figure 2**. Close view of the active site showing the residual positive density modeled as an indole molecule. Four maps are shown: 2fo-fc contoured at the 1 σ level, in grey and the fo-fc at the +3σ, ingreen after addition of the indole molecule; and the respective Omit maps, 2fo-fc at 1 σ in cyan and fo-fc at 3σ, in red. fo-fc maps contoured at the -3σ levelldid not present any residual density and are not shown. In this configuration, the C3 atom of indole is located 2 Å from the closest water molecule, in red, which is 2.6 Å from the mononuclear Fe atom shown in green and at 2.2 Å from the second coordinated water molecule. The figure was produced with BOBSCRIPT [23,24] and Raster3D [25].

### 3.2. Topology of the catalytic pocket

The Phn1 catalytic pocket has a trapezoidal shape, wider at the entrance (close to the solvent) and narrower towards the catalytic Fe atom. The average dimensions are 12 Å in length, 8 Å in height and 5.5 to 6.5 Å in width. The program POCKET [14] was employed to determine the volume of the PhnI substrate binding pocket. The resulting volume corresponds to the unoccupied space determined by a 1.4 Å radius probe following the protein surface. The mononuclear iron and its three ligands, His 207, His 212 and Asp 360, as well as, the coordinated water molecules, were not included for the calculations. For comparison purposes, the volume of the catalytic pocket was also determined for two other dioxygenases in their substrate free form. The smallest pocket size with an estimated

volume of 45 $Å^3$ was determined for biphenyl dioxygenase from *Rhodococcus* sp. Strain RHA1, BPDO-$O_{RHA1}$ (PDB access code 1ULI) [15], while the catalytic pocket volume for naphthalene dioxygenase, NDO-$O_{9816}$ (PDB access code 1NDO) [6], found to be 65 $Å^3$, is significantly larger. The Phn1 catalytic pocket, with its 91 $Å^3$, is the largest reported so far. Compared to the substrate free form of NDO-$O_{9816}$ and BPDO-$O_{RHA1}$ the PhnI catalytic pocket is at least 2 Å longer, wider and higher at the entrance in the solvent region, which would likely explain the ability of this enzyme to oxidize larger PAH substrates with up to 5 rings [5]. The structural determination of the BPDO-$O_{RHA1}$ enzyme in the presence of its substrate showed that the catalytic pocket becomes 1.5 Å longer compared to the substrate-free enzyme [15]. The α-helix, containing residues Leu 274 and Ile 278 (Leu 223 and Ile260 in Phn1), is translated about 1.4 Å towards the solvent upon binding the biphenyl molecule, thereby increasing the pocket length. A similar translation was observed for 2-Oxoquinoline 8-Monooxygenase from *Pseudoomonas putida* 86, OMO-$O_{86}$, where the loop at the entrance of the catalytic pocket (equivalent to loop LI in PhnI) underwent a significant conformational change upon substrate binding [16].

Sequence identity of structurally analogous residues lining the catalytic pocket for different dioxygenases of known crystallographic structure shows that PhnI presents the highest sequence identity with NDO-$O_{9816}$ and the lowest with carbazole-1-9α-dioxygenase from *P. resinovorans* strain CA10 (CARDO-$O_{CA10}$) [17]. The classification presented in Table 1, based on α-subunit sequence homology and substrate specificity, is consistent with current classifications [18]. The PhnI catalytic pocket represented in Fig. 3 can be divided into three regions, distal, central and proximal, depending on the distance to the mononuclear Fe atom (Table 1). Compared to other dioxygenases, the PhnI pocket is rather uniform in shape and does not present any kinks or torsions as was found for example for BPDO-$O_{RHA1}$ [15]. Residues Phe 350 and Phe 404 in the central region are expected to affect the topology of the catalytic pocket and consequently select the shape and form of allowed substrates. Accordingly, the replacement of Phe 352 in NDO-$O_{9816\text{-}4}$ by smaller amino acids significantly altered the regiospecifity of the reaction products [19]. In NDO-$O_{9816\text{-}4}$ Phe 404 is replaced by Ala 407. Further differences, observed at positions 308 and 356, contribute to enlarge the catalytic pocket of PhnI compared to that of NDO-$O_{9816\text{-}4}$. Indeed, Leu 356 in PhnI is replaced by a bulky aromatic residue (Trp or Phe) in naphthalene dioxygenases (Table 1).

| | Group I | | | | Group II | | Group III | |
|---|---|---|---|---|---|---|---|---|
| | **PhnI** | **NDO-$O_{9816-4}$** | **NBDO-$O_{JS765}$** | **NDO-$O_{12038}$** | **CDO-$O_{IPO1}$** | **BPDO-$O_{RHA1}$** | **OMO-$O_{86}$** | **CARDO-$O_{CA10}$** |
| P | Asn200 | Asn201 | Asn199 | Asn209 | Gln227 | Gln217 | Asn215 | Asn177 |
| P | Phe201 | Phe202 | Phe200 | Phe210 | Phe228 | Phe218 | Leu302<br>Gly216 | Leu207<br>Gly178 |
| P | Val202 | Val203 | Val201 | Val211 | Cys229 | Cys219 | Phe217<br>Val296 | Phe179<br>Ile264 |
| C | Gly203 | Gly204 | Gly202 | Gly212 | Ser230 | Ser220 | Gln264<br>Phe217 | -------- |
| C | Gly205 | Ala206 | Gly204 | Ala214 | Met232 | Met222 | Asn219 | Pro181 |
| C | Val208 | Val209 | Val207 | Thr217 | Ala235 | Ala225 | Ile222 | Ile184 |
| C | Leu216 | Leu217 | Leu215 | Val225<br>Ala230 | Val244 | Val234 | Leu238 | Val193<br>Leu200 |
| D LI | Leu223 | Phe224 | Leu222 | Phe293 | Leu284 | Leu274 | Val231 | Ala199 |
| D LI | Leu226 | Leu227 | Leu225 | Phe236 | Leu259 | ------ | Pro239 | Pro201 |
| C | Gly251 | Gly251 | Gly249 | Gly252 | Gly276 | Gly266 | ------- | Ile262<br>Asp229 |
| D LII | Ile253 | Leu253 | Phe251 | Ile254 | Phe278 | Tyr268 | Tyr292<br>Phe267 | -------- |
| D LII | Ile260 | Val260 | Asn258 | Met309 | Ile288 | Ile278 | Trp307 | -------- |
| C | His293 | His295 | Phe293 | Phe293 | Ala321 | Ala311 | Tyr292 | Ala259 |
| P | Asn295 | Asn297 | Asn295 | His295 | His323 | His313 | Thr294 | Ile262 |
| C | Leu305 | Leu307 | Leu305 | Phe307<br>Gly305 | Leu333 | Leu323 | Gln314<br>Val304 | Val272 |
| C | Thr308 | Ser310 | Ser308 | Phe320<br>Phe307 | Ile336 | Ile326 | Trp307 | Phe275 |
| C | Phe350 | Phe352 | Ile350 | Phe362 | Phe378 | Phe368 | Asn362 | Asn330<br>Phe275 |
| C | Leu356 | Trp358 | Trp356 | Phe368 | Tyr384 | Tyr374 | Phe361 | Phe329 |
| C | Phe404 | Ala407 | Phe251<br>Ala405 | Asp418 | Met227<br>Phe273 | Met222<br>Val422 | Leu266<br>Asn219 | --------- |
| **Identity(%)** | | 63 | 47 | 37 | 37 | 32 | 11 | 5 |

**Table 1**. Relative position to the catalytic Fe: D, distal; C, central and P, proximal. (--) no structurally equivalent residues were observed. Substrate free structures were used in the comparison [6,16,7,12,15,17,18].

Residues in the distal region seem to exert a greater influence in selecting the size and shape of allowed substrates. In PhnI, most significant are residues Leu 223 and Leu 226 located on loop LI, and residues Ile 253 and Ile 260 on loop LII. In naphthalene dioxygenases, NDO-$O_{9816}$ and NDO-$O_{12038}$, Leu 223, is substituted by a bulkier and less flexible phenylalanine residue (Table I). The diversity in residues structurally equivalent to Ile 260 in PhnI must relate to the different substrate specificity observed between the members of the naphthalene dioxygenase family.

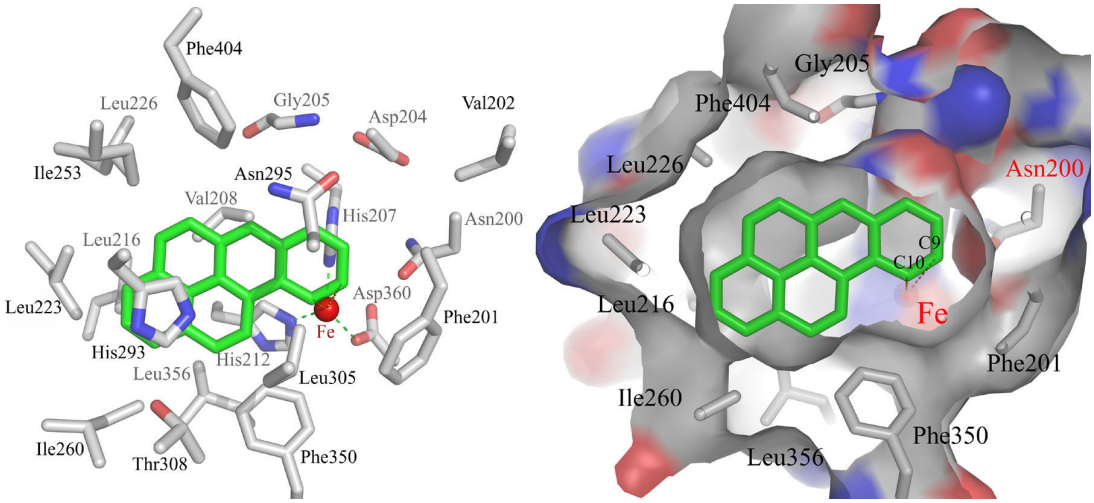



**Figure 3**. The PhnI active site with a BaP substrate modeled in the catalytic pocket. (a) Residues shown in front, relative to the substrate (in green) are labeled in black and residues in the back are labeled in grey. With the exception of Gly 205, only side chains are represented. The mononuclear Fe is represented as a red ball. (b) Surface plot of the PhnI substrate cavity showing the substrate in the same orientation exposing C9 and C10 towards the mononuclear iron.

### 3.3. Substrate specificity

To explore the substrate specificity of PhnI from a structural point of view, 2, 3, 4 and 5 ring PAHs were modeled into the catalytic pocket of the refined structure (PDB access code 2CKF). The structures of the enzyme-substrate complexes described for NDO-$O_{9816}$ and BPDO-$O_{RHA1}$ (PDB access codes 1O7G and 1ULJ), were first used to fit and adjust the position of naphthalene and biphenyl substrates in the PhnI catalytic pocket. As the pocket is narrower in the proximal region (Fig.4), modeling indicated that the two-ring substrates would be locked in a single position consistent with the finding that PhnI hydroxylates naphthalene and biphenyl, respectively, in positions 1,2- and 2,3- [5].

Phenanthrene, a three ring angular molecule, could theoretically be hydroxylated in positions 1,2-, 3,4-, or 9,10-. However, due to steric constraints imposed by the PhnI pocket, only one position, which would bring the C3 and C4 carbon atoms close to the active Fe atom site, is allowed. This analysis corroborates enzymatic assays for which *cis*-3,4- phenanthrene dihydrodiol was the only product detected [5].

The five ring PAH benzo[a]pyrene can only fit into the PhnI catalytic pocket in a single orientation. Minor conformational changes are assumed in modeling indicating that the side chains of Leu 223 and Phe 350 need to be rotated if the substrate is to fit in the pocket. In this orientation, shown in Fig. 4, benzo[a]pyrene would be hydroxylated in position 9,10-. In fact 9,10-*cis*-dihydrodiol- benzo[a]pyrene is the only product observed as the result of enzymatic assays [5]. On the other hand, benzo[a]anthracene (BaA), a four ring PAH, was found to be attacked at positions 1,2-, 8,9- and 10,11- with *cis*-dihydrodiols 1,2- and 10,11- presenting the highest yields. Substrate orientation leading to an hydroxylation on the 1,2- position would require a minimal rearrangement of residues in the catalytic pocket, involving only side chain conformational changes of Leu 223 in loop LI. Hydroxylation on carbons C10 and C11 would occur for a position of the substrate requiring rearrangements of both Leu 223 and Phe 350 side chains. Conformational changes of residues 223, 253 and 404 should take place to allow hydroxylation of BaA at the less favorable 8,9- position. These predictions are in accordance with the catalytic properties of the enzyme, since it was found to produce much less of the 8,9-dihydrodiol [5].

Phe 350 and Phe 404 in the central region contribute to the regio-specificity and to the PhnI pocket shape. Phe 350 is conserved in NDO-$O_{9816\text{-}4}$ and in the naphthalene dioxygenase from *Rhodococcus* sp. strain NCIMB 12038 (NDO-$O_{12038}$) [7], while Phe 404 is replaced by Ala 407 in NDO-$O_{9816\text{-}4}$ and by Asp 418 in NDO-$O_{12038}$. In NDO-$O_{9816\text{-}4}$, mutations of Phe 352 (Phe 350 in PhnI) into a valine and a leucine resulted in altered regio-specificity of the enzyme. These mutations altered the pocket topology, so that other orientations of the biphenyl and phenanthrene substrates were allowed [20]. The pocket size and shape might also influence the effectiveness of the reaction, by locking the substrate in the right position before catalysis. Replacement of Trp 358 by an alanine in NDO-$O_{12038}$ resulted in inefficient transformation of naphthalene, because the bulky side chain of Trp 358 was crucial to maintain the substrate in the right position [19]. It is interesting to note all RHDs of known structure have an aromatic residue at a position equivalent to Trp 358, except PhnI, which has a leucine in the corresponding position.

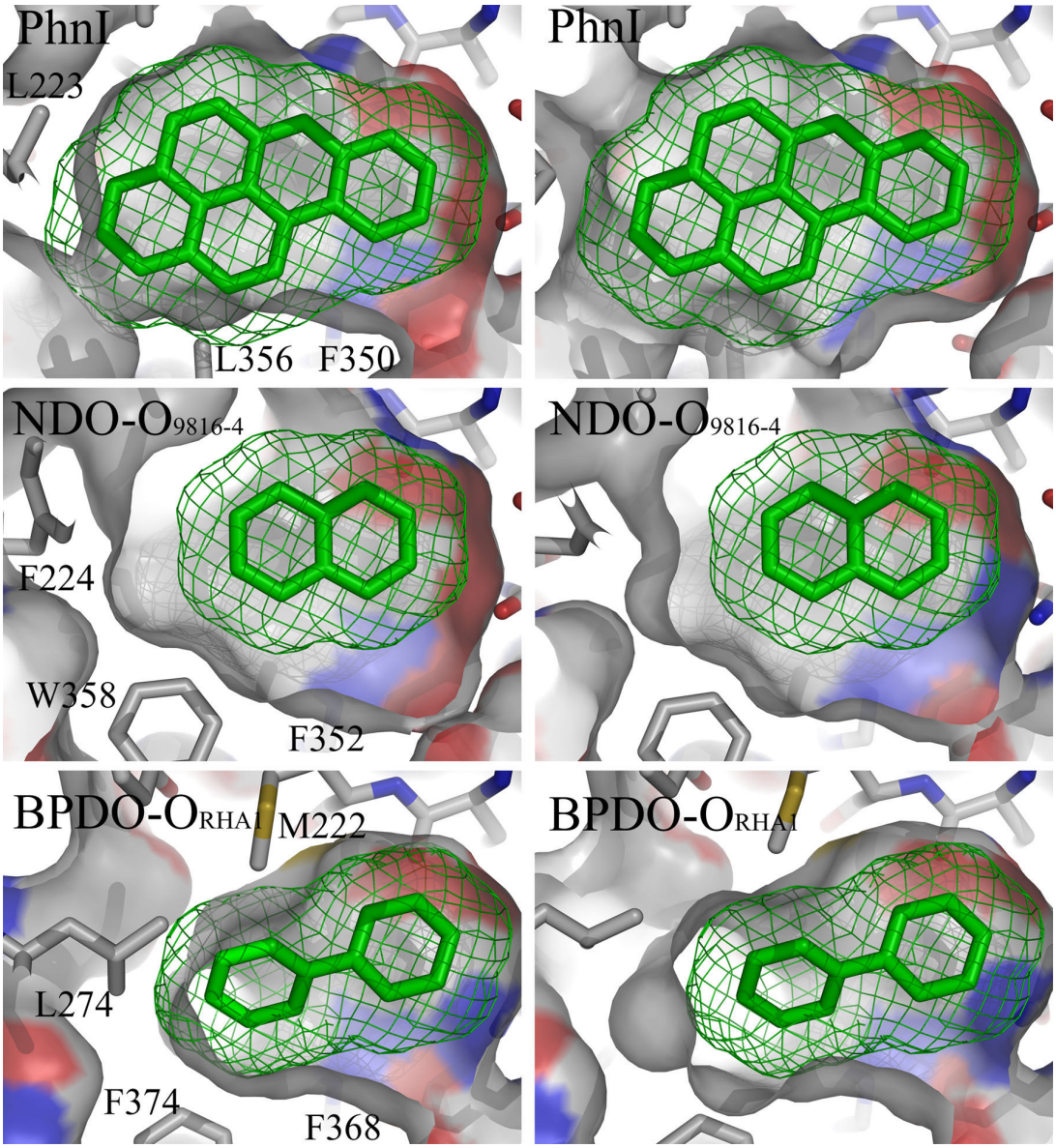



**Figure 4**. The catalytic pocket for PhnI (upper panel), NDO-$O_{9816-4}$ (central panel) and BPDO-$O_{RHA1}$ (lower panel), is shown on the left in its substrate free form and on the right for the substrate bound enzyme complexes. For each of the dioxygenases represented the comparison between the left and right panels shows the changes in the catalytic pocket upon substrate binding. The structures for NDO-$O_{9816-4}$ and BPDO-$O_{RHA1}$ on the right were obtained using the coordinates of the substrate-bound enzyme complexes (PDB access code 1O7G and 1ULJ, respectively). For PhnI, the cavity shown on the right was obtained after rotation of side chains of residues, Leu 223 and Phe 350, taking into account steric constrains due to van der Waals contacts. With these minor conformational changes BaP can only fit in the PhnI pocket in the orientation shown. In this orientation BaP would be hydroxylated in position 9,10- consistent with biochemical assays [5]. The figure was produced with Pymol [22].

## 4. Conclusion

In conclusion, PhnI is endowed with a remarkable broad specificity towards high molecular weight PAHs, which might be explained by the shape and size of its substrate binding pocket. The PhnI pocket was found to be at least 2 Å longer and wider at the entrance a unique feature between dioxygenases of known structures certainly allowing the five-ring benzo[a]pyrene to bind to the catalytic Fe. Modeling of various PAHs showed that Phe 350 in the central region of the pocket is essential for the regio- and substrate specificity, while Leu 223 and Ile260 in the distal region contribute to the specificity of high molecular weight PAHs. Further studies involving replacements of specific residues of the substrate-binding pocket by site-directed mutagenesis should bring new insights into the role of these residues in the catalytic activity of the enzyme.

## Acknowledgements

We would like to thank the staff of the National Synchrotron Light Source, Brookhaven National Laboratory for their continuous support. The NSLS is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-98CH10886. The NIGMS East Coast Structural Biology Facility, the X6A beam line, is funded under contract # GM-0080.

## References


[1] C.S. Butler, J.R., Mason, Structure–function analysis of the bacterial aromatic ring-hydroxylating dioxygenases, Adv. Microb. Physiol. 38 (1997) 47-84.

[2] A.L. Juhasz, 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.

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

[4] 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 *Sphingomonas* strain that degrades various polycyclic aromatic hydrocarbons, App. Environment. Microbiol. 70 (2004) 6714-6725.

[5] Y. Jouanneau, C. Meyer, J. Jakoncic, V. Stojanoff, J. Gaillard, Characterization of a Naphthalene Dioxygenase Endowed with an Exceptionally Broad Substrate Specificity toward Polycyclic Aromatic Hydrocarbons, Biochemistry 45 (2006) 12380-12391.

[6] 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.

[7] L. Gakhar, Z.A. Malik, C.C.R. Allen, D.A. Lipscomb, M.J. Larkin, S. Ramaswamy, Structure and Increased Thermostability of *Rhodococcus* sp. Naphthalene 1,2-Dioxygenase, J. Bacteriol. 197 (2005) 7222-7231.

[8] D.J. Ferraro, L. Gakhar, S. Ramaswamy, Rieske business: Structure–function of Rieske non-heme oxygenases, Biochemical and Biophysical Research Communications 338 (2005) 175–190.

[9] N.E. Chayen, A novel technique to control the rate of vapour diffusion, giving larger protein crystals, J. Appl. Cryst. 30 (1997) 198-202.

[10] M. Allaire, M. Aslantas, A. Berntson, L. Bermann, 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, Synchrotron Radiation News 16 (2003) 20-25.

[11] 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.

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

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

[14] J. Liang, H. Edelsbrunner, C. Woodward (1998) Anatomy of Protein Pockets and Cavities: Measurement of Binding Site Geometry and Implications for Ligand Design, Protein Science, 7, 1884-1897.

[15] 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.

[16] 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.

[17] 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.

[18] Nam, J.W., Nojiri, H., Yoshida, T., Habe, H., Yamane, H., Omori, T. (2001) New Classification System for Oxygenase Components Involved in Ring-Hydroxylating Oxygenations, Biosci. Biotechnol. Biochem. 65, 254-263

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

[20] R.E. Parales, S.M. Resnick, C. Yu, D.R. Boyd, N.D. Sharma, D.T. Gibson, Regioselectivity and Enantioselectivity of Naphthalene Dioxygenase during Arene *cis*-Dihydroxylation: Control by Phenylalanine 352 in the α–Subunit, J. Bacteriol. 182 (2000) 5495-5504.

[21] R. Friemann, M.M. Ivkovic-Jensen, D.J. Lessner, C. 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.

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

[23] R.M. Esnouf, Journal of Molecular Graphics 15 (1997) 132-134.

[24] P.J. Kraulis, J. Appl. Cryst. 24 (1991) 946-950.

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

# Partie 5 : Conclusions et perspectives

## Conclusions

Dans ce chapitre, les travaux ont porté sur la caractérisation complète de la dioxygénase de la bactérie *Sphingomonas* CHY-1. J'ai participé à la purification de l'enzyme et à la détermination de ses propriétés spectroscopiques mais l'essentiel du travail de caractérisation biochimique de la dioxygénase a été réalisé par mes collègues à Grenoble. Ma principale contribution a consisté à mettre au point les conditions de cristallisation, puis à établir par cristallographie la structure 3D de la composante oxygénase de l'enzyme.

### Purification et activité catabolique

Le clonage des gènes *phnA1a* et *phnA2a* à partir du génome de la souche CHY-1 et leur surexpression dans *E. coli* a conduit à l'obtention d'une enzyme, PhnI, capable d'oxyder un large panel de HAP de 2 à 5 cycles (Demanèche *et al*., 2004). L'activité de PhnI *in vivo* est dépendante de la co-expression des gènes *phnA3* et *phnA4* codant une ferrédoxine et une réductase. L'enzyme est donc un complexe constitué des trois protéines PhnI, PhnA3 et PhnA4. Chacune de ces protéines a été purifiée séparément sous forme recombinante à l'aide de systèmes d'expression appropriés. Je ne rappellerai ici que ce qui concerne la composante oxygénase PhnI.

PhnI est une protéine de 215 kDa de type $\alpha_3\beta_3$ semblable aux autres dioxygénases connues. En associant à PhnI les deux protéines purifiées PhnA3 et $Red_{B356}$, nous avons reconstitué un complexe actif *in vitro*, capable d'oxyder les HAP en présence de NADH et d'oxygène. Sur les 16 HAP jugés polluants prioritaires, 9 d'entre eux comportant de 2 à 5 cycles ont été oxydés, ce qui fait de PhnI la dioxygénase ayant la plus large spécificité connue à ce jour. C'est la première enzyme caractérisée capable d'attaquer des HAP à 4 cycles (chrysène, benz[a]anthracène) et à 5 cycles (benzo[a]pyrène). PhnI ne donne le plus souvent qu'un seul produit d'oxydation, sauf pour le fluorène, le fluoranthène et le benz[a]anthracène (Tableau 1). Le benz[a]anthracène donne trois dihydrodiols séparables en CPG-SM, dont deux sont sujets à une seconde hydroxylation donnant lieu à la formation du même produit tetrahydroxylé, le 1,2,10,11-bis-*cis*-dihydrodiol. Ce produit n'est sans doute pas métabolisable par les bactéries. Cependant, il a peu de chances de se

former *in vivo* dans la souche CHY-1, car la réaction de dioxygénation catalysée par PhnI est sans doute couplée à celle catalysée par PDDH, un déshydrogénase qui convertit les dihydrodiols en produits de type catéchol. L'affinité et le turnover de PDDH pour les dihydrodiols produits par PhnI sont tels, que les diols du benz[a]anthracène ne peuvent pas réagir une seconde fois avec PhnI (Jouanneau et Meyer, 2006). Quoiqu'il en soit, la régiosélectivité de PhnI est clairement dépendante du substrat. Pour mieux comprendre cet aspect des propriétés catalytiques de l'enzyme, on peut maintenant s'appuyer sur le modèle structural décrit dans ce chapitre.

Des expériences de cinétique montrent que l'oxydation du naphtalène est étroitement couplée à celle du NADH (stoechiométrie 1 :1), alors que pour tous les autres substrats, notamment les HAP à 4-5 cycles, un certain découplage est observé. Cela nuit au rendement de la réaction, et c'est peut-être une des causes de la relative inefficacité des systèmes bactériens à dégrader les HAP de plus de 4 cycles. Ces observations attendent des explications qui seront sans doute apportées par une description plus détaillée du fonctionnement de la dioxygénase, et par l'étude structurale de PhnI.

| **Substrat** | **Position des OH** |
|---|---|
| **Naphtalène** | 1,2- |
| **Biphényle** | 2,3- |
| **Phénanthrène** | 3,4- |
| **Anthracène** | 1,2- |
| **Fluorène** | 9-, |
| **Fluoranthène** | 8-, 2,3- |
| **Benz[a]anthracène** | 1,2-, 8,9-, 10,11- |
| **Chrysène** | 3,4- |
| **Pyrène** | 4,5- |
| **Benzo[a]pyrène** | 9,10- |

**Table 1**. Position des hydroxyles dans les produits d'oxydation des HAP par PhnI.

### Cristallisation et détermination de la structure de la composante oxygénase

La préparation de PhnI purifiée selon notre protocole standard en trois étapes de chromatographie dont une étape de chromatographie d'affinité, est apparemment homogène d'après l'analyse SDS-PAGE. Cette protéine a été traitée à la thrombine pour éliminer une étiquette polyhistidine greffée en N-terminal de la sous-unité $\alpha$ pour faciliter

la purification. Cependant, l'analyse de la préparation par diffusion dynamique de la lumière a montré la présence de trois espèces, dont une pouvait correspondre à des agrégats de grande taille. Une étape supplémentaire de purification par chromatographie d'exclusion a donc été effectuée avant d'entreprendre les essais de cristallisation. Un criblage de plusieurs centaines de conditions de cristallisation a permis de sélectionner des conditions favorables qui ont ensuite été optimisées par étapes successives. Les meilleurs cristaux testés diffractaient à 1.85 Å de résolution.

Une tentative de détermination des phases expérimentales en utilisant le signal anomal des 9 atomes de fer présents dans la protéine hexamérique $\alpha_3\beta_3$ n'a pas abouti. Une préparation de PhnI dans laquelle les méthionines ont été substituées par des séléno-méthionines a été purifiée, puis cristallisée. Ces cristaux n'ont pas permis la résolution de la structure de PhnI, peut-être parce que la dose de rayonnement X appliqué a endommagé le cristal. La structure a finalement été obtenue par la méthode de remplacement moléculaire. La protéine PhnI montre une similarité de séquence relativement modeste avec les autres oxygénases bactériennes de structure connue. Cependant, le repliement de la chaine polypeptidique de sous-unités α et β ainsi que l'arrangement des sous-unités dans la structure hexamèrique de Phn1 ressemblent beaucoup à ceux décrits pour d'autre dioygénases. PhnI ressemble le plus à la naphtalène dioxygénase de *Pseudomonas* NCIB 9816-4 et à la nitrobenzène dioxygénase de *Comamonas* sp. JS765 tant par le degré de similitude de séquences que par la correspondance des structures (Table 2).

| | **PHN1** | **NDO-R** | **NDO-P** | **NBDO** | **CUDO** | **BPDO** | **OMO** | **CARDO** |
|---|---|---|---|---|---|---|---|---|
| **PHN1** | | **1.3** | **1.2** | **1.2** | **1.4** | **1.2** | **1.7** | **1.7** |
| **NDO-R** | *37.3* | | **1.3** | **1.3** | **1.1** | **1.2** | **1.7** | **none** |
| **NDO-P** | *40.8* | *33.8* | | **0.7** | **1.4** | **1.4** | **1.7** | **none** |
| **NBDO** | *41.4* | *32.1* | *80.0* | | **1.4** | **1.5** | **1.7** | **none** |
| **CUDO** | *37.9* | *36.8* | *32.1* | *31.1* | | **0.8** | **1.7** | **1.7** |
| **BPDO** | *39.3* | *38.0* | *31.6* | *31.2* | *68.7* | | **1.7** | **1.6** |
| **OMO** | *34.3* | *14.3* | *17.8* | *16.8* | *17.6* | *17.6* | | ***1.2*** |
| **CARDO** | *14.3* | *15.2* | *20.2* | *20.2* | *17.4* | *16.8* | *46.6* | |

**Table 2.** Identité des séquences et comparaison structurale des domaines catalytiques des dioxygenases bactériennes dont la structure est connue. Les identités de séquences (%) sont indiquées en italique (matrice utilisée sur le server CLUSTALW: Gonnet 250). Les écarts moyens de repliement des Cα (RMSD en Å) des domaines catalytiques sont indiqués en gras. Les chiffres surlignés en gris indiquent trois paires d'oxygénases structuralement très proches.

Tout comme ses homologues, le site actif est constitué du fer catalytique situé dans le domaine catalytique de la sous-unité α et du centre [2Fe-2S] du domaine Rieske de la sous-unité α voisine, les électrons fournis au centre [2Fe-2S] seraient transférés au fer catalytique par l'intermédiaire d'un acide aspartique totalement conservé (Asp 204 chez Phn1).

**Le site actif**

Les plus grandes différences sont observées au niveau de la cavité hydrophobe qui constitue le site catalytique de l'enzyme. La cavité ou poche catalytique de PhnI est plus volumineuse que celle de toutes les oxygénases caractérisés à ce jour, et c'est pour cette raison que l'enzyme peut oxyder des substrats de grande taille comme le benzo[a]pyrène (BaP). Nos simulations indiquent que le BaP peut en effet se loger dans cette poche selon une orientation unique. Dans cette position, les carbones 9 et 10 du BaP se trouvent au voisinage immédiat du fer catalytique, ce qui permet de prédire que le substrat se fera hydroxyler sur ces deux carbones. La prédiction est confirmée par l'expérience puisque le 9,10-dihydrodiol est le seul produit d'oxydation du BaP par PhnI. L'accès à la poche catalytique est contrôlé par deux boucles L1 et LII ; LI n'a pas pu être entièrement modélisée du fait de sa mobilité dans le cristal. La mobilité de la boucle L1 joue certainement un rôle important dans l'accessibilité du substrat au site actif. L'autre boucle L2 plus lointaine par rapport au site catalytique et moins flexible, jouerait un rôle secondaire dans ce contrôle de l'accès au site actif. Par des simulations numériques, nous avons aussi montré que le benz[a]anthracene (BaA), un HAP à quatre cycles, peut occuper la poche catalytique dans trois orientations différentes. Deux orientations sont plus favorables et prédisent une attaque de la molécule soit sur les carbones en positions 1,2-, soit en positions 10,11-, ce qui correspond aux deux dihydrodiols majeurs produits lors de la réaction catalysée par PhnI avec le BaA comme substrat (Fig.1)

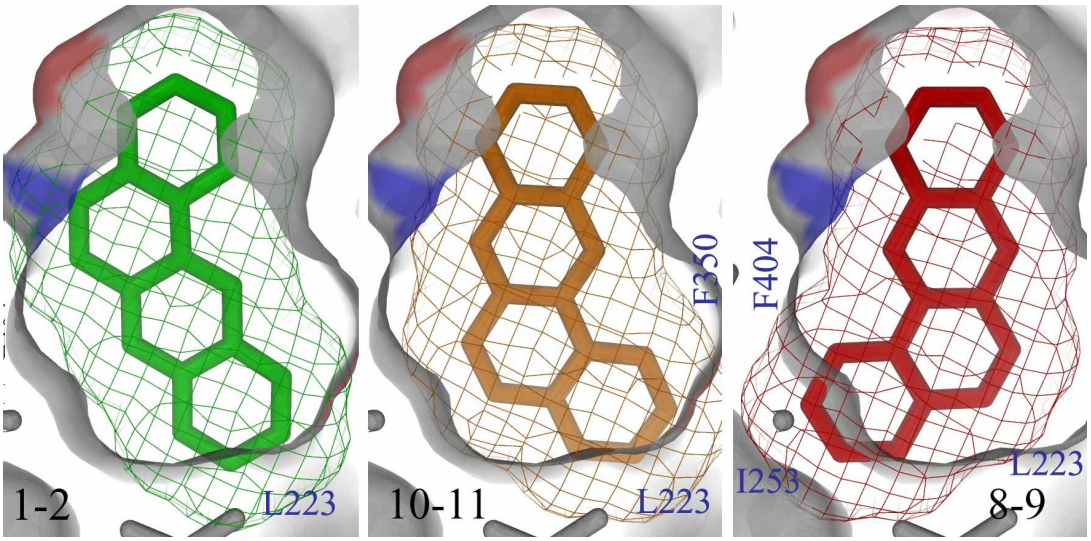



**Figure 1.** Modélisation des trois orientations les plus favorables d'une molécule de BaA dans le site actif de PhnI. La position prédite des hydroxyles sur le produit d'oxydation du BaA est indiquée dans le coin gauche en bas. La troisième orientation (8-9) est la moins probable en raison des interactions stériques du BaA avec les résidus Leu223, Ile251 et Phe 404. Les acides aminés les plus proches du BaA sont indiqués dans chacune des orientations.

La poche catalytique peut être décrite en distinguant trois régions définies en fonction de leur distance par rapport à l'atome de fer catalytique : les régions proximale, centrale et distale. Sur la base d'analyses de la géométrie de la poche et des simulations de complexes enzyme-substrat, on peut avancer quelques hypothèses sur le rôle des acides aminés qui bordent cette poche. Les résidus appartenant à la région proximale sont responsables de la réaction catalytique et participent probablement au positionnement du substrat. Ceux de la région centrale joueraient un rôle dans l'orientation du substrat et contribueraient donc à la regio-spécificité de l'enzyme. Le résidu Phe350 notamment jouerait certainement un rôle important dans l'orientation des HAPs dans la poche. Enfin, la région distale contrôlerait la dimension (longueur) de la poche et donc participerait à la spécificité de l'enzyme vis à vis du substrat.

Nous avons observé de la densité positive à proximité du fer dans deux poches catalytiques sur trois suggérant la présence d'une molécule ressemblant à l'indole. La présence fortuite de cette molécule est plausible puisque PhnI peut oxyder l'indole, ce qui se traduit par la formation d'indigo, qui colore en bleu les cultures de bactéries surproduisant PhnI.

Les trois poches catalytique présentes dans la structure affinée ne sont pas

totalement équivalentes. En effet, si l'une d'entre elles est plus complète et inclut les résidus Leu223 et Leu226, ces deux résidus ne sont pas visibles dans les deux autres en raison de la mobilité de la boucle LI. De même, la position des résidus Leu216 et Ile260 diffère dans chacune des trois poches, en raison de la mobilité des boucles LI et LII. Les positions et orientations des autres acides aminés bordant la poche sont par contre absolument conservées.

Les observations rappelées ci-dessus sont reproductibles. En effet, une seconde structure cristallographique déterminée et affinée à partir d'un second cristal présente des caractéristiques identiques. Une molécule ressemblant à l'indole est observée à proximité de deux des sites actifs, et la boucle LI est invisible dans deux des trois chaines polypeptidiques du domaine catalytique alors qu'elle est quasi complète dans la troisième.

## Perspectives

La purification de la dioxygénase de *Sphingomonas* CHY-1 sous forme recombinante a permis de caractériser cette enzyme aux plans catalytique et structural. La détermination de la structure 3D de PhnI, notamment celle de son site actif, apporte un nouvel éclairage sur le mode d'action de l'enzyme et les mécanismes de reconnaissance de ses substrats. En combinant les approches de biologie moléculaire pour manipuler la séquence protéique et les approches structurales, on peut d'une part étudier le mécanisme catalytique et la sélectivité de l'enzyme, et d'autre part, générer par mutagénèse ou évolution moléculaire dirigée des enzymes ayant de meilleures performances catalytiques.

### Bases moléculaires de la sélectivité des dioxygénases

En se basant sur la structure du site actif de PhnI, on a pu prédire par modélisation la position des HAP substrats dans la poche catalytique, et en déduire la position des carbones du substrat qui sont hydroxylés par la dioxygénase. Ces prédictions doivent être validées expérimentalement en analysant la structure de complexes enzyme-substrat. Le choix des substrats devrait se porter sur des HAP non linéaires comme le phenanthrène, le benz[a]antracène et le benzo[a]pyrène. Récemment, nous avons trouvé que la dioxygénase PhnI était active sur des alkylnaphtalènes, notamment des isomères mono-, di- et

triméthylés. Ces substrats pourraient aussi servir de molécules modèles pour explorer la sélectivité de PhnI, d'autant que leur meilleure solubilité dans l'eau par rapport aux HAP de 3 à 5 cycles pourrait faciliter l'obtention de complexes enzyme-substrat. Ces travaux devraient préciser le rôle de chacun des résidus bordant la poche de fixation du substrat, de mettre en évidence les déformations éventuelles du site catalytique sous l'effet du substrat.

L'importance des résidus de la poche catalytique pourrait être testée par mutagénèse dirigée dans le but :

- d'identifier ceux qui ont un rôle dans la regio- et la stéréospécificité de l'enzyme vis-à-vis de ses substrats.
- d'agrandir la poche pour améliorer l'activité de l'enzyme vis-à-vis des HAP les plus lentement oxydés par PhnI comme le chrysène et le benzo[a]pyrène.
- de faciliter l'accès au site actif en explorant le rôle des résidus faisant partie des boucles mobiles situées à l'entrée de la poche catalytique.

**Amélioration des performances catalytiques par ingénierie moléculaire**

L'augmentation de l'activité catalytique ainsi que l'élargissement de la sélectivité de l'enzyme peuvent aussi être obtenus par des techniques de mutagénèse aléatoire et d'évolution moléculaire dirigée. Les critères de sélection seront les suivants :

(i) l'augmentation de la vitesse de catalyse vis-à-vis de HAP comme le chrysène, le benzanthracène, ou le benzo[a]pyrène

(ii) la faculté d'oxyder des hydrocarbures toxiques et/ou très récalcitrants (HAP de 5 cycles, HAP alkylés, dioxines).

A cette fin, le gène codant la sous-unité alpha de PhnI feront l'objet de mutagenèse dirigée, aléatoire (Error prone PCR) et sera soumis à des expériences d'évolution moléculaire dirigée ou 'DNA shuffling'. Cette technique consiste à réaliser des recombinaisons multiples entre gènes homologues. L'étape clef de ces techniques d'ingénierie moléculaire reste la sélection des protéines mutantes. Pour cela, plusieurs méthodes de criblage de clones surproduisant les dioxygénases recombinantes sont envisagées. L'une d'elles consiste à faire réagir le produit de la deuxième étape

d'oxydation des HAP, un dihydroxy, avec le réactif de Gibbs qui donne un complexe bleu. Ceci suppose que les dioxygénases recombinantes soient exprimées dans une souche co-exprimant une deshydrogénase catalysant l'étape 2. L'enzyme qui catalyse cette réaction dans *Sphingomonas* CHY-1 a été caractérisée et son gène a été cloné (Jouanneau et Meyer, 2006). Dans ces conditions, le criblage pourrait se limiter à un test colorimétrique réalisable directement sur colonies ou bien en plaques de microtitration. Les enzymes recombinantes les plus performantes seront ensuite étudiées en détail aux plans biochimique et structural.

Pour mettre en évidence le bénéfice des améliorations catalytiques sur la biodégradation des HAP, on aura recours à une méthode basée sur la complémentation d'une souche mutante dérivée de CHY-1 dont on a inactivé les gènes codant PhnI (Demanèche *et al.*, 2004). Cette souche, appelée M10-1, est incapable de dégrader les HAP car elle a perdu l'enzyme qui catalyse l'attaque initiale. En y introduisant les gènes codant les dioxygénases recombinantes, on peut non seulement tester *in vivo* l'aptitude de ces enzymes à oxyder les HAP, mais aussi l'impact de l'amélioration des dioxygénases sur la dégradation des HAP dans son ensemble.

# CHAPITRE 3

## Utilisation des rayons X de très haute énergie pour la cristallographie des protéines

# Partie 1 : Introduction

## 1. La radiocristallographie des protéines

L'utilisation du rayonnement synchrotron est de nos jours la méthode standard et de loin la plus utilisée afin de déterminer la structure tridimensionnelle des protéines quand l'obtention de cristaux est possible. On appelle cette technique la radiocristallographie par diffraction des Rayons-X. L'explosion du nombre de structures de protéines connues n'a été permise que par le développement et l'amélioration des sources de rayons X, en particulier des synchrotrons, par l'utilisation de la cryo-cristallographie associée aux détecteurs à résolution spatiale, par le développement de la cristallogenèse des macromolécules, des méthodes/techniques permettant la détermination des phases cristallographiques et enfin par l'utilisation de ressources informatiques accélérant la détermination des structures de protéines.

### 1.1. Description d'une expérience de cristallographie

La proteine, une fois purifiée et caractérisée est soumise à un criblage initial de conditions favorables à l'obtention de cristaux (des centaines, voire des milliers de conditions peuvent être testées, en faisant varier les méthodes et la température). Cette étape est appelée en anglais le screening. Si l'une des conditions testées est favorable (cristal, micro cristal, ou amas de cristaux), un affinement des conditions est alors entrepris afin d'augmenter la taille du cristal et surtout sa qualité. Le cristal est ensuite congelé et stocké dans de l'azote liquide. L'étape de cristallisation est certainement de nos jours l'étape limitante. Les données de diffraction sont ensuite collectées en utilisant soit une source de R-X conventionnelle, soit le rayonnement synchrotron. Les clichés de diffraction sont enregistrés sur des détecteurs à résolution spatiale tels que les CCD (charge-coupled device) actuels. L'ensemble des intensités des taches de diffraction constitue le jeu de données mesurées.

La rotation du cristal dans le faisceau est la méthode la plus utilisée, et consiste simplement à exposer le cristal selon des angles croissants par incréments entre 0 et 360°. Les données sont analysées et traitées et le résultat final consiste à relier les amplitudes aux facteurs de structure du cristal. Les phases nécessaires à la reconstruction du modèle par une transformée de Fourier inverse peuvent être obtenues par le biais de différentes

techniques disponibles (remplacement moléculaire, MAD, SAD, SIRAS etc …). Les étapes expérimentales menant de la protéine purifiée à la structure sont illustrées dans la figure 1

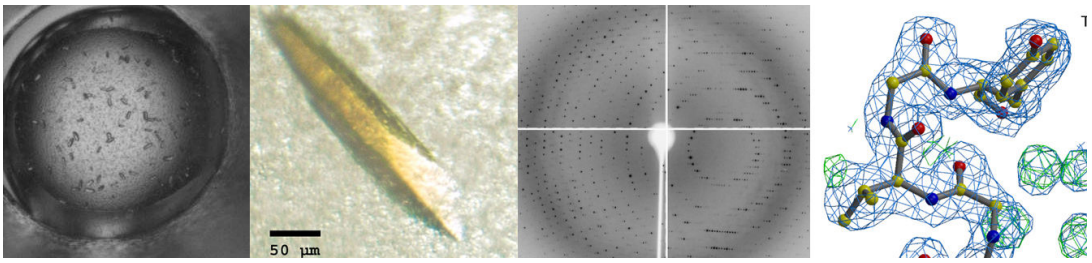



**Figure 1**. Etapes nécessaires à l'obtention de la structure tridimensionnelle d'une proteine. De gauche à droite, criblage (screening), affinement des conditions de cristallisation, cliché de diffraction et finalement carte de densité électronique avec le modèle construit.

### 1.2. Etat des lieux et avancées

Si toutes ces avancées ont permis l'obtention de dizaines de milliers de structures en à peine plus d'une décennie, la radiocristallographie par diffraction des rayons X n'en est pas moins toujours en constante évolution.

En effet, la maîtrise de certaines étapes permet leur automatisation. Lors de l'étape de cristallisation par exemple, le criblage des conditions favorables à l'obtention de cristaux peut être robotisé (Luft *et al*., 2000). L'obtention des phases expérimentales à partir des jeux de données cristallographiques peut être automatisée (Adams *et al*., 2004). Lors de l'expérience même, les étapes du montage et de l'alignement du cristal sur le diffractomètre permettant la collecte des jeux de données sont de nos jours complètement automatisées (Lesli *et al*., 2002, site internet de la ligne X6A). L'essor de la génomique structurale ainsi que la conception de nouveaux médicaments tirent complètement avantage de ces développements.

## 2. Les méthodes de phasage

L'avancée des sources de rayons X, et spécialement des lignes de lumière installées dans les synchrotrons, a entraîné l'élargissement des procédures utilisées afin d'obtenir les phases expérimentales du cristal d'intérêt. La méthode qui a le plus bénéficié de ces avancées est certainement la méthode MAD (Multiwavelength Anomalous Diffraction ou Dispersion), ou en français la dispersion anomale à plusieurs longueurs d'onde. Cette

méthode ne requière qu'un cristal en comparaison de la méthode MIR (Multiple Isomorphous Replacement) ou remplacement isomorphe multiple, exigeant de collecter des jeux de donnée de diffraction sur plusieurs cristaux d'une même protéine mais contenant différents atomes lourds. L'utilisation d'atomes lourds tels que l'uranium, le platine ou le mercure était très courante.

La méthode MAD tire avantage des caractéristiques même du faisceau synchrotron, en termes de modulation de l'énergie des rayons X et d'intensité/brillance du faisceau. La méthode MAD ne nécessite qu'un cristal et les données de diffraction sont collectées à différentes longueurs d'onde. La présence d'un atome lourd engendre des différences d'intensité du signal anomal aux longueurs d'onde utilisées, et en particulier à proximité d'un des seuils d'absorption de l'élément lourd. La contribution anomale (facteur de diffusion anomal) au facteur de diffusion atomique étant maximal au seuil d'absorption, l'intensité du signal anomal f " peut atteindre des valeurs de 3.7 et 10.2 $e^-$ pour les seuils K et LIII des éléments sélénium et mercure, respectivement. Ce sont ces différences qui sont mesurées par la technique MAD.

La technique MIR, quant à elle, exploite les différences d'amplitude produites par la présence des atomes lourds (un atome de mercure représente 80 électrons). La substitution des méthionines par des séléno-méthionines dans les protéines (Hendrickson *et al*., 1990) est l'approche la plus répandue qui permet d'obtenir les phases expérimentales par MAD à proximité du seuil K du sélénium ~ 12.5 keV (0.98 Å). Une variante de la méthode MAD, appelée SAD pour Single wavelength Anomalous Diffraction ou diffraction anomale à une seule longueur d'onde, tire parti des différences d'amplitude des paires de Bijvoet ($F^+$ et $F^-$) en présence de diffuseur anomal dans le cristal (Dauter, 2002). A l'heure actuelle, MAD et SAD sont les deux techniques les plus utilisées pour déterminer les structures 3D de nouvelles protéines.

Toutefois, elles requièrent plus de données et une haute précision des mesures étant donné que le signal à mesurer est faible (quelques $e^-$ seulement), par rapport à la méthode MIR où le signal mesuré correspond au nombre d'électrons de l'atome lourd.

Cette précision est statistiquement atteinte en augmentant la multiplicité des données, et cela est particulièrement critique dans le cas de faibles diffusions anomales, par exemple à des énergies éloignées des seuils d'absorption d'éléments légers. Dans le cas du soufre, élément naturellement présent dans la majorité des protéines, les données de diffraction ne peuvent pas être enregistrées à son seuil K (2.47 keV, 5 Å). Cette énergie est difficilement atteinte par les lignes de lumière synchrotron, mais cependant la méthode

SAD peut être appliquée sur des données obtenues à basse énergie (en général 5 à 7 keV), disponible sur les lignes de lumière standard de cristallographie des macro-molécules (MX).

A ce niveau d'énergie, le facteur de diffusion anomale, f '', du soufre représente environ 1 $e^-$. Selon l'équation du signal anomal décrite par Hendrickson (Hendrickson and Teeter, 1981), pour une protéine constituée de 1000 atomes (environ 130 acides aminés) et contenant 1 atome de soufre, 0.7 % de l'intensité mesurée provient de la diffusion anomale du soufre. Les phases expérimentales utilisant le signal anomal produit par les atomes de soufre ont été obtenues en utilisant le RX à 1.5 et 1.7 Å (Ramagopal *et al.*, 2003) ainsi qu'une source conventionnelle à 2.29 Å (Yang et *al.*, 2003).

Les méthodes MAD et SAD requièrent plus de données, c'est-à-dire un temps d'exposition relativement long du cristal au faisceau de diffraction, ce qui peut engendrer des dommages irréversibles (Rice *et al*., 2000). Ces dommages dus aux rayonnements produisent anisomorphisme, expansion des paramètres de la maille et parfois des effets spécifiques. Quand la méthode MAD échoue, la stratégie la plus courante consiste à collecter un seul jeu de données pour la méthode SAD. Ce jeu de données peut être enregistré à une énergie plus haute que celle du seuil d'absorption, où l'absorption par le cristal est moindre, mais le signal anomal reste exploitable ("high energy remote SAD phasing").

## 3. Les dommages dus aux rayonnements

Ce n'est pas le but de ce mémoire de passer en revue les causes et les effets des rayons X et les dommages qu'ils causent aux cristaux de protéine. Cette thèse ne traite pas de "Radiation Damage" mais propose une nouvelle méthode pour éviter les effets destructeurs des rayons X, basée sur l'utilisation des rayonnements à très haute énergie fournis par les synchrotrons. Dans ce qui suit, j'aborderai brièvement les mécanismes qui provoquent la détérioration des protéines par les rayons X. Pour une analyse plus approfondie, le lecteur se référera à des articles traitant en détail de cette question, parmi lesquels : Gonzales and Nave, 1994, Teng and Moffat, 2000, Weik *et al.,* 2000, O'Neill *et al*., 2002, Murray *et al*., 2004, Nanao and Ravelli, 2006.

### 3.1. L'interaction des R-X avec le cristal

Lors de l'interaction des rayons X (R-X) avec la matière, plusieurs processus peuvent survenir : les R-X peuvent soit diffuser de façon cohérente/élastique (Rayleigh), soit donner lieu à l'observation de la diffraction, de façon incohérente/inélastique (Compton), soit contribuer au bruit de fond (background), soit finalement être absorbés en transférant leur énergie (effet photoélectrique). Dans le cas d'un cristal de protéine de taille moyenne, environ 200 μm et à 12 keV, seulement 7 % des photons vont interagir avec l'échantillon et le reste le traverse sans aucun effet. 88 % des photons qui interagissent avec le cristal sont absorbés via l'effet photoélectrique, 6 % sont diffusés de façon incohérente et 6 % diffusent de façon cohérente et en phase donnent lieu à la diffraction observée (Sanchez del Rio and Dejus, 1998).

### 3.2. Dommages, dose et effets

Lors de l'absorption du photon incident par un cristal via l'effet photoélectrique, le transfert d'énergie s'effectue soit directement dans une molécule de protéine, soit dans une molécule de solvant auquel cas l'effet sera indirect. Ce processus conduit aux dommages primaires et est un phénomène inévitable. Lors de ces événements primaires, des radicaux libres sont générés et vont diffuser dans le cristal entrainant une cascade de réactions impliquant des radicaux secondaires qui vont causer d'autres types de dommages (Murray et Garman, 2002).

La mobilité des radicaux dépend de la température et est fortement réduite à 100 K, température à laquelle la majorité des expériences de cristallographie est exécutée, ce qui augmente le temps de vie du cristal. Cependant, depuis l'avènement des synchrotrons de 3$^{ème}$ génération produisant des faisceaux de R-X ultra intenses (un onduleur de 3$^{ème}$ génération est environ 1000 fois plus brillant qu'un aimant de courbure de seconde génération), l'atténuation des dommages à température basse n'est plus suffisante. Les dommages causés par les rayons dépendent de la dose d'énergie accumulée par le cristal (Sliz *et al*., 2003).

La dose dépend du temps d'exposition aux R-X, de la taille/composition du cristal/solution, de la taille/forme/intensité du faisceau et finalement de l'énergie des R-X. Le recours à des températures plus basses, 40 K et moins, a été proposée, mais aucune amélioration nette du temps de vie du cristal n'a été observée, car les dommages primaires

ne dépendent que de l'énergie absorbée par le cristal (Teng and Moffat, 2000). Henderson a estimé la dose à laquelle le cristal perd la moitié de son intensité initiale de diffraction ; cette dose appelée limite de Henderson est de $2.10^7$ Gray (J. $Kg^{-1}$) à 77 K.

Les dommages causés par les rayons X sont désormais mieux connus et se manifestent par une diminution de l'intensité de diffraction, l'augmentation du volume de la maille cristalline, l'augmentation de la mosaïcité, l'augmentation du facteur de température du cristal, la décarboxylation des acides aspartiques et glutamiques, la rupture progressive des ponts disulfure, la photo-réduction des métaux et des dommages spécifiques aux atomes lourds.

### 3.3. Les alternatives

Nous avons vu qu'une stratégie possible pour limiter la dose reçue par le cristal consiste finalement à collecter le maximum de données avant que la qualité du cristal ne soit trop altérée. Les paramètres influençant la dose reçue par le cristal sont en général connus et peuvent donner lieu à une estimation de la dose limite supportable (Murray *et al*., 2004). Cette information est utile pour optimiser les données recueillies selon l'énergie reçue par le cristal. Il est possible de jouer sur des paramètres qui limitent la dose de rayons tels que la composition de la solution de cristallisation, la taille du cristal, le flux incident et l'énergie du faisceau.

## 4. L'utilisation des hautes énergies

Les rayons à très haute énergie n'ont pour le moment pas été employés pour résoudre des structures de biomolécules, en partie parce qu'il n'existe pas d'installation dédiée à la cristallographie des protéines délivrant des rayons X avec une énergie supérieure à 35 keV. De plus, les détecteurs actuels, très efficaces à 12 keV (énergie moyenne couramment utilisée) ont un rendement inférieur à haute énergie. Comme nous l'avons vu plus tôt, la dose dépend de l'énergie à laquelle les données sont collectées, ce par l'intermédiaire du coefficient de masse énergie-absorption, environ 50 fois plus élevé à moyenne énergie comparée à 55 keV. C'est cette constatation qui nous a conduit à expérimenter une méthode basée sur les rayons X à très haute énergie, bien que dans ce cas, l'intensité de la diffraction, proportionnelle à la section efficace différentielle de la diffusion cohérente, soit plus faible à très haute énergie.

## *Présentation des travaux expérimentaux*

Nous l'avons vu, la détermination de la structure tridimensionnelle d'une nouvelle proteine passe par l'étape nécessaire du phasage, cette étape requière une quantité de données conséquente. Afin de minimiser les dommages dus aux rayonnements, l'utilisation des très hautes énergies est envisagée pour la première fois. Nous nous intéressons ici à la détermination de la structure d'une proteine modèle, le lysozyme de poule.

Dans un premier temps, la proteine modèle a été cristallisée en présence d'un élément présentant un seuil d'absorption à une énergie supérieure à 50 keV. Etant donné qu'aucune installation de cristallographie des protéines n'est capable de délivrer des R-X de plus de 35 keV, les expériences ont été exécutées dans des installations dédiées à science des matériaux. Ces installations délivrent des R-X de très haute à ultra haute énergie, de 50 à environ 200 keV avec une intensité maximale à environ 55 keV ; l'élément lourd de choix est donc holmium avec son seuil K à 55.6 keV. L'enjeu était de savoir si les données recueillies à très hautes énergies étaient exploitables pour résoudre la structure des macromolécules par cristallographie et, en particulier, de réussir à calculer les phases expérimentales nécessaires à la construction du modèle structural.

Dans un deuxième temps, étant donné que l'un des objectifs de ce travail de recherche était de développer une méthode d'analyse structurale des protéines permettant d'éviter les dommages dus aux radiations, les données collectées à 55 keV sont comparées à celles obtenues dans conditions similaires à 12 keV en termes de qualité de diffraction du cristal et d'altération du modèle final.

Ce travail est présenté dans deux articles, le premier décrivant le système expérimental et les résultats de phasage, est publé dans la revue Journal of Applied Crystallography (volume 39, pages 831-841, 2006). Le second traite de la visualisation des dommages dus aux radiations à 55 keV et à 12 keV. Les résultats préliminaires ont été présentés lors du meeting de l'ACA (American Crystallography Association, Hawaii, 2006) et ont été récompensés par le prix Margaret C. Etter Student Lecturer. L'article est actuellement en cours de soummission.

## Références bibliographiques


Adams, P.D., Gopal, K., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Pai, R.K., Read, R.J., Romo, T.D., Sacchettini, J.C., Sauter, N.K., Storoni, L.C. and Terwilliger, T.C. (2004). Recent developments in the *PHENIX* software for automated crystallographic structure determination. *J. Synchrotron Rad*. **11**: 53-55.

Dauter, Z. (2002). One-and-a-half wavelength approach. *Acta Cryst*. **D58** : 1958-1967.

Henderson, R. (1990). Cryo-Protection of Protein Crystals Against Radiation Damage in Electron and X-ray Diffraction. *Proc. R. Soc. London.* **B241**, 6-8.

Hendrickson, W.A. and Teeter, M.M. (1981). Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature (London)*. **290**: 107-113.

Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by Multiwavelength Anomalous Diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J*. **9**: 1665-1672.

Gonzalez, A., and Nave, C. (1994). Radiation Damage in Protein Crystals at Low Temperature. *Acta. Cryst*. **D50**: 874-877.

Leslie, A.G.W., Powell, H.R., Winter, G., Svensson, O., Spruce, D., McSweeney, S., Love, D., Kinder, S., Duke, E. and Nave, C. (2002). Automation of the collection and processing of X-ray diffraction data - a generic approach. *Acta. Cryst*. **D58**: 1924-1928.

Luft, J.R., Wolfley, J., Collins, R., Bianca, B., Weeks, D., Jurisica, I., Rogers, P., Glasgow, J., Fortier, S. and DeTitta, G.T. (2000). High Throughput Macromolecular Crystallization for Structural Genomics. Proceedings of NATO ASI 30th Course - Methods in Macromolecular Crystallography, "Ettore Majorana" Centre for Scientific Culture, Erice, Italy.

Murray, J. and Garman, E. (2002). Investigation of possible free-radical scavengers and metrics for radiation damage in protein cryocrystallography. *J. Synchrotron Rad*. **9**: 347-354.

Murray, J.W., Garman, E.F. and Ravelli, R.B.G. (2004). X-ray absorption by macromolecular crystals: the effects of wavelength and crystal composition on absorbed dose. *J. Appl. Cryst*. **37:** 513-522**.**

Nanao, M.H. and Ravelli, R.G.B. (2006). Phasing macromolecular structures with UV-induced structural changes. *Structure* **14**:791-800.

O'Neill, P., Stevens, D.L. and Garman, E. (2002). Physical and chemical considerations of damage induced in protein crystals by synchrotron radiation: a radiation chemical perspective. *J. Synchrotron Rad.* **9**: 329-332.

Ramagopal, U.A., Dauter, M. and Dauter, Z. (2003). Phasing on anomalous signal of sulfurs: what is the limit? *Acta. Cryst*. **D59**: 1020-1027.

Rice, L.M., Earnest, T.N. and Brunger, A.T. (2000). Single-wavelength anomalous diffraction phasing revisited. *Acta. Cryst*. **D56**: 1413-1420.

Sánchez del Río, M. and Dejus R.J. (1998). XOP: Recent developments. SPIE proceedings 3448.

Sliz, P., Harrison, S.C. and Rosenbaum, G. (2003). How does radiation damage in protein crystals

depend on X-ray dose? *Structure*. **11**: 13-19.

Teng, T.-Y. and Moffat, K. (2000). Primary Radiation Damage of Protein Crystals by an Intense Synchrotron X-ray Beam. *J. Synchrotron Rad*. **7**: 313-317.

Weiss, M.S., Panjikar, S., Mueller-Dickmann, C. and Tucker, P.A. (2005). On the influence of the incident photon energy on the radiation damage in crystalline biological samples. *J. Synchrotron. Rad*. **12**:304-309.

Weik, M., Ravelli, R.B.G., Kryger,G., McSweeney, S., Raves, M.L., Harel, M., Gros, P., Silman, I., Kroon, J. and Sussman, J. (2000). Specific Chemical and Structural Damage to Proteins Produce by Synchrotron Radiation. *Proc. Nat. Acad. Sci. USA*. **97**: 623-628.

Yang, C., Pflugrath, J.W., Courville, D.A., Stence, C.N. and Ferrara, J. D. (2003). Away from the edge: SAD phasing from the sulfur anomalous signal measured in-house with chromium radiation. *Acta. Cryst*. **D59**: 1943-1957.

# Partie 2 : Article 1

## Anomalous Diffraction at Ultra High Energy for Protein Crystallography

Cet article dédié à la premiere experience de cristallographie des macromolécules (MX) à une enregie supérieure à 50 keV est paru en 2006 dans la revue *Journal of Applied Crystallography* (volume 39, pages 831-841). Ces travaux ont été effectués en collaboration avec Zhong Zhong (NSLS) et le groupe de Veijo Honkimaki de l'ESRF. Dans cette experience les données n'ont pas été obtenues dans les meilleures conditions du fait de l'indidponibilité de ligne de lumiere dediée à la MX. De plus, nous décrivons la premiere structure SAD à haute energie, d'autres methodes ont été utilisees avec succès.

# Anomalous Diffraction at Ultra High Energy for Protein Crystallography

**Jean, Jakoncic[a,b], Marco Di Michiel[c], Zhong Zhong[a], Veijo Honkimaki[c], Yves Jouanneau[d] and Vivian Stojanoff[a*]**

[a]Brookhaven National Laboratory, National Synchrotron Light Source, Upton, NY, USA, [b]Universite Joseph Fourrier, Grenoble, France, [c]European Synchrotron Radiation Facility, Grenoble, France and [d]CNRS CEA, Grenoble, France.

**Synopsis**: Ultra high energy X-ray phasing is discussed as a tool for macromolecular crystallography.

## Abstract

Singlewavelength Anomalous Diffraction (SAD), Multiwavelength Anomalous Diffraction (MAD) and Single Isomorphous Replacement with Anomalous Scattering (SIRAS) phasing at ultra high X-ray energy, 55 keV, are used to successfully determine a high-quality and high-resolution experimental electronic density map of Hen Egg White Lysozyme, a model protein. Several combinations between single and three wavelengths with native data was exploited to demonstrate that standard phasing procedures with standard equipment and software can successfully be applied to the 3D crystal structure determination of a macromolecule, even at these very short wavelengths. For the first time a high quality 3D molecular structure is reported from SAD phasing with ultra high energy X-rays. The quality of crystallographic data and experimental electron density maps meet current standards. The 2.7 % anomalous signal from three Ho atoms, at the Ho K-edge, was sufficient to obtain a remarkable electron density and build the first lanthanide structure for HEWL in its entirety.

**Key words**: Ultra High Energy; Phasing; MAD; SAD; SIRAS; HEWL; Holmium

## 1. Introduction

To solve a de-novo protein structure, multiwavelength anomalous diffraction (MAD) (Fourme & Hendrickson, 1990, Hendrickson, 1991) and singlewavelength anomalous diffraction (SAD) (Hendrickson & Teeter, 1981; Wang, 1985) exploits the anomalous scattering phenomena (Cassetta *et al*., 1999; Fourme *et al*., 1999; Hendrickson, 1999; Cianci *et al.*, 2005). Although the use of anomalous centers either naturally occurring in metalloproteins, such as Fe and Mn, or introduced by replacement of light metal centers, ligands or co-factors by lanthanides or organic labels containing Se and Br, has been demonstrated in the early days of synchrotron radiation facilities, the real potential of the MAD and SAD methods only became available in the last decade with significant advances in instrumentation and analytical and computational procedures. One of the major contributions to the actual wide spread acceptance of the MAD method was the introduction of cryogenic protection of crystals allowing for extended crystal lifetime.

With the advent of 3rd generation synchrotrons cryogenic protection of crystals was shown to be insufficient to prevent crystal deterioration due to radiation damage. In fact radiation damage appears to be the major cause of unsuccessful phasing in third generation synchrotrons (Murray & Garman, 2002). To improve protein crystal lifetime in terms of radiation damage at these facilities, experimental procedures and strategies such as, the introduction of scavengers, alternate data collection strategies and development of new phasing procedures are being developed. O'Neil et al. (2002) suggest that the use of glycerol or ascorbic acid as cryo-protectants prevents the formation of free radicals responsible for secondary damages such as S-S breakage and decarboxylation. Alternatively Dauter (2002) suggested the collection of highly redundant single wavelength (SAD) data set while Peterson et al. (1996), Gonzalez (2003) suggested the collection of two wavelength MAD, at the inflection and high energy remote, instead of a three wavelength MAD. Another way to handle the radiation damage issue is to mathematically treat the measured intensities implementing an exponential decay function. Such correction was implemented in SHARP (Schiltz *et al.*, 2004) and XDS (Diederichs, 2006). In the SHARP case, intensities are corrected taking into account the scatterer occupancy and B factor decay improving experimental phases only in the presence of site-specific radiation damage. Recently, Ravelli et al. (2003) introduced a new phasing method, Radiation Damage Induced Phasing (RIP). Experimental phases are determined

from site-specific damage such as disulphide bond breakage or radiolytic dehalogenation of aromatic residues (Banumathi *et al*., 2004; Zwart *et al*., 2004). Ramagopal *et al*. (2005) further developed this method into Radiation Damage Induced Phasing with Anomalous Scattering (RIPAS), combining the anomalous signal of a Hg derivative and the radiation sensitivity of Hg-S bonds at cysteine residues, to successfully determine an unknown structure.

Advantages of short X-ray wavelength (less then 0.33 Å), were first discussed by Helliwell & Fourme (1983) and led to the proposal to extend the wavelength range for protein crystallography beam lines at the European Synchrotron Radiation Facility (ESRF). The report further discussed the likely impact of the high intensity beam on sample lifetime and the need of cryo protection to avoid radiation damage. Arndt (1984) further discussed the impact of lower absorbed dose and the reduced need for absorption corrections at shorter wavelengths. Although the arguments presented in these reports were reiterated later in Helliwell (1992) and reviewed in Helliwell et al. (1993) only a few experiments have been reported at ~35 keV, all at 3rd generation synchrotrons. Single isomorphous replacement with anomalous scattering (SIRAS) at the xenon K-edge was employed successfully by Schiltz and co-workers (Schiltz *et al.*, 1997) to determine the known structure of porcine pancreatic elastase. Recently, Takeda et al (2004) employed MAD phasing at the Xe and I K-edge of hen egg white lysozyme derivatives. High pressure experiments developed by Fourme's group (Fourme *et al.*, 2001, 2003; Girard *et al*., 2004) were performed at the iodine K-edge 37.17 keV, to minimize scattering from the Diamond anvil cell, especially at high Bragg angle, where the Compton energy shift is larger. Possible reasons for the limited use of ultra-high energy X-rays can be the fact that this region is not readily accessible at conventional structural biology beam lines due to technical limitations of the instrumentation. On the other hand the energy range, 5-20 keV, usually available at synchrotron beam lines, allows for the choice of a wide range of anomalous scatters which include the LIII edges of most lanthanides. From this perspective there would be no advantage to collect at ultra-high X-ray energies at the K-edges of these elements as the anomalous power would be similar to Se and Br K-edges.

However as shown in Fig. 1 the total linear mass absorption coefficient calculated for a generic protein sample (with average protein composition) is about 5 $cm^2g^{-1}$ compared to 0.2 $cm^2g^{-1}$ at higher energies leading to potential lower cumulated absorbed dose, in other words less radiation damage at high X-ray energies (Aslantas *et al.*, 2006). This could be of significant advantage in the molecular structure determination of radiation sensitive biological molecules.

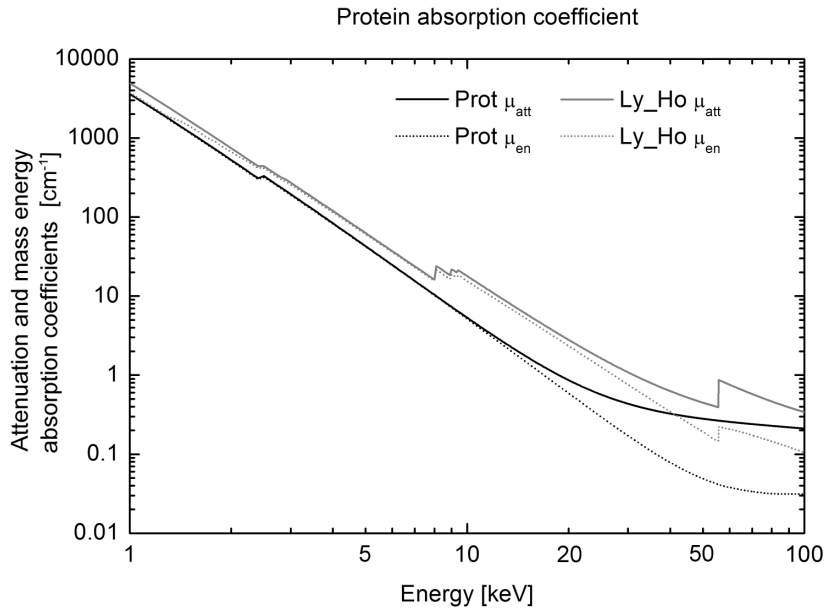



**Figure 1.** Attenuation coefficient ($\mu_{att}$) and energy-absorption coefficient ($\mu_{en}$) for a generic protein crystal sample (Prot) and for a holmium hen egg-white lysozyme (Ly_Ho) derivative (Di Michiel, personnal communication).

To further probe the advantage of ultra-high X-ray energies in structural biology the report below exploits the anomalous signal from Ho scatterers in tetragonal hen egg white lysozyme (HEWL) crystals at 55.6 keV. Experiments were carried out on two high energy beam lines dedicated to the material science community, ID15B at the European Synchrotron Radiation Facility, a third generation machine, and X17B1 at the National Synchrotron Light Source, a second generation facility. Phases for the two experiments were successfully calculated by three separate methods, MAD, SAD and SIRAS, to yield high quality experimental electron density maps for modeling. For the first time a high quality 3D molecular structure is reported from SAD phasing at ultra high X-ray energy. Present results are discussed in light of radiation sensitivity and phasing power, further demonstrating the usefulness of ultra high X-ray energy crystallography.

## 2. Material and Methods

### 2.1. Sample Preparation

Hen egg-white lysozyme (three times crystallized) from Sigma was used without further purification. Crystals were freshly grown 3 to 4 days prior to data collection, at room temperature, by the micro-batch method under paraffin oil (Chayen *et al.,* 2000). For the phasing experiments at ultra high X-ray energies crystals were grown in the presence of holmium. Typically, 2 µL of 100 mg/mL hen egg-white lysozyme in 50 mM sodium acetate pH 4.5 and 50 % V/V Glycerol was mixed with 2 µL of 0.4 M HoCl3 + 1.2 M NaCl in 100 mM sodium acetate pH 4.5; no further cryo-protection was needed. Crystals appeared in a few hours and reached a size of 400 x 400 x 300 µm3 within 2-3 days. Crystals used for the native data collection were grown under the same conditions in the absence of HoCl3.

### 2.2. High Energy X-ray Beam lines

None of the dedicated macromolecular crystallography beam lines are capable to deliver X-rays with energies higher than 35 keV. Therefore experiments were performed at beam lines usually serving the material science community, ID15B, at the European Synchrotron Radiation Facility in Grenoble, France (ESRF) and X17B1 at the National Synchrotron Light Source, Upton, NY, USA. Both beam lines deliver either monochromatic beam or white beam in the high energy range, typically higher than 50 keV, and can accommodate a wide range of experiments; special equipment including cryostream and an area detector can easily be implemented.

The experiment at ID15B (ESRF) was performed during the 2/3 fill mode (200 mA electron beam current); the X-ray beam was provided by an asymmetrical multipole wiggler and focused by a Bragg Silicon (311) crystal monochromator (with energy resolution $dE / E = 5 \times 10^{-4}$). The NSLS high energy beam line, X17B1, is located on a superconducting wiggler. A Si (311) sagitally bent double crystal Laue monochromator (Zhong *et al*., 2001) focuses the beam horizontally on to the sample; vertical beam size is typically of the order of 500 µm. ID15B, located on a 3rd generation light source,

generates typically a beam 200 times brighter than X17B1.

The strong secondary air scattering observed for the first experiments performed at both beam lines made it difficult to accurately and completely measure any anomalous difference from the samples. To minimize these effects in a second experiment and allow the accurate measurement of the anomalous signal the initial beam stop arrangement was reviewed. On ID15B, for the first experiment, a 6 mm in diameter and 25 mm long lead beam stop was positioned on a 0.5 mm aluminum plate right in front of the detector. The sole function of this aluminum plate was to hold the beam stop; absorption is of the order of 3.5 %. The diffraction patterns recorded during this first attempt all comprised a strong background that accounted for 7000 counts in the forward direction and dropped to 1000 counts in the high-resolution range. This background was tremendously reduced after the addition of a second beam stop, 2 mm in diameter and 4 mm long also made out of lead and held on a 0.5 mm Al plate at 1/3 of the sample-detector distance. The combination of both beamstops allowed for the absorption of the direct beam suppressing secondary air scattering of low energy X-rays from the primary beam stop and the 3rd harmonic as well as the remaining beam. With this arrangement the background was 35-fold and 12-fold reduced in the low and high-resolution range, respectively. On X17B1, the initial Pb beam stop, 6.4 mm in diameter and 15 mm long mounted on a 3.2 mm thick Plexiglas plate was positioned 10 cm from the sample, the minimum distance that still allowed sample manipulation given the beam line configuration. The Plexiglas plate absorbed 7 % of the diffracted radiation. Because of its size the beam stop was easy to align and totally absorbed the direct beam, but it did not allow for collection of low-resolution data, beyond 12 Å. This lack of low-resolution data did not allow for proper phasing after the substructure solution was achieved. For the second attempt, the beam stop diameter was halved and tantalum was used for its high stopping power and machining properties.

These changes allowed data to be recorded to at least 20 Å in the lowest resolution range, which ultimately granted the first ultra high energy SAD phasing. Likewise, as it is usual for low energy experiments, the sample was as close as possible to the collimating slits to diminish any scattering upstream to the crystal; beam size and crystal size are shown in Table 1.

| | Source | Flux* ph.s$^{-1}$ 0.1% bw | Beam Size µm$^2$ | Crystal Size µm$^3$ | Energy Resolution dE/E** |
|---|---|---|---|---|---|
| **ID15B** | Wiggler | $10^{12}$ ($3x10^{13}$) | 300x300 | 400x400x300 | $5x10^{-4}$ |
| **X17B1** | Wiggler | $4x10^{9}$ ($10^{11}$) | 300x300 | 400x400x300 | $2x10^{-4}$ |

*Numbers in parentheses refer to total integrated flux calculated with slits wide open.
**White beam slits: 500 µm

**Table 1**. Beam line parameters. The flux is estimated at the sample position, with the used beam-size at 55 keV, at 100 mA for ID15B and at 200 mA for X17B1.

### 2.3. Detectors

Two detectors were used for data collection; at ID15B the MAR Research 345 mm Image Plate (MAR345) and at X17B1 the 165 mm MAR Research CCD (MARCCD) detector. These detectors are routinely utilized at medium energies and are not optimized for high X-ray energies. Both use a phosphor screen to image X-rays but employ different readout techniques. The MARCCD directly converts X-rays to photons in the visible spectra that are detected by a CCD. In comparison, the MAR345 uses a phosphor screen that is excited by X-rays and readout by a laser. The laser reading efficiency is significantly reduced at these high energies as it is depth dependant. Nevertheless, a naive method to characterize the detectors is through their phosphor screen absorption efficiency. The absorption efficiency of both detectors is represented in Fig. 2 for the energy range of 1-100 keV. For the MAR345 a 207 µm thick phosphor screen made from BaFBr:$Eu^{2+}$ (density, 2.86 g.cm$^{-3}$) was considered while for the MARCCD a 45 µm thick phosphor screen made from $Gd_2O_2S$:Tb (density, 2.9 g.cm$^{-3}$) was considered. If both detectors display efficiency greater than 80 % at 12 keV, the energy commonly used for macromolecular crystallography, at 55 keV where this experiment was performed, the MAR345 absorbs 40 % of the photons and the MARCCD 20 %. In principle then the MAR345 would be a better choice for high energy measurements as it presents a better efficiency, better dynamic range and larger size.

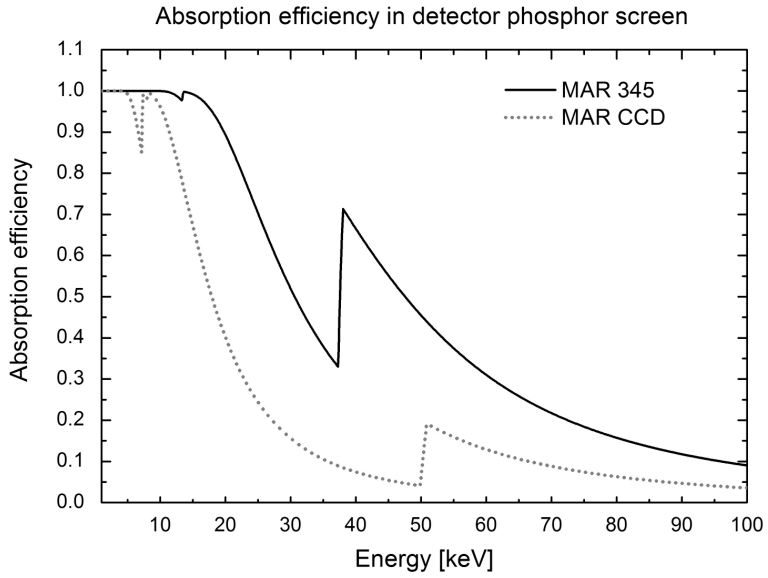



**Figure 2** Phosphor screen absorption efficiency for the MAR345 image plate scanner and MAR CCD detector in the 1-100 keV range. For further discussion see text (Sanchez del Rio and Dejus, 1998).

### 2.4. Energy Scan

For both beam lines the energy was calibrated at the holmium K-edge (55.618 keV) with the crystallization stock solution containing holmium chloride. The three wavelengths selected for the MAD experiment, 0.2229 Å, 0.2227 Å and 0.2200 Å (55.62 keV, 55.68 keV and 56.34 keV) inflection, peak and high remote, respectively, were determined from the absorption spectra measured by the fluorescence method on the solution using an ion chamber. Fig. 3 displays the variation of scattering factors f '' and f ' with energy according to the absorption spectrum recorded on the stock solution at ID15B. The doted arrows show the energies at which the data was actually collected; the plain arrows correspond to the peak and inflection energies if the energies determined by CHOOCH (Evans & Pettiffer, 2001) would have been adopted.

The crystals were directly recovered from the crystallization drop and mounted in a random orientation on the goniometric head. The sample was maintained at a temperature of 100 K by a nitrogen cold stream (Oxford Cryosystem serie 600) throughout the experiment. Data was recorded using the rotation method. On ID15B the sample to detector distance was chosen to allow for data collection up to 0.91 Å resolution at the boarder of the image plate. 180 frames, corresponding to a total of 180 degrees in oscillation range were recorded at the peak energy, followed by 360 frames at the high energy remote and finally 180 at the inflection energy, having in mind the MAD phasing procedure and the maximization of the dispersive signal at the inflection energy in case of

radiation damage. A single wavelength experiment at the peak energy position was carried out on a second crystal for experimental redundancy; this data is not shown. On X17B1 the MARCCD detector was placed at a distance of 400 mm from the sample. The same data collection strategy was followed except that 120 degrees were collected at the inflection energy due to beam time limitations. Exposure time per frame was 4 s on ID15B while 100 s per degree were required at X17B1. Details of the various data collection runs are given in Table 2. An estimate of the absorbed dose for each experiment is presented in Appendix A. In both cases, an additional data set was collected on a native Lysozyme crystal at 0.2227 Å wavelength for phasing purposes.

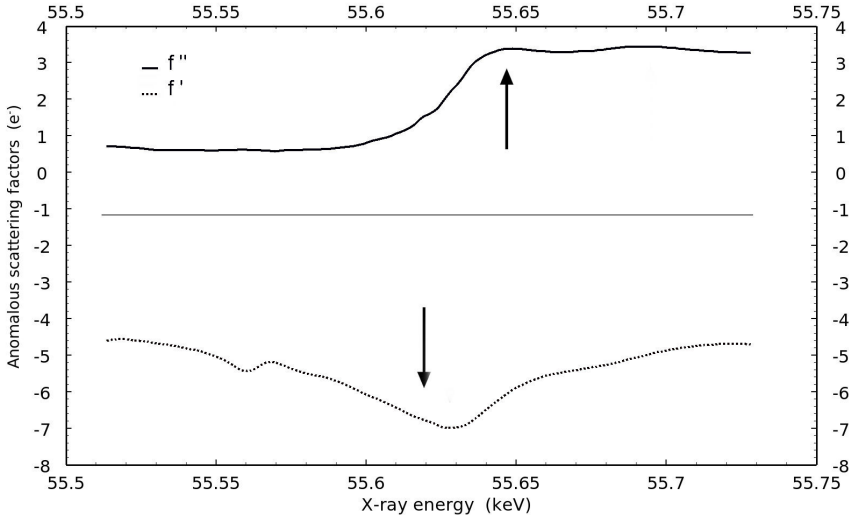



**Figure 3**. Fluorescence energy scan recorded on the holmium chloride stock solution used for crystallization. The spectra was normalized with Io and processed with CHOOCH (Evans & Pettiffer, 2001).The plain arrows indicate the actual energies selected for data collection, 55.62 (0.2229 Å), 55.68 (0.2227 Å); not shown in the figure the energy used for the remote data collection [56.34 (0.22 Å) keV].

### 2.5. Data Processing, Phasing and Refinement

The recorded intensities were integrated, scaled and merged with DENZO/SCALEPACK suite (Otwinowski & Minor, 1997) and were converted into amplitude for further analysis using TRUNCATE (French & Wilson, 1978). This data was used in the determination of experimental phases by three different methods: SIRAS, MAD and SAD. Experimental phases were determined from several combinations between single and three wavelengths with and without consideration of native data. In these studies the SHELX package (Sheldrick & Gould, 1995) was used to locate the heavy atom positions by the powerful dual space recycling principle as implemented in

SHELXD followed by a few density modification cycles with SHELXE. The successful use of other software suites to either just locate the Ho sites or for phasing, such as, Crank (Ness *et al*., 2004) and the CCP4 suite (Collaborative Computational Project, Number 4, 1994) used for SAD phasing at the high energy remote (56.34 keV) and at the peak (55.68 keV), or Hyss (Grosse-Kunstleve & Adams, 2003; Phenix collaboration) used to locate the Ho sites for the three wavelength MAD data and SAD data, further demonstrate the applicability of standard phasing procedures (software) to ultra high X-ray energy data. Further density improvement was performed using DM (Cowtan, 1994) with a 40% solvent content and calculated Matthews's coefficient of 2.06 $Å^3$/Dalton (Matthews, 1968). This modified density served as input to ARPwARP (Perrakis *et al.*, 1999) for automated chain tracing. . The refinement was performed with Refmac (Murshudov *et al.,* 1997). Electron density maps and refined models were inspected with COOT (Emsley & Cowtan, 2004), which also allowed manual building and manual reconstruction of the missing parts.

## 3. Results and Discussion

### 3.1. Crystals and Diffraction Quality

Although previous experiments (Gonzales et al., 1994; Schiltz et al., 1997; Takeda *et al.*, 2004), performed at macromolecular crystallography (MX) beam lines with a fixed and pre-aligned beam stop indicated that background noise did not seriously affected data collection and processing, initial experiments carried out on both beam lines, ID15B and X17B1, did not allow the determination of the molecular structure with any of the phasing methods being considered. In spite of the reasonable diffraction quality of the images and scaling results it was possible to find only two out of the three Ho sites, following the procedure described in section 2.6. However, it was impossible to obtain an interpretable experimental electron density map much less obtain a model for the molecular structure. To improve the signal-to-background ratio and the statistics an additional beam stop was installed on X15B and a smaller beam stop was used on X17B1. The use of this additional beam stop on ID15B led to the reduction of the background by a factor of 10. On X17B1 the use of the smaller beam stop allowed the collection of data in the low resolution shells. These two improved beam stop configurations ultimately allowed for the first structure of a protein crystal at energy higher than 40 keV using current experimental phasing

methods. The data collection statistics presented in Table 2 is the result of a second set of measurements obtained with these improved beam stop configurations.

All the crystals measured belong to the tetragonal space group P43212 with unit cell parameters of the order of a = b = 78 c = 37 Å and exhibit rather low mosaicities. Even if crystals diffracted well beyond 1.3 Å, the diffraction limits reported are a good compromise between exposure time, detector readout and allocated beam time. Fig. 4 displays a section of a typical diffraction pattern obtained at both beam lines.

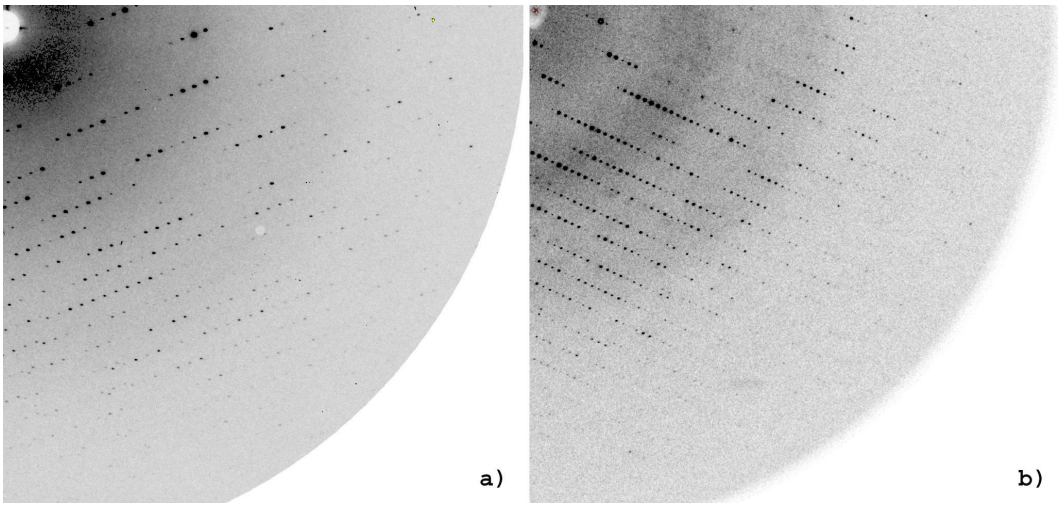



**Figure 4**. Typical diffraction pattern recorded on a: a) MAR CCD 165mm detector at the X17B1 beam line: sample to detector distance 400 mm and wavelength 0.2153 Å; the resolution at the edge is 1.35 Å. b) MAR 345 scanner at the ID15B beam line: sample to detector distance of 850 mm and wavelength 0.219 Å; the resolution at the edge is 1.15 Å.

Data quality, as judged from the statistics (resolution, mosaicity, I/σI, completeness and Rmerge), meets current standards. The values presented in Table 2 are in good agreement with recentlyreported values for HEWL structures determined to similar resolutions on standard MX beam lines at medium energy range (for example, PDB access code 2C8O and 1QIO). The higher Rmerge values observed here are attributed to systematic and experimental errors and might be expected as this experiment was carried out on non-dedicated beam lines. To verify the quality of the data a complete data set was collected at medium energy (12 keV) on a standard MX beam line (Allaire *et al*., 2003) on a crystal grown under the same conditions and of the same size. Similar diffraction resolution with a 4 s exposure per oscillation resulted in similar statistics with an Rmerge of 4.5%, compatible with the values reported in the PDB for the structures mentioned above.

| | ID15B | | | | X17B1 | | | |
|---|---|---|---|---|---|---|---|---|
| **HEWL** | Native | Ho Derivative | | | Native | Ho Derivative | | |
| **Data** | Native | Peak | Remote | Inflection | Native | Peak | Remote | Inflection |
| **Wavelength (Å)** | 0.2227 | 0.2229 | 0.2201 | 0.2227 | 0.2227 | 0.2229 | 0.2201 | 0.2227 |
| **Energy (keV)** | | 55.68 | 56.34 | 55.62 | | 55.68 | 56.34 | 55.62 |
| **Unit Cell (Å) A=b, c** | 78.5 37.0 | 78.5 37.4 | 78.5 37.4 | 78.4 37.4 | 79.2 37.0 | 78.7 37.3 | 78.7 37.3 | 78.6 37.2 |
| **Resolution Limits (Å)** | 25.00-1.40 (1.45-1.40) | 20.00-1.25 (1.29-1.25) | 22.00-1.25 (1.29-1.25) | 20.00-1.25 (1.29-1.25) | 30.0-1.35 (1.40-1.35) | 30.0-1.35 (1.40-1.35) | 30.0-1.35 (1.40-1.35) | 30.0-1.45 (1.50-1.45) |
| **Completeness (%)** | 99.9 (100) | 98.9 (100) | 99.2 (100) | 99.0 (100) | 97.5(97.3) | 99.2(92.4) | 100 (100) | 100 (99.8) |
| **I/σ(I)** | 29.1(5.00) | 21.0 (3.5) | 26.2 (4.3) | 21.0 (3.5) | 14.9 (2.7) | 24.9 (2.1) | 35.4 (3.7) | 19.2 (2.3) |
| **$R_{merge}$* (%)** | 9.2 (48.4) | 10.3(60.0) | 9.8 (63.5) | 9.5 (61.4) | 8.8 (44.2) | 7.3 (75.3) | 6.7 (73.7) | 7.9 (69.3) |
| **Mosaicity (°)** | 0.15 | 0.25 | 0.25 | 0.25 | 0.1 | 0.46 | 0.47 | 0.46 |
| **Total reflections** | 453777 | 453706 | 748407 | 455949 | 191991 | 321800 | 565274 | 177785 |
| **Redundancy** | 10.4 | 7.4 | 12.2 | 7.4 | 4.0 | 6.5 | 11.5 | 4.5 |
| **Unique reflections** | 43620 | 61274 | 61396 | 61272 | 48169 | 49334 | 49081 | 39880 |
| **$\Delta_{ano}/\sigma(\Delta_{ano})$#** | | 1.0(1.7) | 1.0(1.6) | 0.8(3.6) | | 0.9(2.0) | 1.0(1.8) | 0.7(5.0) |
| **#Frames** | 240 | 180 | 360 | 180 | 90 | 180 | 360 | 120 |

Numbers in parenthesis refer to the high resolution shell.

*Rmerge = ΣhΣi |I(h)i - <I(h)>| / ΣhΣi I(h)i, where I(h)i is the ith observation of reflection h and <I(h)> is the mean intensity of that reflection.

#Friedel pairs differences overall value while the number between parenthesis refers to the resolution (Å) where Δano/σ(Δano) cross 1.0.

**Table 2.** Data collection statistics.

The quality of the electron density map after density modification also showed to be of the same quality as those shown in Fig. 5 and Fig. 6. The higher Rmerge value observed for the data collected at high energy can then be traced back to the experimental setup. Most probably the main causes for error are shutter-rotation timing, rotation deviations, etc. Analysis of the diffraction patterns and data processing did not show any visible onset of radiation damage effects.

### 3.2. Phases and Electron Density Maps

The three wavelength data and the native data set collected on each beam line were exploited to test the viability of different standard phasing methods used with medium energy for ultra-high X-ray energies. MAD phasing was performed employing three (Hendrickson, 1981) and two wavelengths (peak and remote; peak and inflection; inflection and remote) (Gonzalez, 2003). Heavy atom positions were determined from the anomalous signal for SAD (peak and remote) and MAD phasing three and two wavelength combinations. For SIRAS phasing the data obtained for the peak and remote wavelengths was combined with the native data while SAD phasing was performed on each of the individual wavelengths, peak and remote. Three strong peaks corresponding to the Ho sites were identified in all the calculated anomalous difference maps. The anomalous difference map obtained from the high energy remote (56.64 keV) data collected on X17B1 is shown in Fig. 5.

The map contoured at the eleven-sigma level indicates the presence of three holmium sites with strong anomalous signal. The phasing quality obtained for the different methods employed is summarized in Table 3; not all the results are shown for both beam lines, as the phasing quality is similar for the same method. According to statistics shown in Table 3, all the methods, SIRAS, MAD and SAD were successful presenting high figure-of-merit (FOM) and connectivity independent of the phasing method. Phasing with the SHELX suites (Sheldrick & Gould, 1995), SHELXD and SHELXE, showed to be as successful as in the medium energy cases (12 keV range) for processing high-resolution data. The use of different software suites, HYSS, PHENIX, CRANK, CCP4 (Ness *et al.,* 2004; Grosse-Kunstleve & Adams, 2003; CCP4, 1994) to phase MAD data was as successful, presenting a high figure-of-merit and good correlation factors. These results are not presented, as the comparison between software suites is not the purpose of this

report. Thus for phasing to succeed at ultra-high X-ray energies similar requirements as found in the medium energy cases (12 keV range) need to be observed.

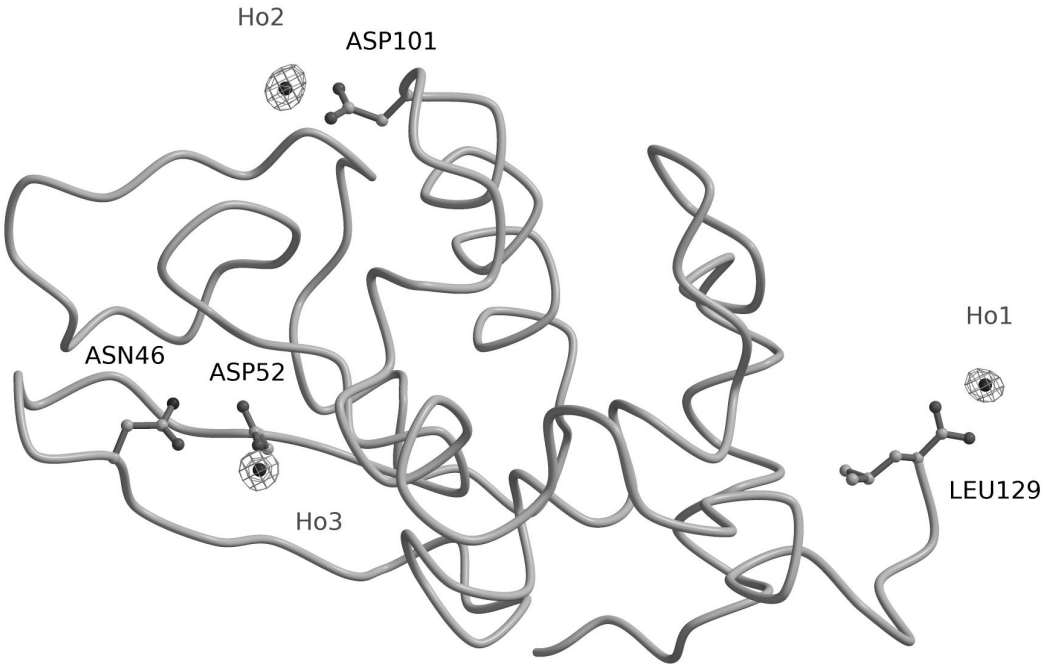



**Figure 5**. Anomalous difference map, contoured at the 11 σ level, after density modification (Cowtan, 1994); SAD phasing of high energy remote, 56.34 keV i.e., 0.22 Å, data collected on X17B1. Main chain trace and ligand residue side chains are shown. The figure is produced with Bobscript (Esnouf, 1997; Kraulis, 1991) and enhanced with RASTER3D (Merrit & Murphy, 1994).

Fig. 6 represents the experimental electron density maps calculated by the three different methods under consideration, SIRAS, MAD, and SAD. The maps after density modification exhibited not only main chains but also most of the side chains and water molecules. Models were built in full using the ARPwARP program suite. They all were at least 95% complete and accurate, as attested by the high correlation coefficients listed in Table 3. Poor electron density was observed only for highly flexible regions including the C and/or N terminal residues and the loop extending from Pro 70 through Ser 72. Missing residues were manually added and two models fully refined and deposited in the Protein Data Bank; one resulting from the SIRAS phasing method using the high energy remote and native data sets collected at 0.22 Å on ID15B (PDB access code: **2CGI**), and the second from SAD phasing at the Ho peak energy collected on X17B1 (PDB access code: **2BPU**).

The SIRAS crystallographic model was refined at 1.35 Å resolution to an R factor of 17.5% and Rfree of 19.5%; Rfree was calculated from 5 % of all the reflections. The model consists of the complete chain, 1 Cl- ion and 215 water molecules. None of the residues were found in the disallowed region of the Ramachandran plot and the low root-mean square deviations (rmsd) from ideal bond lengths and angles values, 0.007 Å and 1.1°, are a confirmation of a good stereo-chemical agreement.

| ***Beamline*** | | ***ID15A*** | | | **X17B1** | | |
|---|---|---|---|---|---|---|---|
| **Method** | | $MAD_{+nat}$ | $SIRAS_{pk}$ | $SIRAS_{rm}$ | $SAD_{pk}$ | $SAD_{rm}$ | $SIRAS_{pk}$ |
| **SHELXD** | Correct/total trials | 14 / 30 | 17 / 30 | 15 / 30 | 19 / 30 | 29 / 30 | 22 / 30 |
| **SHELXE** | CC / contrast | 0.75 / 0.40 | 0.69 / 0.42 | 0.72 / 0.42 | 0.77 / 0.47 | 0.80 / 0.47 | 0.71 / 0.40 |
| | connectivity | 0.89 | 0.9 | 0.88 | 0.89 | 0.89 | 0.88 |
| | fom | 0.81 | 0.79 | 0.79 | 0.75 | 0.77 | 0.8 |
| | mapCC* | 0.83 | 0.76 | 0.8 | 0.89 | 0.85 | 0.77 |
| **Arp** | Res built(AA/wat) | 127 / 162 | 128 / 164 | 128 / 156 | 124 / 162 | 125 / 162 | 126 / 171 |
| | $R / R_{free}$ (%) | 20 / 23 | 20 / 23 | 21 / 25 | 24 / 27 | 25 / 28 | 20 / 24 |
| **Model Refined** | PDB access code | | | 2CGI | 2BPU | | |
| | $R / R_{free}$ (%) | | | 17.5 / 19.5 | 17.1 / 19.5 | | |

**Table 3.** Phasing results. The correlation coefficient (CC) for the maps are calculated using the electron density after density modification with the model built and partially refined as provided by ARPwARP at the model building stage for each method. The refined model PDB access code 2CGI was refined taking into account the native and the Ho remote data collected on ID15B (Table 2).

The SAD model, refined to an R factor of 17.1% and Rfree of 19.5 %, consists of the complete chain, 3 $Ho^{3+}$ ions, 1 $Na^{+}$ ion, 3 $Cl^{-}$ ions and 107 water molecules. As for the SIRAS model no residues were found in the disallowed regions of the Ramachandran plot

and as before the low values found for the stereo-chemical deviations in bond lengths and angles are 0.007 Å and 1.1o. Both models were compared to a native hen egg-white lysozyme structure (PDB access code: 2BLX) determined from medium energy data (wavelength 0.94 Å) and similar diffraction resolution (1.4 Å) (Nanao *et al*., 2005). The rms deviations between the main $C_\alpha$ chains were found to be 0.21 Å and 0.13 Å for the SIRAS and SAD models respectively, with the largest differences being observed at the C terminal and the loop region. The overall B factor, 13.6 Å2 for the SIRAS and 12.8 Å2 for the SAD model is comparable to the native structure used for comparison, with an overall B factor of 10.1 Å2; relative variations of individual B factors are also similar. Identical results were obtained when the K-edge SAD structure was compared to a structure (data not presented) determined by singlewavelength anomalous phasing of data collected at a wavelength of 1 Å on similar crystals on a standard macromolecular crystallography beam line (Allaire *et al*., 2003). The phasing and refinement results found for both structures, SIRAS and SAD, are quite similar to those obtained by Takeda's group (Takeda *et al*., 2004) from MAD phasing data collected on the K-edge of an HEWL xenon (PDB access code: 1VAU) and iodine (PDB access code: 1VAT) derivative. Although the occupancy was quite low for both these derivatives (I and Xe) MAD phasing allowed to distinguish between the protein and solvent regions prior to any density modification, but not to build the structure model. In the present study MAD phasing of the HEWL Ho derivative resulted in an experimental electron density map of sufficient quality to trace the molecular model; most of the main and side chains could be traced automatically. Raw SAD phases, before any density modification from either SHELXE or DM also proved to be sufficient to build a nearly complete model; however, density modification improved the phases and allowed automated construction of most of the model in only 5 cycles in ARPwARP in comparison to the 15 cycles necessary without DM application. The overall correlation coefficients of the maps, fom, determined from SAD phasing for the high energy remote case with and without the application of density modification and the refined model were 0.85 and 0.60 respectively.

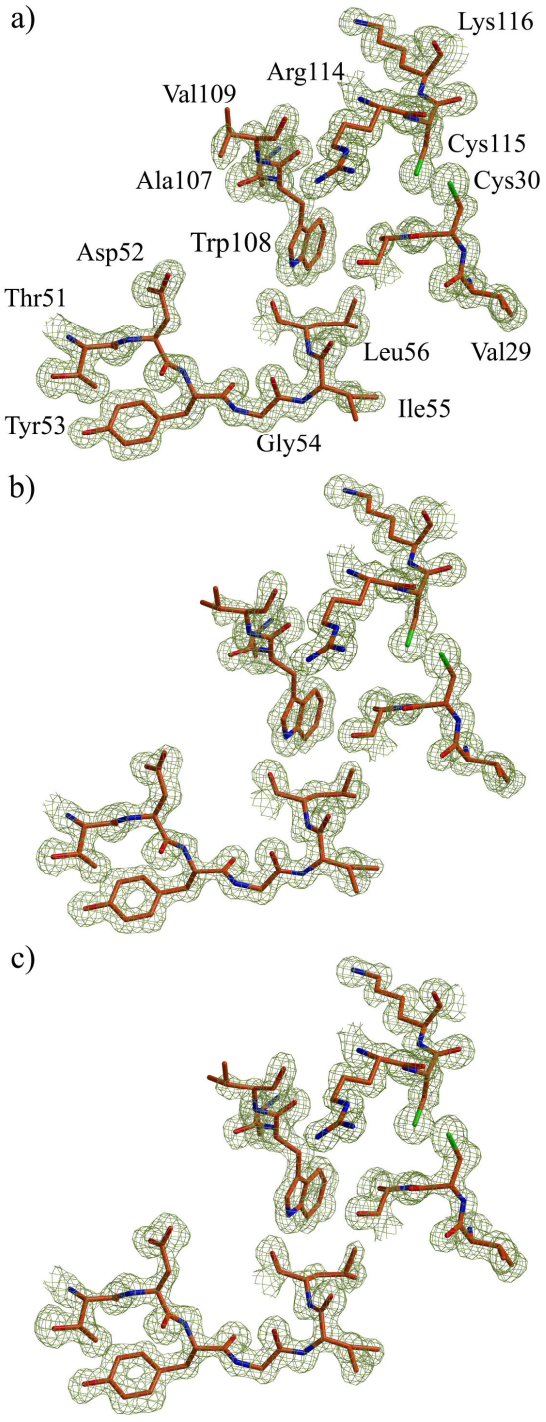



**Figure 6**. Experimental electron density maps. The maps are calculated after phasing and density modification; all maps are contoured at the 1σ level. a) SAD phasing at the peak wavelength (0.2227 Å) using data recorded on X17B1; b) SIRAS phasing with the high energy remote (0.22 Å) and native data recorded at ID15B; residues shown as labelled in a); c) Three wavelength MAD (peak, inflection and high energy remote) phasing combined to native data recorded on ID15B; residues shown as labelled in a). The figure is produced with MOLSCRIPT, BOBSCRIPT and RASTER3D.

### 3.3. The Holmium Sites

Three Ho sites, Ho-1, Ho-2 and Ho-3, were found in the present holmium HEWL derivative as shown in the anomalous difference map in Fig.5. During refinement, site occupancies converged to 0.3, 0.8 and 0.5 for Ho-1, Ho-2 and Ho-3 respectively. The contribution to the anomalous signal at these refined occupancy levels is 2.7 % compared to the expected 3.9 % for fully occupied sites. The holmium site, Ho-1, is located on a crystallographic two-fold axis. Ho-1 binds to the main chain oxygen of residue Leu 129 of the C terminus and the oxygen of one water molecule and the symmetry related residue and water molecule. The Ho-2 site located at the solvent boundary binds to the Asp 101 side chain oxygen and four water molecules. The 3rd site, Ho-3, located in the catalytic cleft binds to Asp 52 side chain OD2 atom and five water molecules; Ho-3 is 7 Å away from the closest symmetry related residue, Arg 21. In general poor electron density was observed around positively charged and hydrophobic side chains in close vicinity to these holmium sites, for example, no density was observed for the side chains for residues Arg 21 and Val 109. As frequently reported in literature holmium sites are usually hexacoordinated often binding to the side chain oxygen atom of acidic amino acids and on occasions to the carbonyl atom from the main chain. The three holmium sites found here are also hexacoordinated by oxygen atoms, but in the present case by oxygen atoms belonging to water molecules.

### 3.4. SAD Phasing and Redundancy

One of the limiting factors in the SAD phasing method is the accurate measurement of intensities that will result (if anomalous scatterers are present with relatively high occupancy) in accurate heavy atom positions and good phase estimates (Dauter et al., 1999). For this purpose, redundancy study was performed on the Ho K-edge peak data recorded at X17B1. Starting with the 180 degrees data set initially recorded; 90, 120 and 150 degrees data was generated at the scaling stage. Table 4 shows the statistics from the scaled data and the phasing - model building steps. All data was scaled to the highest resolution reported for the 180 degrees data, 1.35 Å.

| Number of frames processed | 90 | 120 | 150 | 180 |
|---|---|---|---|---|
| **Completeness (%)** | 92.9 (62.9) | 97.9 (70.9) | 99.0 (83.0) | 99.2 (86.1) |
| **Unique reflections** | 24737 (822) | 26052 (926) | 26365 (1084) | 26419 (1124) |
| **Total reflections** | 147263 | 195059 | 251843 | 315976 |
| **Redundancy** | 6.0 (2.0) | 7.5 (3.5) | 9.6 (4.1) | 12.0 (5.1) |
| **I/σ(I)** | 22.6 (1.6) | 24.2 (1.7) | 28.7 (1.9) | 32.6 (2.1) |
| **Rmerge (%)** | 6.8 (68.9) | 7.1 (73.1) | 7.4 (74.6) | 7.8 (79.9) |
| **FOM_dm*** | 0.72 | 0.76 | 0.77 | 0.78 |
| **MAP_corr&** | 0.58 | 0.72 | 0.75 | 0.74 |
| **RES_built$** | 10 | 107 | 126 | 122 |

*Figure of merit calculated after density modification performed by DM. &The map correlation is calculated between the map after density modification and the deposited structure calculated from the complete peak data devoid of water molecules and ions. $Number of residues built with ARPwARP starting with the structure factors and phases from DM.

**Table 4**. SAD phasing statistics; data sets with different completeness and redundancy were generated from a single data set, for further discussion see text. Numbers in parenthesis refer to the highest resolution shell (1.37-1.35 Å).

Figure 7 displays the four calculated maps after density modification was performed. For the 90 degrees generated data set, with 6 fold redundancy in the whole resolution range and 2 fold redundancy in the highest resolution shell (due to low completeness) it was not possible to build a complete model, whereas for the 120 degrees data, with 7.5 and 3.5 fold redundancies for the full range and highest resolution shell, it was possible to build a 107 amino-acid model, which represents 82 % of the complete model. The last generated data set, 150 degrees, and the 180 degrees data set, were highly redundant over the full range as well as in the last resolution shell. It was possible therefore to build the complete model except for the loop region, residues 69 - 72, and the C and N terminus residues. Similar results were obtained with data collected on a crystal grown in the same condition with the same size at the high energy remote (1.0332 Å, 12 keV, with f'' = 8.7 e-) of the LIII holmium edge where a 120 degrees total oscillation range (4.4 fold redundant and 97 % complete in the full resolution range due to a square detector, ADSC Q210) yielded a 1.35 Å resolution complete model, except for the loop containing residues 71-73, and the C and N terminus residues. Thus, as with low energy data, redundancy (Dauter &

Adamiak, 2001) is a limiting factor for structure determination by SAD phasing at ultra high X-ray energies. However the same amount of data recorded at ultra high (55.6 keV) and medium (12 keV) energy resulted in a similar model with equivalent quality and a lower deposited dose (Table 5).

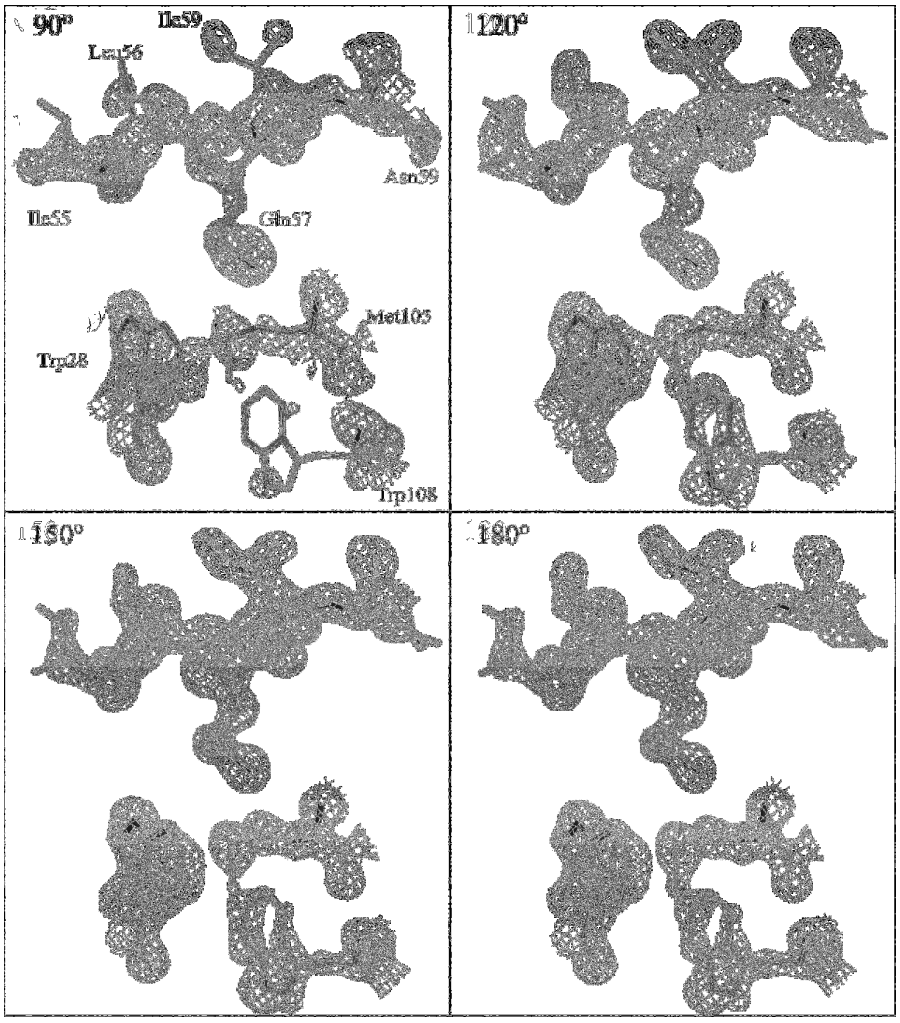



**Figure 7.** The electron density map generated after 10 cycles of density modification performed on the data collected at the peak wavelength of the Ho K edge on beam line X17B1 is shown superimposed to the refined SAD model. The corresponding crystal rotation range within the total amount of data recorded (180°), used to obtain the electron density maps shown, is indicated in the top left corner of each panel. Residues shown are indicated in the top left panel. The resolution of the map is 1.35 Å and the corresponding redundancy is indicated in Table 4. The figure was produced as Fig. 6.

### 3.5. Phasing at Ultra High X-rays Energies

The K-edge anomalous scattering factor for lanthanides is of the same order as the Se K-edge scattering factor making them ideal candidates for anomalous scattering experiments at ultra high X-ray energies. Fig. 8 shows the scattering factors for holmium and selenium in the medium and high energy range.

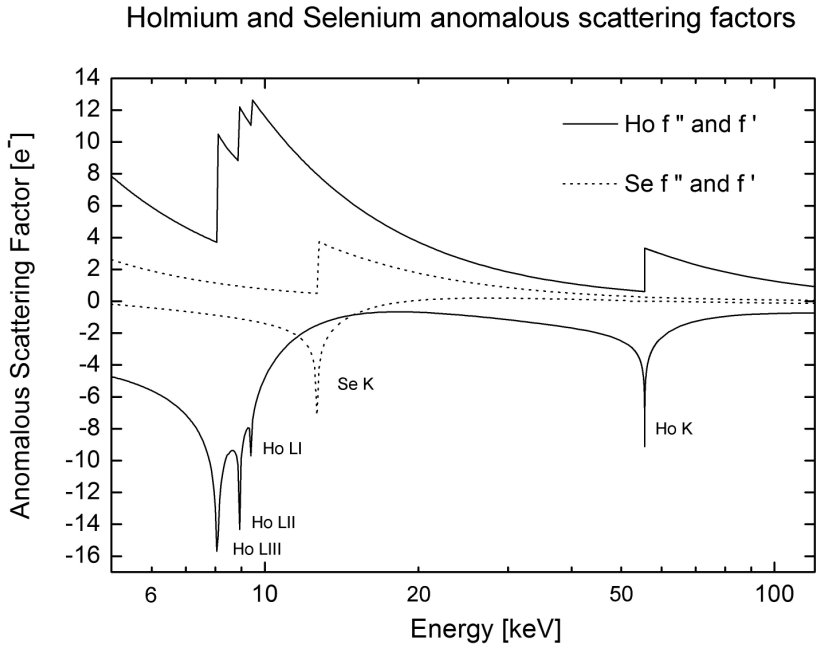



**Figure 8.** Anomalous scattering factors in the 4 - 120 keV range for holmium and selenium. The contribution to the anomalous signal at the K-edge is nearly the same for both elements and other lanthanides. Consequently Ho as other lanthanides is useful in a wide range of energy for phasing purposes (http://lipro.msl.titech.ac.jp/scatfac/scatfac.html).

If Ho f " at its L edges is very attractive (~ 12 e-), when possible it is preferable to collect data at higher energy, i.e. 13 keV, where f " is still high, approximately 7 e-, and the dose deposited on the crystal is expected to be lower due to the lower mass absorption coefficient (30 cm2.gm-1 at 9 keV and 10 cm2.gm-1 at 13 keV; Fig. 1). Therefore, ultra high X-ray energies should further lower the expected dose deposited on the crystal and extend its lifetime. Consequently, Ho, like transition elements and other lanthanides should be useful in a wide energy range for phasing purposes. At the Ho K-edge, the contribution to the anomalous signal is 3.3 e- which would be sufficient (with minimal errors in data) to estimate phases for a 200 amino acid protein with only one fully

occupied holmium atom site with a theoretical Bijvoet ratio of 1.7 %. Several successful phasing experiments were reported for anomalous signals as small as 0.6 % in the medium and low energy ranges for high accuracy, i.e., high redundancy data (Wang, 1985; Ramagopal *et al.*, 2003). Similar contributions to the anomalous signals are expected at the ultra high X-ray energy range for commonly used derivatives such as bromide and mercury. Dauter & Dauter (1999) used the anomalous signal of the bromine K-edge for the determination of experimental phases of a lysozyme molecule. At 50 keV, the expected Bijvoet ratio (<ΔFanom>/<F>), i. e., anomalous contribution, of the six bromine atoms, located in the solvent shell, refined to full occupancy, would be of the order of 0.6%, (calculated with f " = 0.37 e- at 50 keV). Based on error free data, Wang (1985) estimated that for S and $<\Delta F_{anom}>/<F>$ as low as 0.6 % it would be possible to solve a structure by the SAD method. Indeed, Ramagopal and co-workers (Ramagopal *et al*., 2003) confirmed this prediction experimentally concluding that Wang's conjecture was realistic. Recently Ramagopal et al. (2005) have shown that Hg derivatives are quite sensitive to radiation damage. At 55 keV f " = 1.3 e- and the expected contribution to the Bijovoet ratio is of the order of 1 % per Hg atom per 95 residues which would be sufficient for structure solution by the SAD technique.

### 3.6. Potential Future for Ultra High Energy Crystallography

The potential application of ultra-high energy phasing is quite broad. Dedicated beam lines and experimental setups would further reduce systematic errors while the development of systematic data collection strategies (Fourme *et al.*, 2003) will allow significant gain in the signal-to noise ratio. The generated increase in internal data consistency would allow the exploitation of difficult problems such as the measurements of small anomalous signals, or small radiation sensitive or poorly diffracting crystals.

Currently none of the dedicated macromolecular crystallography beam lines is capable to provide X-rays with energies higher than 40 keV. Thus, if crystallographers are interested in collecting data at ultra high energy, new high energy macromolecular beam lines should be planned. New synchrotrons are being built in Europe (SOLEIL, DIAMOND) and existing synchrotrons are being upgraded in the USA (SSRL, ALS). New facilities with ultra bright/ultra focused beam are also planned in USA (NSLSII). These new or upgraded light sources could eventually provide ultra intense/bright/focused high energy X-ray

beams if such a need is made known to the designers. To our knowledge, no such beam line for macromolecular crystallography is planned at any of the facilities listed above.
Current detectors are also not optimized for high energy data collection. It should be relatively straightforward to mount an adequate phosphor layer on a CCD detector, thus optimizing the detector absorption efficiency and reducing the exposure time required per oscillation range further minimizing the radiation damage incurred. With an optimized source, an implemented macromolecular crystallography experimental setup and high energy optimized detectors combined together, it will be possible to collect almost the ideal high quality data (Helliwell *et al*., 1993) without the radiation damage issue, nearly free of absorption errors.

## 4. Conclusion

This study has demonstrated that anomalous scattering at ultra high X-ray energies is a viable alternative to the solution of the phase problem in macromolecular crystallography. For the first time the anomalous signal from the K-edge of a lanthanide was employed to probe several phasing methods, single isomorphous replacement with anomalous scattering (SIRAS), multiwavelength anomalous diffraction (MAD) and singlewavelength anomalous diffraction (SAD) and a high-resolution high-quality 3D molecular model is obtained from single anomalous diffraction phasing (SAD) at ultra high X-ray energies. Two distinct measurements were carried out on ultra high X-ray energy beam lines located at a 3rd generation synchrotron facility, ID15B at the European Synchrotron Radiation Facility, and at a 2nd generation synchrotron facility, X17B1 at the National Synchrotron Light Source. In spite of non-optimized beam line instrumentation, detectors and software, the high quality of the data measured provided experimental electron density maps of comparable quality to the maps obtained in the medium energy regime (12 keV). Dose estimation showed that even with poor detector performances there is a clear advantage to record data at higher energies as the life time of the crystal is extended all while keeping data quality. This study represents one more step towards the measurement of the "ideal data set" (Helliwell et al., 1993; Fourme et al., 2003) and the efficient exploitation of the anomalous signal at ultra high X-ray energies.

## Appendix: Dose estimation

In order to better assess the benefits and drawbacks of conventional data collection at ultra high X-ray energies it is interesting to compare the dose deposited in the sample during data collection at both beam lines. These results are further compared to those found for a similar crystal submitted to data collection at a bending magnet beam line at a 2nd generation synchrotron at medium energy (12 keV). To estimate the dose presented in Table 5 the energy-absorption coefficient (μen) rather than the absorption (μ) or attenuation coefficients (μatt) was considered as at higher energies (> 30keV) Compton Scattering is predominant. Figure 1 shows the mass attenuation as well as the mass energy-absorption coefficients for a generic protein crystal sample and for the refined model protein (Lysozyme, with solvent molecules, Na, Cl and Ho ions) as a function of energy; the mass absorption coefficient not shown, is very close to the attenuation coefficient.

| | **ID15B[$]** | **X17B1[$]** | **X6A[#]** |
|---|---|---|---|
| **Energy (J)** | $8.91\ 10^{-15}$ | $8.91\ 10^{-15}$ | $1.92\ 10^{-15}$ |
| **Flux ($ph.s^{-1}$) [at sample]** | $1.8\ 10^{11}$ | $9\ 10^{9}$ | $9\ 10^{10}$ |
| **T(s)[&] per data set** | 360 | 9000 | 360 |
| **($\mu_{en}/\rho$) ($cm^2g^{-1}$)** | 0.17 | 0.17 | 7.3 |
| **Dose ($J.kg^{-1}$)*** | 98 | 123 | 454 |

$ Data collected at the high energy remote Ho K edge wavelength, 0.22 Å (56.3 keV).
# Data collected at the high energy remote Ho L edge wavelength, 1.03 Å (12.0 keV).
& Total exposure time per data set. Data set is 90 frames of 1o oscillation.
* To obtain the actual dose deposited it is necessary to consider the actual crystal cross section with the beam during data collection.

**Table 5**. Dose estimate for crystals of similar size, and diffraction resolution and intensities, exposed to ultra high X-ray energies (ID15B and X17B1) and to medium X-ray energies (X6A).

Taking in to account the crystal and beam size the dose deposited on the samples is proportional to:

$$D = \left( \frac{\mu_{en}}{\rho} \right) * Fl * E * T * N * 1000 \text{ (Gy)}$$

where Fl the fluence is expressed in ph.s-1.cm-2; (μen/□), the mass energy-absorption coefficient in cm2.g-1; E the energy in J (energy in eV * electron charge = 1.6 10-19 J); T the exposure time in s and N the number of frames recorded per data set. For the three experiments, the two high energy experiments on X15B and X17B1, and for the medium energy experiment on X6A, the same beam size and crystals of approximately the same size were chosen. In this case the above equation can be simplified and the estimated dose is proportional to $\left( \frac{\mu_{en}}{\rho} \right) * Flux * E * T$, with Flux being the flux at the sample determined experimentally by ion chamber at the given energy; the dose is then determined neglecting crystal rotation. The difference between the results obtained for the two high energy experiments is due to the detector type used, the MAR345 is much more efficient for the energy range used here, 55 keV, than the MARCCD and therefore a shorter exposure time is required to reach the same resolution limit and diffraction intensity. It is clear from the estimates shown in Table 5 that to obtain the same diffracting power the dose deposited in the sample at medium energy is significantly higher than that deposited at ultra high X-ray energy.

## Acknowledgement

We would like to thank D. P. Siddons for valuable discussions and T. Buslaps for supporting this project. We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities. We thank R. Greene and the staff of the National Synchrotron Light Source. The NSLS is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-98CH10886. The NIGMS East Coast Structural Biology Research Facility, the X6A beam line, is funded under contract # GM-0080.

## References


Allaire, M., Aslantas, M., Berntson, A.,Berman, L., Cheung, S., Clay, B., Greene, R., Jakoncic, J., Johnson, E., Kao, C. C., Lenhard, A., Pjerov, S., Siddons, D. P.,Stober, W., Venkatagiriyappa, V., Yin, Z. and Stojanoff, V. (2003). *Synchr. Rad. News* **16**, 20-25.

Aslantas, M., Jakoncic, J., Honkimaki, V., Di Michiel, M., Zhong, Z., Siddons, D.P. & Stojanoff, V. (2006). In preparation.

Arndt, U. W. (1984). *J. Appl. Crvst*. **17**, 118-119.

Banumathi, S., Zwart, P., Ramagopal, U. A., Dauter, M. & Dauter, Z. (2004). *Acta Cryst*. **D60**, 1085-1093.

Brodersen, D.E., Etzerodt, M., Madsen, P., Celis, J.E., Thogersen, H.C., Nyborg, J. & Kjeldgaard, M. (1998). *Structure.* **6**, 477-489.

Chayen, N. E. (1998). *Acta Cryst*. **D54**, 8-15.

Cassetta, A., Deacon, M. A., Ealick, S. E., Helliwell, J. R. & Thompson, A. W. (1999). *J. Synchrotron Rad.* **6**, 822-833.

Cianci, M., Helliwell, J. R., Helliwell,M., Kaucic, V., Logar, N. Z., Mali, G. & Tusar, N. N. (2005). Crystallography Reviews. **11,** 245-335.

Collaborative Computational Project, Number 4. (1994). *Acta Cryst*. **D50**, 760-763.

Cowtan, K. (1994). Jnt CCP4/ESF-EAMCB *Newsl. Protein Crystallogr*. **31**, 34-38.

Dauter, Z., Dauter, M., de la Fortelle, E., Bricogne, G. & Sheldrick, G.M. (1999). *J. Mol. Biol*. **289**, 83-92.

Dauter, Z and Dauter, M. (1999). *J. Mol. Biol*. **289**, 93-101.

Dauter, Z. & Adamiak, D.A. (2001). *Acta Cryst*. **D55**, 990-995.

Dauter, Z. (2002). *Curr. Opin. Struct. Biol.* **12**, 674-678.
Diederichs, K. (2006). *Acta Cryst.* **D62**, 96-101.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst*. **D60**, 2126-2132.

Esnouf, R. M. (1997) *Journal of Molecular Graphics*. **15**, 132-134.

Evans, G. & Pettifer, R.F. (2001). *J. Appl. Cryst*. **34**, 82-86.

Fourme, R. & Hendrickson, W. A. (1990). *Synchrotron Radiation and Biophysics*, edited by S. S. Hasnain, 156-175. Chichester: Ellis Horwood.

Fourme, R., Shepard, W., Schiltz, M., Prangé, T., Ramin, M., Kahn, R., de la Fortelle, E. & Bricogne, G. (1999). *J. Synchrotron Rad.* **6**, 834-844.

Fourme, R., Kahn, R., Mezouar, M., Girard, E., Hoerentrup, C., Prange, T. & Ascone, I. (2001). *J. Synchrotron Rad.* **8**, 1149-1156.

Fourme, R., Girard, E., Kahn, R., Ascone, I., Mezouar, M., Dhaussy, A. C., Lin, T. & Johnson, J. E. (2003). *Acta Cryst*. **D59**, 1767-1772.

French, G. S. & Wilson, K. S. (1978). *Acta Cryst*. **A34**, 517-525.

Garman, E. & Murray, J. W. (2003). *Acta Cryst.* **D59**, 1903-1913.

Girard, E., Kahn, R., Ascone, I., Mezouar, M., Dhaussy, A. C., Lin, T., Johnson, J. E. & Fourme, R. (2004). *High Pressure Res*. **24**, 173-182.

Gonzalez, A., Denny, R. and Nave, C. (1994). *Acta Cryst.* **D50**, 276-282.

Gonzalez A. (2003). *Acta Cryst.* **D59**, 315-322.

Grosse-Kunstleve RW & Adams PD. (2003). *Acta Cryst.* **D59**, 1966-1973.

Helliwell, J. R. & Fourme, R. (1983). ESRP Report IRI-4/83, 1-36. CERN, Geneva.

Helliwell, J. R. (1992). *Macromolecular Crystallography with Synchrotron Radiation*. Cambridge University Press.

Helliwell, J. R., Ealick, S., Doing, P., Irving, T. & Szebenyi, M. (1993). *Acta Cryst.* **D49**, 120-129.

Henderson, R. (1990). *Proc. R. Soc. London.* **B241**, 6-8.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107-113.

Hendrickson, W. A., Smith, J. L. & Sheriff, S. (1985). *Methods Enzymol.* **115**, 41-55.

Hendrickson, W. A. (1991). *Science*. **254**, 51-58.

Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845-851.

Kraulis, P. J. (1991). *J. Appl. Cryst*. **24**, 946-950.

Matthews, B. W. (1968). *J. Mol. Biol*. **33**, 491-497.

Merritt, E. A. & Murphy, M. E. (1994). *Acta Crystallogr.* D50, 869-873.

Murray, J.W. & Garman, E.F. (2002). *J. Synchr. Rad.* **9**, 347-354.

Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). *Acta Cryst*. **D53**, 240-255.

Nanao, M.H., Sheldrick, G.M. and Ravelli, R.B.G. (2005). *Acta Cryst*. **D61**, 1227-1237.

Ness, S. R., de Graaff, R. A. G., Abrahams, J. P. & Pannu, N. S. (2004). *Structure*. **12**, 1753-1761.

O'Neill, P., Stevens, D. L. & Garman, F. (2002). *J. Synchrotron Rad.* **9**, 329-332.

Otwinowski, Z. & Minor,W. (1997). *Methods Enzymol.* **276**, 307-326.

Perrakis, A., Morris, R. & Lamzin, V.S. (1999). *Nature Struct. Biol*. **6**, 458-463.

Peterson, M. R., Harrop, S. J., McSweeney, S. M., Leonard, G. A., Thompson, A. W., Hunter, W. N. & Helliwell, J. R. (1996). *J. Synch. Rad.* **3**, 24-34.

Prange, T., Schiltz, M., Pernot, L., Colloc'h, N., Longhi, S., Bourguet, W. & Fourme, R. (1998). *Proteins*. **30**, 61-73.

Ramagopal, U. A., Dauter, M., & Dauter, Z. (2003). *Acta Cryst.* **D59**, 1020-1027.

Ramagopal, U. A., Dauter, Z., Thirumuruhan, R., Fedorov, E. & Almo, S. C. (2005). *Acta Cryst*. **D61** 1289-1298.

Ravelli, R.B.G., Schrøder-Leiros, H.K., Pan, B., Caffrey, M. and McSweeney, S. (2003). *Structure* **11**, 217-224

Ravelli, R.B., Nanao, M.H., Nanao, A., White, S. & McSweeney, S. (2005). *J. Synchrotron Rad.* **12**, 276-84.

Sánchez del Río, M. and Dejus R.J. (1998). XOP: Recent developments. *SPIE proceedings* **3448**.

Schiltz, M., Prange, T. & Fourme, R. (1994). *J. Appl. Cryst*. **27**, 950-960.

Schiltz, M., Kvick, A., Svensson, O. S., Shepard,W., de La Fortelle, E., Prange, T., Kahn, R., Bricogne, G. & Fourme, R. (1997). *J. Synchrotron Rad.* **4**, 287-297.

Schiltz, M., Dumas, P., Ennifar, E., Flensburg, C., Paciorek, W., Vonrhein, C. & Bricogne, G. (2004). *Acta Cryst*. **D60**, 1024-1031.

Sheldrick, G. & Gould, R. O. (1995). Acta Cryst. **B51**, 423-431

Takeda, K., Miyatake, H., Park, S.-Y., Kawamoto, M., Kamiya, N. & Miki, K. (2004). *J. Appl. Cryst.* **37**, 925-933.

Wang, B.-C. (1985). *Methods Enzymol*. **115**, 90-112.

Zhong, Z., Kao, C.C., Siddons, D.P. & Hastings, J. B.(2001). *J. Appl. Cryst*. **34**, 504-509.

Zwart, P. H., Banumathi, S., Dauter, M. & Dauter, Z. (2004). *Acta Cryst*. **D60**, 1958-1963.

Scattering factor at high energy obtained from: http://lipro.msl.titech.ac.jp/scatfac/scatfac.html

# Partie 3 : Article 2

## Protein Crystallography at Ultra High Energy ?

Cet article, en cours de soummission traite de la comparaison des données de diffraction collectées à 56.5 keV avec celles d'un cristal obtenu dans des conditions identiques mais à 12 keV. Nous nous interessons ici particulièrement aux dommages dus aux rayonnements X et à leurs effets.

# Protein Crystallography at Ultra High Energy ?

**Jean Jakoncic[1,4], Marco DiMichiel[2], Zhong Zhong[1], Veijo Honkimaki[2], Peter D. Siddons[1], Yves Jouanneau[3] and Vivian Stojanoff[1]**

[1]National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY 11973, USA. [2]Europen Synchrotron Radiation Facility, Grenoble, France. [3]Centre National de Recherche Scientifique, Grenobl, France. [4]Universite Joseph Fourier, Grenoble, France.

## Abstract

Multiwavelength Anomalous Diffraction (MAD) and Singlewavelength Anomalous Diffraction (SAD) phasing at ultra high energy X-rays, 55 keV, were used to determine a high quality and high resolution experimental electronic density map. The 2.7 % anomalous signal, at the Ho K edge, from three Ho atoms found in the model protein, Hen Egg-White Lysozyme, was sufficient to obtain a high quality electron density and build the 3D molecular model in its integrity using all the phasing methods explored. High energy data were recorded on two beam-lines from a second and a third generation light source. Additional data collected at 56.5 keV in the third generation beamline were compared to data recorded at 12 keV on a second generation synchrotron. The data recorded at ultra high energy yielded to similar resolution with however no or minimal specific and overall decays observed compare to the data collected at medium energy. Advantages and disadvantages on the use of ultra high energy X-rays are discussed in light of radiation damage problems.

**Key words**: Protein crystallography, radiation damage, dose, ultra high energy, absorption.

# 1. Introduction

Macromolecular crystallography is most probably the field that has taken most of the advantages from the intense radiation produced by beamlines in second and third generation synchrotrons. However, with the flux provided in third generation beamlines it is now common to observe radiation damage effects on protein crystals even with the use of cryo-cooling (100 K) during the whole experiment. In fact, radiation damage is the primary cause in phasing failure also taking into account that usually most of the failing experiments are not reported (Murray and Garman, 2002).

Nowadays there are several ways to deal with radiation damage issues; one can minimize it using lower temperature (below 70 K), scavengers (Garman and Owen, 2006) or applying different data collection strategies (2 wavelength MAD, highly redundant SAD) (Gonzales, 2003, Dauter, 2002). The experiment can be prepared knowing all the parameters affecting the dose and therefore applying the best strategy possible (Murray *et al*., 2005). Radiation damage can also be treated either at the data processing step (XDS) (Diederichs, 2006) or in experimental phasing packages (SHARP) (Schiltz *et al.*, 2004) applying exponential decay. Finally radiation damage can be used as implemented in new techniques such as RIP and RIPAS (Ravelli *et al*., 2003, Ramagopal *et al*., 2005).

The total mass-energy absorption coefficient, which depends on the sample in consideration and determines the energy dose deposited on the sample, is energy dependant; for a generic protein sample this parameter is 100 times higher at 12 keV, typical energy used by most of the crystallographers to perform the experiment, compared to 56.5 keV, which is the energy used to perform the experiment here reported. This indicates potential advantage to record diffraction data at higher energy. However, the use of ultra high energy for a macromolecule crystallography (MX) experiment is also not facilitated first because the area detectors currently available are not optimized in this energy range, second there is no dedicated beamline providing X-rays with energy higher than 35 keV and finally if the dose deposited is lower at higher energy the diffracted intensity is diminished.

The use of high and higher energy was first introduced in 1983 (Heliwell *et al*., 1983), but only a few experiments have been performed at energy higher than 25 keV. Experiments were performed in the energy range of 20-35 keV either for phasing purpose, using the anomalous scattering power of Xe and I (Takeda *et al*., 2004), or for high pressure experiment, taking advantage of the lower absorption efficiency of the image

plate detector phosphor screen at energy below the Xe K edge and therefore minimizing scattering from the diamond anvil cell and improving the quality of the data (Fourme *et al*., 2001, Girard *et al*., 2004).

The initial purpose of the experiment described in this report was to further probe ultra high energy X-ray for macromolecular crystallography using a 55 keV X-ray beam provided by two beamlines located in a second and a third generation synchrotrons each equipped with two different detectors. The aim was to explore most of the current phasing methods used at low and medium energies. The previous high energy phasing experiments reported were performed using MAD and SIRAS methods. Here, data collected at high (56.5 keV) and at medium (12 keV) energies allowed for radiation damage monitoring and side by side comparison. The purpose wasn't to compare the data and structure from both energies, but the variation of overall parameters as well as local structural effects due to radiation damage over the course of the experiment. The conclusion will be made only on the difference of variations and will be interpreted with an estimate of the energy dose deposited on each crystal.

## 2. Material and Methods

### 2.1. Data analysis

Details of the experiment at ultra high energy were reported elsewhere (Jakoncic *et al.*, 2006), here only a brief summary is presented. The data used for this study were collected at the beamline ID15B (during a 2 * 1/3 storage ring filling mode) and X6A (Allaire *et al*., 2003) for the ultra high and medium energies respectively. The ultra high energy data were collected at the high energy remote, 1 keV above the Holmium K edge, 56.5 keV and the medium energy data were recorded at 12 keV. It is important to note that both crystals used were grown using the same condition, solutions and under similar techniques, that is the micro-batch method under mineral oil. They were of the same size (400 x 400 x 300 $\mu m^3$) and contained in the same nylon loop size (0.4-0.5 mm). The beam size (300 x 300 $\mu m^2$) was kept constant in both experiments. The area detectors used were the MAR345 image plate and the ADSCQ210 at high and medium energies respectively. Crystals were randomly mounted directly from the crystallization drop as it grown in

solution containing the cryo-protectant (25 % Glycerol) and were kept at 100 K under the cold nitrogen stream.

At 56.5 keV, eight data sets, 120 degrees each, were consecutively recorded. In the initial MAD experiment (Jakoncic *et al*., 2006), first two data sets each 180 degrees, were collected at the peak and inflection energies respectively. To prevent any bias from the anomalous scattering contribution introducing differences in the maps that would be misinterpreted, in this work the data recorded at the peak and inflection wavelengths were not considered. In other words, in total eleven consecutive data sets, 120 degrees each, were collected at ultra high energy and the three first were not used in this work because they were collected at different energies. At 12 keV, twenty data sets, also 120 degrees each were consecutively collected and only the first eleven data sets are considered here. To allow side by side comparison, the three first 12 keV data sets are not considered in the report. At both energies, exposure time and oscillation range were 4 s and 1˚ respectively, which allowed to record data up to the same resolution limit. Particular care was applied to prevent any overload in the low energy experiment while allowing high resolution to be recorded. Among the sequence of data sets recorded at the same wavelength for this experiment (eight at 56 keV and 12 keV) five only are being used for further analysis, the first, third, fifth, seventh and eighth the last; they appear as data set 1, 2, 3, 4 and 5 in this report. The criterion was the resolution limit rather than the dose, or the number of photons diffracted, which is the most important parameter for crystallographers. The data collections parameters are summarized in Table 1.

| | **X6A (12 keV, 1.03 Å)** | **ID15B (56.5 keV, 0.22 Å)** |
|---|---|---|
| **Flux (ph.s-$^{1}$ 0.1% bw mm$^{-1}$)** | 5 $10^{11}$ | $10^{12}$ |
| **Detector efficiency (%)** | 100 | 40 |
| **Exposure time / frame (s)** | 4 | 4 |
| **Detector readout time (s)** | 1 | 100 |
| **Number of data set (120 degrees)** | 11 | 8 |
| **Detector-sample distance (mm)** | 125 | 850 |
| **Exposure time /data set (s)** | 480 | 480 |
| **Total Time /data set (min)** | 10 | 210 |

**Table 1**. Data collection parameters.

## 2.2. Radiation damage evaluation

Radiation damage can be assessed through different ways such as the volume cell expansion, decay of the highest resolution shell signal to noise ratio, increase of the overall temperature factor calculated from a Wilson plot, decarboxylation, photoreduction of anomalous scatterers, as their total cross section is higher, and finally disulfide bond breakage. All these variations and deleterious structural effects will be monitored and visualized to observe any onset of radiation damage at the ultra high and medium energies.

All data were submitted to the same data analysis protocol, using Denzo, Scalepack (Otwinoski and Minor, 1997), truncate (French and Wilson, 1978) cad, scaleit, refmac5 (Murshodov *et al.,* 1997), fft (Collaborative Computational Project, Number 4, 1994) and finally coot (Emsley and Cowtan, 2004) for data integration, scaling, reduction, combination, refinement, difference Fourier map calculation and visualization respectively. The same input scripts were applied and modified in function of the energy and the sequence of the data set.

To assess the influence of radiation damage, two sets of electronic density map were calculated. First, sequential difference Fourier maps ($F_i - F_1$) with i=2-5 were calculated with the phases from the model refined against the first data set. Four difference Fourier maps were then generated for each energy and the peak height was tabulated at each atom position of interest (cysteine, aspartic and glutamate side chain, Ho).

The second type of map calculated was a refined map ($2F_{obs} - F_{calc}$, $Phi_{calc}$) omitting side chain atoms from cysteines involved in disulfide bridge (cysteines were substituted with alanine), from glutamic and aspartic acids, which were also substituted with alanine, one at a time, and finally holmium ions were simply omitted from the model. Negative peaks were expected in the first type of map and positive peaks in the second. Additional analysis, consisting in following the evolution of peak height in the second type of map at amino acids, reported to be mildly or not radiation damage sensitive; these residues are Met 12, Ser 91 and Met 105, and were performed to insure a proper data analysis procedure, with no bias.

## 3. Results

### 3.1. Ultra High Energy Phasing

The three holmium ions present in the HEWL crystal allowed its tridimensional structure determination with all the methods tested using a variety of software packages (Jakoncic *et al*., 2006). The Bijvoet ratio was determined with the refined occupancies of the holmium as well as the anomalous scattering factor f " derived from the absorption spectra, and was 2.7 % with a f " of 3.5 e-. This value is equivalent to the widely employed selenium (Hendrickson *et al.*, 1990) at its K edge, 12.6 keV. MAD, SIRAS and SAD were successfully applied and after density modification resulted in a complete model built. Initialy, 120 degrees collected at either the peak or the remote energies were sufficient to build the model after density modification and they were also sufficient when no density modification at all was performed using the best phases from SHELXE (Sheldrick and Gould, 1995).

In the previous report (Jakoncic *et al*., 2006), the data corresponding to 90 degrees processed resulted in a poor electronic density map with a low connectivity allowing only ten amino acids to be auto traced in ARPwARP (Perrakis *et al*., 1999), the use of the latest version of SHELXCDE tremendously improved the best initial set of phases. This new version allowed building a complete crystallographic model after imrproving phases in DM (Cowtan, 1994) starting with data as low as 45 degrees processed, corresponding to a completeness and overall redundancy of 78 % and 2.2 respectively.

At this stage, there was no sign of radiation damage in all the data inspected. As a note, this experiment was previously attempted two times with no success: this was attributed to the beam-stop configurations used in both beamlines, X17B1 and ID15B, which initially prevented to record low resolution data and authorized strong background scattering.

### 3.2. Data statistics

As previously indicated the criterion applied was the resolution limit achieved and data statistics for the first and fourth data sets are shown in Table 2

| **Energy (keV)** | **12** | | **56.5** | |
|---|---|---|---|---|
| **Data set** | 1st | 4th | 1st | 4th |
| **Unit cell parameters (Å)** | | | | |
| **a = b** | 78.75 | 78.80 | 78.46 | 78.47 |
| **c** | 37.00 | 37.02 | 37.36 | 37.36 |
| **Resolution limits (Å)** | 25.00-1.30<br>1.35-1.30 | 25.00-1.30<br>1.35-1.30 | 25.00-1.25<br>1.29-1.25 | 25.00-1.25<br>1.29-1.25 |
| **Completeness (%)** | 97.7 (83.3) | 97.8 (83.6) | 99.0 (100.0) | 98.9 (100.0) |
| **Multiplicity** | 7.6 (2.5) | 7.6 (2.5) | 9.3 (9.4) | 9.8 (9.9) |
| **I/sigI** | 26.4 (1.9) | 26.0 (1.5) | 22.2 (3.6) | 21.7 (3.6) |
| **$R_{merge}$* (%)** | 6.9 (22) | 5.6 (27.6) | 10.9 (65.4) | 10.0 (63.6) |
| **Total reflexions** | 28556 (2373) | 28633 (2387) | 32584 (3227) | 32490 (3217) |
| **B factor (Å$^2$)** | 12.9 | 14.0 | 10.5 | 10.5 |

The numbers in parenthesis refer to the highest resolution shell, determine when $I/\sigma I > 2$.
*$R_{merge} = \Sigma_h\Sigma_i |I(h)_i - \langle I(h)\rangle| / \Sigma_h\Sigma_i I(h)_i$, where $I(h)_i$ is the ith observation of reflection h and $\langle I(h)\rangle$ is the mean intensity of that reflection.

**Table 2**. Data collection statistics

As shown in Table 1, the exposure times were equal, the detector readout was 100 s during the high energy experiment described here (1 s at 12 keV), which correspond to a total time of 3 h per data set (120 degrees) compared to 12 minutes in the medium energy experiment; at high energy the eight data sets were recorded in a little bit more than a day and less than 2 h at 12 keV. Overall, there was no decay of the beam intensity during the experiment at 12 keV, while decay of the photon flux was inevitable since the data were not collected in the dose mode. Also, at 56.5 keV, the crystal diffracted to higher resolution with a higher I/σI in the highest resolution shell. Hence, due to the experimental setup employed in the high energy experiment, the crystal was exposed slightly more than necessary.

### 3.3. Radiation damage: Overall parameters variation

As previously described (Ravelli *et al*., 2002) the effect of radiation damage can be observed in a sequence of data set with an increase of the overall B factor, mosaicity, unit

cell volume and decrease of the diffraction resolution firstly in the highest resolution shell.

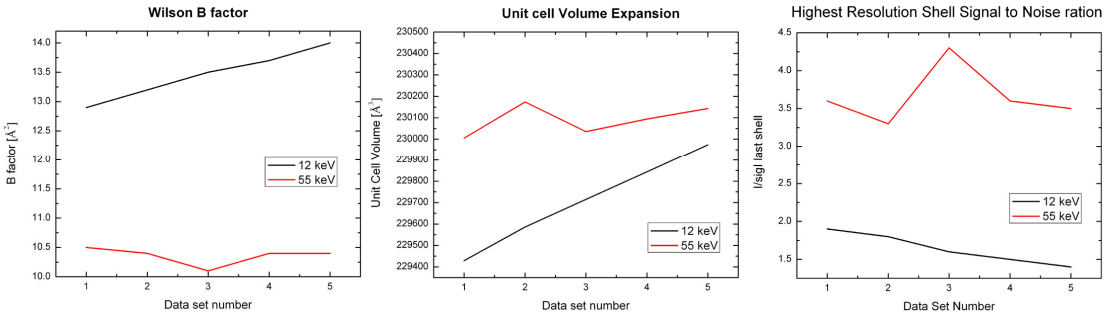



**Figure 1**. Evolution of overall parameters affected by radiation damage at 12 and 56.5 keV.

All parameters from the medium energy data are affected; they monotonously vary over the course of the experiment (Fig. 1). On the other hand, at high energy, variations of these parameters are minimized. The variations are due to the fact that data set were collected at different time of the day and are affected by the photon flux at the time of the data collection. Data set 3, 4, 1, 5 and 2 were collected with decreasing beam intensity: this can be clearly observed in the I/Sigma from the highest resolution shell. Therefore, according to the global parameters, no clear radiation damage was observed at ultra high energy, while all the same parameters indicated sign of radiation damage at 12 keV.

## 3.4. Radiation damage: Structural specific damages

### 3.4.1. Holmium sites

Radiation damage can also be quantitatively assessed following the reduction in peak height of the anomalous scatterers present in the crystal, especially when the crystal is exposed to X-ray at energies very close to an edge, such as the L3 edge of Hg. Specific damage, can then be used when the heavy atoms while being photo reduced (and disappearing) produce isomorphous intensity differences to solve a denovo structure (Ramagopal *et al.*, 2005).

Here, the crystal structure of lysozyme contains three holmium (Ho) sites, Ho1, Ho2 and Ho3 with refined occupancy of 0.8, 0.5 and 0.3. As judged by sequential difference Fourier maps and peak height, there was no detectable decay of the occupancy of any of the three holmium sites. For redundancy, anomalous difference Fourier map were calculated for each of the ten set of data (medium and high energies) and no decrease

of the peak height was observed for all data.

### 3.4.2. Decarboxylation

Decarboxylation of aspartate and glutamate was previously observed and described as a specific damage (Weik *et al.*, 2000); the temperature factor of side chain atoms from these residues increases and their carboxyl group gradually disappears through photo reduction. The method used to follow decarboxylation was the sequential difference Fourier map where the first data set was used as native and each following data set was used to calculate a difference Fourier map; then decarboxylation should be appreciable if occurring in the course of these experiments.

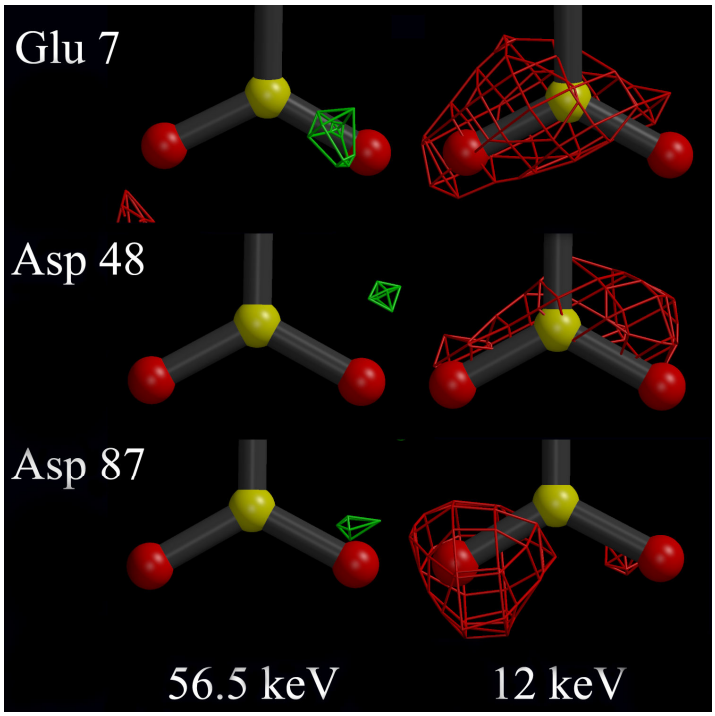



**Figure 2.** Carboxyl groups of Glu 7, Asp 48 and Asp 87 are shown from the top to the bottom; for each carboxyl the left and right sides represent the map calculated at 56.5 and 12 keV respectively. Maps are contoured at 3 sigma level and are the last calculated in sequential difference Fourier analysis. The figure was produced with MOLSCRIPT (Kraulis, 1991), BOBSCRIPT (Esnouf, 1997) and RASTER 3D (Merrit & Murphy, 1994).

As can be observed in the figure 2, decarboxylation didn't occur at 56.5 keV over the period of time the crystal was exposed to X-ray, while electron loss happened at 12 keV

over the same period of time (approximately 1 h exposed to X-rays). Glu 7 and Asp 87 are the most susceptible amino acids to decarboxylation with negative peak height of - 6.2 and - 5.5 sigma level respectively; the peak height at Asp 47 shown in figure 2 is -3.5 sigma. At ultra high energy no peak in the sequential difference Fourier map was observed for any of the carboxyl group with a sigma level cutoff of -3.0. Other Asp and Glu residues are present in HEWL. Some of them are not considered in this study as they bind or they are close to the holmium ions (Asp 101, Asp 52, Asp 35). Asp 119 behaves in the same way than Asp 87, with no detectable peak height at high energy. The two other Asp residues, namely Asp18 and Asp66, didn't show any sign of decarboxylation at 12 and 56.5 keV. In HEWL, most of the Asp and Glu are not involved in crystal contact (with the exception of Asp18) and therefore their fast decarboxylation is not excepted to dramatically contribute to the lost of order in the crystal.

### 3.4.3. Disulfide bridges

HEWL contains four disulfide bridges (S-S) involving all eight cysteine residues present in the structure. Most of the disulfide bridges are located close to the solvent with water molecules approximately 4 Å away and with the exception of the disulfide bridge involving Cys 64 and Cys 80, which is buried. Examination of the sequential difference Fourier maps indicated clear disulfide bond breakage for all bonds at 12 keV while the bond breakage was mitigated at 56.5 keV.

At 56.5 keV, two disulfide bridges (Cys 30-Cys 115 and Cys 6-Cys 127) show only weak and partial negative peak at only one sulfur position with no positive peak in the vicinity of the sulfur, which would indicate a second conformation of the corresponding C-S bond accompanying disulfide bond breakage. The third disulfide bridge between Cys 64 and Cys 80 appears to be more sensitive to radiation with the highest negative peak observed around the sulfur atom in Cys 80. However, no positive peak is observed in the vicinity of the sulfur, and the second S atom involved in this bridge doesn't appear to be sensitive either. Finally, only one sulfur involved in the last S-S bridge between Cys 76 and Cys 94 was shown to be sensitive at the received dose, which is the only C-S bond presenting density for a second conformation, that is Cys 94. However, all peaks observed at high energy are far weaker than the corresponding peaks at 12 keV.

At 12 keV, positive peak associated to negative peak were observed for all cysteine

side chain atom S. The disulfide bridges, according to the associated peak heights, were clearly broken. Also, according to the negative peaks in the sequential difference Fourier map calculation, each disulfide bridges behaved differently. Each single sulfur atom within a disulfide bridge also behaves differently as two different entities. Notably, Cys 30 and Cys 94 appear to be the weakest cysteines in term of peak height decay at 12 keV while Cys 80 and Cys 94 are the weakest at 56.5 keV (highest peak height). At medium energy, the two strongest disulfide bridges seem to be Cys 80-Cys 64 and Cys 6-Cys 127 and at 56.5 keV Cys 6-Cys 127 and Cys 30-Cys 115 are the strongest

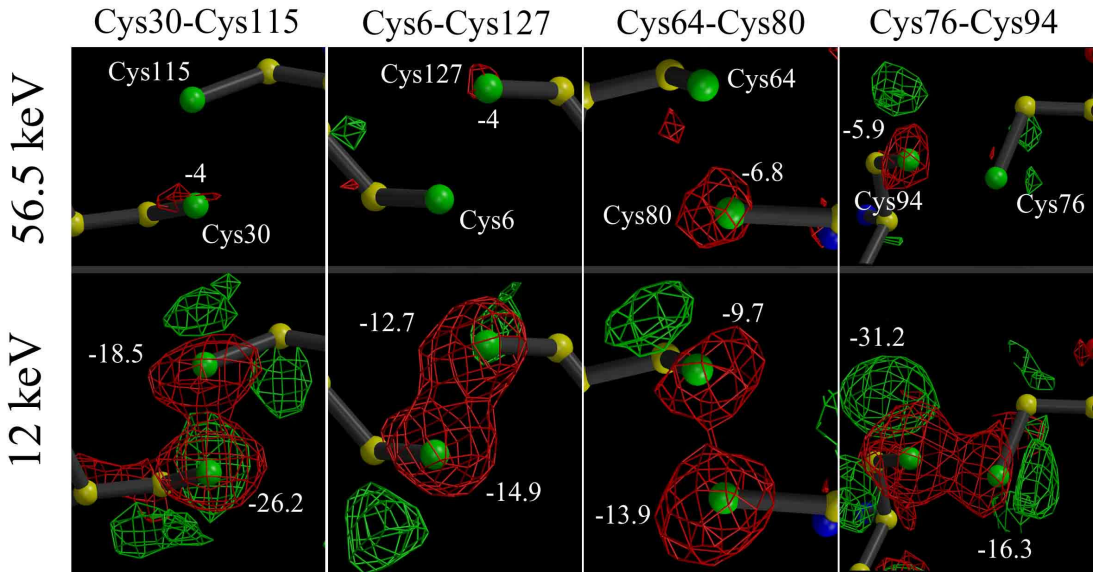



**Figure 3**. Disulfide bond breakage as seen with the last sequential Fourier map. The structure shown is the first structure refined against the first data set at each energy. The maps are contoured at + 3 and – 3 sigma level in green and red respectively. The disulfide bridge is indicated on the top of the figure, each cysteine sulfur atom is indicated in the top view, which correspond to the high energy view. The numbers represent to negative peak height at the sulfur position from the sequential difference Fourier analysis. No number is indicated when the peak height was not detected and smaller than 3 sigma level.

To further study disulfide bond breakage, all cysteine residues were substituted with alanine and difference Fourier maps were calculated for each data set. In that configuration, the decay of the positive peak height was used as the probing criterium. These results clearly indicated a discrepancy between each disulfide bond and also each cysteine involved within a disulfide bond. For instance, in the disulfide bond between Cys 76 and Cys 94, Cys 94 appears to be much more affected than Cys 76 and any other

cysteine within this crystal structure; at 56.5 keV, this is the most sensitive to radiation damage (Fig. 2). The environment of each S-S bond contributes to the observed different strength, nevertheless, the difference observed in a single disulfide bridge remains unclear. At 12 keV additional positive peak could be observed at the corresponding second conformation positions for Cys 94 and 30, the two most sensitive cysteines to radiation damage. At 56.5 keV, no additional positive peak was observed around all the Cys sulfur positions.

The decay of the positive peak (Fig. 4, left panel) indicated that radiation damage could be assessed using this first method at medium energy, while the peak height (Fig. 4., right panel) remained constant in all the data sets for all cysteines at high energy. In summary, Cys 94 is the cysteine that suffers the most and the bridge between Cys 3-Cys 115 the least. Sulfurs in the Cys 64-Cys 80 disulfide bridge, are the only one with occupancy monotonically decreasing over the course of the experiment, at both energies.

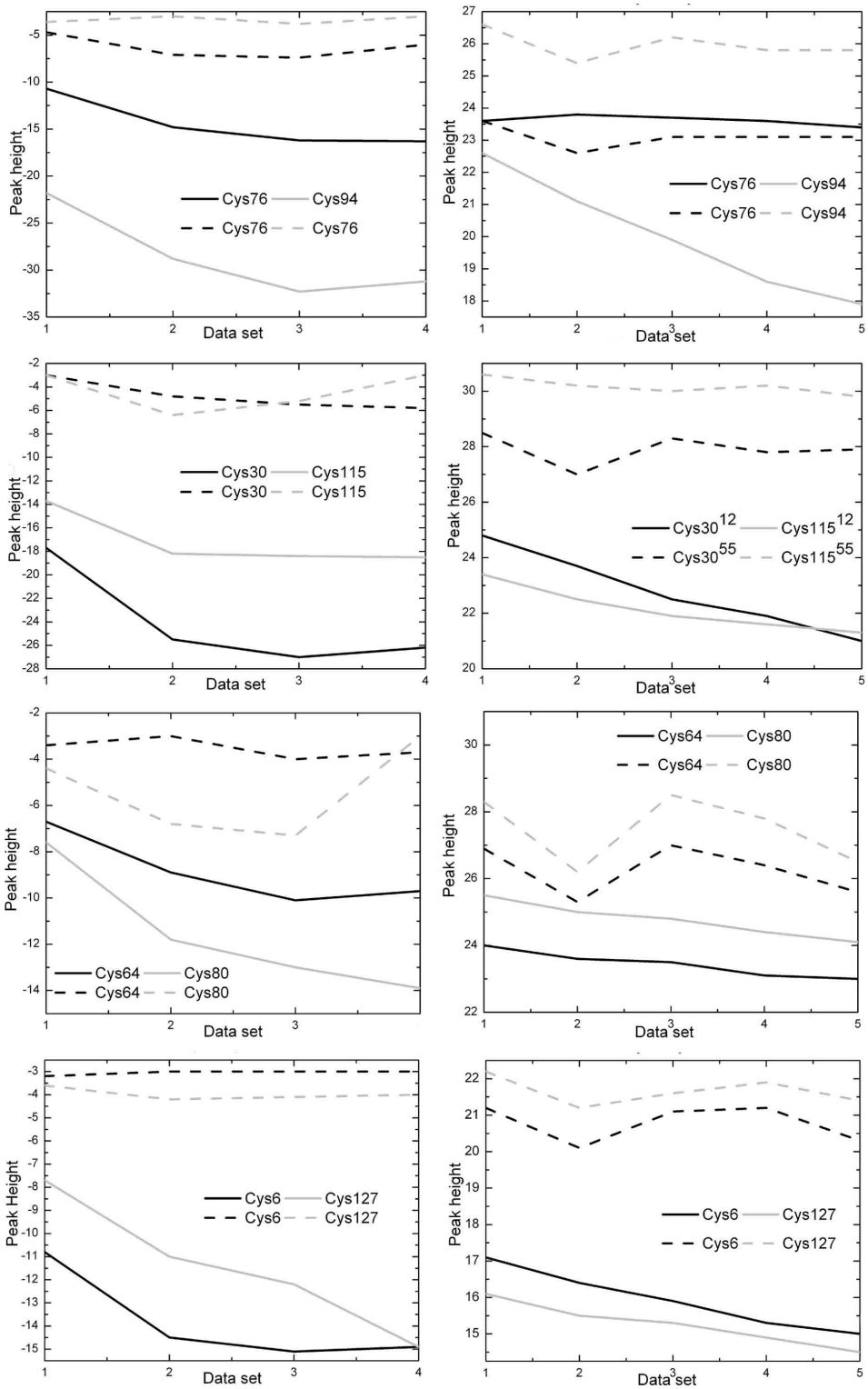



**Figure 4**. Negative peak height in sequential difference Fourier analysis, calculated with the first data set as the reference (left side). Positive peak height in difference Fourier maps calculated with all cysteine residues substituted with alanine (right side). The peak heights are shown in plain and dashed line at medium and ultra high energy respectively.

### 3.5. Reference residues

The most affected amino acids are Cys, Asp and Glu as judged by the relative B-factor increase (Weik *et al*, 2000). Here we observed two methionine residues (Met 12 and 105) and one serine (Ser 91), as they were shown to be mildly and no radiation damage sensitive respectively. The exact same method was applied to assess radiation damage; each amino acid was substituted to an alanine and the positive peak height in the difference Fourier map was used as a radiation damage monitor. Each corresponding positive peak was invariable over the whole set of data processed. This means first that these Met and Ser residues are less sensitive, higher dose is required to break the C-S bond in methionines, secondly that the method used to assess radiation damage is valid. It has to be noted that there was no bias introduced using these data/method, therefore these results are not arising from noise in the data.

### 3.6. Dose estimate

Each data set contains 120 frames, the crystal was exposed 4 s for each frame, crystal, cryo-loop and beam sizes were identical in the two experiments leading to the conclusion that the deposited dose is proportional to the beam fluence and the mass-energy absorption coefficient only as all other parameters influencing the dose were kept constant. The experimentally determined flux at X6A and ID15B are $9\ 10^{10}$ and $2\ 10^{11}$ $ph.s^{-1}$ at the sample respectively; the mass-energy absorption coefficients of the sample are 7.2 and 0.11 $cm^2\ g^{-1}$ at 12 ($1.9\ 10^{-15}$ J) and 56.5 keV ($8.9\ 10^{-15}$ J). The resulting dose is at least 3 to 4 times lower at 56.5 keV compare to 12 keV, thus for the same high resolution limit reached (1.2 Å). According to the difference Fourier analysis, at high energy, the damages observed over the whole course of the experiment (5 data sets) are inferior than damages observed at medium energy between the two first data sets leading to the conclusion that at high energy, the crystal received a dose at least 4 times lower compared to medium energy for similar crystal diffracting to the same resolution limit.

A similar study conducted at 6.2 and 12.4 keV on a cadmium derivative crystal of the porcine pancreatic elastase indicated that for the same dose, no significant difference in radiation damage, according to the photoreduction of the cadmium ion, was observed (Weiss *et al*., 2005). In our study, crystals were not exposed to the same dose, but to the minimum time to allow similar resolution.

## 4. Conclusion

Data collected at ultra high energy (55 keV) were of sufficient quality and allowed the first structure obtained with various phasing methods at energy higher than 40 keV. All methods used at two high energy beamlines located in a second and a third generation synchrotron were successful despite non optimized experimental configuration. At ID15B, when additional data allowed radiation damage quantification and comparison with similar data from medium energy, none of the probes used indicated commencement of radiation damage while onset of radiation damage could be observed in all structural specific sites known to be sensitive as well as in global parameters. Structural specific damages occur at relative low dose, before overall parameters are largely modified and they are proportional to the total energy dose deposited onto the crystal. The most sensitive bonds susceptible to break upon X-ray radiation is S-S bond in disulfide bridges followed by C-O from Asp and Glu. Importantly, at high energy, the dose was estimated to be at least 3-4 times lower for the same resolution limit reached. In this study, we also found that more than the disulfide bond entity a single Cys is to be considered when dealing with radiation damage as the effect is translated in relatively different decay for each S in the S-S bond. The holmium ions with a photoelectric cross section 30 times larger than the sulfur at 12 keV and 300 times at 56 keV were not sensitive to radiation damage at the level of dose exposed over the course of this experiment.

## References


Allaire, M., Aslantas, M., Berntson, A.,Bermann, L., Cheung, S., Clay, B., Greene, R., Jakoncic, J., Johnson, E., Kao, C. C., Lenhard, A., Pjerov, S., Siddons, D. P.,Stober, W., Venkatagiriyappa, V., Yin, Z. and Stojanoff, V. (2003). *Synchrotron Radiation News* **16**, 20-25.

Collaborative Computational Project, Number 4. (1994). *Acta Cryst*. **D50**, 760-763.

Cowtan, K. (1994). Jnt CCP4/ESF-EAMCB *Newsl. Protein Crystallogr*. **31**, 34-38.

Dauter, Z. (2002). *Curr. Opin. Struct. Biol.* **12**:674-678.

Diederichs, K. (2006). *Acta Cryst.* **D62**, 96-101.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst*. **D60**, 2126-2132.

Esnouf, R. M. (1997) *Journal of Molecular Graphics*. **15**, 132-134.

Fourme, R., Kahn, R., Mezouar, M., Girard, E., Hoerentrup, C., Prange, T. & Ascone, I. (2001). *J. Synchrotron Rad.* **8**, 1149-1156.

French, G. S. & Wilson, K. S. (1978). *Acta Cryst*. **A34**, 517-525.

Garman, E.F. and Owen, R.L. (2006). *Acta Cryst*. **D62**:32-47.

Girard, E., Kahn, R., Ascone, I., Mezouar, M., Dhaussy, A. C., Lin, T., Johnson, J. E. & Fourme, R. (2004). *High Pressure Res*. **24**, 173-182.

Gonzalez A. (2003). *Acta Cryst.* **D59**, 315-322.

Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990). *EMBO J.* **9**:1665-1672.

Jakoncic, J., Di Michiel, M., Zhonh, Z., Honkimaki, V., Jouanneau, Y. and Stojanoff, V. (2006). *J. Appl. Cryst*. **39**, 831-841.

Kraulis, P. J. (1991). *J. Appl. Cryst*. **24**, 946-950.

Merritt, E. A. & Murphy, M. E. (1994). *Acta Crystallogr.* D50, 869-873.

Murray, J. and Garman, E. (2002). *J. Synch. Rad*. **9**:347-354.

Murray, J.W., Rudiño-Piñera, E., Owen, R.L., Grininger, M., Ravelli, R.B.G. & Garman. E.F. (2005). *J. Synchrotron Rad*.**12**:268-275.

Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). *Acta Cryst*. **D53**, 240-255.

Otwinowski, Z. & Minor,W. (1997). *Methods Enzymol.* **276**, 307-326.

Perrakis, A., Morris, R. & Lamzin, V.S. (1999). *Nature Struct. Biol*. **6**, 458-463.

Ramagopal, U. A., Dauter, Z., Thirumuruhan, R., Fedorov, E. & Almo, S. C. (2005). *Acta Cryst*. **D61** 1289-1298.

Ravelli, R.B.G., Theveneau, P., McSweeney, S. and Caffrey, M. (2002). *J. Synch. Rad*. **9**:355-360.

Ravelli, R.B.G., Schrøder-Leiros, H.K., Pan, B., Caffrey, M. and McSweeney, S. (2003). *Structure* **11**:217-224

Sheldrick, G. & Gould, R. O. (1995). *Acta Cryst.* **B51**, 423-431.

Schiltz, M., Dumas, P., Ennifar, E., Flensburg, C., Paciorek, W., Vonrhein, C. & Bricogne, G. (2004). *Acta Cryst*. **D60**, 1024-1031.

Takeda, K., Miyatake, H., Park, S.-Y., Kawamoto, M., Kamiya, N. & Miki, K. (2004). *J. Appl. Cryst.* **37**, 925-933.

Weik, M., Ravelli, R.B., Kryger, G., McSweeney, S., Raves, M.L., Harel, M., Gros, P., Silman, I., Kroon, J. and Sussman, J.L. (2000). *Proc. Natl. Acad. Sci. USA* **97**:623–628.

Weiss, M.S., Panjikar, S., Mueller-Dieckmann, C. and Tucke, P.A. (2005). *J. Synchrotron. Rad.* **12**:304-309.

# Partie 4 : Conclusions et perspectives

Si l'utilisation des très hautes énergies a été initialement envisagée par John R. Helliwell (Helliwell *et al.*, 1983) lors de la conception du synchrotron de Grenoble (ESRF), une machine de troisième génération, très peu d'expériences en revanche ont été exécutées. La plupart d'entre elles ont été réalisées à des énergies inferieures à 35 keV et aucune ne traite de dommages dus aux rayonnements.

Dans un premier temps, nous avons démontré qu'il est possible d'utiliser des rayons X à très haute énergie pour déterminer les phases expérimentales d'un cristal de proteine. Nous avons ainsi obtenu la première structure tridimensionnelle d'une proteine à plus de 40 keV. Certes, cette structure est connue, mais notre résultat prouve le potentiel et l'applicabilité des rayons X à très haute énergie en cristallographie des protéines (X-P). L'application des méthodes courantes et l'utilisation des programmes standards nous a permis de déterminer la structure de la proteine modèle, le lysozyme.

Dans un second temps, nous avons comparé les données collectées à 56 keV avec celles d'un cristal similaire dont les données ont été obtenues à 12 keV, avec comme critère la limite de résolution atteinte, et donc indirectement le temps d'exposition. Nous avons démontré que la dose d'énergie absorbée par le cristal est largement inferieure à très haute énergie. Cela ouvre des horizons quant à l'utilisation des très hautes énergies dans le domaine de la cristallographie des protéines. Le critère choisi est la limite de résolution (et donc indirectement $I/\sigma(I)$) et non pas le nombre de photons diffractés ou le nombre de photons absorbés ou tout autre paramètre. En 2005, une étude consistant à comparer des données collectées sur un cristal d'élastase porcine pancréatique (PPE) contenant un ion $Cd^{2+}$ à deux énergies, 12.4 et 6.2 keV, a conduit à la conclusion suivante : l'énergie n'a aucune incidence sur la photoréduction de l'ion $Cd^{2+}$ (Weiss *et al.*, 2005). Toutefois, comparée à notre étude où le facteur $I/\sigma(I)$ est équivalent aux deux énergies employées, ce même facteur est double à 12.4 keV comparé à 6.2 keV. Cela indique qu'à 12.4 keV, le cristal a été surexposé environ quatre fois si l'on considère que le facteur $I/\sigma(I)$ varie avec le temps d'exposition au carré.

Si traditionnellement les expériences de cristallographie des protéines sont exécutées à des énergies allant de 7 à 15 keV, l'énergie optimale/idéale à laquelle le nombre de photons diffractés divisé par le nombre de photons absorbés est maximal sans

tenir en compte l'efficacité du détecteur reste à déterminer. Cette énergie dépend du type d'expérience (MAD, SAD) et de la composition du cristal. A cette énergie et pour un échantillon donné (taille, composition), les dommages sont minimisés tout en maintenant l'intensité de diffraction.

Les perspectives de ce travail sont nombreuses :

- Afin d'apporter un élément de réponse à la question du choix de l'énergie optimale, nous avons commencé une étude théorique basée sur la taille et la composition des échantillons utilisés lors de l'expérience de phasage à 55 keV. Le rapport du nombre de photons diffractés sur le nombre de photons absorbés dépendant de l'angle de diffraction sera déterminé et permettra de selctionner l'énergie optimale.
- L'expérience consistant à déterminer les phases expérimentales d'un cristal de protéine sera répétée en utilisant différentes protéines contenant differents atomes lourds comme par exemple le cadmium et le mercure. Les atomes d'Holmium peuvent être introduits dans un cristal de lysozyme sans aucune difficulté du fait de la présence de chaine latérales acides libres, ce qui n'est pas le cas de toutes les protéines. Ainsi il est nécessaire de tester cette méthode en utilisant d'autres diffuseurs anomaux capables de fixer sur d'autres sites spécifiques (par exemple le mercure sur des cysteines libres). Le but ultime est d'enregistrer les données de diffraction à une énergie ou la diffusion anomale de l'atome lourd est exploitable, sachant qu'une contribution du signal anomal de 0.6 % semble la limite minimale.
- Une protéine dont la structure a été obtenue avec difficulté du fait de la sensibilité de son cristal aux dommages causés par les rayons X sera utilisée en tant qu'échantillon test. Cela permettra à l'avenir de faciliter l'étape de phasage des cristaux particulièrement sensibles.
- Enfin, les conditions dans lesquelles nos expériences ont été effectuées n'étaient pas optimales. En effet les détecteurs ne sont pas adaptés, les instruments utilisés ne sont pas installés de façon permanente. Cela introduit des erreurs inévitables qui sont minimisées dans les installations dédiées. De nouveaux détecteurs de surface seront testés pour leur capacité à enregistrer des données de diffraction à haute énergie. Les autres conditions expérimentales seront aussi améliorées, comme par exemple l'utilisation d'un collimateur positionné à quelques mm de l'échantillon minimisant ainsi le bruit de fond due à la diffusion par l'air. L'utilisation d'un

faisceau focalisé associé à un système optimisé de rotation/obturateur sera une étape supplementaire à la minimisation des erreurs. En effet, les deux installations utilisées lors des expériences acceptent généralement des échantillons de quelques mm à quelques cm, alors que les cristaux de protéines ont en moyenne une taille de 100 μm. Toutes ces améliorations des conditions experimentales permettront de rendre encore plus avantageux l'utilisation des très hautes énergies dans le domaine de la cristallographie.

- Une fois ces étapes effectuées, la structure d'un proteine inconnue sera déterminée en utilisant les phases expérimentales déterminées à très haute énergie et constituera l'étape nécessaire à la généralisation de cette méthode.

# ANNEXES

Dans cette partie j'ai regroupé les abstracts de trois articles auxquels j'ai largement contribué que ce soit lors de la collecte des données, de la détermination de la structure ou de l'affinement du modèle cristallographique.

Le premier article traite de la cristallisation d'un isoforme du cytochrome C bovin, de la résolution de sa structure par la méthode SAD utilisant le signal anomal d'un atome de fer pour 104 acides aminés à 13 keV. Cela afin de minimiser les dommages du aux rayonnements X de moyenne énergie à proximité du seuil K du fer, 7.1 keV. Cet article est en cours de rédaction et sera soumis à la revue *Journal of Crystal Growth* (PDB : **2B4Z**).

Le second article traite de l'étude structurale de VCPB3 contenant les deux domaines tandems V1 et V2 chez amphioxus ou deux structures ont été obtenues par la méthode SAD. Cet article est paru dans le revue *Nature Immunology* (Volume 7, Aout 2006, pages 875-882), (PDB : **1XT5** et **2FBO**)

Enfin le troisième, traite de l'obtention de la structure cristallographique de la 1-4-ß-D-Xylosidase (XO6) de *Geobacillus stearothermophilus,* là encore par le biais de la méthode SAD (PDB **1W91**). Les phases ont été déterminées par la contribution de 64 séléno-méthionines pour une masse moleulaire totale de 450 kDa.

Lors de ma présence au NSLS, j'ai aussi participé au developpement de l'installation X6A dédiée à la cristallographie des macromolécules. Ma participation est ici presentée aussi sous la forme de deux abstracts. Tous deux sont publiés dans la revue *Synchr. Rad. News* (année 2003, volume 16, pages 20-25 et année 2005, volume 18, pages-23-27).

## Abstract 1 (Accepté dans *Proteins: Structure, Function, and Bioinformatics*)

# High Resolution X-ray Crystallographic Structure of Bovine Heart Cytochrome *c* and Its Application to the Design of an Electron Transfer Biosensor

**Nurit Mirkin[1], Jean Jaconcic[2], Vivian Stojanoff[2] and Abel Moreno[3]**

[1]Hunter College, City University of New York, USA. [2]Brookhaven National Laboratory-NSLS, Upton New York, USA. [3]Instituto de Química, UNAM, México D.F., Mexico.

**ABSTRACT**

Cytochrome *c* is one of the most studied proteins probably due to its electron-transfer properties involved in aerobic as well as in anaerobic respiration in living organisms. The advantage of having a red color makes protein purification easier. Particularly, the cytochrome *c* from bovine heart is a small protein $M_r$ 12,230 Da, globular (hydrodynamic diameter of 3.4 nm), soluble in different buffer solutions, and commercially available. In spite of being a quite well-studied protein and relatively easier to manipulate from the biochemical and electrochemical viewpoint, its 3D structure has never been published.

In this work, the purification, crystallization and 3D structural resolution at 1.5Å of one of the isoforms of cytochrome *c* is presented. It is also shown how the presence of isoforms made both the purification and crystallization steps difficult. Finally, a new approach for protein electro-crystallization and design of biosensors is presented.

**Key words**: electrocrystallization, cytochrome C from heart bovine, the gel acupuncture method, atomic force microscopy, X-ray diffraction.

## Abstract 2 (*Nature Immunology* 2006:**7**:875-882)

# Ancient evolutionary origin of diversified variable regions demonstrated by crystal structures of an immune-type receptor in amphioxus

**Jose A Hernandez Prada[1], Robert N Haire[2], Marc Allaire[3], Jean Jakoncic[3], Vivian Stojanoff [3], John P Cannon[2,4], Gary W Litman[2,4,5] and David A Ostrov[1]**

[1]Department of Pathology, Immunology and Laboratory Medicine, University of Florida College of Medicine, Gainesville, Florida 32610, USA. [2]Department of Pediatrics, University of South Florida College of Medicine, Children's Research Institute, St. Petersburg, Florida 33701, USA. [3]Brookhaven National Laboratory National Synchroton Light Source, Upton, NY 11973, USA. [4]Immunology Program, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida 33612, USA. [5]Department of Molecular Genetics, All Children's Hospital, St. Petersburg, Florida 33701, USA.

**ABSTRACT**

Although the origins of genes encoding the rearranging binding receptors remain obscure, it is predicted that their ancestral forms were nonrearranging immunoglobulin-type domains. Variable region–containing chitin-binding proteins (VCBPs) are diversified immune-type molecules found in amphioxus (*Branchiostoma floridae*), an invertebrate that diverged early in deuterostome phylogeny. To study the potential evolutionary relationships between VCBPs and vertebrate adaptive immune receptors, we solved the structures of both a single V-type domain (to 1.15 Å) and a pair of V-type domains (to 1.85 Å) from VCBP3. The deduced structures show integral features of the ancestral variable-region fold as well as unique features of variable-region pairing in molecules that may reflect characteristics of ancestral forms of diversified immune receptors found in modern-day vertebrates.

## Abstract 3 (En cours de soumission)

# Structure determination of the 1-4-ß-D-Xylosidase from Geobacillus stearothermophilus by Seleniomethionine SAD phasing

**Jean Jakoncic[a], Ana Teplytsky[b], GilShoham[c]YuvalShoham[c] andVivian Stojanoff[a]**

[a]Brookhaven National Laboratory, National Synchrotron Light Source, Upton, NY, USA. [b]The Hebrew University of Jerusalem, Jerusalem, Israel. [c]Department of Food Engineereing and Biotechnology, The Technion, Haifa, Israel.

**Synopsis:** SAD phases from 64 Selenium atoms are used for the 3D structure determination of 1-4-ß-D-Xylosidase from *Geobacillus stearothermophilus,* a 450 kDa protein.

**ABSTRACT**

The 3D molecular structure of a mutant of 1-4-ß-D-Xylosidase from *Geobacillus stearothermophilus* was solved using the single anomalous diffraction signal from a 64 Selemethionine tagged and a native protein crystal. The crystals diffracted to 2.8Å and 2.2Å respectively and belonged to space group $P2_12_12_1$ with unit-cell parameters a = 93.67, b = 166.02 and c = 313.03 Å. The asymmetric unit contains 2 tetramers: each monomer comprises 3 different domains, a catalytic domain, a beta-sandwich domain and a small alpha-helical domain. The general fold of the polypeptide chain is very similar to the ß-D-Xylosidase from *Thermoanaerobacterium saccharolyticum*, the first structure solved from the family 39.

**Key words**: SAD, MAD, experimental phases, glycoside hydrolase, xylosidase NCS, tetramer.

## Abstract 4 (*Synchrotron Radiation News* 2003:**16**:20-25)

# The NIGMS Structural Biology Facility at the NSLS

**M.Allaire[1], M.Aslantas[2,1], A.Berntson[1], L.Berman[1], S.Cheung[1], B.Clay[1], R.Greene[1], J.Jakoncic[1], E.Johnson[1], C.C.Kao[1], A.Lenhard[1], S.Pjerov[1], D.P.Siddons[1], W.Stober[1], V.Venkatagiriyappa[1], Z.Yin[1] and V.Stojanoff[1]**

[1]National Synchrontron Light Source, Brookhaven National Laboratory Upton, NY 11973, USA. [2]Hacettepe University, Department of Physics and Engineering, Ankara, Turkey

## ABSTRACT

With the advent of structural genomics and the post-genomic era, there is an increased demand for synchrotron radiation facilities for macromolecular structural biology. Several of the existing facilities are affiliated in ore or more ways with the newly created Centers for Structural Genomics, leaving individual investigators with little to no access to synchrotron radiation. To provide synchrotron access to these small groups, the National Institute of General Medical Sciences (NIGMS) established the X6A facility at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. The purpose is to provide synchrotron access to individual macromolecular crystallography groups and to the biological/biochemical and biophysical community at large. It is the X6A mission to assist expert and non expert crystallographers to provide training to interested individuals in the areas from protein purification to the determination of the molecular atomic coordinates.

The new macromolecular crystallography facility consists of a data collection facility and an associated laboratory for biological sample manipulation. The beam line is located on the X6A bending magnet port of the NSLS X-ray ring. The NSLS is a second generation synchrotron facility; the X-ray ring eventually operates at 2.8 GeV and 300 mA. The X6A was designed for optimal multi-wavelength anomalous diffraction experiments.

## Abstract 5 (*Synchrotron Radiation News,* 2005:**18**:23-27)

# Technical Reports: A Modular Approach to Beam Line Automation: The NIGMS Facility at the NSLS

**M. Allaire[1], A. Berntson, A. Jain, J. Jakoncic[1], C. C. Kao[1], D. R. Siddons[1], I. So[1], V. Venkatagiriyappa[1], Z. Yin[1], V. Tojanoff[1]**

[1]National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY, 11973-5000, USA. [2]Cornell University, Ithaca, NY, 14853, USA

**ABSTRACT**

The last few years have seen an increase in the demand of automation at synchrotron radiation facilities. The main driving forces behind this quest are the Structural Genomics Centers and related projects [1], with their need for large throughput of samples and an increasing number of relatively unskilled users with ever increasing demands.

In order to meet the needs of this diverse community, the structure determination process must be streamlined. A production pipeline for high volume determination of structures requires optimization and automation of current processes in use at synchrotron facilities. The ultimate goal is to arrive at a system that, with little more input than a sample, will provide the researcher with the final molecular structure.